WISCONSIN KNOWLEDGE AND CONCEPTS EXAMINATIONS

FALL 2010 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, a subsidiary of The McGraw-Hill Companies, Inc., 20 Ryan Ranch Road, Monterey, California 93940-5703. Copyright © 2011 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 © 2011 by the Wisconsin Department of Public Instruction

Acknowledgments

The following CTB/McGraw-Hill and DPI staff members are primarily responsible for the content and statistical quality of this report:

CTB/McGraw-Hill:

Christina Schneider, Research Scientist

Marc Julian, Research Manager

Soe Aung, Manager, Statistical Analyst

Kristina Wilson, Senior Research Associate

Kevin Fatica, Research Associate

Russ Wegfehrt, Project Manager, Publishing

Gretchen Schultz, Manager, Content Development

Tracy Keith, Scoring Project Manager

Tammy Bullock, Program Manager

Wisconsin Department of Public Instruction—Office of Educational Accountability:

Lynette Russell, Director

Duane Dorn, Education Consultant

Justin Meyer, Education Consultant

Alison O'Hara, Education Consultant

Viji Somasundaram, Education Consultant

Nick Stroud, Education Consultant

Copyright © 2011 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 2010 Form Selection 9
3.1.2 WKCE Field Test Item Selection 11
3.1.3 Quality Reviews 11
3.2 Description of the WKCE 2010 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 24
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 © 2011 by the Wisconsin Department of Public Instruction

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

6.1 Calibration Sample Data 29
6.1.1 Demographic Comparison of Calibration Sample and Census Data 30

Part 7: Calibration, Equating, and Deriving Scale Scores **33**

7.1 Calibration and Equating Methods 33
7.1.1 Calibration Models 34
7.1.2 Calibration Software 35
7.2 Calibration Results 36
7.2.1 IRT Item Parameters 36
7.2.2 IRT Item Fit 36
7.2.3 Evaluating Anchor Items 38
7.3 Deriving Scale Scores in the WKCE 39
7.3.1 Standard Error of Measurement 43
7.3.2 LOSS and HOSS 44
7.3.3 Test Characteristic Curves 45

Part 8: Test Results **46**

8.1 Classical Item Analysis: Item Level Statistics 46
Flagging for a Positive Distractor Correlation 52
Flagging for the Item-Total Correlation 52
Flagging for p-Value 53
Flagging for Omit Rate 53
Supplemental Tables on Classical Item Analysis 53
8.1.1 Speededness 53
8.2 Raw Score Results 54
Reading 56
Mathematics 56
Language Arts 56
Social Studies 57
Science 57
Subgroup Performance Patterns in Raw Score Results 57
8.3 Summary Statistics for Scale Scores 58
Reading 59
Mathematics 59
Language Arts 59
Social Studies 59
Science 60
Subgroup Performance Patterns in Scale Score Results 60
Gender 60
Race/Ethnicity 60
Socioeconomic Status 61
Disability Status 61
English Language Proficiency 61
Calibration Sample and Census Data Scale Scores 61
8.4 Cut Scores and Performance Level Classifications 62
Reading 63
Mathematics 63
Language Arts 64
Social Studies 64
Science 64
Subgroup Patterns in Performance Level Results 64
8.5 Standard Performance Indicators for Content Standards 65

Copyright © 2011 by the Wisconsin Department of Public Instruction

Reading 67
Mathematics 67
Language Arts 67
Social Studies 67
Science 68
Summary of Student Performance Indicator Results 68
Summary of Student Achievement Results 68

Part 9: Reliability 70

9.1 MEASURES OF INTERNAL CONSISTENCY AND SEM 71
9.2 CLASSIFICATION CONSISTENCY AND ACCURACY 73
9.2.1 Kolen and Kim's Method for Pattern Scoring 75
9.3 INTER-RATER RELIABILITY FOR CR ITEMS AND WRITING PROMPTS 77
Reading 80
Mathematics 81
Writing 81
Summary 82

Part 10: Validity 83

10.1 DIFFERENTIAL ITEM FUNCTIONING 85
Results 91
10.2 CONSTRUCT VALIDITY 92
10.3 TEST INTEGRITY: ERASURE ANALYSIS 94
10.4 STANDARDIZED TEST ADMINISTRATION 94

Part 11: Summary Recommendations 95

Fall 2010 WKCE Technical Report: References 96

Fall 2010 Tables and Figures 100

Appendices

Appendix 1: Fall 2007 Item Selection Check-Off Form 1

Appendix 2: Fall 2006 WKCE Technical Report, Parts 2, 3, 4, 8, and 12 6

Appendix 3: Fall 2005 WKCE Technical Report, Parts 8 and 11 181

Appendix 4: Fall 2007 WKCE Technical Report, Section 2.4 275

Appendix 5: Fall 2010 WKCE Assessment Accommodations Matrix 283

Appendix 6: WKCE Glossary 288

Copyright © 2011 by the Wisconsin Department of Public Instruction

TABLE OF TABLES

PART 2
Table 2-1 Target Reading Test Blueprint: Grades 3–8, 10 101
Table 2-2 Target Mathematics Test Blueprint: Grades 3–8, 10 102
Table 2-3 Target Language Arts Test Blueprint: Grades 4, 8, 10 103
Table 2-4 Target Science Test Blueprint: Grades 4, 8, 10 103
Table 2-5 Target Social Studies Test Blueprint: Grades 4, 8, 10 103
Table 2-6 Actual Reading Test Blueprint: Grades 3–8, 10 104
Table 2-6 Cont'd Actual Reading Test Blueprint: Grades 3–8, 10 105
Table 2-7 Actual Mathematics Test Blueprint: Grades 3–8, 10 106
Table 2-8 Actual Language Arts Test Blueprint: Grades 4, 8, 10 107
Table 2-9 Actual Science Test Blueprint: Grades 4, 8, 10 107
Table 2-10 Actual Social Studies Test Blueprint: Grades 4, 8, 10 107
Table 2-11 Item Development Each Year and Total to Date 108

PART 3
Table 3-1 Fall 2010 Test Configuration 109
Table 3-2 Unique Items Field Tested Each Year and Total to Date 110

PART 4
Table 4-1 Test Accommodations 111

PART 5
Table 5-1 Reading Rubric, Grades 3–8 and 10 122
Table 5-2 Mathematics Rubric, Grades 3–8 and 10 123
Table 5-3 Writing Rubric, Conventions of Written English, Grade 4 124
Table 5-4 Writing Rubric, Composing, Grade 4 125
Table 5-5 Writing Rubric, Conventions of Written English, Grade 8 126
Table 5-6 Writing Rubric, Composing, Grade 8 127
Table 5-7 Writing Rubric, Conventions of Written English, Grade 10 129
Table 5-8 Writing Rubric, Composing, Grade 10 130
Table 5-9 Score Distribution for Reading CR Items 132
Table 5-10 Score Distribution for Mathematics CR Items 133
Table 5-11 Score Distribution for Grades 4, 8, and 10 Writing: Composing Rubric 135
Table 5-12 Percentage Distribution of Scores for Grades 4, 8, and 10 Writing: Composing Rubric 135
Table 5-13 Score Distribution for Grades 4, 8, and 10 Writing: Conventions Rubric 136
Table 5-14 Percentage Distribution of Scores for Grades 4, 8, and 10 Writing: Conventions Rubric 136
Table 5-15 Score Distribution for Grades 4, 8, and 10 Writing: Total Score 137
Table 5-16 Percentage Distribution of Scores for Grades 4, 8, and 10 Writing: Total Score 137

PART 6
Table 6-1 The Current 13 Calibration Districts 138
Table 6-2 Total Number of Students in Census and Calibration Sample Data 139
Table 6-3 Number and Percent of Students in Census and Calibration Sample Data, by Gender 140
Table 6-4 Number and Percent of Students in Census and Calibration Sample Data, by Race/Ethnicity 141
Table 6-5 Number and Percent of Students in Census and Calibration Sample Data, by Socioeconomic Status 142
Table 6-6 Number and Percent of Students in Census and Calibration Sample Data, by Disability 143
Table 6-7 Number and Percent of Students in Census and Calibration Sample Data, by ELP Status 144

Copyright © 2011 by the Wisconsin Department of Public Instruction

TABLE OF TABLES CONTINUED

PART 7

Table 7-1 Item Flagged Based on Yen's Q1 145
Table 7-2 Scoring Table for Reading Grade 3 146
Table 7-3 Scoring Table for Reading Grade 4 147
Table 7-4 Scoring Table for Reading Grade 5 148
Table 7-5 Scoring Table for Reading Grade 6 149
Table 7-6 Scoring Table for Reading Grade 7 150
Table 7-7 Scoring Table for Reading Grade 8 151
Table 7-8 Scoring Table for Reading Grade 10 152
Table 7-9 Scoring Table for Mathematics Grade 3 153
Table 7-10 Scoring Table for Mathematics Grade 4 154
Table 7-11 Scoring Table for Mathematics Grade 5 155
Table 7-12 Scoring Table for Mathematics Grade 6 156
Table 7-13 Scoring Table for Mathematics Grade 7 157
Table 7-14 Scoring Table for Mathematics Grade 8 158
Table 7-15 Scoring Table for Mathematics Grade 10 159
Table 7-16 Scoring Table for Language Arts Grade 4 160
Table 7-17 Scoring Table for Language Arts Grade 8 161
Table 7-18 Scoring Table for Language Arts Grade 10 162
Table 7-19 Scoring Table for Social Studies Grade 4 163
Table 7-20 Scoring Table for Social Studies Grade 8 164
Table 7-21 Scoring Table for Social Studies Grade 10 165
Table 7-22 Scoring Table for Science Grade 4 166
Table 7-23 Scoring Table for Science Grade 8 167
Table 7-24 Scoring Table for Science Grade 10 168
Table 7-25 The Number of Students and Percents at LOSS and HOSS 169

PART 8

Table 8-1 Item Analysis Grade 3 Reading 170
Table 8-2 Item Analysis Grade 4 Reading 173
Table 8-3 Item Analysis Grade 5 Reading 176
Table 8-4 Item Analysis Grade 6 Reading 179
Table 8-5 Item Analysis Grade 7 Reading 182
Table 8-6 Item Analysis Grade 8 Reading 185
Table 8-7 Item Analysis Grade 10 Reading 188
Table 8-8 Item Analysis Grade 3 Mathematics 191
Table 8-9 Item Analysis Grade 4 Mathematics 194
Table 8-10 Item Analysis Grade 5 Mathematics 197
Table 8-11 Item Analysis Grade 6 Mathematics 200
Table 8-12 Item Analysis Grade 7 Mathematics 203
Table 8-13 Item Analysis Grade 8 Mathematics 206
Table 8-14 Item Analysis Grade 10 Mathematics 209
Table 8-15 Item Analysis Grade 4 Language Arts 212
Table 8-16 Item Analysis Grade 8 Language Arts 214
Table 8-17 Item Analysis Grade 10 Language Arts 216
Table 8-18 Item Analysis Grade 4 Social Studies 218
Table 8-19 Item Analysis Grade 8 Social Studies 220
Table 8-20 Item Analysis Grade 10 Social Studies 222
Table 8-21 Item Analysis Grade 4 Science 224
Table 8-22 Item Analysis Grade 8 Science 226
Table 8-23 Item Analysis Grade 10 Science 228
Table 8-24 The Number of Items Flagged 230

Copyright © 2011 by the Wisconsin Department of Public Instruction

TABLE OF TABLES CONTINUED

PART 8 CONT'D

Table 8-25 Raw Score Descriptive Statistics based on Census Data 231
Table 8-26 Raw Score Descriptive Statistics based on Calibration Sample 232
Table 8-27 Raw Score Descriptive Statistics by Gender 233
Table 8-28 Raw Score Descriptive Statistics for Reading by Race/Ethnicity 234
Table 8-29 Raw Score Descriptive Statistics for Mathematics by Race/Ethnicity 235
Table 8-30 Raw Score Descriptive Statistics for Language Arts by Race/Ethnicity 236
Table 8-31 Raw Score Descriptive Statistics for Social Studies by Race/Ethnicity 236
Table 8-32 Raw Score Descriptive Statistics for Science by Race/Ethnicity 237
Table 8-33 Raw Score Descriptive Statistics by Socioeconomic Status 238
Table 8-34 Raw Score Descriptive Statistics by Disability 239
Table 8-35 Raw Score Descriptive Statistics by English Language Proficiency 240
Table 8-36 Scale Score Mean and Standard Deviation for Census and Calibration Sample 241
Table 8-37 Scale Score Descriptive Statistics based on Calibration Sample 242
Table 8-38 Scale Score Descriptive Statistics based on Census Data 243
Table 8-39 Scale Score Descriptive Statistics by Gender 244
Table 8-40 Scale Score Descriptive Statistics for Reading by Race/Ethnicity 245
Table 8-41 Scale Score Descriptive Statistics for Mathematics by Race/Ethnicity 246
Table 8-42 Scale Score Descriptive Statistics for Language Arts by Race/Ethnicity 247
Table 8-43 Scale Score Descriptive Statistics for Social Studies by Race/Ethnicity 248
Table 8-44 Scale Score Descriptive Statistics for Science by Race/Ethnicity 248
Table 8-45 Scale Score Descriptive Statistics by Socioeconomic Status 249
Table 8-46 Scale Score Descriptive Statistics by Disability 250
Table 8-47 Scale Score Descriptive Statistics by English Language Proficiency 251
Table 8-48 Performance Level Cut Scores for all Contents 252
Table 8-49 Percentage of Students in Each Performance Level by Sub-Group (Reading) 253
Table 8-50 Percentage of Students in Each Performance Level by Sub-Group (Mathematics) 256
Table 8-51 Percentage of Students in Each Performance Level by Sub-Group (Language Arts) 259
Table 8-52 Percentage of Students in Each Performance Level by Sub-Group (Social Studies) 260
Table 8-53 Percentage of Students in Each Performance Level by Sub-Group (Science) 261
Table 8-54 Cut Scores and Associated Impact Data for WKCE-CRT Reading 262
Table 8-55 Cut Scores and Associated Impact Data for WKCE-CRT Mathematics 262
Table 8-56 Cut Scores and Associated Impact Data for WKCE-CRT Language Arts 263
Table 8-57 Cut Scores and Associated Impact Data for WKCE-CRT Social Studies 263
Table 8-58 Cut Scores and Associated Impact Data for WKCE-CRT Science 264
Table 8-59 Summary Statistics for Reading Content Standards Raw and SPI Scores 265
Table 8-60 Summary Statistics for Mathematics Content Standards Raw and SPI Scores 267
Table 8-61 Summary Statistics for Language Arts Content Standards Raw and SPI Scores 269
Table 8-62 Summary Statistics for Social Studies Content Standards Raw and SPI Scores 270
Table 8-63 Summary Statistics for Science Content Standards Raw and SPI Scores 271
Table 8-64 SPI Cut Scores 272

PART 9

Table 9-1 Reliability for Total Group and Subgroups Using Cronbach's Alpha 275
Table 9-2 Standard Error of Measurement for Total Group and Subgroups 276
Table 9-3 Cronbach's Alpha Reliability Coefficients for Content Standards 277
Table 9-4 Standard Error of Measurement per Content Standard 278
Table 9-5 Classification Consistency and Classification Accuracy for Reading Grade 3 279
Table 9-6 Classification Consistency and Classification Accuracy for Reading Grade 4 280
Table 9-7 Classification Consistency and Classification Accuracy for Reading Grade 5 281
Table 9-8 Classification Consistency and Classification Accuracy for Reading Grade 6 282
Table 9-9 Classification Consistency and Classification Accuracy for Reading Grade 7 283

Copyright © 2011 by the Wisconsin Department of Public Instruction

TABLE OF TABLES CONTINUED

PART 9 CONT'D

Table 9-10 Classification Consistency and Classification Accuracy for Reading Grade 8 284
Table 9-11 Classification Consistency and Classification Accuracy for Reading Grade 10 285
Table 9-12 Classification Consistency and Classification Accuracy for Mathematics Grade 3 286
Table 9-13 Classification Consistency and Classification Accuracy for Mathematics Grade 4 287
Table 9-14 Classification Consistency and Classification Accuracy for Mathematics Grade 5 288
Table 9-15 Classification Consistency and Classification Accuracy for Mathematics Grade 6 289
Table 9-16 Classification Consistency and Classification Accuracy for Mathematics Grade 7 290
Table 9-17 Classification Consistency and Classification Accuracy for Mathematics Grade 8 291
Table 9-18 Classification Consistency and Classification Accuracy for Mathematics Grade 10 292
Table 9-19 Classification Consistency and Classification Accuracy for Language Arts Grade 4 293
Table 9-20 Classification Consistency and Classification Accuracy for Language Arts Grade 8 294
Table 9-21 Classification Consistency and Classification Accuracy for Language Arts Grade 10 295
Table 9-22 Classification Consistency and Classification Accuracy for Social Studies Grade 4 296
Table 9-23 Classification Consistency and Classification Accuracy for Social Studies Grade 8 297
Table 9-24 Classification Consistency and Classification Accuracy for Social Studies Grade 10 298
Table 9-25 Classification Consistency and Classification Accuracy for Science Grade 4 299
Table 9-26 Classification Consistency and Classification Accuracy for Science Grade 8 300
Table 9-27 Classification Consistency and Classification Accuracy for Science Grade 10 301
Table 9-28 Inter-Rater Reliability, Reading 302
Table 9-29 Inter-Rater Reliability, Mathematics 303
Table 9-30 Inter-Rater Reliability, Writing Prompts 305

PART 10

Table 10-1 Items Flagged for DIF, By Gender 306
Table 10-2 Items Flagged for DIF, By Race/Ethnicity, African American 307
Table 10-3 Items Flagged for DIF, By Race/Ethnicity, Hispanic 308
Table 10-4 Items Flagged for DIF, By Race/Ethnicity, Asian 309
Table 10-5 Items Flagged for DIF, By Race/Ethnicity, American Indian 312
Table 10-6 Items Flagged for DIF, By English Language Proficiency 313
Table 10-7 Items Flagged for DIF, By Disability Status 315
Table 10-8 Correlations among Reading Objectives 316
Table 10-9 Correlations among Mathematics Objectives 317
Table 10-10 Correlations among Language Arts Objectives 318
Table 10-11 Correlations among Social Studies Objectives 318
Table 10-12 Correlations among Science Objectives 319
Table 10-13 Principal Components Analysis 320

Copyright © 2011 by the Wisconsin Department of Public Instruction

TABLE OF FIGURES

PART 7

Figure 7-1 SEM Curves, Reading Grades 3-6 321
Figure 7-2 SEM Curves, Mathematics Grades 3-6 323
Figure 7-3 SEM Curves, Language Arts Grades 4, 8, 10 325
Figure 7-4 SEM Curves, Social Studies Grades 4, 8, 10 326
Figure 7-5 SEM Curves, Science Grades 4, 8, 10 327
Figure 7-6 TCC Curve for Reading Grades 3-8, 10 328
Figure 7-7 TCC Curve for Mathematics Grades 3-8, 10 329
Figure 7-8 TCC Curve for Language Arts Grades 4, 8, 10 330
Figure 7-9 TCC Curve for Social Studies Grades 4, 8, 10 331
Figure 7-10 TCC Curve for Science Grades 4, 8, 10 332

PART 9

Figure 9-1 Reading Indices for Classification Consistency and Classification Accuracy 333
Figure 9-2 Mathematics Indices for Classification Consistency and Classification Accuracy 333
Figure 9-3 Language Arts Indices for Classification Consistency and Classification Accuracy 334
Figure 9-4 Social Studies Indices for Classification Consistency and Classification Accuracy 334
Figure 9-5 Science Indices for Classification Consistency and Classification Accuracy 335

Copyright © 2011 by the Wisconsin Department of Public Instruction

Part 1: Overview

The Fall 2010 Wisconsin Knowledge and Concepts Examinations (WKCE) Technical Report documents the processes and procedures applied in the Fall 2010 WKCE and 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 2010 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/McGraw-Hill. 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 Department of Public Instruction (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 s. 118.30). Based on the NCLB legislation, student performance, reported in terms of proficiency categories, is used to determine the adequate yearly progress 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 (s. 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 © 2011 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 2010 WKCE is the sixth 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/McGraw-Hill and edited and reviewed by Wisconsin teachers and items from CTB/McGraw-Hill's norm-referenced test, *TerraNova*, The Second Edition (*TerraNova*, CTB/McGraw-Hill, 2001). The Fall 2010 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 © 2011 by the Wisconsin Department of Public Instruction

Test Form Development

- Part 3 discusses key development tasks and issues related to creating the Fall 2010 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 2009, totaling 5,524 unique items.
- Selection of the Fall 2010 operational forms was done using the ITEMWIN (Burket, 2000) software using 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 25–November 26, 2010.
- Delivery of materials was handled through the district and school assessment coordinators.
- In 2010, 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 of reliable and valid test scores.
- The scoring rubrics used in handscoring are presented in detail for all content areas with handscored items.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Characteristics of the Calibration Sample

- As detailed in Part 6, the calibration and equating of the Fall 2010 WKCE tests were based on a sample of student response data termed the calibration sample. The school districts used to obtain the calibration sample are identified in the report. These districts were used for the same purposes in previous administrations.
- Part 6 provides a comparison of the demographic characteristics, mean scale scores, and standard deviations of the calibration sample versus the census data. The demographic comparison references five categories: gender, race/ethnicity, socioeconomic status (SES), disability status, and English language proficiency. The comparison shows the calibration sample roughly approximates the 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 2010 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.
- Item-pattern scoring was applied to the Fall 2010 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 the standard error of measurement.
- 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.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Reliability

- Part 9 elaborates on the reliability of the test based on results presented in previous parts of the report.
- Standard error of measurement was assessed for raw scores and scale scores.
- Inter-rater reliability was estimated for all constructed-response items.
- Internal consistency was assessed for all MC and CR items using Cronbach's alpha.
- Classification consistency and accuracy were estimated for performance classification.

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 is presented.
- Erasure analysis, a procedure used to identify high erasure rates, is also discussed.

Summary Recommendations

- Key findings of the Fall 2010 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 2010 administration cover three different phases of the testing cycle: item development, scoring, and psychometric, or measurement-based, research and evaluation.

Copyright © 2011 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 2010 test 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 2010 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 report in Appendix 2. 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 this report). 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 © 2011 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://dpi.wi.gov/oea/wkce.html. The Fall 2006 Technical Report, Section 3.1.1, explains the structure and development of the Assessment Frameworks (see Appendix 2 for reference).

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 which 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 for reference). Tables 2-1 through 2-5 present the target blueprints for the Fall 2010 test. Tables 2-6 through 2-10 present the actual test blueprints showing how the items selected for the Fall 2010 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' 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

Copyright © 2011 by the Wisconsin Department of Public Instruction

selecting the 2008 form, 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 2010 operational [1] forms were selected, reviewed, and approved between 2001 and 2006. 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).

2.4 Item Development and Editing

While historically new items have been developed each year for the WKCE, in 2010 new items were not developed. Table 2-11 shows the number of MC, CR, and total items that have been written up to 2009 for the WKCE.

2.5 Content/Bias Review and Item Alignment

Because there were no field test items on the 2010 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 © 2011 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 2010 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 2010 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 © 2011 by the Wisconsin Department of Public Instruction

CTB assessment editors selected items for the 2010 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.
- Multiple-choice items with *p*-values below 0.30 should be avoided when possible.
- Constructed-response 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 operational test forms for all content areas and grade levels. These 2007 test forms were re-administered in 2010 with no modifications made to the operational selection. 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 selections, as necessary, in order to improve the test characteristic curve or standard error curve.

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 which were used in 2010 were as similar as possible to the 2006 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, the CTB assessment editor 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 2010, DPI reviewed the 2007 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 2010 WKCE.

3.1.2 WKCE Field Test Item Selection

No items were field tested in 2010. 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 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.

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 2010 Tests

The 2010 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 multiple-choice 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 multiple-choice items. Two of the three operational sessions included a constructed-response item. One of the constructed-response items was for the reporting category "Analyzing Text," while the other was for the reporting category "Evaluate and Extend 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 © 2011 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 © 2011 by the Wisconsin Department of Public Instruction

- item selection for the Fall 2007 form that was reused in 2010
- first pages in 2010
- 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 2007, CTB submitted item selection summaries to DPI for the 2007 test form, which were subsequently re-administered in 2010. 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 2007 test forms, it was determined that second pages approvals were not needed for this administration.

Copyright © 2011 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 © 2011 by the Wisconsin Department of Public Instruction

Part 4: Test Administration

In the Fall of 2010, 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 25–November 26, 2010. 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 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 5 and http://dpi.wi.gov/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 accommodation 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

For the Fall 2010 test administration, the State of Wisconsin developed Spanish and Hmong translation scripts for the WKCE. The aim of these scripts is to better help students

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 2010 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 level.

4.3 Test Security

The primary goal of test security is to protect the integrity of the examinations. 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 2010 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 the DPI.

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

Copyright © 2011 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 a highlighter 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 © 2011 by the Wisconsin Department of Public Instruction

Parameters for marking test booklets 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 booklet).
- Do not mark in the skunk lines (the little squares and rectangles across the bottom of each page of the test booklet).
- 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 booklet).
- 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, 2010–2011). 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 © 2011 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 which are available online at: http://dpi.wi.gov/oea/publications.html.

Copyright © 2011 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 2010 test 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 © 2011 by the Wisconsin Department of Public Instruction

Updates—The raw scoring and editing of scanned student data was 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 © 2011 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 2010 administration. The following processing metrics were obtained:

- Number of sheets scanned: 21,506,924
- Number of books scanned/processed: 432,664

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 2010, 204 discrepancies were identified and resolved by emails and telephone calls placed to the districts. These completeness checks occurred from December 2, 2010, to January 14, 2011.

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.

Copyright © 2011 by the Wisconsin Department of Public Instruction

5.2 Scoring of Open-Ended or Constructed-Response Items

Sections 5.2.1 through 5.2.7 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 5.2.7 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 2010 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.

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.

Copyright © 2011 by the Wisconsin Department of Public Instruction

5.2.4 Anchor Papers and Training Papers

In 2010, all training materials, including scoring guides and Reading and Mathematics rubrics, anchor papers, training papers, qualification round papers, and checksets, were from the 2007 operational administration. Prior to the actual scoring in 2007, 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.

<u>Readers</u>

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 and 5.3 and in Part 9.

<u>Team Leaders</u>

Team leaders were selected on the basis of their ability to maintain a high degree of scoring accuracy and consistency, often across multiple subjects and grades. Team leaders were also required to possess good interpersonal and leadership skills in order to be effective when

Copyright © 2011 by the Wisconsin Department of Public Instruction

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

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 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]

Operational items were the same across forms. 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.

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

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 (in "second reads") to obtain indices regarding the consistency and accuracy of 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 © 2011 by the Wisconsin Department of Public Instruction

Part 6: Characteristics of the Calibration Sample

The calibration, equating, and scoring of the Fall 2010 WKCE were 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 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 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.

6.1 Calibration Sample Data

Table 6-1 lists the 13 school districts from which the calibration sample was obtained. In order to maintain comparability across years, the 2010 calibration sample consisted of the same districts that were used in previous years. Prior to the Fall 2007 administration, the calibration sample included 14 districts, but the Ashland School District was dropped from the list in 2007 for logistical reasons. The calibration sample was selected to represent the state student population in terms of demographic composition and student performance. The selection was made based on analyses of the demographic and performance profiles of the districts, as well as recommendations from DPI. The sample was designed to provide a slight overrepresentation of African American, Hispanic, and Asian students in order to ensure that the numbers of students in these ethnic groups were sufficient to support subgroup analyses.

In accordance with federal regulations regarding the capture and reporting of student race and ethnicity information, the DPI changed to the approved federal reporting system for the 2010–11 school year. This results in the following options for students. Students must first identify as either: 1) Hispanic or Latino or 2) Not Hispanic or Latino. Additionally, students must then select one or more of the following: 1) American Indian or Alaska Native, 2) Asian, 3) Black or African American, 4) Native Hawaiian or Other Pacific Islander, and 5) White. The DPI is applying a bridging strategy in order to convert this information back to the existing five categories until further notice (see http://dpi.wi.gov/lbstat/dataracfaq.html for additional information). Given the change in reporting of race and ethnicity information by students and parents and the subsequent bridging of data by the DPI, there is potential for differences within the existing five categories as reported here in comparison to other and prior data aggregations. Where longitudinal differences appear that are likely related to the new coding, a footnote will be applied to alert a reader to the likely reason for the differences.[4]

Table 6-2 shows the number of students in the calibration sample and in the census data for each grade level. The calibration sample included approximately 9% of the total population of tested students at each grade level. Readers should note that the sampling unit in the sample is

[4] The race/ethnicity data collection was revised for the 2010-11 school year. Race/ethnicity differences are likely a result of the changes to the data collection as opposed to actual changes in the population of students in Wisconsin.

Copyright © 2011 by the Wisconsin Department of Public Instruction

the school. Public-school students within selected schools in each of the 13 calibration districts were part of the calibration sample. Students who had invalid records were excluded from the calibration sample. Students are excluded from the calibration sample if all five of the first five items in a content area are omitted, multi-marked (two or more response choices marked), or both.

6.1.1 Demographic Comparison of Calibration Sample and Census Data

The calibration sample has historically been similar to the census data; however, each year CTB performs two analyses to ensure that this trend is maintained. First, the demographic composition of the calibration sample at each grade is examined and compared to that of the census data for the same grade. Demographic checks occur for the following categories: gender, race/ethnicity, socioeconomic status, disability status, and English language proficiency. Next, the mean scale scores of the calibration sample at each grade are compared to the mean scale scores of the census data for the same grade. The results of this latter comparison are provided in Part 8, where results of the Fall 2010 administration are discussed. As detailed in the text that follows and in Part 8, the results of these two analyses show that, in 2010, the demographic makeup of the 13 calibration districts was sufficiently similar to the demographic makeup of Wisconsin students and appropriate to use for the calibrating and equating process. The demographic profiles of the two data sets are similar, but some differences exist. However, because WKCE uses item response theory (IRT) to calibrate and equate student responses and there are sufficient numbers of students at each point along the ability range, the discrepancies between the calibration sample and the population data are not expected to have any appreciable impact on the calibration results.

Side-by-side comparisons of the calibration sample and the census data are presented in Tables 6-3 through 6-7. The two data sources are compared first in terms of gender, followed by race/ethnicity, socioeconomic status (SES), disability status, and finally by English language proficiency (ELP).

The subgroup categories used in Tables 6-3 through 6-7 are also used in subsequent parts of the report. Gender is compared in terms of male and female students. The comparisons in terms of race/ethnicity refer to students who are White, African American, Hispanic, Asian, and American Indian. Socioeconomic status comparisons use the dichotomy “economically disadvantaged” and “not economically disadvantaged.” Disability status comparisons use the dichotomy “not disabled” and “disabled.” Comparisons based on English language proficiency status identify students as “proficient,” or more specifically, “fully English proficient,” and “not proficient,” or “limited English proficient.” Please note that the concept of proficiency used here specifically refers to proficiency in the English language and is not the same as the concept of proficiency used to classify students on the basis of performance levels.

As shown in Table 6-3, there were more male students than female students in both the calibration sample and census data. The differences between the calibration sample and the census data were small. The census contained approximately 51% males and 49% females at each grade. The calibration sample also contained approximately 51% males and 49% females in

Copyright © 2011 by the Wisconsin Department of Public Instruction

all but three grades: In grades 6, 7, and 10 the calibration sample contained 52% males and 48% females.

The composition of the calibration sample and the census data are compared in Table 6-4 on the basis of race/ethnicity. Overall, the two data sources show the same pattern: White students were numerically predominant, by far, in all grades. African American students were the next largest group, followed by Hispanic, Asian, and American Indian students. However, compared to the census data, the calibration sample contained a smaller proportion of White students and a larger proportion of African American, Hispanic, and Asian students. Across all grade levels, White students were 73% to 77% of the student population in the census data, but only 61% to 66% of the calibration sample. This historic overrepresentation of minority students in the calibration sample is done to obtain minority samples large enough to perform all required test analyses.

Calibration sample and census data comparisons for socioeconomic status are shown in Table 6-5. As described previously, students were classified as economically disadvantaged or not economically disadvantaged. The calibration sample and census data show similar profiles: A large proportion of the student population was identified as economically disadvantaged. The calibration sample has a slightly higher percentage of economically disadvantaged students than the census data, but the calibration sample percentages are within 8% of the census data proportions. Both the calibration sample and the census data reflect a general decline in the proportion of students identified as economically disadvantaged as grade level increases. The census data shows that 38% to 44% of students in grades 3–8 were economically disadvantaged, and 34% of students in grade 10 were economically disadvantaged. Similarly, the calibration sample shows that 43% to 48% of students were economically disadvantaged in grades 3–8, and in grade 10, the proportion drops to 37%.

Table 6-6 presents data on disability status. The calibration sample and census data show similar patterns, though the calibration sample has a slightly larger percentage of students identified as having a disability than the census data. Approximately 13% of students in the census data were identified as having a disability. The difference between the census data and the calibration sample was no more than 2% in any grade level.

Table 6-7 shows the percentages of students who were identified as fully English proficient and limited English proficient. The calibration sample and census data show similar patterns: Most students are fully English proficient. However, the percentage of students identified as limited English proficient is larger in the calibration sample than in the census data. Approximately 9% to 15% of students in the calibration sample were identified as limited English proficient, whereas 4% to 8% of students in the census data were identified as limited English proficient. This discrepancy is consistent with the intentional oversampling of Hispanic and Asian students, two groups that may include a disproportionate number of students for whom English is a second language. Both the calibration sample and census data show more students in the lower grades than in the higher grades being identified as limited English proficient, which is consistent with national trends.

Copyright © 2011 by the Wisconsin Department of Public Instruction

In summary, the demographic profiles of the two data sets are similar and consistent with historical trends for the WKCE program. The sample data are sufficiently representative of Wisconsin's student population for the purposes of calibration. The quality of the sample data and its comparability to the entire student population forms an important part of the validity evidence supporting the WKCE program.

Copyright © 2011 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 in the Fall 2010 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 and 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 the Fall 2010 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, are also discussed.

7.1 Calibration and Equating Methods

In the Fall 2010 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 comprise 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 is 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 © 2011 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 were common to both the current 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 are 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 are administered only in grades 4, 8, and 10. Instead, the scales for grades 4, 8, and 10 were constructed in such a way so as 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 administration, which can be found in Appendix 3. The 2005 Technical Report includes a fairly extensive discussion of the scaling methods.

7.1.1 Calibration Models

The three-parameter logistic 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 two-parameter partial credit 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

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

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 © 2011 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 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 also shown that the IRT procedures used here produce accurate vertical scaling (Yen & Burket, 1997).

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 standard error of measurement 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, 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 which 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 © 2011 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 7, item 11 was flagged because the observed Z of 27.32 is larger than the critical Z value of 14.05 based on a sample size of 5,268.

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, Reading grade 6 item 44 was flagged because the observed Z of 17.12 is larger than the critical Z value of 15.76. The misfit in this case may be considered small. The Language Arts grade 10 item 32 was also flagged for a small misfit. In this case the observed Z of 16.33 is slightly higher than the critical Z of 15.13. Although many items in the table show a moderate degree of difference between the obtained Z-statistic and the critical Z-statistic, others, such as the Reading grade 10 item 20, show a much larger difference.

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 WKCE is consistent with, though slightly more than, the number flagged in the year prior. As noted, the difference between the obtained Z-statistic and the critical Z-statistic was

Copyright © 2011 by the Wisconsin Department of Public Instruction

often small or moderate. Items flagged for misfit were reported to CTB Development 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 item characteristic curves 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 are those greater than + 0.15, and moderate differences are 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 item characteristic curves.
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,

Copyright © 2011 by the Wisconsin Department of Public Instruction

improved performance on the item represents true growth in that area. Removing the anchor item may artificially lower test scores.

3. Removal of the item may not significantly alter the content distribution of the anchor set. The distribution of the anchor items across the content standards must remain within 10% of the test blueprint.

No anchor items in 2010 were excluded based upon the criteria specified above.

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.

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

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.

Copyright © 2011 by the Wisconsin Department of Public Instruction

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

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

.

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 multiple-choice 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 © 2011 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.

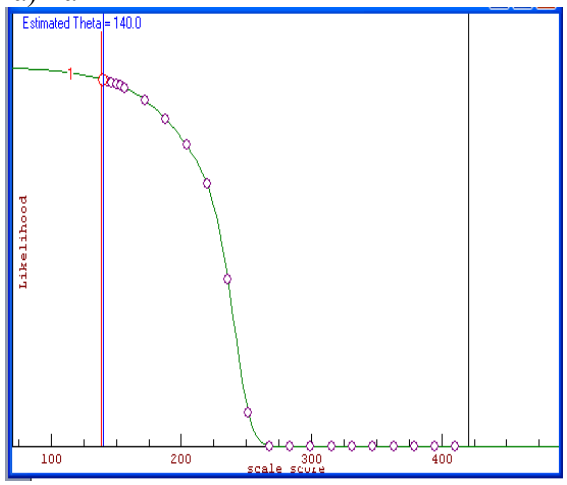
Figure 7-A below shows the probability of each ability estimate (or scale score) for the six examinees. The total scale score range for Language Arts is plotted on the horizontal axis. As indicated by the two vertical lines in the plot, the lower and upper limits of the scale score range are 140 and 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 which 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 which appear to be more likely than the reported score are outside of the scale score range of the test (below the LOSS).

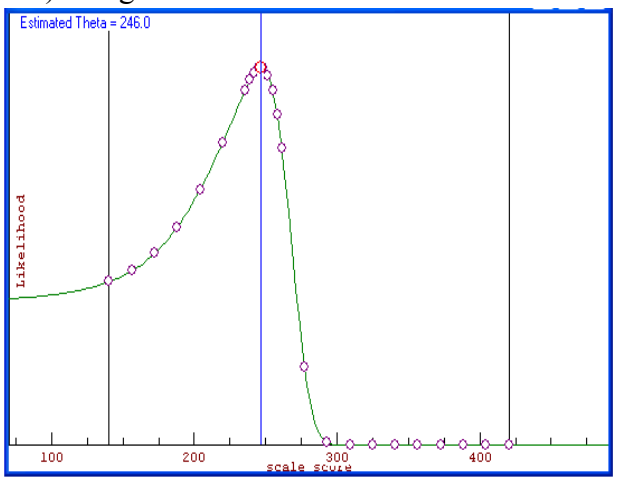
Copyright © 2011 by the Wisconsin Department of Public Instruction

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

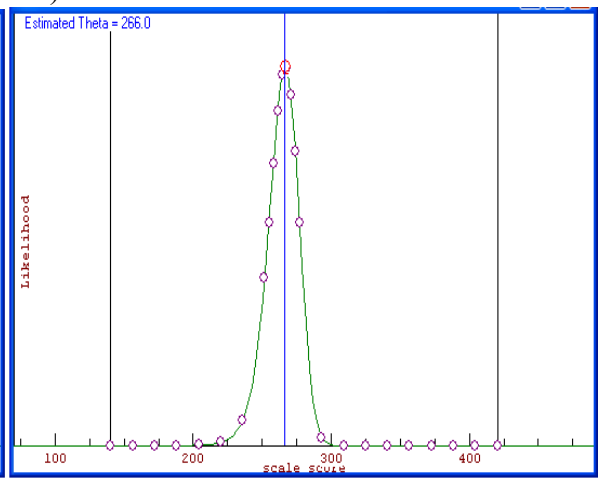
a) Pam



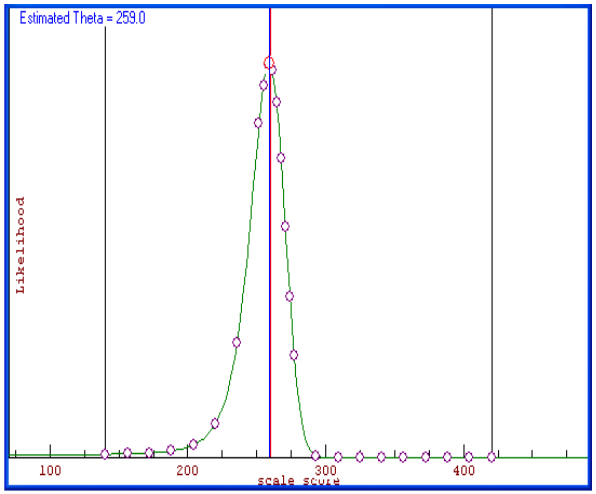
b) Craig



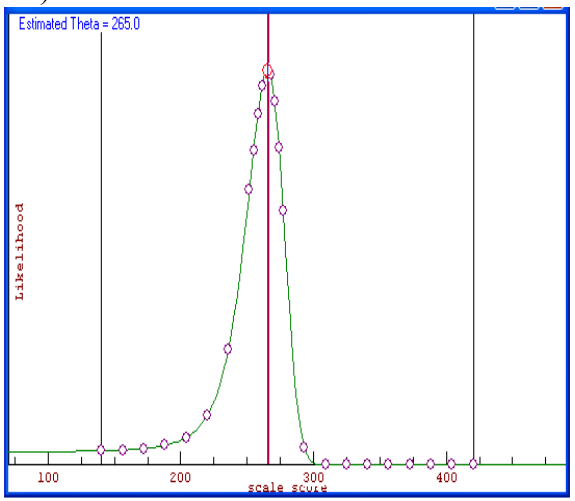
c) Vicki



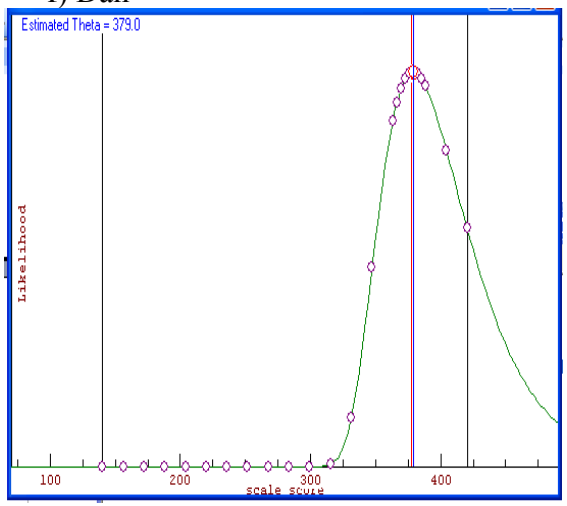
d) Tom



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 © 2011 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 standard error of measurement (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 2010 WKCE 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

Copyright © 2011 by the Wisconsin Department of Public Instruction

extreme scale scores have more measurement error than scores in the middle of the distribution. Scale scores in the high or low extremes of the student distribution are less precise than those in the middle of the distribution because there 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 2010 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 grade 5, 1.3% of students were 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, the percentage at the HOSS in Science grade 8, Social Studies grade 4, and Language Arts grade 8 ranged from 1.44 to 3.36. These tests are shorter than the Mathematics and Reading tests, so it is not unexpected that the percentage of students at the HOSS may be higher on these content areas because a smaller number of items on a test provide less psychometric information about the students. 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 © 2011 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 test characteristic curves 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 © 2011 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 2010 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.[5] The item-total correlation coefficients can range from -1.0 to +1.0. A large positive value (such as 0.40) indicates a strong relationship between a score on an individual item and the total score, with students who earn high scores on the test tending to score higher on the item than students with low scores on the total test. A low

[5] 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 © 2011 by the Wisconsin Department of Public Instruction

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 that 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 of the items.

For the Fall 2010 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 constructed-response items than for multiple-choice items because students who are fairly certain they do not know the answer may be inclined to simply skip the item altogether rather than taking the time to form a response. Items with high omit rates are referred to content specialists for further review in order to ensure there is no unintended

Copyright © 2011 by the Wisconsin Department of Public Instruction

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, 81 items were flagged on the 23 WKCE 2010 operational tests as meeting the investigational criteria bulleted on the previous page. After a review by CTB content experts and DPI, three operational flagged items were suppressed, and therefore, these three items did not contribute to student scores. Of the remaining 78 flagged items that were scored, the number flagged for each of the four criteria is consistent with previous administrations.

Following is a discussion of the operational items that were removed prior to scoring because their item statistics did not meet the established operational item thresholds discussed previously. The suppressed operational items were as follows: Reading grade 3 item 19, Language Arts grade 8 item 28, and Social Studies grade 10 item 41.

Table 8-A shows the number of scored items in the Fall 2010 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 © 2011 by the Wisconsin Department of Public Instruction

Table 8-A. Summary of Flagged Operational Items on the Fall 2010 WKCE

Content	Grade	# of Items Flagged	Number of Flags*			
			Corr <0.15	Distractor Corr >0	Omit >5%	*p*-value <0.30
RD	3	2		2		
	4	2		2		
	5	5	1	4		
	6	3	1	2		
	7	3	1	3		
	8	6	2	5		
	10	4		3	1	
MA	3	4			2	3
	4	3	2			1
	5	2		2		
	6	4	1	1	1	1
	7	4	1	2	1	
	8	9	1	3	2	4
	10	6		2	4	
LA	4	3	1	3		
	8	2	1	1		
	10	4		4		1
SS	4	1		1		
	8	1		1		
	10	3	1	3		1
SC	4	3	2	3		
	8	1			1	
	10	3	1	3		
Total		**78**	**16**	**50**	**12**	**11**

*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 © 2011 by the Wisconsin Department of Public Instruction

Table 8-B. Fall 2010 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
3	RD	27	MC	0.43	0.26	0.45%		+	0.00		
	RD	53	MC	0.50	0.23	1.79%		+	0.00		
4	RD	12	MC	0.57	0.34	0.97%		+	0.01		
	RD	28	MC	0.42	0.23	0.44%		+	0.02		
5	RD	9	MC	0.33	0.19	0.13%		+	0.19		
	RD	19	MC	0.50	0.17	1.35%		+	0.06		
	RD	24	MC	0.43	0.22	0.78%		+	0.02		
	RD	37	MC	0.49	0.12	0.57%	+		.		
	RD	46	MC	0.50	0.18	2.46%		+	0.04		
6	RD	19	MC	0.39	0.16	0.15%		+	0.09		
	RD	20	MC	0.86	0.13	0.19%	+		.		
	RD	26	MC	0.42	0.17	0.61%		+	0.07		
7	RD	34	MC	0.42	0.22	0.61%		+	0.01		
	RD	45	MC	0.45	0.12	1.21%	+	+	0.03		
	RD	51	MC	0.34	0.27	0.51%		+	0.06		
8	RD	7	MC	0.54	0.15	1.13%	+		.		
	RD	33	MC	0.41	0.20	0.42%		+	0.01		
	RD	34	MC	0.61	0.22	0.65%		+	0.03		
	RD	44	MC	0.47	0.14	0.31%	+	+	0.06		
	RD	46	MC	0.68	0.19	0.24%		+	0.02		
	RD	55	MC	0.55	0.25	0.54%		+	0.02		
10	RD	7	CR	0.51	0.50	7.85%			.	+	
	RD	17	MC	0.35	0.32	0.10%		+	0.01		
	RD	18	MC	0.50	0.39	0.21%		+	0.02		
	RD	49	MC	0.51	0.25	0.69%		+	0.06		

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-C. Fall 2010 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	20A	CR	0.48	0.39	5.31%				+	
	MA	20B	CR	0.25	0.44	3.51%					+
	MA	27B	CR	0.28	0.45	5.87%				+	+
	MA	36	MC	0.30	0.36	2.71%					+
4	MA	1	MC	0.97	0.13	0.11%	+				
	MA	11A	CR	0.24	0.34	0.93%					+
	MA	36	MC	0.99	0.13	1.45%	+				
5	MA	38	MC	0.40	0.26	0.29%		+	0.00		
	MA	41	MC	0.39	0.17	0.23%		+	0.01		
6	MA	4	MC	0.47	0.26	0.34%		+	0.03		
	MA	7B	CR	0.45	0.44	5.69%				+	
	MA	22B	CR	0.25	0.45	1.56%					+
	MA	42	MC	0.98	0.14	0.19%	+				
7	MA	12	MC	0.62	0.21	0.80%		+	0.04		
	MA	38A	CR	0.64	0.37	7.43%				+	
	MA	54	MC	0.57	0.44	0.55%		+	0.02		
	MA	55	MC	0.46	0.14	0.70%	+				
8	MA	5	MC	0.47	0.34	0.31%		+	0.05		
	MA	10	CR	0.22	0.51	4.02%					+
	MA	33A	CR	0.31	0.45	7.08%				+	
	MA	33B	CR	0.38	0.62	7.26%				+	
	MA	40	MC	0.64	0.15	1.14%	+	+	0.05		
	MA	45	MC	0.66	0.51	0.18%		+	0.00		
	MA	49A	CR	0.25	0.42	2.51%					+
	MA	49B	CR	0.21	0.42	3.98%					+
	MA	51	MC	0.26	0.34	0.35%					+
10	MA	9	MC	0.45	0.36	0.17%		+	0.00		
	MA	23	CR	0.47	0.65	7.42%				+	
	MA	33	CR	0.34	0.58	15.16%				+	
	MA	36	CR	0.50	0.65	10.56%				+	
	MA	45	MC	0.54	0.43	0.75%		+	0.05		
	MA	46	CR	0.52	0.53	6.15%				+	

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-D. Fall 2010 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	0.41	0.16	0.64%		+	0.00		
	LA	25	MC	0.43	0.19	0.89%		+	0.11		
	LA	30	MC	0.44	0.13	2.44%	+	+	0.09		
8	LA	2	MC	0.90	0.14	0.13%	+				
	LA	26	MC	0.47	0.36	1.02%		+	0.01		
10	LA	13	MC	0.48	0.20	0.26%		+	0.04		
	LA	21	MC	0.41	0.20	0.39%		+	0.04		
	LA	23	MC	0.43	0.32	0.41%		+	0.03		
	LA	27	MC	0.27	0.21	0.62%		+	0.09		+
4	SC	14	MC	0.64	0.07	0.47%	+	+	0.03		
	SC	24	MC	0.46	0.16	0.43%		+	0.04		
	SC	32	MC	0.35	0.11	0.42%	+	+	0.18		
8	SC	36	MC	0.45	0.28	5.11%				+	
10	SC	3	MC	0.43	0.15	0.12%	+	+	0.06		
	SC	37	MC	0.38	0.33	0.81%		+	0.11		
	SC	45	MC	0.36	0.24	1.97%		+	0.04		
4	SS	6	MC	0.60	0.23	0.34%		+	0.05		
8	SS	4	MC	0.39	0.19	0.37%		+	0.03		
10	SS	23	MC	0.49	0.43	0.62%		+	0.00		
	SS	40	MC	0.18	0.12	0.74%	+	+	0.22		+
	SS	43	MC	0.43	0.47	1.27%		+	0.01		

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 0.23. 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.

Flagging for the Item-Total Correlation

Sixteen items were flagged for item-total correlations <0.15, and all of the flagged items were 0.12 or above except for two items (both items are in Science grade 4 with correlations

Copyright © 2011 by the Wisconsin Department of Public Instruction

of 0.07 and 0.11). Although these items, with correlation coefficients ranging from 0.07 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.

Flagging for *p*-Value

Eleven items were flagged for *p*-values <0.30, and all seven of these items had *p*-values between 0.18 and 0.30. While these statistics indicate items that were very difficult, the number of items flagged for difficulty was very small. Only 2 of the 23 test forms had more than one item flagged for difficulty: Three items were flagged for Mathematics grade 3 and four items were flagged for Mathematics grade 8.

Flagging for Omit Rate

Twelve items were flagged for omit rates greater than 5%. All of the items flagged for omit rates were highly discriminating items. With the exception of one item in Mathematics grade 3 that had a borderline *p*-value (0.28), all of the other items flagged for high omit rates had consistently good statistics. All were retained to meet the WKCE blueprints.

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 they were 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.

8.1.1 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

Copyright © 2011 by the Wisconsin Department of Public Instruction

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	99.34%	98.68%	98.74%	99.25%	99.55%	98.25%	96.05%
Mathematics	99.25%	99.37%	99.66%	99.55%	99.37%	99.54%	99.22%
Language Arts		97.45%				98.38%	98.83%
Social Studies		99.34%				99.24%	99.06%
Science		98.90%				99.52%	98.60%

8.2 Raw Score Results

Raw score results based on all students that took the Fall 2010 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 standard error of measurement 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 which 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 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 4, the test difficulty statistic (0.65) was obtained by taking the mean raw score of 38.81 and dividing it by 60.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Note that in Reading grade 3, Language Arts grade 8, and Social Studies grade 10 one item was removed from each test, so the maximum raw scores changed from 60 to 57 (the removed item was a CR item worth 3 points), 30 to 29, and 50 to 49, respectively. Table 8-25 reflects this change in the maximum possible score.

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 which is flatter than a perfectly normal curve, and a positive kurtosis statistic indicates a distribution which 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 Reading grade 3, there is at least one student who failed all items and at least one student who obtained a perfect raw score of 57.

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 8-27 through 8-35.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Although the raw score table (Table 8-25) is based on all students that took the Fall 2010 WKCE assessment, an equivalent table based on the sample of students used for calibration is also provided in Table 8-26. Readers will recall that the calibration sample was determined to be largely representative of the entire student population based on both the demographic composition of the sample and the performance profile. The similarity of the performance profiles of the sample data and the census data are discussed in more detail in the next section addressing scale scores.

Reading

- Test difficulty ranged from 0.63 to 0.69.
- Standard deviations ranged from 10.37 to 11.58 raw score points.
- Alpha was relatively high in every grade (0.90 to 0.93).
- SEM ranged from 2.95 to 3.23.

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 9.03 to 13.30 raw score points.
- Alpha was relatively high in every grade (0.90 to 0.94).
- SEM ranged from 2.89 to 3.37.

Language Arts

- Test difficulty ranged from 0.63 to 0.72.
- Standard deviations ranged from 5.54 to 6.75 raw score points.
- Alpha ranged from 0.84 to 0.86. 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 2.10 to 2.58.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Social Studies

- Test difficulty ranged from 0.66 to 0.77.
- Standard deviations ranged from 6.38 to 9.57 raw score points.
- Alpha ranged from 0.88 to 0.91. This is consistent with prior years and within the expected range for the length of the tests.
- SEM ranged from 2.24 to 2.84.

Science

- Test difficulty ranged from 0.61 to 0.74.
- Standard deviations ranged from 6.57 to 9.26 raw score points.
- Alpha ranged from 0.86 to 0.89. Alpha was lower in grades 4 and 8 and 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 2.43 to 3.01.

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 8-27 through 8-35.

- 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 1.14 points in grade 7 to 2.50 points in grade 8.
- In Mathematics, the raw score differences between genders were very small, ranging from 0.21 point in grade 7 to 0.99 point in grade 8. 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 1.19 points in grade 4 to 2.27 points in grade 10.
- In Social Studies, the raw score differences between genders were very small, ranging from 0.11 point in grade 10 to 0.42 point in grade 4. Small differences like these suggest that the two groups may be best understood as showing similar performance in each grade.

Copyright © 2011 by the Wisconsin Department of Public Instruction

- In Science, male students had a slightly higher mean raw score than female students in each grade level, with differences ranging from 0.31 point in grade 4 to 1.60 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. The only exceptions to this occurred in Mathematics grade 5 where Hispanic students had a slightly higher mean raw score (difference of 0.23) than American Indian students. In all other grades and content areas, 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 less than 1 point in Language Arts, 1.75 points in Social Studies, 2.1 points in Mathematics, and 2.3 points in Science and 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 3.50 points in Language Arts grade 4 to 10.15 points in Mathematics grade 10.

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 4.02 points in Science grade 4 to 14.78 points in Mathematics grade 10.

In every grade and content area, students who were fully English proficient consistently showed a markedly higher mean raw score than students who were limited English proficient. As might be expected, these differences were largest in Reading, where fully English proficient students scored 8.20 to 12.84 points higher (in grades 3 and 10, respectively) than students who were limited English proficient. Mean raw score differences ranged from 3.70 to 13.16 points in Mathematics, 3.00 to 7.00 points in Language Arts, 4.19 to 10.13 points in Social Studies, and 4.42 to 10.00 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 8-38 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

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 8-39 through 8-47.

Reading

- Mean scale score increased by grade level, ranging from 457.90 to 539.04.
- Standard deviations ranged from 40.83 to 60.29 scale score points.
- In each grade level, student scores spanned the full scale score range, from the LOSS to the HOSS.

Mathematics

- Mean scale score increased by grade level, ranging from 431.95 to 562.28.
- Standard deviations ranged from 43.49 to 51.09 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 295.23 to 450.15.
- Standard deviations ranged from 29.93 to 44.58 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 296.16 to 452.77.
- Standard deviations ranged from 27.52 to 44.24 scale score points.
- In each grade level, student scores spanned the full scale score range, from the LOSS to the HOSS.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Science

- Mean scale score increased by grade level, ranging from 298.37 to 454.45.
- Standard deviations ranged from 29.78 to 44.66 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).

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 5.39 to 11.58 scale score points.
- In Mathematics, the differences between genders were very small, from 0.71 scale score point to 3.50 scale score points, with male students scoring slightly higher than female students in all grades except grade 6.
- In Language Arts, female students scored from 6.93 to 14.89 scale score points higher than male students.
- There were only small differences between scale scores by gender in Social Studies, from 0.39 scale score point to 1.85 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 each grade level. The difference ranged from 0.86 to 7.26 scale score points.

Race/Ethnicity

- The scale score results showed some consistent performance differences by ethnicity.
- In almost every grade and content area, White students, as a group, had the highest mean scale score, followed by Asian students, American Indian students, Hispanic students, and African American students, in that order. The only exceptions to this occurred in Mathematics grade 5 where Hispanic students had a slightly higher mean scale score (difference of 0.87) than American Indian students. In all other grades and content areas, American Indian students had a slightly higher mean scale score than Hispanic students.

Copyright © 2011 by the Wisconsin Department of Public Instruction

- 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 about half of the grades and content areas, differences were less than seven scale score points.

Socioeconomic Status

- Economically disadvantaged students, as a group, consistently scored lower than students who were not economically disadvantaged across all grades and content areas. Differences ranged from 17.86 scale score points in Social Studies grade 4 to 43.15 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.

Disability Status

- Students with disabilities and students without disabilities showed consistent and large differences in mean scale score by group. Differences ranged from 17.23 scale score points in Science grade 4 to 72.32 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 16.08 scale score points in Language Arts grade 4 to 70.51 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.

Calibration Sample and Census Data Scale Scores

As indicated in Part 6, the WKCE program used sample data for calibration purposes. The calibration sample was determined to be sufficiently representative of Wisconsin students based on both demographic and performance characteristics. The demographic comparisons

Copyright © 2011 by the Wisconsin Department of Public Instruction

between the calibration sample and the census data were provided in Part 6. The comparison based on scale scores is provided here.

In order to demonstrate the degree of similarity between the calibration sample and the census data, Table 8-36 provides the means and standard deviations (SD) of scale scores based on the calibration sample and the census data in each grade and content area. As can be seen in the table, the scale score difference between the calibration sample and the census data was small across all grades and content areas. The SDs tend to be slightly higher for the calibration sample than for the census data. This reflects the fact that the calibration sample includes a higher percentage of students with extreme scores, which is desirable in order to provide accurate calibration results at the tails of the distribution.

Table 8-37 provides additional information about the calibration sample, including sample size, skewness, and kurtosis, which can also be compared to the summary scale score table based on census data found in Table 8-38. Skewness and kurtosis are generally similar across the two groups, indicating that the shapes of the score distributions for the calibration sample are similar to the census distributions. Overall, these results indicate that the calibration sample and census data had similar characteristics.

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.

CTB performed a special study in which the previous WKCE assessments (those which existed until the Fall 2005 administration) were linked to the current WKCE assessments (those

Copyright © 2011 by the Wisconsin Department of Public Instruction

that began with the Fall 2005 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. Interested readers can also refer to the 2005 Standard Setting Technical Manual, which can be located at http://dpi.wi.gov/oea/publications.html.

Table 8-48 shows the cut scores for each content and grade level. Tables 8-49 through 8-53 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 8-54 through 8-58 provide the scale score ranges that define performance levels together with the percentage of students in each performance level.

Reading

- In terms of the total student population, most students were either *Proficient* or *Advanced* in Reading. Across grade levels, at least 75% of students were either *Proficient* or *Advanced*.
- Approximately 41% or more of the total student population was classified as *Advanced* in Reading.
- Across all grade levels, less than 25% of students were below *Proficient*. The difference ranged from 13% below *Proficient* in grades 6, 7, and 8, to 24% below *Proficient* in grade 10.

Mathematics

- Across grade levels, over 70% of the student population was either *Proficient* or *Advanced* in Mathematics.
- The proportion of students that were *Advanced* climbed by grade, from 32% in grade 3 to 46% in grade 5, and then declined by grade from 44% in grade 6 to 23% in grade 10.
- In grades 3–8, less than 26% of the student population was classified as below *Proficient*, but in grade 10 approximately 28% of students were classified as below *Proficient*.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Language Arts

- At least 62% of the student population was either *Proficient* or *Advanced* in Language Arts.
- In grades 4 and 10, over 73% of students were either *Proficient* or *Advanced*, and in grade 8, 62% of students were *Proficient* or *Advanced*.
- In grades 4 and 10, 23% and 27% of students, respectively, were below *Proficient*, but in grade 8, 38% 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, 83% in grade 8, and 81% in grade 10.
- A large proportion of students were *Advanced*, especially in grade 4: 66% in grade 4, 45% in grade 8, and 51% in grade 10.
- The proportion of students classified as below *Proficient* was 8% in grade 4, 17% in grade 8, and 19% in grade 10.

Science

- Approximately 75% 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 42% in grade 10.
- Approximately 25% 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, in the lowest performance category, and in the highest performance category. In grades 7, 8, and 10, however, there was a slightly higher percentage of male students in the *Advanced* category. In Language Arts, there was a markedly higher percentage of female students than male students that 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

Copyright © 2011 by the Wisconsin Department of Public Instruction

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, though in grade 10, there was a higher percentage of male students who were classified as *Proficient* or *Advanced*. There was also a slightly higher percentage of male students who were classified as *Advanced* in Science grades 4 and 8.

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 a higher percentage 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. The performance differences between American Indian students and Hispanic students were often small, and Hispanic students, as a group, occasionally performed better than American Indian students. The inverse sequence is 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 was a much higher percentage of students who were not economically disadvantaged classified as *Proficient* or above. There was a much higher percentage of students who were economically disadvantaged who were classified in the lowest performance category.

Performance level results showed a similar pattern in comparisons of students who were fully English proficient and limited English proficient. In every grade and content area, there was a higher percentage of students who were fully English proficient classified as *Proficient*, and a much higher percentage of students who were fully English proficient classified as *Advanced.* There was a much lower percentage 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 was a higher percentage of students without disabilities who were classified as *Proficient* or above, and there was a much higher percentage of students without disabilities in the reporting category *Advanced.* There was also a much lower percentage 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

Copyright © 2011 by the Wisconsin Department of Public Instruction

(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 they 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 were 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.*

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 8-59 through 8-63 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 8-64 identifies the SPI cut scores for each content area reporting category and grade level. The results from Tables 8-59 through 8-63 are summarized below.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Reading

Table 8-59 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 66.71%, 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 53.70% to 78.19%. In general, the difference between the lowest and highest mean SPI scores was greatest in grade 5 (20.96%). The difference was smallest in grade 6 (5.39%), and content standard 4 (Evaluates/Extends Text) and content standard 3 (Analyzes Text) were the most difficult standard at all grades.

Mathematics

Table 8-60 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 65.60%, 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 38.42% to 82.69%, with the largest differences observed in grade 3 (where SPI scores ranged from 38.42% to 80.79%). Differences between the highest and lowest mean SPI scores ranged from 13.15% (grade 10) to 42.37% (grade 3). Content standard A (Mathematical Processes) was the most difficult standard in grades 3 through 8. In grade 10, standards D and F (Measurement and Algebraic Relationships) showed lower mean SPI scores, indicating relatively difficult items in the standards.

Language Arts

Table 8-61 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 65.10%, indicating a moderate degree of difficulty. Mean SPI scores ranged from 53.32% to 77.30%, with differences between the highest and lowest mean SPI scores of 18.92% in grade 4, 16.71% in grade 8, and 10.43% in grade 10. 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 in grade 8, and content standard B (Writing) was the most difficult in grade 10.

Social Studies

Table 8-62 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 70.54%. 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 60.95% to 83.23%, with differences between the highest and lowest mean SPI scores of 16.49% in grade 4,

Copyright © 2011 by the Wisconsin Department of Public Instruction

20.87% in grade 8, and 15.72% 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 8-63 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 69.47%. 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 57.25% to 82.13%, with differences between the highest and lowest mean SPI scores of 8.46% in grade 4, 16.32% in grade 8, and 7.66% in grade 10. The mean *p*-values and SPI scores indicated that content standard E (Earth and Space) was the most difficult standard in grades 4 and 8, and content standard F (Life and Environment) was the most difficult in grade 10.

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)

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

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 © 2011 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 2010 WKCE was estimated in four ways:

1. Internal consistency was assessed for all multiple-choice and constructed-response 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 constructed-response 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 and 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 2010 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.

Combined, Cronbach's alpha, SEM, classification consistency, classification accuracy, and inter-rater reliability provide several forms of evidence bearing on the reliability of the

Copyright © 2011 by the Wisconsin Department of Public Instruction

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). Standard error of measurement (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 0.90 to 0.93 across grades for Reading, from 0.89 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 0.85 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 WKCE grade 4 Mathematics test, the grades 4 and 8 Social Studies tests, and all of the tests in Language Arts and Science 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 57 item test in Mathematics grade 4 would need to be increased to 58 items, the current 30-item tests in Language Arts in grades 4 and 8 and the 32-item test in grade 10 would need to be increased in length to 60, 45, and 54 items, respectively; the current 40-item tests in Science grades 4 and 8 and the 50-item test in grade 10 would need to be increased to 64 , 58, and 59 items,

Copyright © 2011 by the Wisconsin Department of Public Instruction

respectively; and the current 36- and 40-item tests in Social Studies in grades 4 and 8, respectively, would need to be increased to 51 and 52 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 and between economically disadvantaged and not disadvantaged students were within 0.03 of one another for all grades and content areas. The difference between disabled and not disabled was within 0.05 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 Language Arts grade 4 (where the reliability for Asian students was 0.04 higher than the reliability for American Indian students), Mathematics grade 10 (where the reliability for Asian students was 0.05 higher than the reliability for African American students, and the reliability for White students was 0.04 higher than the reliability for African American Students), and Science grade 10 (where the reliability for Asian students was 0.04 higher than the reliability for African American students). The greatest differences were between fully English proficient and limited English proficient students, with consistently lower reliability among limited English proficient students.

Table 9-2 presents the raw score SEM for the total population and for the subgroups described above. These values provide important information for raw score interpretation since we can expect that an individual's obtained score will fall within two standard errors of his or her true score approximately 95% of the time. Although there were some observable differences in SEM for the different subgroups, all differences were within one-half of a score point. The SEMs for 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 © 2011 by the Wisconsin Department of Public Instruction

Reading,[6] 6 to 15 items for Mathematics, 5 to 20 items for Language Arts, 5 to 13 items for Social Studies, and 4 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 0.43 (for standard 4 in grade 4, with 7 items) to 0.84 (for standard 3 in grade 3, with 19 items).

- For Mathematics, reliability indices by content standard ranged from 0.49 (for standard C in grade 5, with 10 items) to 0.76 (for standard D in grade 10, with 10 items, and standard F in grade 5, with 11 items).

- For Language Arts, reliability indices by content standard ranged from 0.36 (for standard F in grade 8, with 7 items) to 0.80 (for standard B in grade 8, with 18 items).

- For Social Studies, reliability indices by content standard ranged from 0.46 (for standard C in grade 4, with 7 items) to 0.71 (for standard B in grade 10, with 12 items, and for standard C in grade 10, with 12 items).

- For Science, reliability indices by content standard ranged from 0.23 (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 cornerstones of the NCLB Act (2002) is the measurement of Adequate Yearly Progress (AYP) for states with respect to the percentage of students at or above the academic performance standards established by the states. Because of a heavy emphasis on moving all students to levels of academic achievement at or above each state's self-defined *Proficient* category by year 2014, the consistency and accuracy of the classification of students into these

[6] Note that content standard D 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 © 2011 by the Wisconsin Department of Public Instruction

performance categories is of particular interest. The following section describes how the consistency and accuracy of these classifications were evaluated and it 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 © 2011 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 © 2011 by the Wisconsin Department of Public Instruction

Table 9-B
Classification Table for One Cut Point (C_1)[7]

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

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

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

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

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, 0.04 represents the probability of belonging to *Minimal Performance* in both the first and second administrations. The 0.05 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 0.81, chance probability is 0.35, kappa (*k*) is 0.70, and classification accuracy is 0.87. The values for "Cut 1"

[7] 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 © 2011 by the Wisconsin Department of Public Instruction

were obtained by applying only the first cut score. There are two levels whenever only one cut is applied (i.e., performance levels above and below the cut). It is clear that the values for P, 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 0.94, kappa was 0.81, and classification accuracy was 0.96. 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 0.91, and kappa was greater than or equal to 0.75 for all Reading and Mathematics tests. For Language Arts, the lowest P associated with the *Proficient* cut was 0.87, and the lowest kappa was 0.65. In Social Studies, the lowest P associated with the *Proficient* cut was 0.91, and the lowest kappa was 0.70. For Science, the lowest P was 0.90 and the lowest kappa was 0.70. 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 0.70 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 0.60 for all grades in Reading and Mathematics. The kappa for Language Arts, Social Studies, and Science were all greater than 0.50. 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 constructed-response 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.

Copyright © 2011 by the Wisconsin Department of Public Instruction

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

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

where:

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

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

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

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

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

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

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

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

Copyright © 2011 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 © 2011 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 47, the perfect agreement, adjacent agreement, discrepant agreement, and codes are 63%, 31%, 4%, 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 0.71.

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—6 items (46%)
- 2% discrepant ratings—3 items (23%)
- 3% discrepant ratings—3 items (23%)
- 4% discrepant ratings—1 item (8%)

The percentage of responses with condition codes ranged from 1% to 7% across all items; the percentage exceeded 4% for only two items (15% of the 13 items). The mean intraclass correlation, averaged across all items, was 0.86. Intraclass correlations ranged from 0.77 to 0.91, and weighted kappa ranged from 0.53 to 0.82.

Copyright © 2011 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 each of the 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—17 items (37%)
- 1% discrepant ratings—21 items (46%)
- 2% discrepant ratings—3 items (7%)
- 3% discrepant ratings—3 items (7%)
- 4% discrepant ratings—2 items (4%)

The percentage of responses with condition codes ranged from 1% to 13% across all items; the percentage exceeded 4% for only 10 items (22% of the 46 items). The mean intraclass correlation, averaged across all items, was 0.96. Intraclass correlations ranged from 0.89 to 1.0, and weighted kappa ranged from 0.78 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 58% to 63% on the Composing Rubric and from 83% to 95% on the Conventions Rubric. Adjacent agreement ranged from 33% to 35% on the Composing Rubric and from 4% to 13% on the Conventions Rubric. The percentage of discrepant (i.e., nonadjacent) ratings for the Writing prompt in grades 4, 8, and 10 ranged from 1% 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 0.86 to 0.93, and weighted kappa ranged from 0.73 to 0.85.

Copyright © 2011 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 © 2011 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 © 2011 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 (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 the 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 and compared the demographic composition of the sample data to the Wisconsin student population. It showed that the calibration sample was sufficiently representative of the Wisconsin student population, providing a foundation for the subsequent analyses. Part 6 thereby demonstrated adherence to AERA/APA/NCME (1999) standards 1.5, 1.13, 2.4, 4.7, and 6.1.

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 WKCE 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 2010 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 2010 WKCE

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 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

Copyright © 2011 by the Wisconsin Department of Public Instruction

English proficient. Students identified as fully English proficient are the reference group, and those identified as limited English proficient are the focal group.

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:

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

Copyright © 2011 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 © 2011 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 © 2011 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 constructed-response 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 © 2011 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

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

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 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, 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 standardized mean difference produced a flag.

In Table 10-1, looking at all items and all grades and content areas, 23 items were flagged for gender DIF. More items were flagged in the Reading and Mathematics tests than in the other content areas. This is expected given that more grades are tested in Reading and Mathematics. The number of flagged items in Reading and Mathematics relative to the other content areas should be understood in this context. Note that of the eight items flagged by Linn-Harnisch, all eight of these items indicate that the DIF favors (rather than disadvantages) female students. Note also that all of the items flagged by the LH method are CR items, while the MH method flagged both MC and CR items.

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 economically 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 is 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.

Copyright © 2011 by the Wisconsin Department of Public Instruction

The Fall 2010 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 2010 WKCE flagged for DIF were notated as such in the classical item analyses and in the item pool so that content experts will be able to reevaluate these items in future item selection activities. Items with DIF (particularly items flagged for strong DIF) are avoided in future selections.

10.2 Construct Validity

Construct validity can be defined as the extent to which tests measure the skills or constructs they intend to measure, and it is the central concept underlying the Fall 2010 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 0.56 to 0.84 in Reading, from 0.45 to 0.74 in Mathematics, from 0.38 to 0.70 in Language Arts, from 0.47 to 0.69 in Social Studies, and from 0.32 to 0.68 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

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 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 3.97 to 7.55. 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 5.24 to 7.55 in Reading, from 4.76 to 7.14 in Mathematics, from 3.97 to 5.04 in Language Arts, from 4.39 to 5.19 in Social Studies, and from 4.14 to 5.82 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.

Copyright © 2011 by the Wisconsin Department of Public Instruction

10.3 Test Integrity: Erasure Analysis

The Fall 2010 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. A high rate of erasures can identify situations in which test integrity needs 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 © 2011 by the Wisconsin Department of Public Instruction

Part 11: Summary Recommendations

Results and key findings of the Fall 2010 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 © 2011 by the Wisconsin Department of Public Instruction

Fall 2010 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 © 2011 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 © 2011 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 © 2011 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 © 2011 by the Wisconsin Department of Public Instruction

Copyright © 2011 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 © 2011 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 © 2011 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 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.

Table 2-5
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

Copyright © 2011 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		6	
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		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	6		7		6		6		5		3		2	
2.2	Demonstrates understanding of literal meaning by identifying stated information in informational text	10		6		7		5		5		8		5	
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		11	
3.2	Analyzes informational text.	8		7		11		6		8		6		3	
3.3	Analyzes author's use of language in literary and informational text.	5		5		4		4		4		3		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	1		1		2		5		2		4		5	
4.2	Evaluates and extends informational text	1		2		4		1		5		4		6	
4.3	Evaluates and extends author's use of language in literary and informational text	2		2		2		5		4		3		3	

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 2-6 Cont'd
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
	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 © 2011 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 © 2011 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 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.

Table 2-10
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

Copyright © 2011 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 © 2011 by the Wisconsin Department of Public Instruction

Table 3-1
Fall 2010 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	40						40	40
	8	40						40	40
	10	50						50	50

* An item was dropped in Reading grade 3, Language Arts grade 8, and Social Studies grade 10. 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 © 2011 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														
Reading	242	12	24	2	27	2	40	4	40	4	40	4	413	28
Math	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														
Reading	294	12	24	2	32	3	40	4	40	4	40	4	470	29
Math	231	29	15	2	32	4	34	4	28	8	40	4	380	51
Language Arts	0	0	0	6	0	0	0	0	0	0	0	0	0	6
Science	0	0	0	0	40	0	0	0	0	0	0	0	40	0
Social Studies	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	525	41	39	10	104	7	74	8	68	12	80	8	890	86
Grade 5														
Reading	235	14	24	2	28	2	29	6	40	4	40	4	396	32
Math	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														
Reading	259	14	24	1	33	3	35	5	40	4	40	4	431	31
Math	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														
Reading	259	14	24	1	17	2	35	4	40	4	40	4	415	29
Math	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														
Reading	274	14	24	1	33	4	32	5	40	4	40	4	443*	32
Math	234	33	15	2	40	4	32	4	32	5	32	4	385	52
Language Arts	0	0	0	6	0	0	0	0	0	0	0	0	0	6
Science	0	0	0	0	40	0	0	0	0	0	0	0	40	0
Social Studies	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	508	47	39	9	113	8	64	9	72	9	72	8	868	90
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	10	0	0	0	0	0	0	0	10	0
Social Studies	0	0	0	0	0	0	0	0	0	0	0	0	0	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	5025	499

Copyright © 2011 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	58856	Reading	830	1.41
		59168	Math	1087	1.84
	4	59960	Reading	926	1.54
		60136	Math	1047	1.74
		60056	Science	422	0.70
		59912	Language Arts	419	0.70
		60184	Writing	566	0.94
		59984	Social Studies	407	0.68
	5	60424	Reading	717	1.19
		60560	Math	771	1.27
	6	59976	Reading	471	0.79
		60128	Math	522	0.87
	7	60288	Reading	317	0.53
		60392	Math	337	0.56
	8	60280	Reading	254	0.42
		60400	Math	254	0.42
		60280	Science	148	0.25
		60184	Language Arts	145	0.24
		60536	Writing	298	0.49
		60168	Social Studies	141	0.23
	10	64672	Reading	106	0.16
		64696	Math	78	0.12
		64432	Science	46	0.07
		64192	Language Arts	55	0.09
		65160	Writing	120	0.18
		64296	Social Studies	49	0.08

Copyright © 2011 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	58856	Reading	6454	10.97
		59168	Math	6765	11.43
	4	59960	Reading	7357	12.27
		60136	Math	7549	12.55
		60056	Science	7271	12.11
		59912	Language Arts	7180	11.98
		60184	Writing	7075	11.76
		59984	Social Studies	7201	12.00
	5	60424	Reading	7419	12.28
		60560	Math	7601	12.55
	6	59976	Reading	6656	11.10
		60128	Math	6850	11.39
	7	60288	Reading	6237	10.35
		60392	Math	6316	10.46
	8	60280	Reading	5885	9.76
		60400	Math	6017	9.96
		60280	Science	5676	9.42
		60184	Language Arts	5812	9.66
		60536	Writing	5748	9.49
		60168	Social Studies	5660	9.41
	10	64672	Reading	4311	6.67
		64696	Math	4434	6.85
		64432	Science	4158	6.45
		64192	Language Arts	4185	6.52
		65160	Writing	4193	6.43
		64296	Social Studies	4143	6.44

Copyright © 2011 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	58856	Reading	0	0.00
		59168	Math	6002	10.14
	4	59960	Reading	0	0.00
		60136	Math	6688	11.12
		60056	Science	6535	10.88
		59912	Language Arts	6005	10.02
		60184	Writing	6131	10.19
		59984	Social Studies	6453	10.76
	5	60424	Reading	0	0.00
		60560	Math	6253	10.33
	6	59976	Reading	0	0.00
		60128	Math	5273	8.77
	7	60288	Reading	0	0.00
		60392	Math	4900	8.11
	8	60280	Reading	0	0.00
		60400	Math	4460	7.38
		60280	Science	4459	7.40
		60184	Language Arts	4220	7.01
		60536	Writing	4358	7.20
		60168	Social Studies	4396	7.31
	10	64672	Reading	0	0.00
		64696	Math	2698	4.17
		64432	Science	2748	4.27
		64192	Language Arts	2670	4.16
		65160	Writing	2869	4.40
		64296	Social Studies	2720	4.23

Copyright © 2011 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	58856	Reading	0	0.00
		59168	Math	913	1.54
	4	59960	Reading	0	0.00
		60136	Math	873	1.45
		60056	Science	817	1.36
		59912	Language Arts	0	0.00
		60184	Writing	575	0.96
		59984	Social Studies	761	1.27
	5	60424	Reading	0	0.00
		60560	Math	640	1.06
	6	59976	Reading	0	0.00
		60128	Math	383	0.64
	7	60288	Reading	0	0.00
		60392	Math	232	0.38
	8	60280	Reading	0	0.00
		60400	Math	247	0.41
		60280	Science	216	0.36
		60184	Language Arts	0	0.00
		60536	Writing	182	0.30
		60168	Social Studies	197	0.33
	10	64672	Reading	0	0.00
		64696	Math	159	0.25
		64432	Science	134	0.21
		64192	Language Arts	0	0.00
		65160	Writing	129	0.20
		64296	Social Studies	129	0.20

Copyright © 2011 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	58856	Reading	0	0.00
		59168	Math	0	0.00
	4	59960	Reading	0	0.00
		60136	Math	182	0.30
		60056	Science	175	0.29
		59912	Language Arts	0	0.00
		60184	Writing	144	0.24
		59984	Social Studies	164	0.27
	5	60424	Reading	0	0.00
		60560	Math	0	0.00
	6	59976	Reading	0	0.00
		60128	Math	0	0.00
	7	60288	Reading	0	0.00
		60392	Math	0	0.00
	8	60280	Reading	0	0.00
		60400	Math	75	0.12
		60280	Science	67	0.11
		60184	Language Arts	0	0.00
		60536	Writing	45	0.07
		60168	Social Studies	52	0.09
	10	64672	Reading	0	0.00
		64696	Math	28	0.04
		64432	Science	31	0.05
		64192	Language Arts	0	0.00
		65160	Writing	27	0.04
		64296	Social Studies	27	0.04

Copyright © 2011 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	58856	Reading	0	0.00
		59168	Math	0	0.00
	4	59960	Reading	0	0.00
		60136	Math	356	0.59
		60056	Science	344	0.57
		59912	Language Arts	0	0.00
		60184	Writing	0	0.00
		59984	Social Studies	310	0.52
	5	60424	Reading	0	0.00
		60560	Math	0	0.00
	6	59976	Reading	0	0.00
		60128	Math	0	0.00
	7	60288	Reading	0	0.00
		60392	Math	0	0.00
	8	60280	Reading	0	0.00
		60400	Math	147	0.24
		60280	Science	128	0.21
		60184	Language Arts	0	0.00
		60536	Writing	0	0.00
		60168	Social Studies	113	0.19
	10	64672	Reading	0	0.00
		64696	Math	69	0.11
		64432	Science	56	0.09
		64192	Language Arts	0	0.00
		65160	Writing	0	0.00
		64296	Social Studies	49	0.08

Copyright © 2011 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	58856	Reading	0	0.00
		59168	Math	13	0.02
	4	59960	Reading	0	0.00
		60136	Math	4	0.01
		60056	Science	5	0.01
		59912	Language Arts	4	0.01
		60184	Writing	2	0.00
		59984	Social Studies	1	0.00
	5	60424	Reading	0	0.00
		60560	Math	6	0.01
	6	59976	Reading	0	0.00
		60128	Math	3	0.01
	7	60288	Reading	0	0.00
		60392	Math	4	0.01
	8	60280	Reading	0	0.00
		60400	Math	2	0.00
		60280	Science	2	0.00
		60184	Language Arts	1	0.00
		60536	Writing	7	0.01
		60168	Social Studies	1	0.00
	10	64672	Reading	0	0.00
		64696	Math	19	0.03
		64432	Science	23	0.04
		64192	Language Arts	21	0.03
		65160	Writing	1	0.00
		64296	Social Studies	22	0.03

Copyright © 2011 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	58856	Reading	0	0.00
		59168	Math	30	0.05
	4	59960	Reading	0	0.00
		60136	Math	20	0.03
		60056	Science	20	0.03
		59912	Language Arts	19	0.03
		60184	Writing	19	0.03
		59984	Social Studies	22	0.04
	5	60424	Reading	0	0.00
		60560	Math	39	0.06
	6	59976	Reading	0	0.00
		60128	Math	17	0.03
	7	60288	Reading	0	0.00
		60392	Math	15	0.02
	8	60280	Reading	0	0.00
		60400	Math	12	0.02
		60280	Science	11	0.02
		60184	Language Arts	11	0.02
		60536	Writing	12	0.02
		60168	Social Studies	11	0.02
	10	64672	Reading	0	0.00
		64696	Math	18	0.03
		64432	Science	16	0.02
		64192	Language Arts	14	0.02
		65160	Writing	15	0.02
		64296	Social Studies	15	0.02

Copyright © 2011 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	58856	Reading	1587	2.70
		59168	Math	1813	3.06
	4	59960	Reading	1907	3.18
		60136	Math	2020	3.36
		60056	Science	1828	3.04
		59912	Language Arts	1799	3.00
		60184	Writing	1803	3.00
		59984	Social Studies	1811	3.02
	5	60424	Reading	1936	3.20
		60560	Math	2045	3.38
	6	59976	Reading	1608	2.68
		60128	Math	1684	2.80
	7	60288	Reading	1422	2.36
		60392	Math	1449	2.40
	8	60280	Reading	1436	2.38
		60400	Math	1463	2.42
		60280	Science	1395	2.31
		60184	Language Arts	1357	2.25
		60536	Writing	1381	2.28
		60168	Social Studies	1375	2.29
	10	64672	Reading	619	0.96
		64696	Math	630	0.97
		64432	Science	587	0.91
		64192	Language Arts	573	0.89
		65160	Writing	628	0.96
		64296	Social Studies	582	0.91

Copyright © 2011 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	58856	Reading	0	0.00
		59168	Math	0	0.00
	4	59960	Reading	17	0.03
		60136	Math	17	0.03
		60056	Science	17	0.03
		59912	Language Arts	17	0.03
		60184	Writing	17	0.03
		59984	Social Studies	17	0.03
	5	60424	Reading	15	0.02
		60560	Math	15	0.02
	6	59976	Reading	2	0.00
		60128	Math	2	0.00
	7	60288	Reading	0	0.00
		60392	Math	0	0.00
	8	60280	Reading	14	0.02
		60400	Math	14	0.02
		60280	Science	14	0.02
		60184	Language Arts	14	0.02
		60536	Writing	14	0.02
		60168	Social Studies	14	0.02
	10	64672	Reading	0	0.00
		64696	Math	0	0.00
		64432	Science	0	0.00
		64192	Language Arts	0	0.00
		65160	Writing	0	0.00
		64296	Social Studies	0	0.00

Copyright © 2011 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	58856	Reading	0	0.00
		59168	Math	0	0.00
	4	59960	Reading	0	0.00
		60136	Math	0	0.00
		60056	Science	0	0.00
		59912	Language Arts	0	0.00
		60184	Writing	0	0.00
		59984	Social Studies	0	0.00
	5	60424	Reading	0	0.00
		60560	Math	0	0.00
	6	59976	Reading	0	0.00
		60128	Math	0	0.00
	7	60288	Reading	0	0.00
		60392	Math	0	0.00
	8	60280	Reading	0	0.00
		60400	Math	0	0.00
		60280	Science	0	0.00
		60184	Language Arts	0	0.00
		60536	Writing	0	0.00
		60168	Social Studies	0	0.00
	10	64672	Reading	0	0.00
		64696	Math	0	0.00
		64432	Science	0	0.00
		64192	Language Arts	0	0.00
		65160	Writing	0	0.00
		64296	Social Studies	0	0.00

Copyright © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 ***									
	31	58803	5721	24108	27054	411	1393	29	37	50
4	13	59914	4366	21762	28998	4267	496	1	16	8
	47	59914	11369	23578	20518	3504	903	3	10	29
5	13	60412	4015	19332	29166	7157	701	3	3	35
	47	60412	8040	20793	28417	2420	709	3	5	25
6	18	59939	4919	31324	21381	1752	529	15	9	10
	24	59939	13048	22775	21519	2056	512	11	14	4
7	31	60236	2844	21663	31084	3831	744	0	1	69
	50	60236	18701	19316	18102	3332	779	1	1	4
8	19	60245	3189	20724	32306	3452	554	1	6	13
	56	60245	7764	10918	27962	12462	1053	1	6	79
10	7	64637	12064	19115	21344	8152	3912	3	11	36
	52	64637	3809	27980	23367	6885	2555	4	10	27

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

Copyright © 2011 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		59152	27562	7593	23563	421	0	12	1
	20	A	59152	28470	27074	0	3602	1	3	2
	20	B	59152	36947	11830	8712	1571	11	56	25
	27	A	59152	19672	37913	0	1524	12	12	19
	27	B	59152	35425	6734	14161	2741	22	53	16
	39	A	59152	34602	21820	0	2724	5	1	0
	39	B	59152	35937	10741	11211	1168	25	52	18
4	11	A	60111	46511	13167	0	408	14	4	7
	11	B	60111	28580	22548	7323	1550	30	31	49
	22		60111	3352	15200	41208	303	18	11	19
	25	A	60111	31512	28327	0	271	0	0	1
	25	B	60111	29095	20988	8669	1301	10	38	10
	39	A	60111	24914	34946	0	246	0	3	2
	39	B	60111	33627	7386	17998	1048	3	40	9
5	12		60525	17138	5101	37681	577	8	2	18
	20	A	60525	8521	51645	0	326	22	3	8
	20	B	60525	36114	12472	10998	877	35	21	8
	27	A	60525	23808	36209	0	498	9	0	1
	27	B	60525	13914	5926	39824	826	10	22	3
	46	A	60525	37465	22771	0	277	6	3	3
	46	B	60525	38893	2843	18279	487	9	11	3
6	7	A	60108	26730	31297	0	2080	0	1	0
	7	B	60108	32023	1663	23840	2558	2	16	6
	22	A	60108	5281	54575	0	245	2	3	2
	22	B	60108	30697	27802	950	616	9	26	8
	39	A	60108	31949	27620	0	537	0	1	1
	39	B	60108	25968	15633	17558	921	3	23	2
	51		60108	23710	22581	13435	381	0	1	0
7	7	A	60371	17774	42166	0	429	1	1	0
	7	B	60371	18106	10037	31355	862	3	6	2
	19	A	60371	27903	31799	0	646	2	18	3
	19	B	60371	29909	20736	8359	1339	3	20	5
	38	A	60371	22283	32886	0	5200	2	0	0
	38	B	60371	22110	30124	5856	2220	31	15	15
	52		60371	27936	20706	11173	552	0	0	4

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

Copyright © 2011 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		60355	41151	9802	7891	1501	8	1	1
	19	A	60355	16210	43519	0	626	0	0	0
	19	B	60355	16134	11893	30642	1645	6	17	18
	33	A	60355	40300	16458	0	3595	0	1	1
	33	B	60355	25097	23260	8831	3115	16	20	16
	49	A	60355	43564	15774	0	1004	5	3	5
	49	B	60355	44166	4446	10198	1516	5	13	11
10	23		64648	23390	22446	15408	3355	4	26	19
	33		64648	31428	15025	10589	7588	1	7	10
	36		64648	22811	14211	22261	5337	1	11	16
	46		64648	24845	14759	22023	2986	2	22	11

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

Copyright © 2011 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	3042	63	464	1179	1048	212	17	27	0	0	31
	Rater 2	3042	62	491	1221	1013	182	14	26	0	0	32
	Diff*	0	1	-27	-42	35	30	3	1	0	0	-1
8	Rater 1	3071	23	464	981	1131	363	34	30	0	0	44
	Rater 2	3071	23	444	1016	1127	356	34	30	0	0	40
	Diff*	0	0	20	-35	4	7	0	0	0	0	4
10	Rater 1	3266	113	498	876	1174	370	97	121	0	0	17
	Rater 2	3266	111	505	886	1166	359	109	116	0	0	14
	Diff*	0	2	-7	-10	8	11	-12	5	0	0	3

* 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 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	3042	2.08	15.26	38.76	34.46	6.96	0.56	0.88	0.00	0.00	1.02
	Rater 2	3042	2.04	16.14	40.14	33.30	5.98	0.46	0.86	0.00	0.00	1.06
8	Rater 1	3071	0.74	15.10	31.94	36.82	11.82	1.10	0.98	0.00	0.00	1.44
	Rater 2	3071	0.74	14.46	33.08	36.70	11.60	1.10	0.98	0.00	0.00	1.30
10	Rater 1	3266	3.46	15.24	26.82	35.94	11.32	2.96	3.70	0.00	0.00	0.52
	Rater 2	3266	3.40	15.46	27.12	35.70	11.00	3.34	3.56	0.00	0.00	0.42

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

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 5-13
Score Distribution 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	3042	58	2918	38	27	0	1
	Rater 2	3042	59	2923	33	26	0	1
	Diff*	0	-1	-5	5	1	0	0
8	Rater 1	3071	34	2971	35	30	0	1
	Rater 2	3071	46	2961	33	30	0	1
	Diff*	0	-12	10	2	0	0	0
10	Rater 1	3266	26	2743	376	121	0	0
	Rater 2	3266	30	2761	359	116	0	0
	Diff*	0	-4	-18	17	5	0	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	3042	1.90	95.92	1.24	0.88	0.00	0.04
	Rater 2	3042	1.94	96.08	1.08	0.86	0.00	0.04
8	Rater 1	3071	1.10	96.74	1.14	0.98	0.00	0.04
	Rater 2	3071	1.50	96.42	1.08	0.98	0.00	0.04
10	Rater 1	3266	0.80	83.98	11.52	3.70	0.00	0.00
	Rater 2	3266	0.92	84.54	11.00	3.56	0.00	0.00

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

Copyright © 2011 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	3042	28	0	65	46	450	1174	1042	206	21	10
	Rater 2	3042	27	1	60	53	479	1210	1013	169	20	10
	Diff*	0	1	-1	5	-7	-29	-36	29	37	1	0
8	Rater 1	3071	31	2	55	22	459	973	1126	364	16	23
	Rater 2	3071	31	1	46	36	439	1001	1121	351	32	13
	Diff*	0	0	1	9	-14	20	-28	5	13	-16	10
10	Rater 1	3266	121	1	38	94	494	874	1146	156	246	96
	Rater 2	3266	116	2	32	96	502	881	1147	149	240	101
	Diff*	0	5	-1	6	-2	-8	-7	-1	7	6	-5

* 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	3042	0.92	2.14	1.52	14.80	38.60	34.26	6.78	0.70	0.32	0.00
	Rater 2	3042	0.88	0.04	1.98	1.74	15.74	39.78	33.30	5.56	0.66	0.32
8	Rater 1	3071	1.00	0.06	1.80	0.72	14.94	31.68	36.66	11.86	0.52	0.74
	Rater 2	3071	1.00	0.04	1.50	1.18	14.30	32.60	36.50	11.42	1.04	0.42
10	Rater 1	3266	3.70	0.04	1.16	2.88	15.12	26.76	35.08	4.78	7.54	2.94
	Rater 2	3266	3.56	0.06	0.98	2.94	15.38	26.98	35.12	4.56	7.34	3.10

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 6-1
The Current 13 Calibration Districts

	District Name
1	BUTTERNUT
2	KENOSHA
3	LA CROSSE
4	MADISON
5	PLATTEVILLE
6	RICHLAND
7	SHEBOYGAN
8	SHOREWOOD
9	VERONA
10	WABENO
11	WATERTOWN
12	WAUSAU
13	WAUWATOSA

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 6-2
Total Number of Students in Census and Calibration Sample Data

Grade	Census	Calibration Sample Data
3	59269	4653
4	60227	5309
5	60639	4767
6	60229	5949
7	60539	5294
8	60602	5449
10	65377	5836

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 6-3
Number and Percent of Students in Census and Calibration Sample Data, by Gender*

Grade	Census		Calibration Sample Data	
	M	F	M	F
3	30246	29021	2383	2270
4	30878	29349	2726	2583
5	31040	29596	2444	2323
6	30882	29346	3068	2881
7	30904	29634	2736	2558
8	30970	29629	2788	2661
10	33365	32009	3022	2814

Grade	Census		Calibration Sample Data	
	M	F	M	F
3	51.03	48.97	51.21	48.79
4	51.27	48.73	51.35	48.65
5	51.19	48.81	51.27	48.73
6	51.28	48.72	51.57	48.43
7	51.05	48.95	51.68	48.32
8	51.11	48.89	51.17	48.83
10	51.04	48.96	51.78	48.22

* Note that students of unspecified gender or race/ethnicity are not counted and percentages may not total 100% due to rounding.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 6-4
Number and Percent of Students in Census and Calibration Sample Data, by Race/Ethnicity*

Grade	Census					Calibration Sample Data				
	W	B	H	A	AI	W	B	H	A	AI
3	43013	6531	6335	2391	994	2969	559	612	474	39
4	44085	6433	6232	2490	987	3377	634	730	526	42
5	44943	6599	5709	2424	961	3040	597	574	517	39
6	44852	6504	5579	2316	976	3723	844	844	486	52
7	45538	6485	5287	2219	1005	3234	802	727	474	57
8	45889	6442	5020	2292	955	3394	786	713	500	56
10	50513	6445	4855	2513	1045	3875	676	675	542	68

Grade	Census					Calibration Sample Data				
	W	B	H	A	AI	W	B	H	A	AI
3	72.58	11.02	10.69	4.03	1.68	63.81	12.01	13.15	10.19	0.84
4	73.20	10.68	10.35	4.13	1.64	63.61	11.94	13.75	9.91	0.79
5	74.12	10.88	9.42	4.00	1.58	63.77	12.52	12.04	10.85	0.82
6	74.47	10.80	9.26	3.85	1.62	62.58	14.19	14.19	8.17	0.87
7	75.23	10.71	8.73	3.67	1.66	61.09	15.15	13.73	8.95	1.08
8	75.73	10.63	8.28	3.78	1.58	62.29	14.42	13.08	9.18	1.03
10	77.27	9.86	7.43	3.84	1.60	66.40	11.58	11.57	9.29	1.17

* Note that students of unspecified gender or race/ethnicity are not counted and percentages may not total 100% due to rounding.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 6-5
Number and Percent of Students in Census and Calibration Sample Data, by Socioeconomic Status (SES)*

Grade	Census		Calibration Sample Data	
	Economically Disadvantaged	Not Economically Disadvantaged	Economically Disadvantaged	Not Economically Disadvantaged
3	26098	33171	2061	2592
4	26118	34109	2406	2903
5	25636	35003	2060	2707
6	24693	35536	2704	3245
7	24306	36233	2523	2771
8	23314	37288	2472	2977
10	22391	42986	2177	3659

Grade	Census		Calibration Sample Data	
	Economically Disadvantaged	Not Economically Disadvantaged	Economically Disadvantaged	Not Economically Disadvantaged
3	44.03	55.97	44.29	55.71
4	43.37	56.63	45.32	54.68
5	42.28	57.72	43.21	56.79
6	41.00	59.00	45.45	54.55
7	40.15	59.85	47.66	52.34
8	38.47	61.53	45.37	54.63
10	34.25	65.75	37.30	62.70

* Note that percentages may not total 100% due to rounding.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 6-6
Number and Percent of Students in Census and Calibration Sample Data, by Disability*

Grade	Census		Calibration Sample Data	
	Disabled	Not Disabled	Disabled	Not Disabled
3	7551	51718	517	4136
4	8141	52086	710	4599
5	8076	52563	627	4140
6	7889	52340	776	5173
7	7853	52686	753	4541
8	7640	52962	798	4651
10	7901	57476	765	5071

Grade	Census		Calibration Sample Data	
	Disabled	Not Disabled	Disabled	Not Disabled
3	12.74	87.26	11.11	88.89
4	13.52	86.48	13.37	86.63
5	13.32	86.68	13.15	86.85
6	13.10	86.90	13.04	86.96
7	12.97	87.03	14.22	85.78
8	12.61	87.39	14.64	85.36
10	12.09	87.91	13.11	86.89

* Note that percentages may not total 100% due to rounding.

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 6-7
Number and Percent of Students in Census and Calibration Sample Data, by ELP Status*

Grade	Census		Calibration Sample Data	
	Fully English Proficient	Limited English Proficient	Fully English Proficient	Limited English Proficient
3	54485	4784	4054	599
4	55604	4623	4534	775
5	56709	3930	4170	597
6	56815	3414	5232	717
7	57352	3187	4636	658
8	57519	3083	4828	621
10	62844	2533	5301	535

Grade	Census		Calibration Sample Data	
	Fully English Proficient	Limited English Proficient	Fully English Proficient	Limited English Proficient
3	91.93	8.07	87.13	12.87
4	92.32	7.68	85.40	14.60
5	93.52	6.48	87.48	12.52
6	94.33	5.67	87.95	12.05
7	94.74	5.26	87.57	12.43
8	94.91	5.09	88.60	11.40
10	96.13	3.87	90.83	9.17

* Note that percentages may not total 100% due to rounding.

Copyright © 2011 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
RD	10	20	MC	5745	55.72	15.32
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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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	69	29	498	14
1	350	69	30	502	14
2	350	69	**31**	**507**	**14**
3	350	69	32	511	14
4	350	69	33	516	14
5	350	69	34	521	14
6	350	69	35	525	14
7	350	69	36	530	14
8	350	69	37	535	14
9	350	69	38	540	14
10	350	69	39	544	14
11	350	69	40	549	14
12	377	46	41	554	14
13	395	35	**42**	**559**	**14**
14	408	29	43	565	14
15	418	25	44	570	14
16	427	22	45	576	15
17	435	20	46	582	15
18	442	18	47	588	15
19	448	17	48	595	16
20	454	17	49	603	17
21	**459**	**16**	50	612	19
22	465	15	51	622	21
23	470	15	52	635	24
24	475	15	53	651	28
25	479	15	54	674	36
26	484	14	55	715	55
27	489	14	56	820	141
28	493	14			

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

Copyright © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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 © 2011 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	663	1.13	640	19	0.03
	4	280	487	0.81	650	14	0.02
	5	290	779	1.29	690	4	0.01
	6	300	438	0.73	730	3	0.01
	7	310	330	0.55	780	1	0.00
	8	330	308	0.51	790	13	0.02
	10	350	638	0.99	820	77	0.12
MA	3	220	63	0.11	630	203	0.34
	4	240	59	0.10	650	116	0.19
	5	270	78	0.13	680	144	0.24
	6	310	91	0.15	700	19	0.03
	7	330	82	0.14	710	87	0.14
	8	350	444	0.74	730	61	0.10
	10	410	724	1.12	750	254	0.39
LA	4	140	228	0.38	420	218	0.36
	8	250	530	0.88	520	2021	3.36
	10	290	350	0.55	630	18	0.03
SS	4	170	153	0.26	400	1362	2.27
	8	230	318	0.53	530	561	0.93
	10	240	253	0.39	620	129	0.20
SC	4	170	96	0.16	440	282	0.47
	8	230	149	0.25	560	870	1.44
	10	240	307	0.48	610	45	0.07

Copyright © 2011 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	0.58	0.19	0.15%				
2	MC	0.67	0.41	0.19%				
3	MC	0.87	0.51	0.15%				
4	MC	0.66	0.52	0.41%				
5	MC	0.62	0.41	0.50%				
6	MC	0.53	0.40	0.56%				
7	MC	0.77	0.55	1.10%				
8	MC	0.86	0.54	0.39%				
9	MC	0.76	0.48	1.06%				
10	MC	0.81	0.39	0.76%				
11	MC	0.67	0.56	1.10%				
12	MC	0.77	0.50	0.99%				
13	MC	0.85	0.49	1.32%				
14	MC	0.69	0.49	1.69%				
15	MC	0.80	0.44	2.07%				
16	MC	0.74	0.45	3.63%				
17	MC	0.42	0.43	1.77%				
18	MC	0.66	0.46	2.03%				
19 **	CR							
20	MC	0.60	0.45	0.17%				
21	MC	0.85	0.43	0.24%				
22	MC	0.77	0.53	0.69%				
23	MC	0.56	0.35	0.43%				
24	MC	0.72	0.42	0.58%				
25	MC	0.64	0.50	1.34%				

* 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.
** Item dropped from scoring.

Copyright © 2011 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	0.53	0.26%				
27	MC	0.43	0.26	0.45%		+		
28	MC	0.80	0.51	0.86%				
29	MC	0.73	0.56	0.45%				
30	MC	0.68	0.37	0.78%				
31	CR	0.47	0.44	2.27%				
32	MC	0.98	0.25	1.49%				
33	MC	0.82	0.45	1.77%				
34	MC	0.86	0.47	2.12%				
35	MC	0.87	0.46	1.79%				
36	MC	0.70	0.36	2.10%				
37	MC	0.72	0.35	2.27%				
38	MC	0.46	0.41	2.51%				
39	MC	0.76	0.46	0.26%				
40	MC	0.67	0.49	0.35%				
41	MC	0.52	0.39	0.45%				
42	MC	0.76	0.57	0.73%				
43	MC	0.82	0.48	1.77%				
44	MC	0.59	0.49	0.45%				
45	MC	0.82	0.44	0.52%				
46	MC	0.87	0.54	0.52%				
47	MC	0.55	0.26	1.27%				
48	MC	0.84	0.43	0.52%				
49	MC	0.60	0.33	0.73%				
50	MC	0.84	0.48	0.76%				

* 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 © 2011 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	0.82%				
52	MC	0.79	0.53	0.95%				
53	MC	0.50	0.23	1.79%		+		
54	MC	0.64	0.49	0.89%				
55	MC	0.74	0.56	1.69%				
56	MC	0.54	0.44	1.06%				

* 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 © 2011 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	0.79	0.53					
2	MC	0.59	0.54	0.06%				
3	MC	0.61	0.30	0.19%				
4	MC	0.84	0.53	0.57%				
5	MC	0.72	0.44					
6	MC	0.65	0.48	0.34%				
7	MC	0.81	0.51	0.11%				
8	MC	0.85	0.45	0.98%				
9	MC	0.72	0.57	0.15%				
10	MC	0.72	0.57	0.51%				
11	MC	0.89	0.52	0.34%				
12	MC	0.57	0.34	0.97%		+		
13	CR	0.52	0.48	1.15%				
14	MC	0.77	0.41	0.70%				
15	MC	0.67	0.24	1.40%				
16	MC	0.68	0.33	0.93%				
17	MC	0.84	0.28	1.53%				
18	MC	0.88	0.43	1.17%				
19	MC	0.40	0.38	0.11%				
20	MC	0.79	0.46	0.23%				
21	MC	0.57	0.23	0.19%				
22	MC	0.68	0.27	0.42%				
23	MC	0.86	0.37	1.00%				
24	MC	0.51	0.21	0.11%				
25	MC	0.58	0.42	0.19%				

* 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 © 2011 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	0.52	0.32%				
27	MC	0.72	0.43	0.13%				
28	MC	0.42	0.23	0.44%		+		
29	MC	0.42	0.32	0.36%				
30	MC	0.48	0.27	0.38%				
31	MC	0.35	0.26	0.78%				
32	MC	0.56	0.45	0.38%				
33	MC	0.70	0.24	0.45%				
34	MC	0.80	0.53	0.40%				
35	MC	0.54	0.44	0.47%				
36	MC	0.76	0.55	0.98%				
37	MC	0.86	0.45	0.08%				
38	MC	0.58	0.32	0.30%				
39	MC	0.55	0.28	0.40%				
40	MC	0.65	0.60	0.27%				
41	MC	0.88	0.42	0.13%				
42	MC	0.86	0.48	0.28%				
43	MC	0.47	0.34	0.45%				
44	MC	0.78	0.46	0.47%				
45	MC	0.66	0.51	0.87%				
46	MC	0.36	0.20	1.68%				
47	CR	0.42	0.52	2.16%				
48	MC	0.58	0.43	0.45%				
49	MC	0.72	0.61	0.66%				
50	MC	0.68	0.34	0.66%				

* 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 © 2011 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	0.68	0.49	0.57%				
52	MC	0.76	0.50	0.81%				
53	MC	0.59	0.44	0.81%				
54	MC	0.71	0.57	0.81%				
55	MC	0.61	0.41	0.93%				
56	MC	0.65	0.46	1.21%				

* 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 © 2011 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	0.87	0.42	0.04%				
2	MC	0.88	0.40	0.02%				
3	MC	0.81	0.47	0.04%				
4	MC	0.68	0.43	0.29%				
5	MC	0.33	0.38	0.53%				
6	MC	0.67	0.33	0.27%				
7	MC	0.80	0.49	0.46%				
8	MC	0.81	0.47	1.20%				
9	MC	0.33	0.19	0.13%		+		
10	MC	0.65	0.47	0.13%				
11	MC	0.58	0.37	0.32%				
12	MC	0.86	0.51	0.17%				
13	CR	0.55	0.42	1.05%				
14	MC	0.89	0.36	0.97%				
15	MC	0.78	0.39	1.05%				
16	MC	0.50	0.32	1.24%				
17	MC	0.72	0.40	1.20%				
18	MC	0.85	0.43	1.33%				
19	MC	0.50	0.17	1.35%		+		
20	MC	0.69	0.35	0.08%				
21	MC	0.39	0.39	0.08%				
22	MC	0.63	0.30	0.34%				
23	MC	0.70	0.46	0.50%				
24	MC	0.43	0.22	0.78%		+		
25	MC	0.71	0.38	0.13%				

* 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 © 2011 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	0.52	0.22	0.17%				
27	MC	0.70	0.37	0.50%				
28	MC	0.50	0.42	0.06%				
29	MC	0.58	0.52	0.15%				
30	MC	0.76	0.46	0.17%				
31	MC	0.73	0.49	0.21%				
32	MC	0.73	0.33	0.44%				
33	MC	0.70	0.49	0.65%				
34	MC	0.65	0.46	0.88%				
35	MC	0.62	0.48	0.21%				
36	MC	0.46	0.23	0.25%				
37	MC	0.49	0.12	0.57%	+			
38	MC	0.61	0.30	1.37%				
39	MC	0.96	0.36	0.08%				
40	MC	0.84	0.46	0.59%				
41	MC	0.92	0.47	0.19%				
42	MC	0.88	0.44	0.08%				
43	MC	0.87	0.44	0.29%				
44	MC	0.83	0.49	0.29%				
45	MC	0.84	0.36	1.11%				
46	MC	0.50	0.18	2.46%		+		
47	CR	0.48	0.53	1.20%				
48	MC	0.89	0.43	0.21%				
49	MC	0.62	0.48	0.46%				
50	MC	0.59	0.30	0.50%				

* 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 © 2011 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	0.58	0.39	0.38%				
52	MC	0.75	0.54	0.27%				
53	MC	0.68	0.53	0.32%				
54	MC	0.78	0.46	2.42%				
55	MC	0.69	0.60	0.32%				
56	MC	0.65	0.38	1.01%				

* 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 © 2011 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	0.52	0.23	0.08%				
2	MC	0.90	0.42	0.08%				
3	MC	0.61	0.47	0.10%				
4	MC	0.79	0.51	0.15%				
5	MC	0.69	0.22	0.56%				
6	MC	0.67	0.45	0.08%				
7	MC	0.45	0.27	0.08%				
8	MC	0.85	0.42	2.67%				
9	MC	0.57	0.18	3.60%				
10	MC	0.35	0.23	0.17%				
11	MC	0.80	0.34	0.27%				
12	MC	0.68	0.32	0.51%				
13	MC	0.46	0.36	0.42%				
14	MC	0.82	0.34	0.19%				
15	MC	0.80	0.36	0.64%				
16	MC	0.71	0.45	0.22%				
17	MC	0.78	0.41	0.25%				
18	CR	0.44	0.41	1.23%				
19	MC	0.39	0.16	0.15%		+		
20	MC	0.86	0.13	0.19%	+			
21	MC	0.68	0.48	0.46%				
22	MC	0.34	0.17	0.32%				
23	MC	0.54	0.34	0.74%				
24	CR	0.40	0.57	1.23%				
25	MC	0.52	0.32	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 © 2011 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	0.42	0.17	0.61%		+		
27	MC	0.54	0.47	0.42%				
28	MC	0.74	0.28	2.30%				
29	MC	0.48	0.32	0.44%				
30	MC	0.62	0.38	0.81%				
31	MC	0.69	0.42	1.37%				
32	MC	0.45	0.33	2.69%				
33	MC	0.45	0.34	3.49%				
34	MC	0.62	0.42	1.42%				
35	MC	0.88	0.45	2.37%				
36	MC	0.79	0.46	1.79%				
37	MC	0.83	0.50	1.64%				
38	MC	0.83	0.54	0.15%				
39	MC	0.66	0.45	0.25%				
40	MC	0.69	0.26	0.36%				
41	MC	0.71	0.40	0.25%				
42	MC	0.84	0.50	0.49%				
43	MC	0.68	0.34	2.22%				
44	MC	0.65	0.18	0.17%				
45	MC	0.76	0.43	0.30%				
46	MC	0.88	0.38	1.81%				
47	MC	0.78	0.41	2.69%				
48	MC	0.62	0.33	0.19%				
49	MC	0.77	0.48	0.37%				
50	MC	0.41	0.37	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 © 2011 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	0.35	0.93%				
52	MC	0.58	0.40	0.47%				
53	MC	0.64	0.41	0.71%				
54	MC	0.95	0.40	0.46%				
55	MC	0.86	0.45	1.20%				
56	MC	0.51	0.38	0.95%				

* 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 © 2011 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	0.72	0.55	0.15%				
2	MC	0.74	0.31	0.21%				
3	MC	0.73	0.37	0.09%				
4	MC	0.57	0.39	0.08%				
5	MC	0.68	0.30	0.06%				
6	MC	0.89	0.42	0.13%				
7	MC	0.30	0.29	0.21%				
8	MC	0.87	0.51	0.28%				
9	MC	0.64	0.25	0.78%				
10	MC	0.73	0.33	1.04%				
11	MC	0.72	0.29	2.09%				
12	MC	0.77	0.34	0.42%				
13	MC	0.34	0.22	0.57%				
14	MC	0.73	0.39	0.82%				
15	MC	0.61	0.34	0.93%				
16	MC	0.63	0.31	0.99%				
17	MC	0.47	0.34	1.01%				
18	MC	0.81	0.41	0.17%				
19	MC	0.62	0.41	0.30%				
20	MC	0.46	0.40	0.49%				
21	MC	0.46	0.32	0.47%				
22	MC	0.79	0.45	0.70%				
23	MC	0.45	0.40	0.78%				
24	MC	0.94	0.34	0.13%				
25	MC	0.74	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 © 2011 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	0.60	0.49	0.93%				
27	MC	0.46	0.36	1.86%				
28	MC	0.77	0.31	2.18%				
29	MC	0.65	0.38	0.15%				
30	MC	0.71	0.33	0.21%				
31	CR	0.53	0.40	1.90%				
32	MC	0.93	0.36	0.47%				
33	MC	0.82	0.39	0.59%				
34	MC	0.42	0.22	0.61%		+		
35	MC	0.76	0.31	0.59%				
36	MC	0.86	0.44	0.63%				
37	MC	0.67	0.44	0.17%				
38	MC	0.62	0.26	0.28%				
39	MC	0.83	0.50	0.40%				
40	MC	0.90	0.44	1.02%				
41	MC	0.79	0.49	1.29%				
42	MC	0.58	0.39	0.17%				
43	MC	0.65	0.33	0.53%				
44	MC	0.50	0.28	0.30%				
45	MC	0.45	0.12	1.21%	+	+		
46	MC	0.83	0.46	0.27%				
47	MC	0.60	0.38	0.51%				
48	MC	0.69	0.35	0.59%				
49	MC	0.41	0.25	1.27%				
50	CR	0.35	0.46	2.01%				

* 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 © 2011 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	0.34	0.27	0.51%		+		
52	MC	0.71	0.48	0.51%				
53	MC	0.74	0.35	0.57%				
54	MC	0.79	0.46	0.85%				
55	MC	0.93	0.44	0.57%				
56	MC	0.82	0.48	0.66%				

* 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 © 2011 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	0.26	0.06%				
2	MC	0.83	0.43	0.11%				
3	MC	0.79	0.36	0.06%				
4	MC	0.87	0.28	0.20%				
5	MC	0.82	0.42	0.09%				
6	MC	0.80	0.36	0.17%				
7	MC	0.54	0.15	1.13%	+			
8	MC	0.81	0.45	1.40%				
9	MC	0.66	0.33	0.07%				
10	MC	0.68	0.38	1.09%				
11	MC	0.73	0.50	0.17%				
12	MC	0.91	0.41	0.07%				
13	MC	0.80	0.48	0.28%				
14	MC	0.92	0.31	0.18%				
15	MC	0.53	0.33	0.50%				
16	MC	0.62	0.36	0.24%				
17	MC	0.78	0.41	0.31%				
18	MC	0.73	0.27	0.46%				
19	CR	0.52	0.47	1.59%				
20	MC	0.72	0.32	0.28%				
21	MC	0.82	0.49	0.20%				
22	MC	0.56	0.36	0.26%				
23	MC	0.83	0.47	0.11%				
24	MC	0.78	0.53	0.35%				
25	MC	0.72	0.56	0.74%				

* 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 © 2011 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	0.44	0.31	1.11%				
27	MC	0.54	0.45	0.20%				
28	MC	0.58	0.35	0.70%				
29	MC	0.62	0.18	1.35%				
30	MC	0.62	0.51	0.24%				
31	MC	0.77	0.39	0.41%				
32	MC	0.84	0.55	0.26%				
33	MC	0.41	0.20	0.42%		+		
34	MC	0.61	0.22	0.65%		+		
35	MC	0.63	0.37	0.30%				
36	MC	0.83	0.49	0.39%				
37	MC	0.77	0.38	0.44%				
38	MC	0.79	0.49	0.30%				
39	MC	0.74	0.42	0.28%				
40	MC	0.71	0.41	0.20%				
41	MC	0.67	0.46	0.11%				
42	MC	0.53	0.37	1.18%				
43	MC	0.77	0.54	0.15%				
44	MC	0.47	0.14	0.31%	+	+		
45	MC	0.78	0.35	0.44%				
46	MC	0.68	0.19	0.24%		+		
47	MC	0.64	0.31	0.20%				
48	MC	0.81	0.54	0.41%				
49	MC	0.52	0.29	0.39%				
50	MC	0.89	0.48	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 © 2011 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	0.85	0.44	0.22%				
52	MC	0.78	0.31	0.22%				
53	MC	0.57	0.46	0.59%				
54	MC	0.50	0.36	0.31%				
55	MC	0.55	0.25	0.54%		+		
56	CR	0.60	0.52	2.86%				

* 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 © 2011 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	0.88	0.36	0.02%				
2	MC	0.92	0.36	0.02%				
3	MC	0.65	0.32	0.19%				
4	MC	0.81	0.44	0.10%				
5	MC	0.45	0.38	0.10%				
6	MC	0.88	0.38	0.90%				
7	CR	0.51	0.50	7.85%			+	
8	MC	0.47	0.36	0.12%				
9	MC	0.76	0.51	0.28%				
10	MC	0.57	0.30	0.14%				
11	MC	0.72	0.40	0.35%				
12	MC	0.77	0.48	0.17%				
13	MC	0.76	0.47	0.24%				
14	MC	0.82	0.50	0.49%				
15	MC	0.88	0.49	0.40%				
16	MC	0.39	0.47	0.33%				
17	MC	0.35	0.32	0.10%		+		
18	MC	0.50	0.39	0.21%		+		
19	MC	0.92	0.33	0.23%				
20	MC	0.71	0.42	0.19%				
21	MC	0.74	0.53	0.42%				
22	MC	0.47	0.50	0.54%				
23	MC	0.91	0.49	0.23%				
24	MC	0.81	0.48	0.26%				
25	MC	0.80	0.44	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 © 2011 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	0.82	0.50	0.23%				
27	MC	0.76	0.47	0.45%				
28	MC	0.72	0.51	1.91%				
29	MC	0.81	0.39	2.08%				
30	MC	0.55	0.42	1.44%				
31	MC	0.46	0.47	2.01%				
32	MC	0.81	0.52	0.35%				
33	MC	0.84	0.47	0.35%				
34	MC	0.82	0.38	0.94%				
35	MC	0.65	0.38	0.94%				
36	MC	0.70	0.40	0.52%				
37	MC	0.61	0.39	0.47%				
38	MC	0.60	0.44	0.64%				
39	MC	0.45	0.34	0.56%				
40	MC	0.72	0.48	0.57%				
41	MC	0.61	0.34	0.43%				
42	MC	0.84	0.50	2.26%				
43	MC	0.73	0.40	0.45%				
44	MC	0.93	0.45	0.47%				
45	MC	0.59	0.40	0.49%				
46	MC	0.47	0.40	0.52%				
47	MC	0.75	0.58	0.63%				
48	MC	0.57	0.36	0.87%				
49	MC	0.51	0.25	0.69%		+		
50	MC	0.67	0.51	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 © 2011 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	0.50	0.26	0.64%				
52	CR	0.53	0.58	4.62%				

* 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 © 2011 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	0.95	0.29	0.24%				
2	MC	0.94	0.37	0.19%				
3	MC	0.84	0.45	0.24%				
4	MC	0.74	0.32	0.22%				
5	CR	0.52	0.50	0.88%				
6	MC	0.85	0.40	0.26%				
7	MC	0.67	0.42	0.71%				
8	MC	0.88	0.35	0.39%				
9	MC	0.90	0.30	1.25%				
10	MC	0.85	0.30	1.03%				
11	MC	0.93	0.30	0.73%				
12	MC	0.64	0.34	0.77%				
13	MC	0.82	0.40	1.31%				
14	MC	0.97	0.25	1.44%				
15	MC	0.89	0.24	0.30%				
16	MC	0.78	0.40	0.34%				
17	MC	0.61	0.28	0.45%				
18	MC	0.90	0.38	0.32%				
19	MC	0.74	0.34	0.62%				
20A	CR	0.48	0.39	5.31%			+	
20B	CR	0.25	0.44	3.51%				+
21	MC	0.87	0.46	0.65%				
22	MC	0.73	0.28	0.47%				
23	MC	0.68	0.41	0.56%				
24	MC	0.81	0.24	0.93%				

* 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 © 2011 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	0.65	0.45	1.18%				
26	MC	0.88	0.39	1.53%				
27A	CR	0.65	0.50	3.29%				
27B	CR	0.28	0.45	5.87%			+	+
28	MC	0.77	0.41	2.35%				
29	MC	0.57	0.35	3.10%				
30	MC	0.77	0.39	3.18%				
31	MC	0.87	0.42	0.32%				
32	MC	0.75	0.47	0.99%				
33	MC	0.90	0.38	0.41%				
34	MC	0.60	0.32	0.28%				
35	MC	0.87	0.31	0.80%				
36	MC	0.30	0.36	2.71%				+
37	MC	0.94	0.15	0.82%				
38	MC	0.93	0.25	0.82%				
39A	CR	0.39	0.50	4.35%				
39B	CR	0.32	0.54	2.82%				
40	MC	0.90	0.33	0.37%				
41	MC	0.94	0.32	0.41%				
42	MC	0.61	0.38	0.56%				
43	MC	0.67	0.40	0.43%				
44	MC	0.93	0.26	0.47%				
45	MC	0.79	0.44	0.58%				
46	MC	0.69	0.25	0.69%				
47	MC	0.71	0.45	0.84%				

* 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 © 2011 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	0.39	1.01%				
49	MC	0.74	0.43	1.27%				
50	MC	0.94	0.29	1.16%				

* 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 © 2011 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	0.13	0.11%	+			
2	MC	0.89	0.22	0.11%				
3	MC	0.87	0.30	0.19%				
4	MC	0.91	0.39	0.40%				
5	MC	0.88	0.41	0.11%				
6	MC	0.80	0.43	0.53%				
7	MC	0.75	0.40	0.66%				
8	MC	0.92	0.34	0.11%				
9	MC	0.62	0.46	0.45%				
10	MC	0.97	0.31	0.34%				
11A	CR	0.24	0.34	0.93%				+
11B	CR	0.32	0.36	3.55%				
12	MC	0.69	0.54	0.89%				
13	MC	0.78	0.52	1.00%				
14	MC	0.83	0.45	1.10%				
15	MC	0.94	0.33	0.09%				
16	MC	0.71	0.40	0.34%				
17	MC	0.72	0.32	0.28%				
18	MC	0.85	0.37	0.43%				
19	MC	0.73	0.29	0.11%				
20	MC	0.98	0.15	0.13%				
21	MC	0.75	0.45	0.38%				
22	CR	0.83	0.22	0.57%				
23	MC	0.72	0.30	0.21%				
24	MC	0.84	0.34	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 © 2011 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	0.48	0.55	0.57%				
25B	CR	0.31	0.42	2.38%				
26	MC	0.67	0.35	0.36%				
27	MC	0.91	0.32	0.42%				
28	MC	0.83	0.37	0.34%				
29	MC	0.78	0.41	0.49%				
30	MC	0.45	0.41	0.55%				
31	MC	0.91	0.27	0.19%				
32	MC	0.94	0.34	1.11%				
33	MC	0.94	0.28	0.19%				
34	MC	0.77	0.35	0.28%				
35	MC	0.88	0.30	0.30%				
36	MC	0.99	0.13	1.45%	+			
37	MC	0.54	0.36	0.32%				
38	MC	0.76	0.51	3.87%				
39A	CR	0.61	0.44	0.45%				
39B	CR	0.40	0.50	2.21%				
40	MC	0.60	0.40	0.25%				
41	MC	0.90	0.22	0.26%				
42	MC	0.71	0.29	0.23%				
43	MC	0.77	0.38	0.34%				
44	MC	0.46	0.28	0.40%				
45	MC	0.58	0.48	0.38%				
46	MC	0.65	0.42	0.77%				
47	MC	0.60	0.47	0.64%				

* 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 © 2011 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	0.79	0.29	0.57%				
49	MC	0.90	0.40	0.59%				
50	MC	0.87	0.32	0.62%				

* 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 © 2011 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	0.90	0.30	0.08%				
2	MC	0.86	0.38	0.13%				
3	MC	0.70	0.35	0.15%				
4	MC	0.62	0.48	0.57%				
5	MC	0.94	0.23	0.19%				
6	MC	0.80	0.53	0.25%				
7	MC	0.44	0.60	0.40%				
8	MC	0.74	0.42	0.48%				
9	MC	0.51	0.25	0.63%				
10	MC	0.68	0.53	0.55%				
11	MC	0.64	0.38	0.36%				
12	CR	0.68	0.58	1.13%				
13	MC	0.54	0.43	0.46%				
14	MC	0.67	0.44	0.65%				
15	MC	0.90	0.28	0.08%				
16	MC	0.64	0.31	0.19%				
17	MC	0.76	0.37	0.08%				
18	MC	0.75	0.36	0.42%				
19	MC	0.64	0.20	0.21%				
20A	CR	0.87	0.23	0.48%				
20B	CR	0.33	0.44	1.26%				
21	MC	0.40	0.42	0.13%				
22	MC	0.87	0.29	0.11%				
23	MC	0.71	0.53	0.08%				
24	MC	0.96	0.22	0.08%				

* 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 © 2011 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	0.72	0.44	0.17%				
26	MC	0.57	0.42	0.40%				
27A	CR	0.64	0.47	0.61%				
27B	CR	0.75	0.46	1.07%				
28	MC	0.81	0.34	0.19%				
29	MC	0.90	0.26	0.23%				
30	MC	0.50	0.43	0.63%				
31	MC	0.67	0.30	0.36%				
32	MC	0.94	0.23	0.55%				
33	MC	0.96	0.27	0.55%				
34	MC	0.64	0.47	0.38%				
35	MC	0.94	0.32	0.50%				
36	MC	0.70	0.34	0.17%				
37	MC	0.60	0.38	0.23%				
38	MC	0.40	0.26	0.29%		+		
39	MC	0.91	0.28	0.15%				
40	MC	0.89	0.26	0.17%				
41	MC	0.39	0.17	0.23%		+		
42	MC	0.92	0.20	0.13%				
43	MC	0.82	0.31	0.25%				
44	MC	0.76	0.20	0.21%				
45	MC	0.88	0.41	0.48%				
46A	CR	0.39	0.48	0.34%				
46B	CR	0.36	0.47	0.78%				
47	MC	0.88	0.34	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 © 2011 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	0.83	0.51	0.40%				
49	MC	0.63	0.31	0.38%				
50	MC	0.92	0.35	0.29%				
51	MC	0.57	0.46	0.82%				
52	MC	0.84	0.25	0.15%				
53	MC	0.84	0.34	0.32%				
54	MC	0.99	0.16	0.32%				
55	MC	0.81	0.48	0.29%				

* 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 © 2011 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	0.22%				
2	MC	0.87	0.42	0.10%				
3	MC	0.67	0.40	0.42%				
4	MC	0.47	0.26	0.34%		+		
5	MC	0.87	0.39	0.30%				
6	MC	0.55	0.44	0.25%				
7A	CR	0.55	0.39	4.74%				
7B	CR	0.45	0.44	5.69%			+	
8	MC	0.69	0.41	0.77%				
9	MC	0.90	0.41	0.59%				
10	MC	0.67	0.36	1.18%				
11	MC	0.87	0.36	1.04%				
12	MC	0.51	0.31	1.38%				
13	MC	0.54	0.37	1.92%				
14	MC	0.87	0.51	1.78%				
15	MC	0.81	0.50	1.82%				
16	MC	0.54	0.44	0.50%				
17	MC	0.75	0.39	0.24%				
18	MC	0.67	0.40	0.27%				
19	MC	0.91	0.35	0.22%				
20	MC	0.90	0.34	0.37%				
21	MC	0.53	0.40	0.47%				
22A	CR	0.90	0.34	0.66%				
22B	CR	0.25	0.45	1.56%				+
23	MC	0.65	0.29	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 © 2011 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	0.63	0.38	0.35%				
25	MC	0.38	0.31	0.50%				
26	MC	0.80	0.40	0.62%				
27	MC	0.68	0.42	0.49%				
28	MC	0.63	0.50	0.61%				
29	MC	0.80	0.33	0.89%				
30	MC	0.48	0.38	0.25%				
31	MC	0.80	0.41	0.24%				
32	MC	0.56	0.47	0.27%				
33	MC	0.96	0.19	0.17%				
34	MC	0.79	0.32	0.22%				
35	MC	0.97	0.21	0.40%				
36	MC	0.71	0.47	0.42%				
37	MC	0.83	0.22	0.64%				
38	MC	0.92	0.30	0.56%				
39A	CR	0.43	0.57	1.56%				
39B	CR	0.41	0.56	2.62%				
40	MC	0.59	0.48	0.62%				
41	MC	0.67	0.44	0.32%				
42	MC	0.98	0.14	0.19%	+			
43	MC	0.73	0.30	0.29%				
44	MC	0.67	0.56	0.32%				
45	MC	0.77	0.47	0.42%				
46	MC	0.70	0.55	0.57%				
47	MC	0.80	0.44	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 © 2011 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	0.78	0.47	0.91%				
49	MC	0.60	0.42	0.72%				
50	MC	0.58	0.47	1.04%				
51	CR	0.41	0.52	1.03%				
52	MC	0.44	0.28	0.40%				
53	MC	0.78	0.26	0.47%				
54	MC	0.57	0.36	0.71%				
55	MC	0.54	0.48	0.71%				

* 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 © 2011 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	0.72	0.45	0.21%				
2	MC	0.91	0.40	0.13%				
3	MC	0.62	0.52	0.11%				
4	MC	0.73	0.24	0.17%				
5	MC	0.76	0.41	0.17%				
6	MC	0.73	0.37	0.25%				
7A	CR	0.70	0.51	1.25%				
7B	CR	0.61	0.58	2.20%				
8	MC	0.73	0.31	0.49%				
9	MC	0.59	0.48	0.61%				
10	MC	0.67	0.52	0.89%				
11	MC	0.61	0.58	1.16%				
12	MC	0.62	0.21	0.80%		+		
13	MC	0.54	0.57	1.10%				
14	MC	0.90	0.33	0.87%				
15	MC	0.78	0.41	1.00%				
16	MC	0.64	0.24	0.23%				
17	MC	0.56	0.28	0.30%				
18	MC	0.86	0.50	0.15%				
19A	CR	0.54	0.40	1.61%				
19B	CR	0.33	0.52	3.26%				
20	MC	0.98	0.21	0.23%				
21	MC	0.49	0.25	0.42%				
22	MC	0.73	0.29	0.34%				
23	MC	0.56	0.46	0.32%				

* 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 © 2011 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	0.86	0.49	0.23%				
25	MC	0.40	0.43	0.38%				
26	MC	0.92	0.22	0.61%				
27	MC	0.72	0.42	0.44%				
28	MC	0.82	0.49	0.42%				
29	MC	0.85	0.24	0.38%				
30	MC	0.89	0.32	0.11%				
31	MC	0.51	0.38	0.44%				
32	MC	0.65	0.47	0.34%				
33	MC	0.58	0.34	0.38%				
34	MC	0.40	0.29	0.42%				
35	MC	0.68	0.56	0.49%				
36	MC	0.66	0.36	0.32%				
37	MC	0.95	0.24	0.30%				
38A	CR	0.64	0.37	7.43%			+	
38B	CR	0.38	0.40	4.70%				
39	MC	0.67	0.42	0.27%				
40	MC	0.89	0.36	0.49%				
41	MC	0.75	0.41	0.34%				
42	MC	0.91	0.32	0.61%				
43	MC	0.61	0.44	0.27%				
44	MC	0.98	0.24	0.32%				
45	MC	0.39	0.43	0.36%				
46	MC	0.89	0.39	0.23%				
47	MC	0.83	0.32	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 © 2011 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	0.87	0.33	0.21%				
49	MC	0.47	0.52	0.51%				
50	MC	0.75	0.60	0.91%				
51	MC	0.52	0.35	0.93%				
52	CR	0.34	0.56	1.40%				
53	MC	0.82	0.30	0.57%				
54	MC	0.57	0.44	0.55%		+		
55	MC	0.46	0.14	0.70%	+			

* 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 © 2011 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	0.46	0.20	0.13%				
2	MC	0.54	0.48	0.17%				
3	MC	0.71	0.51	0.11%				
4	MC	0.55	0.53	0.41%				
5	MC	0.47	0.34	0.31%		+		
6	MC	0.88	0.41	0.20%				
7	MC	0.64	0.52	0.13%				
8	MC	0.54	0.35	0.18%				
9	MC	0.51	0.42	0.18%				
10	CR	0.22	0.51	4.02%				+
11	MC	0.86	0.44	0.33%				
12	MC	0.43	0.35	0.83%				
13	MC	0.68	0.52	0.70%				
14	MC	0.90	0.36	0.61%				
15	MC	0.59	0.29	0.63%				
16	MC	0.75	0.39	0.09%				
17	MC	0.39	0.36	0.37%				
18	MC	0.73	0.24	0.24%				
19A	CR	0.75	0.39	0.83%				
19B	CR	0.62	0.47	3.37%				
20	MC	0.77	0.39	0.20%				
21	MC	0.31	0.38	0.17%				
22	MC	0.38	0.41	0.33%				
23	MC	0.86	0.38	0.24%				
24	MC	0.96	0.24	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 © 2011 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	0.42	0.44%				
26	MC	0.65	0.39	0.72%				
27	MC	0.60	0.43	0.44%				
28	MC	0.73	0.38	0.66%				
29	MC	0.41	0.42	0.15%				
30	MC	0.49	0.34	0.11%				
31	MC	0.54	0.37	0.26%				
32	MC	0.50	0.38	0.53%				
33A	CR	0.31	0.45	7.08%			+	
33B	CR	0.38	0.62	7.26%			+	
34	MC	0.52	0.38	0.35%				
35	MC	0.65	0.34	0.68%				
36	MC	0.82	0.46	0.59%				
37	MC	0.40	0.46	0.42%				
38	MC	0.64	0.35	0.85%				
39	MC	0.70	0.30	1.18%				
40	MC	0.64	0.15	1.14%	+	+		
41	MC	0.88	0.36	0.76%				
42	MC	0.67	0.35	0.79%				
43	MC	0.78	0.30	0.29%				
44	MC	0.78	0.37	0.24%				
45	MC	0.66	0.51	0.18%		+		
46	MC	0.41	0.28	0.41%				
47	MC	0.81	0.30	0.44%				
48	MC	0.45	0.44	0.46%				

* 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 © 2011 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	0.25	0.42	2.51%				+
49B	CR	0.21	0.42	3.98%				+
50	MC	0.74	0.38	0.26%				
51	MC	0.26	0.34	0.35%				+
52	MC	0.60	0.53	0.41%				
53	MC	0.55	0.51	0.52%				
54	MC	0.85	0.51	0.31%				
55	MC	0.67	0.26	0.39%				

* 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 © 2011 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	0.83	0.45	0.12%				
2	MC	0.67	0.39	0.17%				
3	MC	0.50	0.34	0.43%				
4	MC	0.57	0.50	0.24%				
5	MC	0.53	0.43	0.17%				
6	MC	0.61	0.46	0.21%				
7	MC	0.43	0.43	0.19%				
8	MC	0.69	0.48	0.21%				
9	MC	0.45	0.36	0.17%		+		
10	MC	0.49	0.36	0.16%				
11	MC	0.51	0.48	0.09%				
12	MC	0.73	0.44	0.50%				
13	MC	0.81	0.36	0.30%				
14	MC	0.50	0.44	0.36%				
15	MC	0.64	0.43	0.42%				
16	MC	0.80	0.50	0.43%				
17	MC	0.69	0.63	0.31%				
18	MC	0.71	0.47	0.26%				
19	MC	0.50	0.41	0.76%				
20	MC	0.56	0.50	0.36%				
21	MC	0.74	0.44	0.21%				
22	MC	0.43	0.55	0.36%				
23	CR	0.47	0.65	7.42%			+	
24	MC	0.71	0.39	0.89%				
25	MC	0.75	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 © 2011 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	0.56	0.58	1.06%				
27	MC	0.52	0.43	1.44%				
28	MC	0.50	0.52	0.52%				
29	MC	0.80	0.44	0.45%				
30	MC	0.60	0.40	0.47%				
31	MC	0.64	0.22	0.35%				
32	MC	0.49	0.42	0.43%				
33	CR	0.34	0.58	15.16%			+	
34	MC	0.71	0.42	0.61%				
35	MC	0.94	0.25	0.54%				
36	CR	0.50	0.65	10.56%			+	
37	MC	0.62	0.32	0.36%				
38	MC	0.59	0.42	0.59%				
39	MC	0.66	0.50	0.49%				
40	MC	0.47	0.49	0.47%				
41	MC	0.54	0.64	0.64%				
42	MC	0.59	0.49	0.52%				
43	MC	0.49	0.54	0.64%				
44	MC	0.75	0.46	0.68%				
45	MC	0.54	0.43	0.75%		+		
46	CR	0.52	0.53	6.15%			+	
47	MC	0.56	0.46	0.64%				
48	MC	0.86	0.39	0.61%				
49	MC	0.72	0.41	0.66%				
50	MC	0.78	0.47	0.87%				

* 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 © 2011 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	0.38	0.42	0.90%				
52	MC	0.43	0.49	0.92%				
53	MC	0.52	0.37	1.11%				
54	MC	0.74	0.53	0.87%				

* 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 © 2011 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	0.42	0.32	0.15%				
2	MC	0.93	0.29	0.19%				
3	MC	0.80	0.30	0.53%				
4	MC	0.93	0.27	0.51%				
5	MC	0.41	0.16	0.64%		+		
6	MC	0.84	0.36	0.23%				
7	MC	0.60	0.38	0.27%				
8	MC	0.84	0.28	0.28%				
9	MC	0.83	0.41	0.30%				
10	MC	0.47	0.32	0.97%				
11	MC	0.91	0.38	0.27%				
12	MC	0.95	0.30	0.25%				
13	MC	0.79	0.44	0.23%				
14	MC	0.66	0.42	0.45%				
15	MC	0.53	0.24	0.27%				
16	MC	0.68	0.46	1.69%				
17	MC	0.67	0.49	0.25%				
18	MC	0.71	0.47	2.56%				
19	MC	0.47	0.35	0.97%				
20	MC	0.71	0.43	0.45%				
21	MC	0.73	0.44	0.51%				
22	MC	0.64	0.42	1.61%				
23	MC	0.53	0.35	0.85%				
24	MC	0.59	0.43	0.66%				
25	MC	0.43	0.19	0.89%		+		

* 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 © 2011 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	0.55	0.41	1.16%				
27	MC	0.68	0.43	1.16%				
28	MC	0.37	0.30	1.72%				
29	MC	0.78	0.42	2.07%				
30	MC	0.44	0.13	2.44%	+	+		

* 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 © 2011 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	0.66	0.17	0.11%				
2	MC	0.90	0.14	0.13%	+			
3	MC	0.69	0.31	0.33%				
4	MC	0.84	0.39	0.06%				
5	MC	0.73	0.38	0.35%				
6	MC	0.54	0.30	0.33%				
7	MC	0.88	0.37	0.39%				
8	MC	0.65	0.45	0.26%				
9	MC	0.81	0.39	0.48%				
10	MC	0.65	0.39	0.11%				
11	MC	0.84	0.36	0.17%				
12	MC	0.72	0.46	0.61%				
13	MC	0.50	0.35	0.15%				
14	MC	0.81	0.50	0.20%				
15	MC	0.88	0.40	0.19%				
16	MC	0.88	0.36	0.26%				
17	MC	0.66	0.43	0.28%				
18	MC	0.85	0.50	1.02%				
19	MC	0.81	0.42	0.44%				
20	MC	0.60	0.56	3.24%				
21	MC	0.90	0.48	0.63%				
22	MC	0.62	0.43	1.56%				
23	MC	0.70	0.46	1.26%				
24	MC	0.78	0.42	0.72%				
25	MC	0.71	0.45	1.91%				

* 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 © 2011 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	0.47	0.36	1.02%		+		
27	MC	0.74	0.35	1.89%				
28 **	MC							
29	MC	0.69	0.41	1.65%				
30	MC	0.65	0.45	2.06%				

* 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 © 2011 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	0.73	0.41	0.02%				
2	MC	0.82	0.42	0.05%				
3	MC	0.64	0.27	0.18%				
4	MC	0.69	0.30	1.39%				
5	MC	0.82	0.37	0.28%				
6	MC	0.69	0.36	0.25%				
7	MC	0.41	0.21	0.39%				
8	MC	0.81	0.36	0.14%				
9	MC	0.60	0.31	0.05%				
10	MC	0.68	0.41	0.11%				
11	MC	0.78	0.45	1.27%				
12	MC	0.64	0.26	0.28%				
13	MC	0.48	0.20	0.26%		+		
14	MC	0.79	0.52	0.18%				
15	MC	0.59	0.40	0.21%				
16	MC	0.70	0.35	0.18%				
17	MC	0.88	0.41	0.41%				
18	MC	0.70	0.29	0.42%				
19	MC	0.64	0.48	0.28%				
20	MC	0.51	0.36	0.53%				
21	MC	0.41	0.20	0.39%		+		
22	MC	0.66	0.46	0.32%				
23	MC	0.43	0.32	0.41%		+		
24	MC	0.30	0.22	0.51%				
25	MC	0.75	0.53	0.90%				

* 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 © 2011 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**
26	MC	0.71	0.43	0.44%				
27	MC	0.27	0.21	0.62%		+		+
28	MC	0.84	0.41	0.67%				
29	MC	0.66	0.40	0.95%				
30	MC	0.74	0.50	1.00%				
1A**	WR	0.57	0.56	2.50%				
1B**	WR	0.70	0.37	2.50%				

* 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.
** Writing prompt items are included here. The Writing raw score contributes to the scale score for Language Arts in grade 10.

Copyright © 2011 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	0.41	0.02%				
2	MC	0.92	0.33	0.06%				
3	MC	0.96	0.31	0.08%				
4	MC	0.51	0.43	0.09%				
5	MC	0.84	0.20	0.13%				
6	MC	0.60	0.23	0.34%		+		
7	MC	0.80	0.39	0.47%				
8	MC	0.88	0.36	0.13%				
9	MC	0.82	0.40	0.72%				
10	MC	0.75	0.25	0.11%				
11	MC	0.84	0.48	0.36%				
12	MC	0.61	0.48	0.40%				
13	MC	0.95	0.30	0.17%				
14	MC	0.97	0.23	1.52%				
15	MC	0.84	0.38	0.23%				
16	MC	0.52	0.30	1.08%				
17	MC	0.81	0.33	0.28%				
18	MC	0.92	0.32	0.30%				
19	MC	0.97	0.28	0.93%				
20	MC	0.66	0.48	1.16%				
21	MC	0.56	0.44	0.61%				
22	MC	0.74	0.47	0.42%				
23	MC	0.86	0.42	0.57%				
24	MC	0.94	0.28	0.23%				
25	MC	0.92	0.38	0.53%				

* 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 © 2011 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	0.74	0.49	1.67%				
27	MC	0.95	0.36	0.25%				
28	MC	0.82	0.39	0.44%				
29	MC	0.65	0.41	0.32%				
30	MC	0.68	0.41	0.45%				
31	MC	0.66	0.39	0.59%				
32	MC	0.76	0.33	0.42%				
33	MC	0.59	0.35	0.49%				
34	MC	0.72	0.42	1.25%				
35	MC	0.88	0.41	0.34%				
36	MC	0.63	0.48	1.02%				
37	MC	0.91	0.45	0.51%				
38	MC	0.75	0.47	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 © 2011 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	0.95	0.33	0.04%				
2	MC	0.93	0.31	0.41%				
3	MC	0.67	0.41	0.19%				
4	MC	0.39	0.19	0.37%		+		
5	MC	0.79	0.33	0.37%				
6	MC	0.96	0.34	0.39%				
7	MC	0.80	0.28	0.46%				
8	MC	0.92	0.35	0.22%				
9	MC	0.58	0.41	0.41%				
10	MC	0.83	0.48	1.00%				
11	MC	0.73	0.41	1.85%				
12	MC	0.76	0.39	0.20%				
13	MC	0.60	0.34	0.61%				
14	MC	0.76	0.46	1.06%				
15	MC	0.85	0.43	0.94%				
16	MC	0.76	0.56	0.30%				
17	MC	0.69	0.48	1.09%				
18	MC	0.80	0.34	0.31%				
19	MC	0.51	0.38	0.43%				
20	MC	0.53	0.36	0.24%				
21	MC	0.57	0.41	1.00%				
22	MC	0.93	0.36	0.28%				
23	MC	0.73	0.36	0.37%				
24	MC	0.59	0.40	0.46%				
25	MC	0.83	0.38	0.81%				

* 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 © 2011 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	0.61	0.37	3.50%				
27	MC	0.66	0.45	0.57%				
28	MC	0.69	0.52	0.54%				
29	MC	0.61	0.42	1.17%				
30	MC	0.47	0.26	0.43%				
31	MC	0.66	0.35	1.44%				
32	MC	0.81	0.50	0.48%				
33	MC	0.77	0.46	1.17%				
34	MC	0.82	0.46	0.80%				
35	MC	0.66	0.39	0.69%				
36	MC	0.49	0.29	0.43%				
37	MC	0.53	0.22	0.70%				
38	MC	0.58	0.36	0.63%				
39	MC	0.48	0.40	0.72%				
40	MC	0.65	0.41	0.85%				

* 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 © 2011 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	0.80	0.48	0.11%				
2	MC	0.97	0.30	0.55%				
3	MC	0.67	0.32	0.12%				
4	MC	0.52	0.34	0.19%				
5	MC	0.66	0.49	0.23%				
6	MC	0.43	0.38	0.09%				
7	MC	0.55	0.38	0.21%				
8	MC	0.61	0.27	0.35%				
9	MC	0.58	0.40	0.46%				
10	MC	0.80	0.35	0.49%				
11	MC	0.53	0.38	0.58%				
12	MC	0.48	0.27	0.11%				
13	MC	0.55	0.28	0.14%				
14	MC	0.78	0.38	0.19%				
15	MC	0.35	0.31	0.11%				
16	MC	0.67	0.47	0.44%				
17	MC	0.55	0.39	0.28%				
18	MC	0.45	0.23	0.32%				
19	MC	0.88	0.40	0.05%				
20	MC	0.87	0.37	0.32%				
21	MC	0.72	0.45	0.69%				
22	MC	0.79	0.50	0.16%				
23	MC	0.49	0.43	0.62%		+		
24	MC	0.69	0.52	0.28%				
25	MC	0.97	0.26	0.12%				

* 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 © 2011 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	0.70	0.43	0.72%				
27	MC	0.78	0.46	1.91%				
28	MC	0.82	0.45	0.76%				
29	MC	0.46	0.16	0.56%				
30	MC	0.63	0.41	0.60%				
31	MC	0.79	0.45	0.60%				
32	MC	0.62	0.43	1.59%				
33	MC	0.63	0.40	1.08%				
34	MC	0.78	0.46	0.67%				
35	MC	0.63	0.36	1.62%				
36	MC	0.78	0.46	0.90%				
37	MC	0.64	0.40	0.58%				
38	MC	0.76	0.50	0.60%				
39	MC	0.82	0.42	0.69%				
40	MC	0.18	0.12	0.74%	+	+		+
41 **	MC							
42	MC	0.66	0.36	0.72%				
43	MC	0.43	0.47	1.27%		+		
44	MC	0.68	0.52	0.76%				
45	MC	0.52	0.53	0.71%				
46	MC	0.80	0.51	0.69%				
47	MC	0.84	0.47	0.95%				
48	MC	0.67	0.48	1.34%				
49	MC	0.49	0.43	0.76%				
50	MC	0.87	0.51	1.32%				

* 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 © 2011 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	0.22	0.02%				
2	MC	0.86	0.27	0.08%				
3	MC	0.90	0.43	0.49%				
4	MC	0.80	0.36	0.23%				
5	MC	0.93	0.16	0.06%				
6	MC	0.81	0.30	0.59%				
7	MC	0.93	0.35	0.17%				
8	MC	0.71	0.40	0.25%				
9	MC	0.80	0.36	0.32%				
10	MC	0.94	0.22	0.30%				
11	MC	0.75	0.38	0.26%				
12	MC	0.77	0.34	0.57%				
13	MC	0.68	0.31	0.15%				
14	MC	0.64	0.07	0.47%	+	+		
15	MC	0.89	0.48	2.78%				
16	MC	0.85	0.37	3.16%				
17	MC	0.81	0.46	0.17%				
18	MC	0.54	0.40	0.32%				
19	MC	0.53	0.44	0.28%				
20	MC	0.66	0.37	0.26%				
21	MC	0.66	0.40	0.53%				
22	MC	0.82	0.48	0.70%				
23	MC	0.88	0.34	0.30%				
24	MC	0.46	0.16	0.43%		+		
25	MC	0.86	0.30	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 © 2011 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	0.56	0.37	0.34%				
27	MC	0.70	0.34	0.43%				
28	MC	0.92	0.39	0.64%				
29	MC	0.71	0.31	0.74%				
30	MC	0.80	0.44	0.53%				
31	MC	0.48	0.35	1.70%				
32	MC	0.35	0.11	0.42%	+	+		
33	MC	0.70	0.43	0.77%				
34	MC	0.96	0.24	1.08%				
35	MC	0.79	0.51	1.02%				
36	MC	0.68	0.39	0.45%				
37	MC	0.46	0.35	0.43%				
38	MC	0.48	0.31	0.51%				
39	MC	0.50	0.45	0.81%				
40	MC	0.50	0.38	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 © 2011 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	0.89	0.37	0.13%				
2	MC	0.97	0.30	0.06%				
3	MC	0.96	0.25	0.09%				
4	MC	0.89	0.26	0.57%				
5	MC	0.92	0.37	0.28%				
6	MC	0.52	0.28	0.35%				
7	MC	0.93	0.32	0.24%				
8	MC	0.78	0.34	0.42%				
9	MC	0.58	0.37	0.59%				
10	MC	0.87	0.40	0.42%				
11	MC	0.95	0.30	0.94%				
12	MC	0.96	0.28	1.09%				
13	MC	0.78	0.34	0.26%				
14	MC	0.67	0.37	0.29%				
15	MC	0.71	0.40	0.28%				
16	MC	0.77	0.43	0.35%				
17	MC	0.75	0.34	0.35%				
18	MC	0.89	0.26	0.15%				
19	MC	0.72	0.45	0.52%				
20	MC	0.74	0.44	0.35%				
21	MC	0.74	0.36	0.39%				
22	MC	0.73	0.41	0.18%				
23	MC	0.87	0.41	0.39%				
24	MC	0.70	0.49	0.61%				
25	MC	0.83	0.47	0.18%				

* 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 © 2011 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	0.61	0.36	0.28%				
27	MC	0.66	0.46	4.29%				
28	MC	0.74	0.50	0.24%				
29	MC	0.79	0.41	0.39%				
30	MC	0.75	0.27	0.98%				
31	MC	0.46	0.33	1.62%				
32	MC	0.60	0.43	0.42%				
33	MC	0.82	0.35	0.50%				
34	MC	0.69	0.44	0.48%				
35	MC	0.55	0.27	0.37%				
36	MC	0.45	0.28	5.11%			+	
37	MC	0.67	0.50	0.66%				
38	MC	0.65	0.33	0.88%				
39	MC	0.60	0.21	0.70%				
40	MC	0.59	0.32	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 © 2011 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	0.74	0.41	0.05%				
2	MC	0.90	0.34	0.02%				
3	MC	0.43	0.15	0.12%	+	+		
4	MC	0.43	0.16	3.94%				
5	MC	0.66	0.37	0.21%				
6	MC	0.84	0.40	0.12%				
7	MC	0.73	0.43	0.23%				
8	MC	0.83	0.39	0.12%				
9	MC	0.89	0.35	0.19%				
10	MC	0.66	0.32	0.44%				
11	MC	0.68	0.44	0.70%				
12	MC	0.43	0.27	0.28%				
13	MC	0.85	0.43	0.32%				
14	MC	0.75	0.28	0.16%				
15	MC	0.49	0.31	0.56%				
16	MC	0.51	0.31	0.19%				
17	MC	0.64	0.36	0.23%				
18	MC	0.58	0.29	0.33%				
19	MC	0.78	0.42	0.33%				
20	MC	0.71	0.43	0.26%				
21	MC	0.84	0.26	0.25%				
22	MC	0.43	0.42	0.62%				
23	MC	0.49	0.45	0.32%				
24	MC	0.40	0.24	0.42%				
25	MC	0.59	0.46	0.19%				

* 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 © 2011 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	0.42	0.30	0.37%				
27	MC	0.68	0.41	0.37%				
28	MC	0.68	0.50	0.42%				
29	MC	0.35	0.28	0.42%				
30	MC	0.69	0.44	0.46%				
31	MC	0.40	0.17	0.69%				
32	MC	0.70	0.18	0.35%				
33	MC	0.31	0.29	0.44%				
34	MC	0.62	0.35	0.49%				
35	MC	0.72	0.47	0.53%				
36	MC	0.71	0.43	0.67%				
37	MC	0.38	0.33	0.81%		+		
38	MC	0.68	0.34	0.39%				
39	MC	0.71	0.50	1.21%				
40	MC	0.69	0.37	0.69%				
41	MC	0.44	0.44	0.72%				
42	MC	0.61	0.40	0.40%				
43	MC	0.46	0.51	0.56%				
44	MC	0.74	0.42	0.53%				
45	MC	0.36	0.24	1.97%		+		
46	MC	0.56	0.46	0.55%				
47	MC	0.61	0.40	0.72%				
48	MC	0.49	0.35	0.48%				
49	MC	0.89	0.23	1.50%				
50	MC	0.70	0.45	1.60%				

* 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 © 2011 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		2		
	4		2		
	5	1	4		
	6	1	2		
	7	1	3		
	8	2	5		
	10		3	1	
MA	3			2	3
	4	2			1
	5		2		
	6	1	1	1	1
	7	1	2	1	
	8	1	3	2	4
	10		2	4	
LA	4	1	3		
	8	1	1		
	10		4		1
SS	4		1		
	8		1		
	10	1	3		1
SC	4	2	3		
	8			1	
	10	1	3		

Copyright © 2011 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	58810	39.09	0.69	11.35	-0.68	-0.45	0	57	57	0.93	2.97
	4	59903	38.81	0.65	11.07	-0.61	-0.42	0	60	60	0.92	3.18
	5	60406	39.40	0.66	10.69	-0.62	-0.37	0	60	60	0.91	3.16
	6	59935	38.67	0.64	9.95	-0.60	-0.27	0	60	60	0.90	3.19
	7	60238	39.01	0.65	9.86	-0.62	-0.19	0	60	60	0.90	3.19
	8	60243	41.59	0.69	9.98	-0.75	0.01	0	60	60	0.90	3.15
	10	64632	37.09	0.66	10.55	-0.58	-0.38	0	56	56	0.92	3.03
Mathematics	3	59160	40.13	0.70	9.40	-0.62	-0.05	1	57	57	0.90	2.98
	4	60100	40.92	0.72	8.84	-0.68	0.07	1	57	57	0.89	2.91
	5	60520	42.93	0.69	10.57	-0.50	-0.37	0	62	62	0.91	3.24
	6	60105	41.14	0.66	11.27	-0.42	-0.59	1	62	62	0.92	3.23
	7	60372	41.08	0.66	11.09	-0.40	-0.56	1	62	62	0.92	3.20
	8	60352	36.02	0.58	11.57	-0.10	-0.70	0	62	62	0.92	3.37
	10	64645	34.05	0.59	12.91	-0.06	-1.07	0	58	58	0.94	3.28
Language Arts	4	59838	19.66	0.66	5.29	-0.37	-0.55	0	30	30	0.82	2.25
	8	60120	21.08	0.73	5.47	-0.68	-0.21	0	29	29	0.85	2.11
	10	64195	24.84	0.64	6.56	-0.63	-0.07	0	39	39	0.84	2.59
Social Studies	4	59940	29.13	0.77	6.30	-0.96	0.54	2	38	38	0.87	2.26
	8	60131	28.06	0.70	7.15	-0.52	-0.41	2	40	40	0.87	2.53
	10	64251	32.06	0.65	9.06	-0.43	-0.56	0	49	49	0.90	2.86
Science	4	60007	29.01	0.73	6.33	-0.67	-0.04	3	40	40	0.85	2.45
	8	60265	29.88	0.75	6.53	-0.76	0.11	1	40	40	0.86	2.42
	10	64375	31.03	0.62	8.86	-0.30	-0.60	0	50	50	0.88	3.01

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-26
Raw Score Descriptive Statistics based on Calibration Sample

Content	Grade	N Count	Mean	Test Difficulty	SD	Skewness	Kurtosis	Min Obtained	Max Obtained	Max Possible	Alpha	SEM
Reading	3	4627	39.29	0.69	11.55	-0.67	-0.48	2	57	57	0.93	2.95
	4	5282	38.80	0.65	11.58	-0.58	-0.53	3	60	60	0.93	3.17
	5	4751	39.83	0.66	10.79	-0.62	-0.38	4	59	60	0.92	3.13
	6	5908	38.03	0.63	10.37	-0.48	-0.48	8	58	60	0.90	3.23
	7	5272	38.63	0.64	10.39	-0.58	-0.28	5	60	60	0.90	3.21
	8	5411	41.23	0.69	10.61	-0.69	-0.22	0	60	60	0.91	3.17
	10	5758	37.13	0.66	11.05	-0.54	-0.53	4	56	56	0.92	3.04
Mathematics	3	4646	40.22	0.71	9.54	-0.59	-0.09	4	57	57	0.90	2.98
	4	5294	41.14	0.72	9.03	-0.66	-0.03	6	57	57	0.90	2.89
	5	4757	43.87	0.71	10.48	-0.55	-0.36	10	62	62	0.91	3.22
	6	5942	40.75	0.66	11.53	-0.37	-0.65	6	62	62	0.92	3.24
	7	5278	40.87	0.66	11.39	-0.37	-0.61	1	62	62	0.92	3.21
	8	5425	35.70	0.58	11.95	-0.07	-0.78	5	62	62	0.92	3.37
	10	5760	33.98	0.59	13.30	0.00	-1.12	2	58	58	0.94	3.28
Language Arts	4	5276	19.72	0.66	5.54	-0.37	-0.66	3	30	30	0.84	2.24
	8	5395	21.02	0.72	5.67	-0.69	-0.27	2	29	29	0.86	2.10
	10	5677	24.66	0.63	6.75	-0.59	-0.17	2	39	39	0.85	2.58
Social Studies	4	5279	29.26	0.77	6.38	-0.90	0.33	3	38	38	0.88	2.24
	8	5399	27.72	0.69	7.49	-0.47	-0.59	2	40	40	0.89	2.54
	10	5667	32.17	0.66	9.57	-0.43	-0.68	3	49	49	0.91	2.84
Science	4	5290	28.85	0.72	6.57	-0.62	-0.19	6	40	40	0.86	2.45
	8	5424	29.59	0.74	6.84	-0.71	-0.09	1	40	40	0.87	2.43
	10	5684	30.70	0.61	9.26	-0.24	-0.72	1	50	50	0.89	3.01

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-27
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	29976	38.02	0.67	11.79	0.93	28832	40.20	0.71	10.76	0.93
	4	30690	37.97	0.63	11.35	0.92	29213	39.70	0.66	10.71	0.91
	5	30901	38.63	0.64	10.97	0.92	29502	40.20	0.67	10.32	0.91
	6	30709	37.89	0.63	10.17	0.90	29225	39.49	0.66	9.64	0.89
	7	30722	38.45	0.64	10.19	0.90	29516	39.59	0.66	9.47	0.89
	8	30758	40.37	0.67	10.34	0.91	29485	42.87	0.71	9.43	0.89
	10	32944	36.22	0.65	10.88	0.92	31686	37.99	0.68	10.12	0.91
Mathematics	3	30182	40.25	0.71	9.46	0.90	28976	40.02	0.70	9.33	0.90
	4	30804	41.13	0.72	8.88	0.89	29296	40.70	0.71	8.80	0.89
	5	30969	43.14	0.70	10.71	0.91	29548	42.72	0.69	10.43	0.90
	6	30813	41.01	0.66	11.48	0.92	29291	41.28	0.67	11.05	0.91
	7	30808	41.18	0.66	11.32	0.92	29564	40.97	0.66	10.85	0.91
	8	30839	36.50	0.59	11.85	0.92	29513	35.51	0.57	11.24	0.91
	10	32942	34.52	0.60	13.11	0.94	31701	33.55	0.58	12.68	0.93
Language Arts	4	30652	19.08	0.64	5.39	0.82	29186	20.27	0.68	5.10	0.81
	8	30686	20.17	0.70	5.71	0.86	29434	22.04	0.76	5.02	0.84
	10	32669	23.73	0.61	6.79	0.85	31524	26.00	0.67	6.10	0.83
Social Studies	4	30717	28.93	0.76	6.43	0.87	29223	29.35	0.77	6.16	0.87
	8	30696	28.18	0.70	7.39	0.88	29434	27.93	0.70	6.90	0.86
	10	32703	32.11	0.66	9.51	0.91	31546	32.00	0.65	8.57	0.89
Science	4	30745	29.17	0.73	6.46	0.86	29262	28.85	0.72	6.19	0.84
	8	30779	30.07	0.75	6.81	0.88	29486	29.68	0.74	6.23	0.85
	10	32786	31.82	0.64	9.30	0.90	31587	30.21	0.60	8.30	0.87

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-28
Raw Score Descriptive Statistics for Reading by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Reading	W	3	42907	41.17	0.72	10.51	0.93
		4	43972	40.83	0.68	10.22	0.91
		5	44875	41.37	0.69	9.81	0.90
		6	44756	40.54	0.68	9.09	0.88
		7	45399	40.69	0.68	9.10	0.88
		8	45769	43.25	0.72	9.14	0.89
		10	50192	38.93	0.70	9.64	0.91
	AA	3	6483	31.23	0.55	11.63	0.92
		4	6405	30.73	0.51	11.43	0.91
		5	6566	31.21	0.52	10.95	0.90
		6	6428	30.96	0.52	10.09	0.89
		7	6432	31.62	0.53	10.18	0.89
		8	6334	33.77	0.56	10.72	0.90
		10	6195	28.08	0.50	10.76	0.91
	H	3	6084	33.70	0.59	11.16	0.92
		4	6083	33.53	0.56	10.81	0.90
		5	5622	34.32	0.57	10.58	0.90
		6	5495	33.72	0.56	9.93	0.89
		7	5213	34.57	0.58	9.89	0.89
		8	4929	37.22	0.62	10.15	0.89
		10	4739	31.33	0.56	10.75	0.91
	A	3	2343	38.30	0.67	11.04	0.93
		4	2460	38.02	0.63	11.21	0.92
		5	2382	38.28	0.64	10.92	0.91
		6	2288	37.04	0.62	10.55	0.90
		7	2193	37.73	0.63	10.22	0.90
		8	2265	40.76	0.68	10.21	0.90
		10	2486	34.96	0.62	11.01	0.92
	AI	3	988	35.29	0.62	11.28	0.92
		4	983	35.86	0.60	10.90	0.91
		5	958	35.58	0.59	10.56	0.90
		6	966	34.94	0.58	9.68	0.88
		7	997	35.99	0.60	9.91	0.89
		8	945	38.78	0.65	9.66	0.88
		10	1015	33.41	0.60	10.23	0.91

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-29
Raw Score Descriptive Statistics for Mathematics by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Mathematics	W	3	42945	41.81	0.73	8.59	0.88
		4	43998	42.49	0.75	8.02	0.87
		5	44865	44.78	0.72	9.81	0.90
		6	44784	43.14	0.70	10.44	0.91
		7	45434	43.07	0.69	10.29	0.91
		8	45776	38.06	0.61	10.88	0.91
		10	50185	36.42	0.63	12.19	0.93
	AA	3	6505	32.44	0.57	10.18	0.91
		4	6413	33.68	0.59	9.71	0.90
		5	6574	34.50	0.56	10.67	0.90
		6	6466	31.67	0.51	11.07	0.91
		7	6443	31.60	0.51	10.62	0.90
		8	6345	25.91	0.42	10.31	0.89
		10	6178	21.94	0.38	10.07	0.89
	H	3	6326	36.75	0.64	9.01	0.88
		4	6217	37.44	0.66	8.69	0.88
		5	5703	38.57	0.62	10.03	0.89
		6	5567	36.25	0.58	10.77	0.90
		7	5274	35.89	0.58	10.52	0.90
		8	4996	30.56	0.49	10.46	0.89
		10	4768	26.18	0.45	11.16	0.91
	A	3	2387	41.23	0.72	9.04	0.89
		4	2487	41.69	0.73	8.74	0.89
		5	2416	43.82	0.71	10.68	0.91
		6	2312	42.67	0.69	11.12	0.92
		7	2215	41.94	0.68	11.16	0.92
		8	2284	36.86	0.59	11.94	0.92
		10	2496	33.61	0.58	13.27	0.94
	AI	3	992	36.95	0.65	9.11	0.89
		4	985	37.91	0.67	8.69	0.88
		5	959	38.34	0.62	10.10	0.89
		6	974	36.54	0.59	10.66	0.90
		7	1002	37.17	0.60	10.57	0.90
		8	950	32.17	0.52	10.87	0.90
		10	1013	28.26	0.49	11.44	0.91

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-30
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	43935	20.49	0.68	5.02	0.81
		8	45715	21.84	0.75	5.18	0.84
		10	50019	25.98	0.67	6.01	0.82
	AA	4	6385	16.05	0.53	5.30	0.79
		8	6285	17.58	0.61	5.79	0.84
		10	6040	18.85	0.48	6.66	0.83
	H	4	6077	17.83	0.59	5.04	0.78
		8	4919	18.91	0.65	5.37	0.82
		10	4667	21.29	0.55	6.47	0.83
	A	4	2458	19.53	0.65	5.32	0.82
		8	2252	21.11	0.73	5.17	0.83
		10	2472	24.26	0.62	6.52	0.84
	AI	4	983	18.05	0.60	5.11	0.79
		8	948	19.21	0.66	5.52	0.83
		10	992	22.21	0.57	6.17	0.81

Table 8-31
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	43956	30.30	0.80	5.63	0.85
		8	45724	29.31	0.73	6.59	0.86
		10	50060	33.59	0.69	8.42	0.89
	AA	4	6392	24.20	0.64	7.24	0.88
		8	6287	21.94	0.55	7.29	0.85
		10	6035	24.31	0.50	8.85	0.88
	H	4	6147	26.45	0.70	6.36	0.85
		8	4930	24.79	0.62	6.92	0.85
		10	4675	27.09	0.55	8.83	0.88
	A	4	2463	28.56	0.75	6.31	0.87
		8	2244	27.77	0.69	6.98	0.87
		10	2479	30.65	0.63	9.10	0.90
	AI	4	982	27.31	0.72	6.54	0.87
		8	944	25.83	0.65	7.19	0.86
		10	997	28.83	0.59	8.59	0.88

Copyright © 2011 by the Wisconsin Department of Public Instruction

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

Content	Race/Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Science	W	4	43959	30.26	0.76	5.74	0.83
		8	45766	31.11	0.78	5.86	0.84
		10	50105	32.72	0.65	8.14	0.87
	AA	4	6396	23.77	0.59	6.67	0.84
		8	6316	23.97	0.60	6.98	0.85
		10	6076	22.38	0.45	8.02	0.84
	H	4	6184	26.10	0.65	6.14	0.82
		8	4958	26.77	0.67	6.52	0.84
		10	4698	25.66	0.51	8.36	0.86
	A	4	2484	28.27	0.71	6.39	0.85
		8	2280	29.25	0.73	6.56	0.86
		10	2489	29.58	0.59	8.97	0.88
	AI	4	984	27.51	0.69	6.19	0.83
		8	944	28.06	0.70	6.76	0.86
		10	1002	28.00	0.56	8.07	0.85

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-33
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	25727	34.58	0.61	11.60	0.93	33083	42.60	0.75	9.80	0.92
	4	25885	34.18	0.57	11.26	0.91	34018	42.34	0.71	9.53	0.90
	5	25467	34.84	0.58	10.84	0.91	34939	42.72	0.71	9.25	0.89
	6	24493	34.11	0.57	10.06	0.89	35442	41.82	0.70	8.55	0.87
	7	24107	34.79	0.58	10.05	0.89	36131	41.82	0.70	8.65	0.87
	8	23072	37.15	0.62	10.34	0.90	37171	44.35	0.74	8.68	0.88
	10	21943	31.99	0.57	10.73	0.91	42689	39.71	0.71	9.44	0.90
Mathematics	3	26032	36.51	0.64	9.61	0.90	33128	42.98	0.75	8.18	0.87
	4	26048	37.41	0.66	9.16	0.89	34052	43.60	0.76	7.57	0.86
	5	25574	38.56	0.62	10.54	0.90	34946	46.14	0.74	9.38	0.89
	6	24621	36.14	0.58	11.16	0.91	35484	44.61	0.72	9.98	0.90
	7	24217	36.06	0.58	10.84	0.90	36155	44.44	0.72	9.94	0.90
	8	23154	30.69	0.50	10.90	0.90	37198	39.34	0.63	10.70	0.90
	10	21947	27.34	0.47	11.76	0.92	42698	37.50	0.65	12.09	0.93
Language Arts	4	25850	17.67	0.59	5.22	0.80	33988	21.18	0.71	4.81	0.80
	8	22979	18.91	0.65	5.61	0.84	37141	22.43	0.77	4.92	0.83
	10	21655	21.63	0.55	6.64	0.83	42540	26.48	0.68	5.87	0.82
Social Studies	4	25938	26.64	0.70	6.67	0.87	34002	31.03	0.82	5.26	0.84
	8	22986	24.84	0.62	7.23	0.86	37145	30.05	0.75	6.34	0.85
	10	21671	27.68	0.56	9.01	0.89	42580	34.29	0.70	8.24	0.89
Science	4	25983	26.48	0.66	6.51	0.84	34024	30.94	0.77	5.45	0.82
	8	23085	27.02	0.68	6.88	0.86	37180	31.66	0.79	5.62	0.83
	10	21759	26.69	0.53	8.74	0.87	42616	33.25	0.66	8.06	0.87

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-34
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	7435	29.06	0.51	12.32	0.93	51375	40.54	0.71	10.43	0.92
	4	8037	28.38	0.47	12.01	0.92	51866	40.43	0.67	9.99	0.90
	5	8003	28.49	0.47	11.17	0.91	52403	41.07	0.68	9.57	0.90
	6	7773	27.86	0.46	10.20	0.89	52162	40.28	0.67	8.85	0.87
	7	7740	28.02	0.47	10.22	0.89	52498	40.63	0.68	8.70	0.87
	8	7496	29.87	0.50	10.40	0.89	52747	43.26	0.72	8.73	0.88
	10	7676	25.59	0.46	10.30	0.90	56956	38.64	0.69	9.58	0.90
Mathematics	3	7515	34.49	0.61	10.28	0.91	51645	40.96	0.72	8.97	0.89
	4	8088	34.24	0.60	10.20	0.91	52012	41.96	0.74	8.14	0.88
	5	8030	33.63	0.54	11.15	0.90	52490	44.36	0.72	9.73	0.89
	6	7841	30.05	0.48	11.28	0.91	52264	42.81	0.69	10.29	0.90
	7	7810	29.56	0.48	10.78	0.90	52562	42.79	0.69	10.07	0.90
	8	7544	24.61	0.40	10.19	0.89	52808	37.65	0.61	10.81	0.90
	10	7661	21.02	0.36	9.78	0.88	56984	35.80	0.62	12.26	0.93
Language Arts	4	8031	16.05	0.54	5.21	0.78	51807	20.22	0.67	5.07	0.81
	8	7486	15.78	0.54	5.61	0.82	52634	21.84	0.75	5.01	0.83
	10	7520	17.89	0.46	6.22	0.80	56675	25.77	0.66	6.03	0.82
Social Studies	4	8057	25.12	0.66	6.96	0.87	51883	29.76	0.78	5.96	0.86
	8	7485	21.64	0.54	7.37	0.86	52646	28.97	0.72	6.64	0.86
	10	7524	23.41	0.48	8.83	0.88	56727	33.21	0.68	8.45	0.89
Science	4	8071	25.54	0.64	6.88	0.85	51936	29.55	0.74	6.07	0.84
	8	7508	24.03	0.60	7.30	0.86	52757	30.72	0.77	5.97	0.84
	10	7563	23.12	0.46	8.48	0.86	56812	32.08	0.64	8.36	0.87

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-35
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	4472	31.51	0.55	10.42	0.90	54338	39.71	0.70	11.19	0.93
	4	4446	30.80	0.51	10.00	0.88	55457	39.46	0.66	10.90	0.92
	5	3796	30.68	0.51	9.40	0.86	56610	39.98	0.67	10.51	0.91
	6	3308	28.83	0.48	8.64	0.84	56627	39.24	0.65	9.72	0.89
	7	3093	30.04	0.50	8.82	0.85	57145	39.49	0.66	9.68	0.89
	8	2968	32.48	0.54	9.00	0.85	57275	42.06	0.70	9.80	0.90
	10	2435	24.73	0.44	8.65	0.85	62197	37.57	0.67	10.32	0.91
Mathematics	3	4777	36.73	0.64	8.89	0.88	54383	40.43	0.71	9.38	0.90
	4	4610	36.64	0.64	8.67	0.87	55490	41.27	0.72	8.77	0.89
	5	3921	36.62	0.59	9.66	0.87	56599	43.37	0.70	10.49	0.91
	6	3404	33.29	0.54	9.95	0.88	56701	41.61	0.67	11.17	0.92
	7	3173	32.77	0.53	9.49	0.87	57199	41.54	0.67	10.99	0.92
	8	3064	27.13	0.44	9.19	0.86	57288	36.50	0.59	11.49	0.91
	10	2496	21.40	0.37	8.47	0.84	62149	34.56	0.60	12.80	0.93
Language Arts	4	4440	16.89	0.56	4.78	0.75	55398	19.89	0.66	5.26	0.82
	8	2952	17.08	0.59	4.75	0.75	57168	21.29	0.73	5.42	0.85
	10	2411	18.11	0.46	5.52	0.75	61784	25.11	0.64	6.46	0.84
Social Studies	4	4524	25.26	0.66	6.15	0.84	55416	29.45	0.77	6.21	0.87
	8	2955	21.88	0.55	5.95	0.77	57176	28.38	0.71	7.07	0.87
	10	2409	22.31	0.46	7.08	0.80	61842	32.44	0.66	8.91	0.90
Science	4	4577	24.93	0.62	5.92	0.79	55430	29.35	0.73	6.25	0.85
	8	3022	24.10	0.60	5.92	0.78	57243	30.19	0.75	6.42	0.86
	10	2439	21.40	0.43	6.51	0.75	61936	31.41	0.63	8.72	0.88

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-36
Scale Score Mean and Standard Deviation for Census and Calibration Sample

Content	Grade	Calibration Districts Mean	Census Mean	Diff = Calibration Districts - Census	Calibration Districts Standard Deviation	Census Standard Deviation	Diff = Calibration Districts - Census
Reading	3	459.28	457.90	1.38	42.65	40.83	1.82
	4	477.91	477.62	0.29	48.67	44.91	3.76
	5	486.52	484.07	2.45	46.50	45.71	0.79
	6	501.86	504.40	-2.54	50.53	47.17	3.36
	7	514.87	516.48	-1.61	50.17	45.85	4.32
	8	531.40	532.78	-1.38	55.29	51.10	4.19
	10	539.84	539.04	0.80	64.65	60.29	4.36
Mathematics	3	433.01	431.95	1.06	45.23	43.49	1.74
	4	471.98	469.93	2.05	45.58	44.14	1.44
	5	502.57	498.20	4.37	48.19	48.55	-0.36
	6	520.87	522.54	-1.67	51.44	49.29	2.15
	7	536.26	537.25	-0.99	46.03	44.04	1.99
	8	546.21	547.32	-1.11	53.23	51.09	2.14
	10	562.42	562.28	0.14	49.87	46.46	3.41
Language Arts	4	296.02	295.23	0.79	31.70	29.93	1.77
	8	397.47	397.55	-0.08	46.43	44.58	1.85
	10	449.23	450.15	-0.92	44.25	42.75	1.50
Social Studies	4	297.55	296.16	1.39	29.34	27.52	1.82
	8	395.69	397.30	-1.61	42.62	39.92	2.70
	10	453.52	452.77	0.75	48.48	44.24	4.24
Science	4	298.66	298.37	0.29	31.79	29.78	2.01
	8	401.46	402.37	-0.91	44.21	42.08	2.13
	10	453.02	454.45	-1.43	47.92	44.66	3.26

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-37
Scale Score Descriptive Statistics based on Calibration Sample

Content	Grade	N Count	Mean	SD	Skewness	Kurtosis	Min	Max	LOSS	HOSS
Reading	3	4627	459.28	42.65	-0.77	3.94	270	640	270	640
	4	5282	477.91	48.67	-0.84	2.49	280	650	280	650
	5	4751	486.52	46.50	-0.90	3.01	290	678	290	690
	6	5908	501.86	50.53	-0.88	2.28	300	648	300	730
	7	5272	514.87	50.17	-0.77	2.23	310	780	310	780
	8	5411	531.40	55.29	-0.42	1.44	330	790	330	790
	10	5758	539.84	64.65	-0.18	0.95	350	820	350	820
Mathematics	3	4646	433.01	45.23	0.13	2.02	220	630	220	630
	4	5294	471.98	45.58	-0.16	1.46	240	650	240	650
	5	4757	502.57	48.19	-0.11	1.06	270	680	270	680
	6	5942	520.87	51.44	-0.30	1.10	310	700	310	700
	7	5278	536.26	46.03	-0.30	1.30	330	710	330	710
	8	5425	546.21	53.23	-0.54	1.25	350	730	350	730
	10	5760	562.42	49.87	-0.12	1.54	410	750	410	750
Language Arts	4	5276	296.02	31.70	-0.23	2.74	140	420	140	420
	8	5395	397.47	46.43	0.11	1.20	250	520	250	520
	10	5677	449.23	44.25	-0.28	1.07	290	630	290	630
Social Studies	4	5279	297.55	29.34	0.89	3.79	170	400	170	400
	8	5399	395.69	42.62	-0.26	1.77	230	530	230	530
	10	5667	453.52	48.48	-0.57	2.00	240	620	240	620
Science	4	5290	298.66	31.79	0.48	2.98	170	440	170	440
	8	5424	401.46	44.21	0.25	2.06	230	560	230	560
	10	5684	453.02	47.92	-0.67	2.36	240	610	240	610

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-38
Scale Score Descriptive Statistics based on Census Data

Content	Grade	N Count	Mean	SD	Skewness	Kurtosis	Min	Max	LOSS	HOSS
Reading	3	58810	457.90	40.83	-1.04	4.38	270	640	270	640
	4	59903	477.62	44.91	-0.91	2.78	280	650	280	650
	5	60406	484.07	45.71	-1.00	3.26	290	690	290	690
	6	59935	504.40	47.17	-0.90	2.48	300	730	300	730
	7	60238	516.48	45.85	-0.73	2.03	310	780	310	780
	8	60243	532.78	51.10	-0.48	1.75	330	790	330	790
	10	64632	539.04	60.29	-0.30	1.02	350	820	350	820
Mathematics	3	59160	431.95	43.49	-0.01	1.93	220	630	220	630
	4	60100	469.93	44.14	-0.28	1.83	240	650	240	650
	5	60520	498.20	48.55	-0.26	1.59	270	680	270	680
	6	60105	522.54	49.29	-0.30	1.07	310	700	310	700
	7	60372	537.25	44.04	-0.25	1.21	330	710	330	710
	8	60352	547.32	51.09	-0.62	1.59	350	730	350	730
	10	64645	562.28	46.46	-0.34	1.44	410	750	410	750
Language Arts	4	59838	295.23	29.93	-0.28	3.65	140	420	140	420
	8	60120	397.55	44.58	0.17	1.42	250	520	250	520
	10	64195	450.15	42.75	-0.24	1.06	290	630	290	630
Social Studies	4	59940	296.16	27.52	0.62	4.37	170	400	170	400
	8	60131	397.30	39.92	-0.19	1.93	230	530	230	530
	10	64251	452.77	44.24	-0.56	2.39	240	620	240	620
Science	4	60007	298.37	29.78	0.34	3.22	170	440	170	440
	8	60265	402.37	42.08	0.32	2.61	230	560	230	560
	10	64375	454.45	44.66	-0.68	2.42	240	610	240	610

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-39
Scale Score Descriptive Statistics by Gender

Content	Grade	Male					Female				
		N Count	Mean	SD	Min	Max	N Count	Mean	SD	Min	Max
Reading	3	29976	454.03	42.99	270	640	28832	461.92	38.05	270	640
	4	30690	474.20	46.64	280	650	29213	481.21	42.72	280	650
	5	30901	480.80	47.88	290	690	29502	487.50	43.06	290	690
	6	30709	500.69	48.85	300	695	29225	508.29	45.01	300	730
	7	30722	513.84	48.18	310	745	29516	519.23	43.13	310	780
	8	30758	527.11	52.74	330	790	29485	538.69	48.64	330	790
	10	32944	534.17	62.33	350	820	31686	544.10	57.66	350	820
Mathematics	3	30182	432.65	44.04	220	630	28976	431.23	42.90	220	630
	4	30804	471.43	44.74	240	650	29296	468.34	43.44	240	650
	5	30969	499.59	50.14	270	680	29548	496.76	46.79	270	680
	6	30813	522.20	50.67	310	700	29291	522.91	47.79	310	700
	7	30808	537.97	45.53	330	710	29564	536.50	42.42	330	710
	8	30839	549.03	53.21	350	730	29513	545.53	48.71	350	730
	10	32942	563.92	48.03	410	750	31701	560.57	44.70	410	750
Language Arts	4	30652	291.85	30.76	140	420	29186	298.78	28.61	140	420
	8	30686	390.26	45.06	250	520	29434	405.15	42.79	250	520
	10	32669	443.29	43.62	290	630	31524	457.27	40.62	290	630
Social Studies	4	30717	295.26	27.91	170	400	29223	297.11	27.07	170	400
	8	30696	397.96	42.19	230	530	29434	396.62	37.40	230	530
	10	32703	452.96	47.21	240	620	31546	452.57	40.92	240	620
Science	4	30745	298.78	31.18	170	440	29262	297.92	28.23	170	440
	8	30779	404.06	45.45	230	560	29486	400.60	38.17	230	560
	10	32786	458.01	48.18	240	610	31587	450.75	40.36	240	610

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-40
Scale Score Descriptive Statistics for Reading by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
Reading	W	3	42907	464.86	37.79	270	640
		4	43972	485.41	41.31	280	650
		5	44875	492.07	41.55	290	690
		6	44756	513.11	42.58	300	730
		7	45399	524.27	42.03	310	780
		8	45769	541.05	47.43	330	790
		10	50192	549.24	55.55	350	820
	AA	3	6483	430.81	45.44	270	584
		4	6405	445.64	49.47	280	638
		5	6566	450.11	51.39	290	605
		6	6428	468.23	51.83	300	617
		7	6432	482.24	49.63	310	659
		8	6334	493.68	53.61	330	711
		10	6195	488.88	61.14	350	736
	H	3	6084	440.49	39.36	270	584
		4	6083	457.38	43.65	280	605
		5	5622	463.89	45.10	290	690
		6	5495	481.79	46.78	300	648
		7	5213	495.87	45.34	310	659
		8	4929	510.44	49.69	330	776
		10	4739	506.78	60.02	350	820
	A	3	2343	456.24	38.42	270	584
		4	2460	475.94	44.30	280	650
		5	2382	480.62	46.45	290	690
		6	2288	497.17	50.98	300	660
		7	2193	510.68	47.79	310	674
		8	2265	529.20	53.23	330	776
		10	2486	527.92	64.00	350	820
	AI	3	988	444.51	40.72	270	584
		4	983	466.46	43.20	280	607
		5	958	469.39	44.10	290	621
		6	966	487.35	45.29	300	611
		7	997	503.48	46.06	310	650
		8	945	519.71	46.84	330	721
		10	1015	518.66	55.71	350	704

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-41
Scale Score Descriptive Statistics for Mathematics by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
Mathematics	W	3	42945	439.45	40.91	220	630
		4	43998	477.44	40.89	240	650
		5	44865	506.28	45.41	270	680
		6	44784	531.12	45.67	310	700
		7	45434	544.87	40.98	330	710
		8	45776	556.50	45.73	350	730
		10	50185	570.82	42.20	410	750
	AA	3	6505	398.21	44.55	220	630
		4	6413	435.46	47.55	240	650
		5	6574	460.82	49.65	270	680
		6	6466	481.61	50.27	310	678
		7	6443	500.50	43.72	330	677
		8	6345	499.91	56.55	350	671
		10	6178	516.82	46.10	410	750
	H	3	6326	416.01	38.00	220	630
		4	6217	453.02	40.54	240	650
		5	5703	479.03	43.59	270	680
		6	5567	501.53	45.06	310	700
		7	5274	517.41	39.67	330	710
		8	4996	524.32	48.03	350	730
		10	4768	535.26	43.22	410	750
	A	3	2387	437.21	44.83	220	630
		4	2487	473.90	44.78	240	650
		5	2416	503.25	52.26	270	680
		6	2312	529.79	50.60	310	700
		7	2215	541.92	46.81	330	710
		8	2284	551.77	54.17	350	730
		10	2496	562.15	49.53	410	750
	AI	3	992	417.64	38.56	266	551
		4	985	455.36	41.46	240	593
		5	959	478.16	43.92	270	638
		6	974	502.90	45.22	321	645
		7	1002	522.14	40.16	339	667
		8	950	531.65	48.75	350	730
		10	1013	543.91	41.14	410	750

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-42
Scale Score Descriptive Statistics for Language Arts by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
Language Arts	W	4	43935	299.72	28.28	140	420
		8	45715	403.38	43.58	250	520
		10	50019	457.19	39.92	290	630
	AA	4	6385	275.16	32.17	140	420
		8	6285	370.52	42.79	250	520
		10	6040	413.09	42.20	290	564
	H	4	6077	285.13	27.73	140	420
		8	4919	380.05	39.37	250	520
		10	4667	428.22	39.86	290	597
	A	4	2458	295.52	30.43	140	420
		8	2252	398.73	42.96	250	520
		10	2472	446.55	43.56	290	630
	AI	4	983	286.65	28.04	140	386
		8	948	383.62	40.53	250	520
		10	992	433.31	37.73	290	598

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-43
Scale Score Descriptive Statistics for Social Studies by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
Social Studies	W	4	43956	300.71	26.53	170	400
		8	45724	404.00	37.10	230	530
		10	50060	460.09	40.45	240	620
	AA	4	6392	277.27	27.19	170	400
		8	6287	364.21	41.62	230	530
		10	6035	415.03	48.42	240	584
	H	4	6147	285.24	22.95	170	400
		8	4930	380.05	36.82	230	530
		10	4675	429.87	42.96	240	620
	A	4	2463	294.43	27.53	170	400
		8	2244	396.41	39.24	230	530
		10	2479	446.06	46.24	240	620
	AI	4	982	288.38	25.45	170	400
		8	944	385.52	37.60	230	530
		10	997	438.03	41.03	240	620

Table 8-44
Scale Score Descriptive Statistics for Science by Race/Ethnicity

Content	Race/Ethnicity	Grade	N Count	Mean	SD	Min	Max
Science	W	4	43959	303.83	28.19	170	440
		8	45766	409.43	39.84	230	560
		10	50105	462.73	40.02	240	610
	AA	4	6396	275.68	28.53	170	440
		8	6316	368.31	39.70	230	560
		10	6076	411.19	47.83	240	610
	H	4	6184	285.14	25.95	170	440
		8	4958	383.79	36.92	230	560
		10	4698	428.56	44.18	240	600
	A	4	2484	295.75	31.05	170	440
		8	2280	399.91	43.85	230	560
		10	2489	447.66	46.32	240	610
	AI	4	984	291.62	28.02	170	440
		8	944	391.53	40.47	230	560
		10	1002	440.78	38.94	240	604

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-45
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	25727	442.53	41.98	270	640	33083	469.84	35.60	270	640
	4	25885	459.59	46.11	280	650	34018	491.33	38.70	280	650
	5	25467	465.46	47.06	290	661	34939	497.63	39.53	290	690
	6	24493	483.33	48.36	300	634	35442	518.96	40.34	300	730
	7	24107	497.21	46.78	310	714	36131	529.34	40.41	310	780
	8	23072	510.60	51.00	330	790	37171	546.54	46.08	330	790
	10	21943	510.54	59.37	350	820	42689	553.69	55.32	350	820
Mathematics	3	26032	415.43	41.78	220	630	33128	444.93	40.29	220	630
	4	26048	452.73	43.73	240	650	34052	483.08	39.71	240	650
	5	25574	478.53	46.82	270	680	34946	512.60	44.58	270	680
	6	24621	500.92	47.74	310	679	35484	537.55	44.55	310	700
	7	24217	517.75	41.96	330	710	36155	550.31	40.42	330	710
	8	23154	524.11	51.67	350	730	37198	561.76	45.05	350	730
	10	21947	538.90	45.28	410	750	42698	574.29	42.29	410	750
Language Arts	4	25850	284.41	29.35	140	420	33988	303.46	27.67	140	420
	8	22979	380.46	41.77	250	520	37141	408.12	42.96	250	520
	10	21655	429.90	41.26	290	630	42540	460.46	39.70	290	630
Social Studies	4	25938	286.03	25.15	170	400	34002	303.89	26.72	170	400
	8	22986	380.01	38.67	230	530	37145	408.00	36.81	230	530
	10	21671	432.16	44.73	240	620	42580	463.26	40.10	240	620
Science	4	25983	286.94	28.12	170	440	34024	307.09	28.01	170	440
	8	23085	385.22	40.12	230	560	37180	413.02	39.69	230	560
	10	21759	433.26	45.99	240	610	42616	465.26	39.84	240	610

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-46
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	7435	421.89	51.85	270	584	51375	463.11	36.12	270	640
	4	8037	434.73	56.27	280	650	51866	484.26	38.85	280	650
	5	8003	437.00	57.53	290	622	52403	491.26	38.90	290	690
	6	7773	452.40	56.75	300	623	52162	512.15	40.17	300	730
	7	7740	466.14	52.82	310	624	52498	523.91	39.65	310	780
	8	7496	475.64	55.62	330	669	52747	540.90	44.87	330	790
	10	7676	475.31	60.14	350	706	56956	547.63	54.92	350	820
Mathematics	3	7515	406.81	45.75	220	630	51645	435.61	41.92	220	630
	4	8088	437.20	50.67	240	650	52012	475.02	40.74	240	650
	5	8030	456.08	53.23	270	656	52490	504.65	44.40	270	680
	6	7841	474.24	52.57	310	678	52264	529.79	44.46	310	700
	7	7810	491.94	45.87	330	677	52562	543.98	39.56	330	710
	8	7544	493.45	56.77	350	730	52808	555.01	45.27	350	730
	10	7661	511.74	46.49	410	681	56984	569.07	42.05	410	750
Language Arts	4	8031	275.24	31.89	140	420	51807	298.33	28.39	140	420
	8	7486	358.29	42.24	250	520	52634	403.13	42.03	250	520
	10	7520	407.71	38.32	290	605	56675	455.79	40.06	290	630
Social Studies	4	8057	280.94	26.44	170	400	51883	298.52	26.92	170	400
	8	7485	362.49	42.97	230	530	52646	402.25	36.90	230	530
	10	7524	410.95	48.29	240	620	56727	458.32	40.55	240	620
Science	4	8071	283.45	30.25	170	440	51936	300.68	29.03	170	440
	8	7508	368.15	44.08	230	560	52757	407.24	39.45	230	560
	10	7563	414.19	49.36	240	610	56812	459.81	41.13	240	610

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-47
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	4472	433.95	37.28	270	640	54338	459.87	40.49	270	640
	4	4446	447.70	40.89	280	650	55457	480.01	44.35	280	650
	5	3796	449.78	42.10	290	601	56610	486.37	45.02	290	690
	6	3308	459.54	44.28	300	590	56627	507.02	46.00	300	730
	7	3093	475.73	42.35	310	638	57145	518.69	45.00	310	780
	8	2968	487.96	43.99	330	659	57275	535.10	50.37	330	790
	10	2435	471.19	51.55	350	612	62197	541.70	59.04	350	820
Mathematics	3	4777	415.37	37.46	220	630	54383	433.41	43.68	220	630
	4	4610	448.89	40.23	240	650	55490	471.67	44.00	240	650
	5	3921	470.67	42.61	270	656	56599	500.11	48.36	270	680
	6	3404	489.11	42.89	310	679	56701	524.55	48.93	310	700
	7	3173	505.81	36.23	330	635	57199	539.00	43.78	330	710
	8	3064	508.82	46.91	350	663	57288	549.37	50.49	350	730
	10	2496	517.08	39.92	410	634	62149	564.09	45.78	410	750
Language Arts	4	4440	280.34	26.72	140	390	55398	296.42	29.86	140	420
	8	2952	367.44	32.20	250	520	57168	399.10	44.58	250	520
	10	2411	409.10	34.18	290	521	61784	451.76	42.25	290	630
Social Studies	4	4524	281.01	20.86	170	400	55416	297.40	27.63	170	400
	8	2955	365.33	32.21	230	530	57176	398.95	39.59	230	530
	10	2409	407.40	40.34	240	540	61842	454.54	43.43	240	620
Science	4	4577	280.31	24.96	170	430	55430	299.86	29.66	170	440
	8	3022	368.93	32.56	230	505	57243	404.13	41.79	230	560
	10	2439	407.04	40.41	240	527	61936	456.32	43.79	240	610

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-48
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	394	430	466	396	440	489	401	444	497	418	457	514	434	467	523	445	480	539	456	503	555
Mathematics	392	407	452	421	438	484	445	463	505	464	485	532	480	504	555	483	513	573	516	541	595
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 © 2011 by the Wisconsin Department of Public Instruction

Table 8-49
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	2930	4.98	3.49	6.42	3.21	14.02	7.86	3.88	7.69	4.66	8.86	21.14	2.64	8.75	2.06
	B	8291	14.10	12.03	16.08	10.49	27.24	24.19	15.41	19.03	12.91	28.53	30.53	11.72	21.70	8.18
	P	20945	35.62	35.49	35.74	33.37	39.18	44.08	41.74	42.81	34.60	47.92	31.76	36.17	40.94	31.48
	A	26644	45.31	48.99	41.77	52.92	19.56	23.87	38.97	30.47	47.82	14.69	16.57	49.46	28.61	58.29
Total		58810	100	28832	29976	42907	6483	6084	2343	988	54338	4472	7435	51375	25727	33083
4	M	2358	3.94	2.86	4.96	2.43	12.19	6.21	3.41	4.78	3.62	7.83	18.99	1.60	7.15	1.49
	B	7413	12.38	11.13	13.56	8.57	27.31	22.42	14.76	17.09	11.10	28.25	31.37	9.43	20.29	6.35
	P	24376	40.69	40.78	40.61	38.91	42.72	49.35	43.17	47.30	39.79	51.89	34.61	41.63	46.61	36.19
	A	25756	43.00	45.23	40.87	50.09	17.78	22.01	38.66	30.82	45.48	12.03	15.03	47.33	25.95	55.97
Total		59903	100	29213	30690	43972	6405	6083	2460	983	55457	4446	8037	51866	25885	34018
5	M	2159	3.57	2.44	4.66	2.07	11.39	6.31	3.36	4.70	3.25	8.46	18.57	1.28	6.57	1.39
	B	6886	11.40	10.28	12.46	7.90	26.35	19.82	13.52	18.06	10.35	27.05	30.70	8.45	18.86	5.96
	P	26454	43.79	44.18	43.42	41.70	47.82	52.60	48.24	51.88	42.96	56.22	39.31	44.48	50.74	38.73
	A	24907	41.23	43.10	39.46	48.33	14.44	21.27	34.89	25.37	43.44	8.27	11.42	45.79	23.83	53.91
Total		60406	100	29502	30901	44875	6566	5622	2382	958	56610	3796	8003	52403	25467	34939

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-49 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	2419	4.04	2.87	5.15	2.25	13.27	7.06	5.11	5.59	3.57	11.97	20.71	1.55	7.66	1.53
	B	5438	9.07	8.22	9.88	5.94	20.82	18.54	12.06	15.22	7.92	28.78	27.74	6.29	16.09	4.22
	P	24875	41.50	41.00	41.99	38.91	49.18	50.77	44.62	50.62	40.86	52.60	40.36	41.67	49.88	35.71
	A	27203	45.39	47.91	42.98	52.90	16.74	23.62	38.20	28.57	47.65	6.65	11.19	50.48	26.36	58.54
Total		59935	100	29225	30709	44756	6428	5495	2288	966	56627	3308	7773	52162	24493	35442
7	M	2448	4.06	2.86	5.22	2.34	12.83	7.62	4.92	5.62	3.60	12.61	21.30	1.52	7.63	1.68
	B	5182	8.60	7.64	9.53	6.11	19.79	14.92	10.90	11.74	7.84	22.70	27.34	5.84	14.76	4.49
	P	23627	39.22	40.20	38.29	36.33	47.99	49.88	43.87	48.75	38.36	55.25	38.55	39.32	47.42	33.75
	A	28981	48.11	49.31	46.96	55.23	19.39	27.58	40.31	33.90	50.20	9.44	12.80	53.32	30.19	60.07
Total		60238	100	29516	30722	45399	6432	5213	2193	997	57145	3093	7740	52498	24107	36131
8	M	2752	4.57	2.98	6.09	2.75	14.90	8.12	4.50	4.87	4.14	12.90	24.25	1.77	8.70	2.00
	B	4913	8.16	6.71	9.55	5.71	19.80	14.04	10.20	12.80	7.43	22.24	25.76	5.65	14.09	4.47
	P	23928	39.72	39.20	40.22	37.22	47.22	50.40	43.00	46.88	38.88	56.00	38.78	39.85	48.00	34.58
	A	28650	47.56	51.12	44.14	54.31	18.08	27.45	42.30	35.45	49.56	8.86	11.21	52.72	29.21	58.94
Total		60243	100	29485	30758	45769	6334	4929	2265	945	57275	2968	7496	52747	23072	37171

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-49 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	5581	8.64	6.55	10.65	5.20	27.55	18.02	11.67	11.72	7.70	32.44	36.69	4.85	16.70	4.49
	B	10155	15.71	14.24	17.13	12.51	29.91	26.31	21.48	23.84	14.79	39.30	31.25	13.62	24.86	11.01
	P	21740	33.64	35.31	32.03	34.05	28.96	34.31	34.07	37.44	33.96	25.26	22.72	35.11	35.27	32.80
	A	27156	42.02	43.90	40.20	48.24	13.58	21.35	32.78	27.00	43.54	3.00	9.34	46.42	23.16	51.71
Total		64632	100	31686	32944	50192	6195	4739	2486	1015	62197	2435	7676	56956	21943	42689

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-50
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	9409	15.90	15.82	15.98	10.72	41.37	25.09	12.15	23.89	15.14	24.58	35.26	13.09	26.05	7.93
	B	5857	9.90	10.17	9.64	8.56	13.96	14.21	9.76	13.91	9.53	14.09	13.80	9.33	13.30	7.23
	P	25110	42.44	42.98	41.93	43.24	34.99	44.29	43.40	42.54	42.12	46.12	36.35	43.33	42.93	42.06
	A	18784	31.75	31.03	32.45	37.47	9.68	16.41	34.69	19.66	33.20	15.22	14.58	34.25	17.72	42.78
Total		59160	100	28976	30182	42945	6505	6326	2387	992	54383	4777	7515	51645	26032	33128
4	M	6927	11.53	11.91	11.16	7.20	33.78	19.11	10.21	15.13	10.69	21.56	34.03	8.03	20.02	5.03
	B	5342	8.89	9.44	8.36	7.12	15.48	13.48	9.85	13.50	8.35	15.42	14.12	8.08	12.88	5.84
	P	24856	41.36	42.17	40.59	41.35	36.60	46.24	38.92	48.22	41.05	45.12	34.92	42.36	44.00	39.33
	A	22975	38.23	36.48	39.89	44.33	14.14	21.17	41.01	23.15	39.92	17.90	16.94	41.54	23.10	49.80
Total		60100	100	29296	30804	43998	6413	6217	2487	985	55490	4610	8088	52012	26048	34052
5	M	7097	11.73	11.61	11.83	7.74	32.87	18.43	9.64	18.46	10.99	22.42	38.41	7.65	20.15	5.56
	B	5224	8.63	9.17	8.12	6.86	15.07	14.33	8.44	13.87	8.11	16.09	14.78	7.69	12.83	5.56
	P	20463	33.81	35.62	32.09	32.64	34.94	40.19	33.86	43.07	33.21	42.46	30.07	34.38	38.49	30.39
	A	27736	45.83	43.60	47.96	52.76	17.11	27.06	48.05	24.61	47.69	19.03	16.74	50.28	28.53	58.49
Total		60520	100	29548	30969	44865	6574	5703	2416	959	56599	3921	8030	52490	25574	34946

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-50 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	6400	10.65	9.77	11.48	6.49	33.00	18.12	8.17	16.43	9.87	23.62	40.33	6.20	19.47	4.53
	B	5506	9.16	9.25	9.07	7.26	16.49	15.39	8.00	14.99	8.56	19.18	17.32	7.94	14.16	5.69
	P	21910	36.45	37.18	35.76	35.83	35.66	41.69	35.47	42.81	36.02	43.60	29.35	37.52	40.46	33.67
	A	26289	43.74	43.80	43.68	50.42	14.85	24.79	48.36	25.77	45.55	13.60	13.01	48.35	25.91	56.11
Total		60105	100	29291	30813	44784	6466	5567	2312	974	56701	3404	7841	52264	24621	35484
7	M	5314	8.80	8.48	9.11	5.27	28.01	15.51	7.27	13.47	8.12	21.08	37.30	4.57	16.27	3.80
	B	7032	11.65	11.95	11.36	8.86	23.79	19.59	12.19	17.07	10.90	25.06	23.51	9.89	18.34	7.16
	P	27095	44.88	46.47	43.36	45.26	38.68	48.37	44.60	49.70	44.77	46.80	31.23	46.91	47.50	43.13
	A	20931	34.67	33.10	36.17	40.61	9.51	16.53	35.94	19.76	36.20	7.06	7.96	38.64	17.89	45.91
Total		60372	100	29564	30808	45434	6443	5274	2215	1002	57199	3173	7810	52562	24217	36155
8	M	5589	9.26	9.05	9.47	5.27	32.37	15.97	8.32	14.00	8.53	23.01	37.41	5.24	17.45	4.16
	B	7223	11.97	12.32	11.63	9.39	22.88	20.64	12.00	17.68	11.20	26.37	24.63	10.16	18.99	7.59
	P	28685	47.53	49.61	45.54	48.79	37.38	49.26	46.15	49.05	47.66	45.04	31.87	49.77	48.26	47.07
	A	18855	31.24	29.02	33.37	36.55	7.36	14.13	33.54	19.26	32.61	5.58	6.10	34.83	15.29	41.17
Total		60352	100	29513	30839	45776	6345	4996	2284	950	57288	3064	7544	52808	23154	37198

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-50 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	9265	14.33	14.47	14.20	8.81	45.40	30.05	15.26	21.92	13.11	44.75	51.76	9.30	27.16	7.74
	B	8984	13.90	14.48	13.34	11.58	23.62	22.57	16.63	22.01	13.36	27.36	21.62	12.86	21.25	10.12
	P	31223	48.30	49.53	47.11	51.85	27.63	40.25	43.75	47.38	49.17	26.64	23.53	51.63	42.47	51.30
	A	15173	23.47	21.52	25.35	27.76	3.35	7.13	24.36	8.69	24.36	1.24	3.09	26.21	9.13	30.84
Total		64645	100	31701	32942	50185	6178	4768	2496	1013	62149	2496	7661	56984	21947	42698

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-51
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	3188	5.33	3.66	6.91	3.39	15.76	8.29	4.64	7.73	4.97	9.84	16.24	3.64	9.18	2.39
	B	10807	18.06	15.62	20.38	14.58	33.69	25.26	19.57	23.81	17.13	29.66	32.64	15.80	26.66	11.52
	P	26724	44.66	44.71	44.61	44.92	38.87	48.17	44.34	49.64	44.30	49.21	39.95	45.39	46.03	43.62
	A	19119	31.95	36.00	28.09	37.11	11.68	18.28	31.45	18.82	33.61	11.28	11.18	35.17	18.12	42.47
Total		59838	100	29186	30652	43935	6385	6077	2458	983	55398	4440	8031	51807	25850	33988
8	M	9022	15.01	9.96	19.84	11.56	33.22	23.42	12.57	22.57	14.17	31.30	46.51	10.53	24.78	8.96
	B	13564	22.56	20.33	24.70	20.30	29.90	31.19	25.75	30.59	21.68	39.63	29.87	21.52	29.41	18.33
	P	20379	33.90	35.89	31.98	35.08	26.73	31.92	35.70	30.17	34.32	25.78	17.51	36.23	30.79	35.82
	A	17155	28.54	33.81	23.48	33.05	10.15	13.48	25.98	16.67	29.84	3.29	6.10	31.72	15.02	36.90
Total		60120	100	29434	30686	45715	6285	4919	2252	948	57168	2952	7486	52634	22979	37141
10	M	5770	8.99	5.77	12.09	5.78	28.77	17.27	8.29	12.90	8.28	27.04	33.54	5.73	17.02	4.90
	B	11559	18.01	14.95	20.95	14.43	34.62	29.61	23.18	29.44	17.04	42.72	36.93	15.50	28.37	12.73
	P	33926	52.85	55.11	50.66	56.12	32.48	46.05	50.44	50.00	53.75	29.78	27.38	56.23	46.62	56.02
	A	12940	20.16	24.16	16.30	23.67	4.12	7.07	18.08	7.66	20.93	0.46	2.15	22.55	7.99	26.35
Total		64195	100	31524	32669	50019	6040	4667	2472	992	61784	2411	7520	56675	21655	42540

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-52
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	1189	1.98	1.64	2.31	0.97	8.20	2.72	1.75	2.75	1.88	3.25	5.62	1.42	3.67	0.70
	B	3471	5.79	5.31	6.25	3.62	16.19	9.81	6.01	9.37	5.34	11.34	12.78	4.70	10.00	2.58
	P	15859	26.46	25.89	26.99	22.11	40.36	40.05	30.86	34.32	24.83	46.42	41.79	24.08	36.67	18.67
	A	39421	65.77	67.15	64.45	73.29	35.25	47.42	61.39	53.56	67.95	38.99	39.80	69.80	49.66	78.05
Total		59940	100	29223	30717	43956	6392	6147	2463	982	55416	4524	8057	51883	25938	34002
8	M	2856	4.75	4.07	5.40	2.59	17.74	8.22	3.79	7.10	4.38	11.98	19.76	2.62	9.24	1.97
	B	7576	12.60	12.50	12.70	9.21	28.84	21.14	14.39	19.49	11.61	31.68	29.00	10.27	20.96	7.42
	P	22678	37.71	40.11	35.42	36.46	38.14	46.11	41.93	41.74	37.19	47.85	36.65	37.87	43.11	34.37
	A	27021	44.94	43.32	46.49	51.74	15.29	24.54	39.88	31.67	46.82	8.49	14.59	49.25	26.68	56.24
Total		60131	100	29434	30696	45724	6287	4930	2244	944	57176	2955	7485	52646	22986	37145
10	M	8651	13.46	11.50	15.36	8.92	38.89	26.76	15.69	19.66	12.29	43.63	44.25	9.38	25.03	7.58
	B	3760	5.85	5.92	5.79	4.81	10.59	9.63	7.38	8.02	5.55	13.62	12.15	5.02	9.26	4.12
	P	18890	29.40	32.31	26.59	28.15	31.71	35.64	34.29	36.71	29.16	35.66	27.72	29.62	34.70	26.70
	A	32950	51.28	50.27	52.26	58.12	18.81	27.98	42.64	35.61	53.00	7.10	15.88	55.98	31.01	61.60
Total		64251	100	31546	32703	50060	6035	4675	2479	997	61842	2409	7524	56727	21671	42580

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-53
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	2587	4.31	3.96	4.64	2.23	15.37	7.50	4.47	4.98	3.90	9.26	10.30	3.38	7.83	1.62
	B	10623	17.70	17.64	17.77	13.28	35.94	28.28	19.81	24.90	16.51	32.18	30.01	15.79	27.04	10.57
	P	34940	58.23	59.57	56.95	60.29	44.37	57.68	58.41	58.94	58.45	55.54	51.25	59.31	55.90	60.00
	A	11857	19.76	18.83	20.64	24.20	4.32	6.53	17.31	11.18	21.14	3.02	8.45	21.52	9.23	27.80
Total		60007	100	29262	30745	43959	6396	6184	2484	984	55430	4577	8071	51936	25983	34024
8	M	4896	8.12	7.30	8.91	4.74	27.39	14.26	7.72	12.08	7.41	21.74	29.02	5.15	15.16	3.75
	B	8485	14.08	14.51	13.67	10.76	28.01	24.36	17.81	18.75	13.08	33.06	26.16	12.36	22.07	9.12
	P	27677	45.93	49.38	42.61	47.00	36.83	46.69	47.19	47.46	46.17	41.36	35.00	47.48	45.35	46.28
	A	19207	31.87	28.80	34.81	37.50	7.77	14.68	27.28	21.72	33.35	3.84	9.82	35.01	17.41	40.85
Total		60265	100	29486	30779	45766	6316	4958	2280	944	57243	3022	7508	52757	23085	37180
10	M	9322	14.48	14.03	14.91	8.82	46.36	30.80	17.76	19.66	13.12	49.04	45.25	10.39	27.63	7.77
	B	6638	10.31	11.78	8.90	8.35	17.99	17.50	14.66	17.27	9.81	23.04	17.20	9.39	15.42	7.70
	P	21467	33.35	37.61	29.24	34.21	25.81	32.91	33.71	37.33	33.72	23.82	25.08	34.45	33.72	33.16
	A	26948	41.86	36.58	46.95	48.63	9.84	18.80	33.87	25.75	43.35	4.10	12.47	45.77	23.23	51.37
Total		64375	100	31587	32786	50105	6076	4698	2489	1002	61936	2439	7563	56812	21759	42616

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-54
Cut Scores and Associated Impact Data for WKCE-CRT Reading

Grade	Score Range				Impact Data				
	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
3	270-393	394-429	430-465	466-640	4.98	14.10	35.61	45.31	80.92
4	280-395	396-439	440-488	489-650	3.94	12.38	40.69	43.00	83.69
5	290-400	401-443	444-496	497-690	3.57	11.40	43.79	41.23	85.03
6	300-417	418-456	457-513	514-730	4.04	9.07	41.50	45.39	86.89
7	310-433	434-466	467-522	523-780	4.06	8.60	39.22	48.11	87.33
8	330-444	445-479	480-538	539-790	4.57	8.16	39.72	47.56	87.28
10	350-455	456-502	503-554	555-820	8.64	15.71	33.64	42.02	75.65

Table 8-55
Cut Scores and Associated Impact Data for WKCE-CRT Mathematics

Grade	Score Range				Impact Data				
	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
3	220-391	392-406	407-451	452-630	15.90	9.90	42.44	31.75	74.20
4	240-420	421-437	438-483	484-650	11.53	8.89	41.36	38.23	79.59
5	270-444	445-462	463-504	505-680	11.73	8.63	33.81	45.83	79.64
6	310-463	464-484	485-531	532-700	10.65	9.16	36.45	43.74	80.19
7	330-479	480-503	504-554	555-710	8.80	11.65	44.88	34.67	79.55
8	350-482	483-512	513-572	573-730	9.26	11.97	47.53	31.24	78.77
10	410-515	516-540	541-594	595-750	14.33	13.90	48.30	23.47	71.77

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-56
Cut Scores and Associated Impact Data for WKCE-CRT 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	5.33	18.06	44.66	31.95	76.61
8	250-357	358-384	385-417	418-520	15.01	22.56	33.90	28.53	62.43
10	290-392	393-427	428-483	484-630	8.99	18.01	52.85	20.16	73.01

Table 8-57
Cut Scores and Associated Impact Data for WKCE-CRT 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	1.98	5.79	26.46	65.77	92.23
8	230-333	334-363	364-402	403-530	4.75	12.60	37.71	44.94	82.65
10	240-407	408-419	420-454	455-620	13.46	5.85	29.40	51.28	80.68

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-58
Cut Scores and Associated Impact Data for WKCE-CRT 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	4.31	17.70	58.23	19.76	77.99
8	230-348	349-374	375-418	419-560	8.12	14.08	45.93	31.87	77.80
10	240-410	411-428	429-465	466-610	14.48	10.31	33.35	41.86	75.21

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-59
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	58810	1	Determines Meaning	12	0	12	8.00	0.67	2.73	66.51	20.42
	58810	2	Understands Text	17	0	17	12.70	0.75	3.73	74.69	20.73
	58810	3	Analyzes Text	21	1	24	15.81	0.66	4.88	66.13	19.64
	58810	4	Evaluates/Extends Text	4	0	4	2.58	0.64	1.19	63.96	21.65
4	59903	1	Determines Meaning	11	0	11	7.96	0.72	2.45	72.19	20.12
	59903	2	Understands Text	17	0	17	11.79	0.69	3.59	69.33	19.36
	59903	3	Analyzes Text	21	1	24	14.75	0.61	4.72	61.68	18.78
	59903	4	Evaluates/Extends Text	5	1	8	4.31	0.54	1.65	53.70	15.68
5	60406	1	Determines Meaning	11	0	11	7.52	0.68	2.63	68.17	21.63
	60406	2	Understands Text	15	0	15	11.75	0.78	2.96	78.19	18.38
	60406	3	Analyzes Text	20	1	23	13.83	0.60	4.20	60.31	17.06
	60406	4	Evaluates/Extends Text	8	1	11	6.30	0.57	2.23	57.23	16.54
6	59935	1	Determines Meaning	11	0	11	7.57	0.69	2.32	68.28	18.25
	59935	2	Understands Text	14	0	14	8.99	0.64	2.82	64.16	17.29
	59935	3	Analyzes Text	18	1	21	13.15	0.63	3.68	62.89	16.14
	59935	4	Evaluates/Extends Text	11	1	14	8.97	0.64	2.56	64.12	15.87
7	60238	1	Determines Meaning	11	0	11	7.97	0.72	2.15	72.29	16.87
	60238	2	Understands Text	14	0	14	9.42	0.67	2.70	67.39	16.71
	60238	3	Analyzes Text	18	1	21	12.51	0.60	3.83	59.96	16.72
	60238	4	Evaluates/Extends Text	11	1	14	9.11	0.65	2.62	64.92	15.96
8	60243	1	Determines Meaning	11	0	11	8.17	0.74	2.28	73.84	18.11
	60243	2	Understands Text	14	0	14	10.25	0.73	2.76	73.39	17.62
	60243	3	Analyzes Text	18	1	21	13.79	0.66	3.49	65.83	14.98
	60243	4	Evaluates/Extends Text	11	1	14	9.38	0.67	2.88	67.42	18.31

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-59 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	64632	1	Determines Meaning	7	0	7	5.26	0.75	1.50	74.48	17.03
	64632	2	Understands Text	7	0	7	4.43	0.63	1.84	63.58	21.81
	64632	3	Analyzes Text	22	1	25	15.10	0.60	5.00	60.81	18.99
	64632	4	Evaluates/Extends Text	14	1	17	12.30	0.72	3.44	72.43	19.02

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-60
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	59160	A	Mathematical Processes	3	3	9	3.40	0.38	2.25	38.42	22.55
	59160	B	Number Operations	11	1	12	9.10	0.76	2.43	75.92	17.90
	59160	C	Geometry	9	1	10	8.15	0.82	1.65	80.79	13.57
	59160	D	Measurement	8	0	8	5.96	0.75	1.62	73.78	15.13
	59160	E	Statistics/Probability	7	1	8	5.74	0.72	1.83	71.80	19.54
	59160	F	Algebraic Relationships	8	1	10	7.77	0.78	1.91	78.24	15.71
4	60100	A	Mathematical Processes	3	3	9	4.26	0.47	2.05	47.76	19.08
	60100	B	Number Operations	11	0	11	9.17	0.83	1.85	82.69	14.40
	60100	C	Geometry	8	1	10	7.74	0.77	1.76	77.10	13.14
	60100	D	Measurement	8	1	9	6.83	0.76	1.76	75.57	16.58
	60100	E	Statistics/Probability	7	1	8	5.13	0.64	1.69	63.21	17.33
	60100	F	Algebraic Relationships	9	1	10	7.80	0.78	2.05	79.00	17.94
5	60520	A	Mathematical Processes	3	3	9	5.24	0.58	2.10	59.23	19.22
	60520	B	Number Operations	11	0	11	8.69	0.79	1.91	78.70	14.43
	60520	C	Geometry	9	1	10	7.97	0.80	1.59	78.95	12.04
	60520	D	Measurement	9	1	11	7.26	0.66	2.47	66.39	19.29
	60520	E	Statistics/Probability	9	1	10	6.20	0.62	2.24	62.26	19.14
	60520	F	Algebraic Relationships	10	1	11	7.57	0.69	2.63	68.81	21.70

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-60 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	60105	A	Mathematical Processes	3	3	9	4.33	0.48	2.17	48.35	20.40
	60105	B	Number Operations	12	0	12	7.81	0.65	2.61	65.20	18.81
	60105	C	Geometry	9	1	10	7.88	0.79	1.91	77.67	15.68
	60105	D	Measurement	9	1	11	6.95	0.63	2.43	63.86	19.42
	60105	E	Statistics/Probability	8	1	9	6.15	0.68	2.17	68.08	20.41
	60105	F	Algebraic Relationships	10	1	11	8.02	0.73	2.28	73.80	17.99
7	60372	A	Mathematical Processes	3	3	9	4.43	0.49	2.18	50.03	20.75
	60372	B	Number Operations	12	0	12	8.29	0.69	2.64	68.80	19.27
	60372	C	Geometry	10	2	12	8.33	0.69	2.42	69.97	16.85
	60372	D	Measurement	9	0	9	6.67	0.74	1.72	73.61	15.95
	60372	E	Statistics/Probability	8	1	10	6.61	0.66	2.02	65.43	17.53
	60372	F	Algebraic Relationships	9	1	10	6.75	0.67	2.26	68.72	19.12
8	60352	A	Mathematical Processes	3	3	9	4.30	0.48	2.30	48.36	21.66
	60352	B	Number Operations	7	0	7	3.85	0.55	1.91	55.20	22.46
	60352	C	Geometry	8	1	9	6.26	0.70	1.90	69.44	16.80
	60352	D	Measurement	11	1	12	5.93	0.49	2.64	49.40	19.40
	60352	E	Statistics/Probability	8	1	9	5.43	0.60	2.03	59.88	18.49
	60352	F	Algebraic Relationships	14	1	16	10.26	0.64	3.19	63.91	18.16
10	64645	A	Mathematical Processes	7	1	9	5.31	0.59	2.38	59.29	23.97
	64645	B	Number Operations	7	0	7	4.65	0.66	1.89	66.12	23.13
	64645	C	Geometry	8	1	10	5.93	0.59	2.61	59.74	22.59
	64645	D	Measurement	9	1	11	5.80	0.53	3.01	52.97	24.84
	64645	E	Statistics/Probability	9	0	9	5.64	0.63	2.07	62.59	19.21
	64645	F	Algebraic Relationships	10	1	12	6.72	0.56	2.95	55.98	22.16

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-61
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	59838	B	Writing	19	0	19	13.71	0.72	3.52	72.24	17.95
	59838	D	Language	5	0	5	2.80	0.56	1.32	55.57	18.21
	59838	F	Research and Inquiry	6	0	6	3.16	0.53	1.47	53.32	18.25
8	60120	B	Writing	18	0	18	13.94	0.77	3.47	77.30	18.60
	60120	D	Language	6	0	6	3.57	0.59	1.68	60.59	23.74
	60120	F	Research and Inquiry	5	0	5	3.58	0.72	1.17	71.00	16.20
10	64195	B	Writing	15	2	24	14.64	0.61	3.86	61.05	15.45
	64195	D	Language	9	0	9	6.41	0.71	2.15	71.48	20.89
	64195	F	Research and Inquiry	6	0	6	3.80	0.63	1.45	63.31	17.68

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-62
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	59940	A	Geography	9	0	9	7.22	0.80	1.85	79.87	18.02
	59940	B	History	8	0	8	6.68	0.84	1.46	83.23	15.47
	59940	C	Political Science	7	0	7	4.63	0.66	1.57	66.74	17.06
	59940	D	Economics	7	0	7	5.34	0.76	1.29	76.59	14.62
	59940	E	Behavioral Science	7	0	7	5.26	0.75	1.65	75.54	20.11
8	60131	A	Geography	10	0	10	8.31	0.83	1.85	82.49	16.41
	60131	B	History	11	0	11	7.22	0.66	2.33	66.66	18.07
	60131	C	Political Science	6	0	6	3.93	0.65	1.54	65.51	19.82
	60131	D	Economics	8	0	8	5.57	0.70	1.80	69.21	18.70
	60131	E	Behavioral Science	5	0	5	3.03	0.61	1.35	61.62	20.56
10	64251	A	Geography	9	0	9	5.65	0.63	1.78	62.37	15.20
	64251	B	History	12	0	12	7.68	0.64	2.68	64.88	19.59
	64251	C	Political Science	12	0	12	7.28	0.61	2.74	60.95	20.35
	64251	D	Economics	8	0	8	5.27	0.66	1.91	65.77	20.27
	64251	E	Behavioral Science	8	0	8	6.17	0.77	1.67	76.67	17.53

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-63
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	60007	A/B	Connections & Nature of Sci	8	0	8	5.65	0.71	1.68	71.03	17.24
	60007	C	Science Inquiry	7	0	7	4.90	0.70	1.55	70.56	18.02
	60007	D	Physical Science	6	0	6	4.56	0.76	0.96	77.15	9.86
	60007	E	Earth and Space	6	0	6	4.21	0.70	1.28	68.69	14.44
	60007	F	Life and Environment	6	0	6	4.39	0.73	1.37	71.62	17.72
	60007	G/H	Appl & Social Perspectives	7	0	7	5.30	0.76	1.60	76.12	18.92
8	60265	A/B	Connections & Nature of Sci	7	0	7	4.92	0.70	1.65	70.45	19.38
	60265	C	Science Inquiry	8	0	8	6.43	0.80	1.49	80.34	15.25
	60265	D	Physical Science	6	0	6	4.57	0.76	1.16	75.79	14.07
	60265	E	Earth and Space	6	0	6	3.95	0.66	1.45	65.81	18.66
	60265	F	Life and Environment	6	0	6	4.27	0.71	1.37	71.42	16.74
	60265	G/H	Appl & Social Perspectives	7	0	7	5.75	0.82	1.40	82.13	16.68
10	64375	A/B	Connections & Nature of Sci	10	0	10	6.49	0.65	2.22	64.91	19.44
	64375	C	Science Inquiry	9	0	9	5.44	0.60	2.00	60.81	18.29
	64375	D	Physical Science	8	0	8	5.06	0.63	1.86	62.29	18.43
	64375	E	Earth and Space	6	0	6	3.51	0.59	1.31	57.25	13.93
	64375	F	Life and Environment	6	0	6	3.41	0.57	1.28	59.64	15.69
	64375	G/H	Appl & Social Perspectives	11	0	11	7.11	0.65	2.61	64.51	20.89

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-64
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	0-28	0-26	0-24	0-31	0-38	0-36	0-47
		2	29-45	27-51	25-41	32-44	39-52	37-50	48-64
		3	46-72	52-80	42-77	45-73	53-75	51-78	65-80
		4	73-100	81-100	78-100	74-100	76-100	79-100	81-100
	Standard 2 Understands Text	1	0-31	0-29	0-34	0-30	0-33	0-38	0-29
		2	32-55	30-47	35-57	31-42	34-45	39-51	30-46
		3	56-83	48-77	58-86	43-68	46-70	52-77	47-70
		4	84-100	78-100	87-100	69-100	71-100	78-100	71-100
	Standard 3 Analyzes Text	1	0-27	0-25	0-26	0-29	0-27	0-36	0-31
		2	28-47	26-40	27-40	30-42	28-38	37-47	32-46
		3	48-72	41-67	41-66	43-66	39-62	48-68	47-66
		4	73-100	68-100	67-100	67-100	63-100	69-100	67-100
	Standard 4 Evaluates/Extends Text	1	0-29	0-23	0-28	0-31	0-33	0-31	0-40
		2	30-42	24-38	29-38	32-44	34-45	32-44	41-61
		3	43-68	39-57	39-61	45-68	46-67	45-71	62-80
		4	69-100	58-100	62-100	69-100	68-100	72-100	81-100
MA	Standard A Mathematical Processes	1	0-15	0-24	0-36	0-21	0-20	0-18	0-29
		2	16-20	25-30	37-43	22-28	21-29	19-29	30-43
		3	21-45	31-52	44-61	29-51	30-59	30-59	44-80
		4	46-100	53-100	62-100	52-100	60-100	60-100	81-100
	Standard B Number Operations	1	0-57	0-65	0-61	0-39	0-39	0-25	0-36
		2	58-66	66-73	62-68	40-46	40-50	26-32	37-51
		3	67-87	74-89	69-82	47-69	51-79	33-68	52-86
		4	88-100	90-100	83-100	70-100	80-100	69-100	87-100
	Standard C Geometry	1	0-68	0-61	0-65	0-56	0-46	0-45	0-30
		2	69-74	62-67	66-70	57-64	47-56	46-56	31-44
		3	75-88	68-82	71-81	65-83	57-78	57-79	45-78
		4	89-100	83-100	82-100	84-100	79-100	80-100	79-100
	Standard D Measurement	1	0-59	0-53	0-40	0-37	0-50	0-24	0-25
		2	60-65	54-61	41-49	38-45	51-59	25-31	26-32
		3	66-82	62-83	50-71	46-68	60-81	32-57	33-75
		4	83-100	84-100	72-100	69-100	82-100	58-100	76-100
	Standard E Statistics/Probability	1	0-50	0-42	0-36	0-40	0-39	0-34	0-40
		2	51-59	43-47	37-44	41-49	40-50	35-43	41-49
		3	60-83	48-67	45-66	50-73	51-72	44-69	50-78
		4	84-100	68-100	67-100	74-100	73-100	70-100	79-100
	Standard F Algebraic Relationships	1	0-63	0-55	0-38	0-49	0-38	0-36	0-29
		2	64-71	56-65	39-48	50-59	39-51	37-49	30-39
		3	72-87	66-89	49-74	60-79	52-79	50-73	40-74
		4	88-100	90-100	75-100	80-100	80-100	74-100	75-100

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 8-64 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 © 2011 by the Wisconsin Department of Public Instruction

Table 8-64 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		0-42				0-38	0-42
		2		43-57				39-50	43-47
		3		58-80				51-75	48-59
		4		81-100				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 © 2011 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	0.93	0.93	0.92	0.92	0.93	0.92	0.93	0.90	0.93	0.92	0.93	0.92
	4	0.92	0.91	0.92	0.91	0.91	0.90	0.92	0.91	0.92	0.88	0.92	0.90	0.91	0.90
	5	0.91	0.91	0.92	0.90	0.90	0.90	0.91	0.90	0.91	0.86	0.91	0.90	0.91	0.89
	6	0.90	0.89	0.90	0.88	0.89	0.89	0.90	0.88	0.89	0.84	0.89	0.87	0.89	0.87
	7	0.90	0.89	0.90	0.88	0.89	0.89	0.90	0.89	0.89	0.85	0.89	0.87	0.89	0.87
	8	0.90	0.89	0.91	0.89	0.90	0.89	0.90	0.88	0.90	0.85	0.89	0.88	0.90	0.88
	10	0.92	0.91	0.92	0.91	0.91	0.91	0.92	0.91	0.91	0.85	0.90	0.90	0.91	0.90
Mathematics	3	0.90	0.90	0.90	0.88	0.91	0.88	0.89	0.89	0.90	0.88	0.91	0.89	0.90	0.87
	4	0.89	0.89	0.89	0.87	0.90	0.88	0.89	0.88	0.89	0.87	0.91	0.88	0.89	0.86
	5	0.91	0.90	0.91	0.90	0.90	0.89	0.91	0.89	0.91	0.87	0.90	0.89	0.90	0.89
	6	0.92	0.91	0.92	0.91	0.91	0.90	0.92	0.90	0.92	0.88	0.91	0.90	0.91	0.90
	7	0.92	0.91	0.92	0.91	0.90	0.90	0.92	0.90	0.92	0.87	0.90	0.90	0.90	0.90
	8	0.92	0.91	0.92	0.91	0.89	0.89	0.92	0.90	0.91	0.86	0.89	0.90	0.90	0.90
	10	0.94	0.93	0.94	0.93	0.89	0.91	0.94	0.91	0.93	0.84	0.88	0.93	0.92	0.93
Language Arts	4	0.82	0.81	0.82	0.81	0.79	0.78	0.82	0.79	0.82	0.75	0.78	0.81	0.80	0.80
	8	0.85	0.84	0.86	0.84	0.84	0.82	0.83	0.83	0.85	0.75	0.82	0.83	0.84	0.83
	10	0.84	0.83	0.85	0.82	0.83	0.83	0.84	0.81	0.84	0.75	0.80	0.82	0.83	0.82
Social Studies	4	0.87	0.87	0.87	0.85	0.88	0.85	0.87	0.87	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	0.86	0.87	0.77	0.86	0.86	0.86	0.85
	10	0.90	0.89	0.91	0.89	0.88	0.88	0.90	0.88	0.90	0.80	0.88	0.89	0.89	0.89
Science	4	0.85	0.84	0.86	0.83	0.84	0.82	0.85	0.83	0.85	0.79	0.85	0.84	0.84	0.82
	8	0.86	0.85	0.88	0.84	0.85	0.84	0.86	0.86	0.86	0.78	0.86	0.84	0.86	0.83
	10	0.88	0.87	0.90	0.87	0.84	0.86	0.88	0.85	0.88	0.75	0.86	0.87	0.87	0.87

Copyright © 2011 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	2.92	3.02	2.88	3.27	3.21	3.02	3.16	2.94	3.29	3.30	2.92	3.17	2.80
	4	3.18	3.16	3.19	3.10	3.42	3.38	3.22	3.30	3.15	3.46	3.40	3.13	3.34	3.04
	5	3.16	3.14	3.17	3.08	3.43	3.36	3.21	3.31	3.13	3.47	3.44	3.10	3.34	3.01
	6	3.19	3.17	3.20	3.13	3.40	3.36	3.25	3.32	3.17	3.47	3.44	3.15	3.34	3.08
	7	3.19	3.17	3.20	3.13	3.39	3.34	3.26	3.28	3.17	3.47	3.44	3.15	3.32	3.09
	8	3.15	3.09	3.18	3.07	3.46	3.36	3.22	3.28	3.13	3.53	3.50	3.08	3.35	3.01
	10	3.03	2.99	3.04	2.96	3.26	3.20	3.12	3.14	3.01	3.34	3.27	2.98	3.18	2.93
Mathematics	3	2.98	2.99	2.98	2.94	3.14	3.08	2.96	3.06	2.97	3.09	3.14	2.96	3.08	2.89
	4	2.91	2.92	2.89	2.85	3.11	3.04	2.88	3.01	2.89	3.07	3.11	2.87	3.03	2.79
	5	3.24	3.25	3.22	3.17	3.43	3.39	3.20	3.39	3.22	3.43	3.44	3.19	3.37	3.12
	6	3.23	3.23	3.22	3.17	3.40	3.39	3.18	3.38	3.21	3.45	3.40	3.19	3.37	3.12
	7	3.20	3.22	3.19	3.14	3.40	3.35	3.17	3.31	3.19	3.41	3.39	3.16	3.34	3.09
	8	3.37	3.38	3.35	3.35	3.41	3.44	3.35	3.43	3.36	3.44	3.40	3.36	3.43	3.32
	10	3.28	3.32	3.24	3.25	3.32	3.37	3.29	3.36	3.28	3.37	3.32	3.26	3.36	3.22
Language Arts	4	2.25	2.21	2.29	2.21	2.40	2.35	2.25	2.34	2.24	2.39	2.42	2.23	2.35	2.17
	8	2.11	2.03	2.17	2.05	2.31	2.27	2.12	2.25	2.09	2.38	2.40	2.06	2.26	2.00
	10	2.59	2.51	2.64	2.53	2.78	2.70	2.63	2.70	2.58	2.77	2.77	2.54	2.71	2.50
Social Studies	4	2.26	2.25	2.27	2.18	2.54	2.44	2.30	2.39	2.24	2.49	2.48	2.22	2.43	2.11
	8	2.53	2.55	2.51	2.46	2.80	2.72	2.56	2.66	2.51	2.84	2.80	2.49	2.70	2.42
	10	2.86	2.87	2.85	2.80	3.09	3.04	2.92	3.00	2.85	3.15	3.11	2.82	3.02	2.77
Science	4	2.45	2.45	2.44	2.37	2.70	2.62	2.49	2.56	2.43	2.68	2.63	2.42	2.59	2.32
	8	2.42	2.44	2.40	2.34	2.73	2.63	2.47	2.54	2.40	2.77	2.73	2.37	2.61	2.30
	10	3.01	3.05	2.97	2.96	3.18	3.17	3.06	3.14	3.00	3.23	3.18	2.98	3.14	2.94

Copyright © 2011 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		0.83	0.84	0.48				0.93
	4	0.73		0.79	0.81	0.43				0.92
	5	0.74		0.78	0.77	0.56				0.91
	6	0.66		0.69	0.73	0.66				0.90
	7	0.65		0.68	0.74	0.65				0.90
	8	0.68		0.73	0.71	0.68				0.90
	10	0.55		0.65	0.82	0.80				0.92
Mathematics	3	0.61		0.72	0.57	0.55	0.65	0.58		0.90
	4	0.54		0.67	0.52	0.60	0.58	0.70		0.89
	5	0.52		0.62	0.49	0.66	0.65	0.76		0.91
	6	0.56		0.69	0.64	0.69	0.71	0.70		0.92
	7	0.59		0.72	0.68	0.60	0.63	0.67		0.92
	8	0.63		0.63	0.59	0.68	0.61	0.75		0.92
	10	0.72		0.69	0.71	0.76	0.62	0.74		0.94
Language Arts	4			0.77		0.44		0.44		0.82
	8			0.80		0.61		0.36		0.85
	10			0.72		0.69		0.47		0.84
Social Studies	4	0.68		0.61	0.46	0.49	0.64			0.87
	8	0.68		0.64	0.55	0.60	0.47			0.87
	10	0.52		0.71	0.71	0.65	0.62			0.90
Science	4		0.56		0.55	0.23	0.38	0.50	0.63	0.85
	8		0.57		0.57	0.41	0.48	0.46	0.60	0.86
	10		0.65		0.58	0.57	0.33	0.36	0.72	0.88

Copyright © 2011 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		1.54	1.95	0.86				2.97
	4	1.27		1.65	2.06	1.25				3.18
	5	1.34		1.39	2.01	1.48				3.16
	6	1.35		1.57	1.91	1.49				3.19
	7	1.27		1.53	1.95	1.55				3.19
	8	1.29		1.43	1.88	1.63				3.15
	10	1.01		1.09	2.12	1.54				3.03
Mathematics	3	1.41		1.29	1.08	1.09	1.08	1.24		2.98
	4	1.39		1.06	1.22	1.11	1.10	1.12		2.91
	5	1.45		1.18	1.14	1.44	1.33	1.29		3.24
	6	1.44		1.45	1.15	1.35	1.17	1.25		3.23
	7	1.40		1.40	1.37	1.09	1.23	1.30		3.20
	8	1.40		1.16	1.22	1.49	1.27	1.60		3.37
	10	1.26		1.05	1.41	1.47	1.28	1.50		3.28
Language Arts	4			1.69		0.99		1.10		2.25
	8			1.55		1.05		0.94		2.11
	10			2.04		1.20		1.06		2.59
Social Studies	4	1.05		0.91	1.15	0.92	0.99			2.26
	8	1.05		1.40	1.03	1.14	0.98			2.53
	10	1.23		1.44	1.48	1.13	1.03			2.86
Science	4		1.11		1.04	0.84	1.01	0.97	0.97	2.45
	8		1.08		0.98	0.89	1.05	1.01	0.89	2.42
	10		1.31		1.30	1.22	1.07	1.02	1.38	3.01

Copyright © 2011 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	0.04	0.01	0.00	0.00	0.05
Basic Proficient	0.02	0.09	0.03	0.00	0.14
Proficient	0.00	0.03	0.28	0.05	0.35
Advanced	0.00	0.00	0.05	0.40	0.45
Sum	0.06	0.14	0.36	0.45	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.97	0.94	0.90	0.81
Probability of Chance	0.90	0.69	0.51	0.35
Kappa (k)	0.70	0.81	0.80	0.70
Classification Accuracy	0.98	0.96	0.93	0.87

Copyright © 2011 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	0.03	0.02	0.00	0.00	0.05
Basic Proficient	0.01	0.08	0.02	0.00	0.12
Proficient	0.00	0.03	0.32	0.05	0.40
Advanced	0.00	0.00	0.05	0.38	0.43
Sum	0.04	0.13	0.40	0.43	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.97	0.95	0.89	0.81
Probability of Chance	0.92	0.72	0.51	0.36
Kappa (k)	0.68	0.81	0.78	0.70
Classification Accuracy	0.98	0.96	0.92	0.86

Copyright © 2011 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	0.03	0.01	0.00	0.00	0.04
Basic Proficient	0.01	0.07	0.02	0.00	0.11
Proficient	0.00	0.03	0.35	0.06	0.43
Advanced	0.00	0.00	0.06	0.35	0.41
Sum	0.04	0.12	0.43	0.41	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.97	0.94	0.88	0.80
Probability of Chance	0.92	0.74	0.52	0.37
Kappa (k)	0.67	0.79	0.76	0.68
Classification Accuracy	0.98	0.96	0.92	0.86

Copyright © 2011 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	0.03	0.01	0.00	0.00	0.05
Basic Proficient	0.01	0.05	0.02	0.00	0.09
Proficient	0.00	0.03	0.32	0.05	0.40
Advanced	0.00	0.00	0.07	0.39	0.46
Sum	0.05	0.09	0.42	0.45	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.97	0.94	0.87	0.79
Probability of Chance	0.91	0.76	0.50	0.38
Kappa (k)	0.69	0.76	0.75	0.66
Classification Accuracy	0.98	0.96	0.91	0.85

Copyright © 2011 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	0.03	0.01	0.00	0.00	0.05
Basic Proficient	0.01	0.05	0.02	0.00	0.08
Proficient	0.00	0.03	0.29	0.06	0.38
Advanced	0.00	0.00	0.07	0.42	0.49
Sum	0.05	0.09	0.39	0.48	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.97	0.95	0.88	0.80
Probability of Chance	0.91	0.77	0.50	0.39
Kappa (k)	0.69	0.76	0.75	0.67
Classification Accuracy	0.98	0.96	0.92	0.86

Copyright © 2011 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	0.04	0.01	0.00	0.00	0.05
Basic Proficient	0.01	0.05	0.02	0.00	0.08
Proficient	0.00	0.03	0.30	0.05	0.38
Advanced	0.00	0.00	0.07	0.42	0.49
Sum	0.05	0.09	0.39	0.47	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.97	0.95	0.88	0.80
Probability of Chance	0.90	0.77	0.50	0.39
Kappa (k)	0.73	0.77	0.76	0.68
Classification Accuracy	0.98	0.96	0.91	0.86

Copyright © 2011 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	0.07	0.02	0.00	0.00	0.09
Basic Proficient	0.02	0.10	0.04	0.00	0.16
Proficient	0.00	0.03	0.23	0.06	0.33
Advanced	0.00	0.00	0.06	0.36	0.42
Sum	0.09	0.16	0.33	0.43	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.96	0.93	0.88	0.77
Probability of Chance	0.83	0.63	0.51	0.32
Kappa (k)	0.75	0.81	0.75	0.66
Classification Accuracy	0.97	0.95	0.92	0.84

Copyright © 2011 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	0.13	0.02	0.01	0.00	0.16
Basic Proficient	0.03	0.04	0.04	0.00	0.10
Proficient	0.01	0.04	0.31	0.06	0.42
Advanced	0.00	0.00	0.07	0.26	0.32
Sum	0.17	0.10	0.42	0.32	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.93	0.91	0.87	0.73
Probability of Chance	0.72	0.61	0.56	0.31
Kappa (k)	0.75	0.76	0.70	0.60
Classification Accuracy	0.95	0.94	0.91	0.81

Copyright © 2011 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	0.10	0.02	0.01	0.00	0.13
Basic Proficient	0.02	0.03	0.03	0.00	0.09
Proficient	0.01	0.03	0.29	0.07	0.40
Advanced	0.00	0.00	0.06	0.33	0.38
Sum	0.13	0.09	0.39	0.39	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.92	0.87	0.74
Probability of Chance	0.78	0.66	0.52	0.33
Kappa (k)	0.72	0.75	0.73	0.62
Classification Accuracy	0.96	0.94	0.91	0.81

Copyright © 2011 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	0.10	0.02	0.01	0.00	0.12
Basic Proficient	0.02	0.03	0.03	0.00	0.09
Proficient	0.01	0.03	0.23	0.06	0.32
Advanced	0.00	0.00	0.06	0.40	0.46
Sum	0.13	0.09	0.33	0.46	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.92	0.88	0.76
Probability of Chance	0.78	0.66	0.50	0.34
Kappa (k)	0.73	0.76	0.77	0.64
Classification Accuracy	0.96	0.94	0.92	0.83

Copyright © 2011 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	0.09	0.02	0.00	0.00	0.11
Basic Proficient	0.02	0.03	0.04	0.00	0.09
Proficient	0.00	0.03	0.26	0.05	0.35
Advanced	0.00	0.00	0.06	0.38	0.44
Sum	0.11	0.09	0.37	0.44	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.95	0.92	0.89	0.77
Probability of Chance	0.80	0.68	0.51	0.34
Kappa (k)	0.75	0.76	0.78	0.65
Classification Accuracy	0.97	0.95	0.92	0.83

Copyright © 2011 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	0.07	0.02	0.00	0.00	0.09
Basic Proficient	0.02	0.06	0.03	0.00	0.11
Proficient	0.00	0.03	0.35	0.05	0.44
Advanced	0.00	0.00	0.05	0.30	0.36
Sum	0.10	0.11	0.44	0.35	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.95	0.93	0.89	0.78
Probability of Chance	0.83	0.67	0.54	0.34
Kappa (k)	0.73	0.80	0.77	0.67
Classification Accuracy	0.97	0.95	0.93	0.85

Copyright © 2011 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	0.08	0.02	0.00	0.00	0.10
Basic Proficient	0.02	0.06	0.03	0.00	0.11
Proficient	0.00	0.04	0.38	0.05	0.47
Advanced	0.00	0.00	0.05	0.27	0.32
Sum	0.10	0.12	0.46	0.33	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.95	0.93	0.90	0.79
Probability of Chance	0.82	0.66	0.56	0.34
Kappa (k)	0.73	0.78	0.78	0.68
Classification Accuracy	0.96	0.94	0.93	0.84

Copyright © 2011 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	0.12	0.03	0.00	0.00	0.15
Basic Proficient	0.03	0.07	0.03	0.00	0.13
Proficient	0.00	0.04	0.40	0.04	0.48
Advanced	0.00	0.00	0.04	0.21	0.24
Sum	0.15	0.13	0.47	0.24	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.92	0.93	0.79
Probability of Chance	0.74	0.60	0.63	0.33
Kappa (k)	0.75	0.81	0.80	0.69
Classification Accuracy	0.95	0.94	0.95	0.84

Copyright © 2011 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	0.02	0.09	0.06	0.00	0.17
Proficient	0.00	0.06	0.28	0.08	0.43
Advanced	0.00	0.00	0.08	0.26	0.34
Sum	0.07	0.18	0.42	0.33	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.95	0.87	0.84	0.67
Probability of Chance	0.87	0.63	0.55	0.33
Kappa (k)	0.56	0.65	0.65	0.51
Classification Accuracy	0.96	0.91	0.89	0.77

Copyright © 2011 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	0.11	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	0.08	0.23	0.31
Sum	0.16	0.22	0.33	0.30	

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	0.74	0.53	0.58	0.27
Kappa (k)	0.69	0.72	0.66	0.53
Classification Accuracy	0.94	0.91	0.89	0.74

Copyright © 2011 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	0.07	0.03	0.00	0.00	0.10
Basic Proficient	0.03	0.11	0.05	0.00	0.19
Proficient	0.00	0.05	0.38	0.06	0.50
Advanced	0.00	0.00	0.06	0.16	0.22
Sum	0.10	0.18	0.50	0.22	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.94	0.90	0.87	0.71
Probability of Chance	0.83	0.60	0.66	0.34
Kappa (k)	0.66	0.74	0.63	0.56
Classification Accuracy	0.96	0.92	0.91	0.79

Copyright © 2011 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	0.18	0.06	0.26
Advanced	0.00	0.00	0.06	0.59	0.65
Sum	0.02	0.06	0.27	0.65	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.99	0.96	0.88	0.82
Probability of Chance	0.96	0.85	0.55	0.50
Kappa (k)	0.67	0.70	0.73	0.64
Classification Accuracy	0.99	0.97	0.91	0.87

Copyright © 2011 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	0.03	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	0.39	0.46
Sum	0.05	0.13	0.37	0.46	

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)	0.62	0.72	0.71	0.60
Classification Accuracy	0.97	0.94	0.90	0.82

Copyright © 2011 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	0.11	0.02	0.02	0.00	0.14
Basic Proficient	0.02	0.02	0.03	0.00	0.06
Proficient	0.01	0.03	0.18	0.06	0.27
Advanced	0.00	0.00	0.06	0.46	0.52
Sum	0.14	0.06	0.28	0.52	

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	0.76	0.68	0.50	0.37
Kappa (k)	0.73	0.73	0.75	0.61
Classification Accuracy	0.96	0.94	0.91	0.82

Copyright © 2011 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	0.12	0.05	0.00	0.18
Proficient	0.00	0.05	0.44	0.06	0.55
Advanced	0.00	0.00	0.06	0.16	0.22
Sum	0.05	0.18	0.55	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	0.75
Probability of Chance	0.91	0.64	0.66	0.38
Kappa (k)	0.63	0.72	0.65	0.59
Classification Accuracy	0.98	0.93	0.91	0.82

Copyright © 2011 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	0.85	0.71
Probability of Chance	0.84	0.65	0.55	0.33
Kappa (k)	0.67	0.71	0.67	0.56
Classification Accuracy	0.96	0.93	0.90	0.79

Copyright © 2011 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	0.11	0.03	0.01	0.00	0.15
Basic Proficient	0.03	0.04	0.04	0.00	0.11
Proficient	0.01	0.04	0.21	0.07	0.33
Advanced	0.00	0.00	0.06	0.36	0.42
Sum	0.16	0.10	0.31	0.43	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All cuts
Classification Consistency (P)	0.92	0.90	0.87	0.72
Probability of Chance	0.74	0.62	0.51	0.32
Kappa (k)	0.69	0.75	0.73	0.59
Classification Accuracy	0.94	0.93	0.91	0.78

Copyright © 2011 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												
3	31	3	79	17	1	3	0.91	0.82	1.35	6096	780	2446	2820	50
4	13	3	68	30	1	1	0.85	0.70	1.54	6330	508	2369	2990	463
4	47	3	63	31	4	2	0.85	0.69	1.23	6330	1417	2426	2095	392
5	13	3	62	34	3	1	0.82	0.65	1.63	6304	508	2042	3032	722
5	47	3	73	24	1	1	0.89	0.77	1.40	6304	944	2144	2986	230
6	18	3	61	35	2	2	0.77	0.53	1.32	6194	579	3270	2162	183
6	24	3	67	30	1	2	0.87	0.75	1.20	6194	1422	2328	2245	199
7	31	3	68	29	1	2	0.83	0.67	1.56	6236	416	2326	3100	394
7	50	3	67	29	2	2	0.89	0.77	1.08	6236	2017	2018	1864	337
8	19	3	67	31	1	1	0.84	0.67	1.56	6350	444	2298	3229	379
8	56	3	62	33	3	2	0.88	0.75	1.70	6350	954	1244	2908	1,244
10	7	3	57	33	3	7	0.88	0.77	1.32	6648	1721	1929	2160	838
10	52	3	63	30	2	5	0.87	0.74	1.46	6648	733	2847	2347	721

* The sum of the modes of agreement and codes may not equal exactly 100% due to rounding.
** Item dropped from scoring. See Part 8 for more information.

Copyright © 2011 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	92	7	1	1	0.97	0.95	0.91	6096	2913	843	2340
3	20A	1	90	2	3	5	0.98	0.95	0.46	6096	3319	2777	
3	20B	2	84	12	1	2	0.93	0.87	0.50	6096	3949	1253	894
3	27A	1	95	2	0	3	0.98	0.96	0.64	6096	2221	3875	
3	27B	2	79	13	4	4	0.91	0.83	0.59	6096	3973	679	1444
3	39A	1	93	1	1	5	0.99	0.97	0.36	6096	3901	2195	
3	39B	2	87	9	2	2	0.94	0.89	0.55	6096	3896	1071	1129
4	11A	1	98	1	0	1	0.98	0.96	0.21	6330	4974	1356	
4	11B	2	79	16	2	3	0.90	0.79	0.60	6330	3274	2309	747
4	22	2	96	3	1	1	0.97	0.95	1.61	6330	452	1596	4282
4	25A	1	99	1	0	1	0.99	0.99	0.45	6330	3465	2865	
4	25B	2	82	13	3	3	0.89	0.78	0.62	6330	3235	2250	845
4	39A	1	99	1	0	1	1.00	0.99	0.59	6330	2614	3716	
4	39B	2	87	9	1	2	0.96	0.92	0.73	6330	3621	822	1887
5	12	2	96	3	0	1	0.99	0.98	1.32	6304	1861	558	3885
5	20A	1	99	0	0	1	0.99	0.98	0.84	6304	997	5307	
5	20B	2	86	11	1	2	0.94	0.88	0.56	6304	3903	1244	1157
5	27A	1	98	1	0	1	1.00	0.99	0.59	6304	2565	3739	
5	27B	2	85	12	2	1	0.93	0.87	1.41	6304	1550	617	4137
5	46A	1	99	1	0	0	0.99	0.98	0.37	6304	3987	2317	
5	46B	2	89	7	3	1	0.94	0.88	0.63	6304	4166	296	1842
6	7A	1	94	1	1	4	0.99	0.98	0.51	6194	3062	3132	
6	7B	2	87	4	4	5	0.95	0.91	0.81	6194	3618	165	2411
6	22A	1	99	1	0	1	0.99	0.96	0.90	6194	606	5588	
6	22B	2	91	7	1	1	0.93	0.86	0.48	6194	3314	2814	66

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

Copyright © 2011 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	97	2	0	1	0.98	0.97	0.46	6194	3370	2824	
6	39B	2	90	7	1	2	0.97	0.93	0.83	6194	2867	1538	1789
6	51	2	94	5	0	1	0.98	0.95	0.81	6194	2520	2353	1321
7	7A	1	98	1	0	1	1.00	0.99	0.70	6236	1888	4348	
7	7B	2	89	9	1	2	0.97	0.93	1.21	6236	1981	961	3294
7	19A	1	97	1	0	2	0.99	0.98	0.52	6236	2992	3244	
7	19B	2	90	6	1	3	0.96	0.93	0.60	6236	3321	2066	849
7	38A	1	88	2	1	9	0.98	0.97	0.53	6236	2912	3324	
7	38B	2	91	4	1	4	0.97	0.94	0.69	6236	2489	3182	565
7	52	2	96	2	0	1	0.99	0.98	0.71	6236	2944	2138	1154
8	10	2	89	7	0	3	0.96	0.92	0.40	6350	4603	951	796
8	19A	1	97	1	0	2	0.99	0.98	0.72	6350	1765	4585	
8	19B	2	87	9	1	3	0.97	0.93	1.19	6350	1952	1238	3160
8	33A	1	86	7	1	6	0.91	0.81	0.27	6350	4637	1713	
8	33B	2	82	11	1	6	0.94	0.88	0.67	6350	3012	2445	893
8	49A	1	97	0	0	2	1.00	0.99	0.26	6350	4716	1634	
8	49B	2	89	7	1	3	0.95	0.91	0.41	6350	4802	517	1031
10	23	2	79	14	1	7	0.94	0.87	0.80	6648	2805	2361	1482
10	33	2	80	6	1	13	0.97	0.93	0.54	6648	4103	1525	1020
10	36	2	77	13	1	9	0.95	0.90	0.88	6648	2962	1503	2183
10	46	2	84	9	1	6	0.96	0.92	0.90	6648	2880	1561	2207

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

Copyright © 2011 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	61	35	2	2	0.89	0.79	3.21	6330	134	140	1004	2489	2124	408	31
4	1B	3	95	4	0	1	0.86	0.73	1.97	6330	67	126	6065	72			
8	1A	6	63	33	1	2	0.92	0.83	3.40	6350	159	52	933	2066	2325	746	69
8	1B	3	95	4	0	1	0.87	0.73	1.98	6350	72	85	6122	71			
10	1A	6	58	35	3	4	0.93	0.85	3.31	6648	284	231	1041	1798	2356	732	206
10	1B	3	83	13	0	4	0.88	0.75	2.03	6648	253	59	5598	738			

* 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 © 2011 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	0.25		CC		BB
RD	5	52	MC	0.01	-0.05	-4.30	0.04	0.00	4.45		-1.52			-C
RD	6	18	CR	0.09	0.00	7.55	0.00	-0.09	-7.53	0.25				CC
RD	6	56	MC	0.00	-0.07	-8.20	0.06	-0.01	6.79		-1.53			-C
RD	7	31	CR	0.10	0.00	7.61	0.01	-0.11	-7.68	0.26		CC		CC
RD	8	19	CR	0.10	0.00	7.96	0.03	-0.12	-8.42	0.27		CC	-CC	CC
RD	8	21	MC	0.01	-0.05	-3.89	0.03	0.00	5.53		-1.58			-C
RD	8	23	MC	0.00	-0.06	-5.15	0.05	0.00	6.13		-1.94			-C
RD	8	31	MC	0.00	-0.06	-6.67	0.05	0.00	6.77		-1.78			-C
RD	8	50	MC	0.03	0.00	4.18	0.00	-0.03	-2.32		1.51			C
RD	8	56	CR	0.14	0.00	8.81	0.00	-0.14	-9.23	0.27		CC	-CC	CC
RD	10	7	CR	0.13	0.00	7.60	0.00	-0.13	-8.26	0.25		CC	-CC	BB
RD	10	22	MC	0.02	-0.07	-6.42	0.06	-0.03	5.37		-1.51			-C
RD	10	37	MC	0.00	-0.07	-8.31	0.07	0.00	7.86		-1.81			-C
RD	10	38	MC	0.00	-0.06	-7.43	0.05	0.00	6.76		-1.58			-C
RD	10	52	CR	0.11	0.00	8.78	0.00	-0.11	-9.54	0.26		CC	-CC	CC
MA	5	17	MC	0.00	-0.05	-6.51	0.05	0.00	5.81		-1.64			-C
MA	6	30	MC	0.00	-0.07	-8.51	0.06	0.00	7.49		-1.63			-C
MA	7	10	MC	0.00	-0.06	-6.28	0.05	0.00	6.19		-1.58			-C
LA	4	4	MC	0.03	0.00	5.49	0.00	-0.02	-3.46		2.17			C
LA	4	1A	CR	0.15	0.00	8.87	0.00	-0.16	-9.41	0.31		CC	-CC	CC
LA	8	1A	CR	0.12	-0.02	5.94	0.05	-0.15	-7.92	0.21		CC	-CC	BB
SC	8	12	MC	0.02	0.00	3.87	0.02	-0.01	-3.36		1.64			C

Copyright © 2011 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	8	1	MC	0.04	0.00	2.63		1.68		C
MA	7	44	MC	0.03	-0.02	-0.54		-2.08		-C
LA	8	1A	CR	0.08	-0.18	-3.60	-0.07		-CC	
LA	10	26	MC	0.04	-0.07	-3.07		-1.58		-C
SS	8	22	MC	0.04	-0.04	-2.84		-1.81		-C

Copyright © 2011 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	3	32	MC	0.03	0.00	2.89		2.02		C
RD	7	24	MC	0.01	-0.01	-1.43		-1.82		-C
RD	8	1	MC	0.03	0.00	2.76		2.18		C
RD	8	21	MC	0.00	-0.06	-3.72		-1.97		-C
RD	10	7	CR	0.16	-0.06	3.12	0.18		CC	BB
MA	4	1	MC	0.02	-0.01	2.54		1.91		C
MA	5	24	MC	0.03	-0.01	2.08		1.55		C
MA	6	7B	CR	0.15	0.00	5.17	0.23		CC	BB
MA	6	37	MC	0.05	-0.11	-6.21		-1.62		-C
MA	8	54	MC	0.02	-0.06	-2.06		-1.60		-C
LA	4	11	MC	0.00	-0.03	-2.06		-1.64		-C
LA	10	26	MC	0.03	-0.11	-5.28		-2.18		-C
SC	4	1	MC	0.01	-0.01	-0.28		-1.93		-C
SC	4	34	MC	0.03	0.00	3.14		1.71		C
SC	8	11	MC	0.01	-0.05	-1.56		-1.77		-C
SS	8	1	MC	0.01	-0.04	-0.75		-1.67		-C
SS	10	10	MC	0.03	-0.11	-5.34		-1.70		-C

Copyright © 2011 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	0.21		CC	BB
RD	4	7	MC	0.03	-0.12	-2.61		-1.57		-C
RD	4	13	CR	0.15	0.00	4.95	0.24		CC	BB
RD	5	39	MC	0.03	-0.01	1.90		1.94		C
RD	5	47	CR	0.13	0.00	4.06	0.22		CC	BB
RD	6	18	CR	0.15	0.00	5.02	0.26		CC	CC
RD	6	24	CR	0.14	0.00	4.40	0.17		CC	
RD	6	29	MC	0.04	-0.15	-6.08		-1.85	-C	-C
RD	6	37	MC	0.01	-0.05	-2.31		-1.58		-C
RD	6	38	MC	0.00	-0.08	-4.65		-1.97		-C
RD	6	43	MC	0.08	-0.13	-5.83		-1.47	-C	-B
RD	6	44	MC	0.15	-0.10	4.91		1.67	C	C
RD	6	55	MC	0.08	-0.01	3.11		1.54		C
RD	7	24	MC	0.02	-0.03	-1.76		-2.43		-C
RD	7	31	CR	0.17	-0.01	4.88	0.28		CC	CC
RD	8	1	MC	0.03	-0.01	2.23		2.20		C
RD	8	8	MC	0.01	-0.08	-3.71		-1.66		-C
RD	8	19	CR	0.21	-0.07	6.00	0.29		CC	CC
RD	8	21	MC	0.01	-0.07	-3.16		-1.96		-C
RD	8	56	CR	0.26	-0.09	5.65	0.27		CC	CC
RD	10	7	CR	0.18	0.00	4.86	0.23		CC	BB
RD	10	12	MC	0.01	-0.11	-4.41		-1.61		-C
RD	10	21	MC	0.08	-0.03	4.01		1.55		C
RD	10	52	CR	0.21	0.00	7.44	0.30		CC	CC
MA	3	16	MC	0.00	-0.13	-7.73		-2.53	-C	-C
MA	3	21	MC	0.01	-0.06	-1.63		-1.74		-C
MA	4	27	MC	0.01	-0.04	-3.35		-1.53		-C

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 10-4 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
MA	5	17	MC	0.08	0.00	4.35		1.60		C
MA	5	33	MC	0.02	-0.01	2.04		2.36		C
MA	5	48	MC	0.05	0.00	3.85		1.55		C
MA	5	55	MC	0.01	-0.09	-4.13		-1.73		-C
MA	6	24	MC	0.00	-0.12	-5.85		-1.48	-C	-B
MA	6	36	MC	0.00	-0.10	-4.93		-1.82		-C
MA	6	39B	CR	0.17	-0.07	3.71	0.17		CC	
MA	6	47	MC	0.00	-0.08	-5.16		-1.66		-C
MA	7	21	MC	0.04	-0.15	-4.30		-1.23	-C	-B
MA	7	44	MC	0.01	-0.04	-2.86		-4.22		-C
MA	8	23	MC	0.08	-0.02	4.13		2.59		C
MA	8	37	MC	0.00	-0.10	-4.80		-1.66		-C
MA	8	54	MC	0.01	-0.13	-2.97		-2.47		-C
MA	10	4	MC	0.08	-0.14	-6.01		-1.79	-C	-C
MA	10	33	CR	0.15	-0.02	5.40	0.20		CC	BB
MA	10	54	MC	0.10	0.00	5.54		1.93		C
LA	4	12	MC	0.03	-0.02	2.04		1.53		C
LA	4	13	MC	0.00	-0.15	-8.51		-2.67	-C	-C
LA	4	1A	CR	0.14	0.00	3.60	0.17		CC	BB
LA	8	1A	CR	0.29	-0.03	6.18	0.30		CC	CC
LA	10	26	MC	0.00	-0.21	-12.53		-3.67	-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		-4.01		-C
SC	4	14	MC	0.10	-0.14	-5.62		-1.03	-C	-B
SC	4	28	MC	0.05	0.00	3.55		2.14		C
SC	4	34	MC	0.04	-0.01	2.53		1.93		C
SC	8	4	MC	0.05	-0.01	2.63		1.53		C

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 10-4 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
SC	10	34	MC	0.00	-0.12	-6.43		-1.57	-C	-C
SS	4	14	MC	0.04	-0.03	1.84		2.03		C
SS	4	19	MC	0.03	0.00	1.81		2.30		C
SS	4	37	MC	0.03	-0.02	2.29		1.50		C
SS	8	3	MC	0.00	-0.13	-6.97		-2.16	-C	-C
SS	8	37	MC	0.12	0.00	5.53		1.58	C	C
SS	10	25	MC	0.01	-0.14	-1.13		-2.33		-C
SS	10	28	MC	0.02	-0.10	-4.60		-1.62		-C

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 10-5
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			-C	
RD	8	56	CR	0.11	-0.44	-3.18	-0.42		-CC	
RD	10	52	CR	0.00	-0.24	-3.12	-0.35		-CC	
MA	6	12	MC	0.06	-0.26	-3.58			-C	
MA	7	13	MC	0.22	-0.19	-2.84			-C	
MA	7	41	MC	0.19	-0.19	-3.10			-C	
MA	7	53	MC	0.08	-0.25	-2.79			-C	
MA	8	4	MC	0.10	-0.24	-1.26		-5.08		-C
LA	8	21	MC	0.01	-0.17	-0.80		-8.06		-C
SS	8	7	MC					-5.23		-C
SS	10	6	MC	0.21	-0.11	2.74		0.45	C	
SS	10	10	MC	0.21	-0.02	2.74		5.25	C	
SS	10	44	MC	0.02	-0.16	-2.59		-2.97	-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 © 2011 by the Wisconsin Department of Public Instruction

Table 10-6
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		-1.58	-C		-C
RD	4	13	CR	0.16	-0.10	4.95	0.04	-0.05	-2.38	0.19		CC		BB
RD	5	13	CR	0.17	-0.16	3.42	0.06	-0.10	-1.62	0.17		CC		
RD	5	47	CR	0.13	-0.06	3.63	0.03	-0.03	-2.17	0.16		CC		
RD	6	4	MC	0.01	-0.10	-5.15	0.02	0.00	2.64		-1.68			-C
RD	6	18	CR	0.13	-0.06	4.53	0.02	-0.04	-1.89	0.18		CC		BB
RD	6	24	CR	0.13	-0.03	4.30	0.03	-0.04	-2.31	0.14		CC		
RD	6	38	MC	0.00	-0.07	-5.14	0.02	0.00	3.22		-1.74			-C
RD	6	44	MC	0.15	-0.17	6.18	0.02	-0.05	-2.52		1.66	C		C
RD	7	24	MC	0.01	-0.03	-2.88	0.02	0.00	0.80		-1.85			-C
RD	7	31	CR	0.15	0.00	5.31	0.03	-0.05	-2.16	0.22		CC		BB
RD	8	1	MC	0.04	-0.03	3.22	0.01	0.00	0.41		1.53			C
RD	8	19	CR	0.13	0.00	4.78	0.07	-0.03	-2.09	0.18		CC		BB
RD	8	21	MC	0.02	-0.12	-6.01	0.02	0.00	3.39		-1.88	-C		-C
RD	8	56	CR	0.23	-0.09	5.32	0.02	-0.05	-2.28	0.22		CC		BB
RD	10	7	CR	0.21	-0.03	4.68	0.01	-0.05	-2.12	0.22		CC		BB
RD	10	52	CR	0.12	-0.02	3.79	0.05	-0.04	-1.92	0.16		CC		
MA	3	16	MC	0.00	-0.13	-7.81	0.02	-0.01	3.20		-1.98	-C		-C
MA	3	21	MC	0.03	-0.09	-4.59	0.02	0.00	3.46		-1.68			-C
MA	4	20	MC	0.02	0.00	2.65	0.00	0.00	-1.21		2.25			C
MA	5	18	MC	0.00	-0.11	-5.71	0.03	-0.03	1.80		-1.60	-C		-C
MA	6	7B	CR	0.20	-0.03	5.19	0.04	-0.04	-2.53	0.22		CC		BB
MA	8	37	MC	0.00	-0.12	-6.88	0.02	-0.01	1.75		-2.27	-C		-C
MA	10	4	MC	0.04	-0.16	-5.99	0.02	-0.03	1.48		-1.76	-C		-C
LA	4	13	MC	0.02	-0.10	-5.75	0.02	-0.01	3.82		-1.52			-C

Copyright © 2011 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	4	1A	CR	0.13	0.00	4.08	0.05	-0.07	-2.22	0.17		CC		
LA	8	1A	CR	0.18	-0.07	4.11	0.11	-0.05	-3.13	0.19		CC		BB
LA	10	22	MC	0.03	-0.12	-4.92	0.02	-0.02	1.53		-1.42	-C		-B
LA	10	26	MC	0.00	-0.21	-10.30	0.02	0.00	3.29		-2.75	-C		-C
LA	10	1A	CR	0.21	0.00	4.83	0.07	-0.06	-2.50	0.18		CC		BB
SC	4	1	MC	0.01	-0.05	-3.05	0.01	0.00	2.79		-3.36			-C
SC	4	34	MC	0.05	-0.01	4.48	0.01	-0.01	-1.05		2.22			C
SC	8	2	MC	0.04	-0.01	3.05	0.01	0.00	-0.19		1.59			C
SS	8	3	MC	0.01	-0.14	-6.27	0.02	-0.01	1.97		-1.48	-C		-B
SS	10	28	MC	0.01	-0.13	-5.67	0.03	-0.01	1.89		-1.91	-C		-C

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 10-7
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	-BB
RD	7	31	CR	0.05	-0.03	1.68	0.15	-0.14	-5.01	-0.16			-CC	
RD	8	19	CR	0.06	-0.02	2.02	0.03	-0.21	-6.13	-0.28			-CC	-CC
RD	8	56	CR	0.06	-0.10	2.31	0.00	-0.19	-6.67	-0.22			-CC	-BB
RD	10	52	CR	0.05	-0.02	0.88	0.03	-0.14	-4.37	-0.14			-CC	
MA	3	2	MC	0.01	0.00	2.21	0.01	-0.07	-1.39		-1.66			-C
MA	3	29	MC	0.01	-0.04	-2.87	0.15	0.00	6.95		1.95		C	C
MA	5	54	MC	0.00	0.00	-0.50	0.03	-0.01	2.01		2.15			C
LA	4	11	MC	0.02	0.00	4.78	0.00	-0.05	-5.37		-1.89			-C
LA	4	12	MC	0.02	0.00	1.84	0.01	-0.05	-1.34		-1.94			-C
LA	4	1A	CR	0.10	-0.07	2.77	0.00	-0.30	-8.56	-0.37			-CC	-CC
LA	4	1B	CR	0.02	-0.01	0.73	0.04	-0.02	-1.73	-0.28				-CC
LA	8	1A	CR	0.09	-0.04	1.06	0.09	-0.28	-6.30	-0.24			-CC	-BB
LA	10	1A	CR	0.11	-0.05	0.34	0.20	-0.27	-3.38	-0.16			-CC	
SC	8	2	MC	0.01	0.00	1.62	0.00	-0.02	-1.58		-3.24			-C
SC	8	3	MC	0.02	0.00	0.31	0.01	-0.02	-1.23		-1.76			-C

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 10-8
Correlations among Reading Objectives

Grade	CS	1	2	3
3	2	0.75		
	3	0.76	0.84	
	4	0.56	0.59	0.61
4	2	0.73		
	3	0.75	0.79	
	4	0.58	0.61	0.63
5	2	0.74		
	3	0.75	0.76	
	4	0.65	0.67	0.68
6	2	0.65		
	3	0.69	0.71	
	4	0.64	0.66	0.71
7	2	0.65		
	3	0.69	0.72	
	4	0.63	0.65	0.70
8	2	0.64		
	3	0.67	0.70	
	4	0.66	0.70	0.72
10	2	0.57		
	3	0.66	0.72	
	4	0.65	0.69	0.79

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 10-9
Correlations among Mathematics Objectives

Grade	CS	A	B	C	D	E
	B	0.61				
	C	0.51	0.59			
3	D	0.49	0.57	0.51		
	E	0.62	0.64	0.58	0.55	
	F	0.53	0.66	0.54	0.51	0.58
	B	0.53				
	C	0.50	0.48			
4	D	0.59	0.61	0.51		
	E	0.55	0.51	0.45	0.55	
	F	0.62	0.65	0.53	0.64	0.54
	B	0.56				
	C	0.52	0.50			
5	D	0.57	0.62	0.52		
	E	0.64	0.58	0.53	0.61	
	F	0.65	0.67	0.54	0.69	0.65
	B	0.63				
	C	0.57	0.58			
6	D	0.62	0.69	0.61		
	E	0.68	0.63	0.60	0.64	
	F	0.67	0.66	0.56	0.65	0.61
	B	0.64				
	C	0.64	0.63			
7	D	0.61	0.68	0.61		
	E	0.63	0.67	0.63	0.64	
	F	0.66	0.68	0.61	0.63	0.64
	B	0.58				
	C	0.62	0.53			
8	D	0.64	0.64	0.56		
	E	0.65	0.58	0.56	0.61	
	F	0.66	0.65	0.63	0.65	0.65
	B	0.71				
	C	0.71	0.66			
10	D	0.72	0.68	0.71		
	E	0.69	0.65	0.65	0.68	
	F	0.72	0.68	0.72	0.74	0.68

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 10-10
Correlations among Language Arts Objectives

Grade	CS	B	D
4	D	0.52	
	F	0.52	0.38
8	D	0.66	
	F	0.51	0.44
10	D	0.70	
	F	0.57	0.55

Table 10-11
Correlations among Social Studies Objectives

Grade	CS	A	B	C	D
4	B	0.65			
	C	0.52	0.50		
	D	0.55	0.53	0.47	
	E	0.64	0.61	0.53	0.55
8	B	0.61			
	C	0.53	0.56		
	D	0.59	0.61	0.54	
	E	0.51	0.54	0.50	0.52
10	B	0.61			
	C	0.59	0.69		
	D	0.59	0.66	0.68	
	E	0.55	0.60	0.62	0.61

Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 10-12
Correlations among Science Objectives

Grade	CS	A/B	C	D	E	F
4	C	0.57				
	D	0.35	0.35			
	E	0.43	0.42	0.32		
	F	0.52	0.52	0.35	0.45	
	G/H	0.58	0.57	0.37	0.45	0.54
8	C	0.58				
	D	0.48	0.46			
	E	0.51	0.48	0.44		
	F	0.51	0.48	0.44	0.47	
	G/H	0.59	0.60	0.47	0.48	0.51
10	C	0.60				
	D	0.60	0.57			
	E	0.45	0.42	0.42		
	F	0.47	0.43	0.42	0.34	
	G/H	0.68	0.62	0.59	0.45	0.48

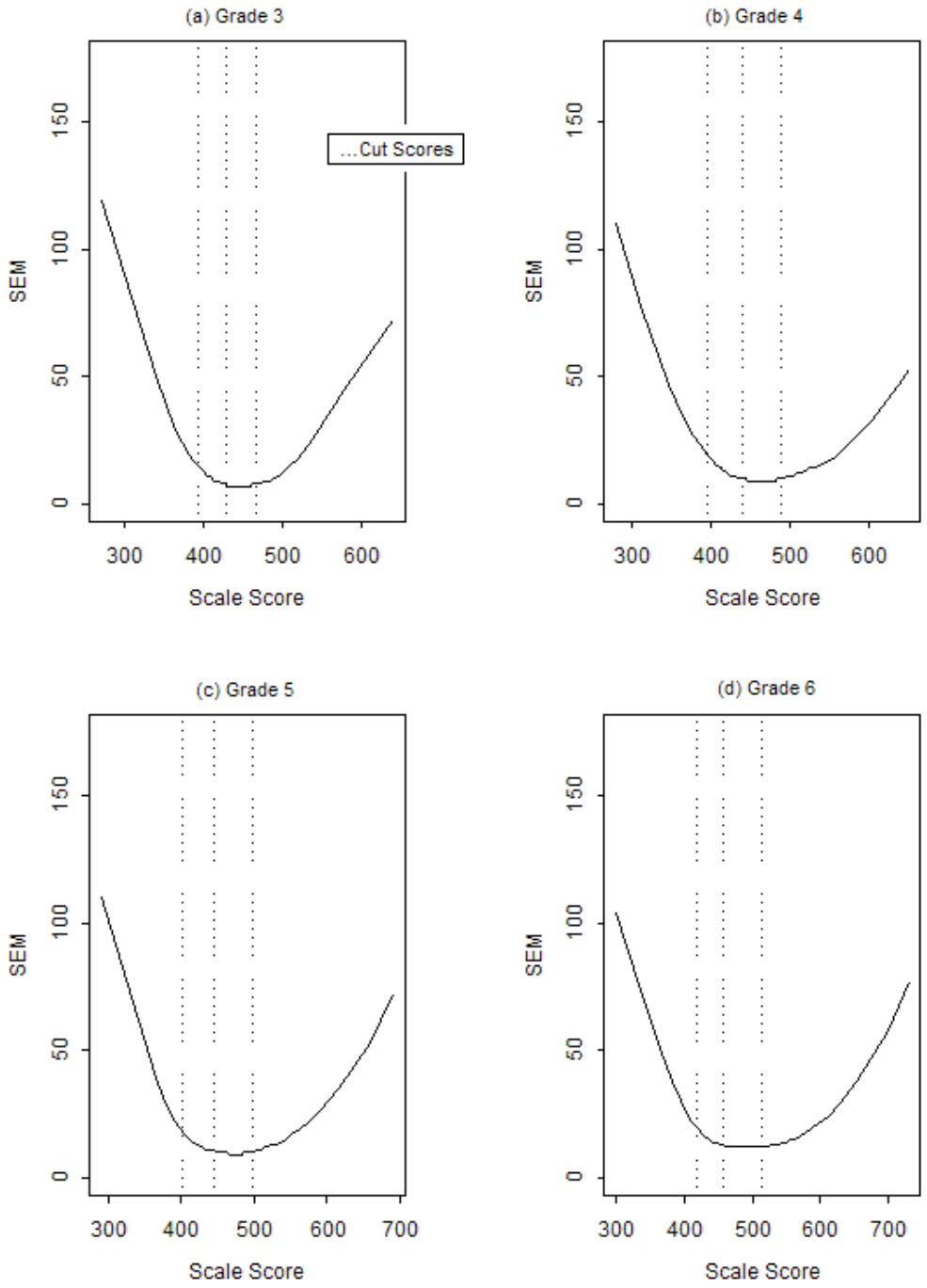
Copyright © 2011 by the Wisconsin Department of Public Instruction

Table 10-13
Principal Components Analysis

Content Area	Grade	First Eigenvalue	Second Eigenvalue	Ratio of First Two Eigenvalues
Reading	3	12.513	1.658	7.545
	4	11.280	1.528	7.383
	5	10.780	1.671	6.452
	6	9.346	1.636	5.713
	7	9.230	1.763	5.236
	8	9.761	1.477	6.607
	10	10.891	1.942	5.610
Mathematics	3	9.192	1.856	4.952
	4	8.633	1.815	4.757
	5	9.867	1.884	5.237
	6	10.852	1.749	6.206
	7	10.857	1.791	6.063
	8	10.515	1.881	5.591
	10	12.659	1.774	7.136
Language Arts	4	5.160	1.299	3.972
	8	5.920	1.199	4.938
	10	6.070	1.205	5.037
Social Studies	4	7.036	1.601	4.394
	8	7.280	1.607	4.531
	10	9.075	1.749	5.189
Science	4	6.422	1.551	4.140
	8	6.678	1.561	4.276
	10	8.082	1.389	5.819

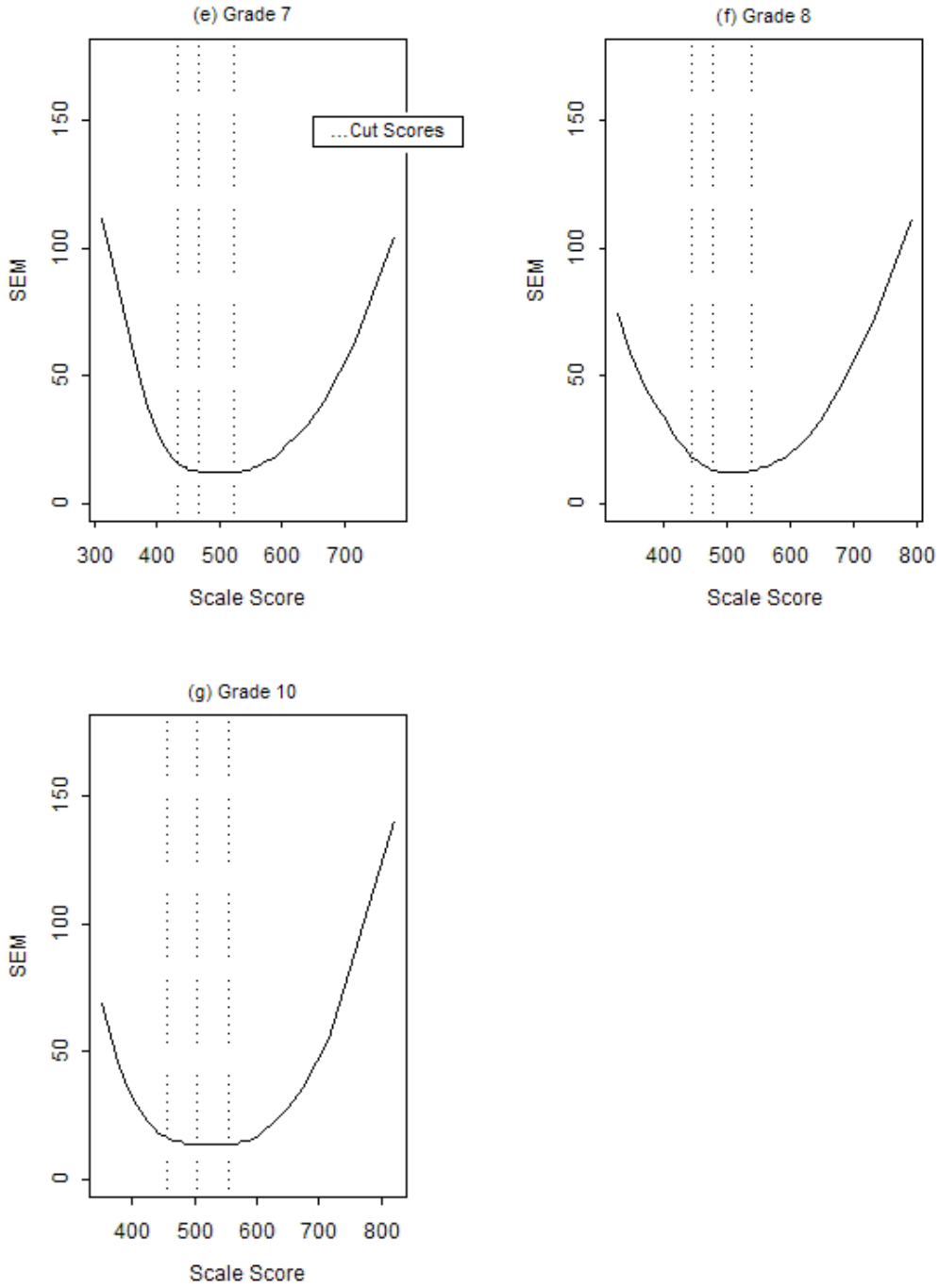
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-1
SEM Curves, Reading Grades 3-6



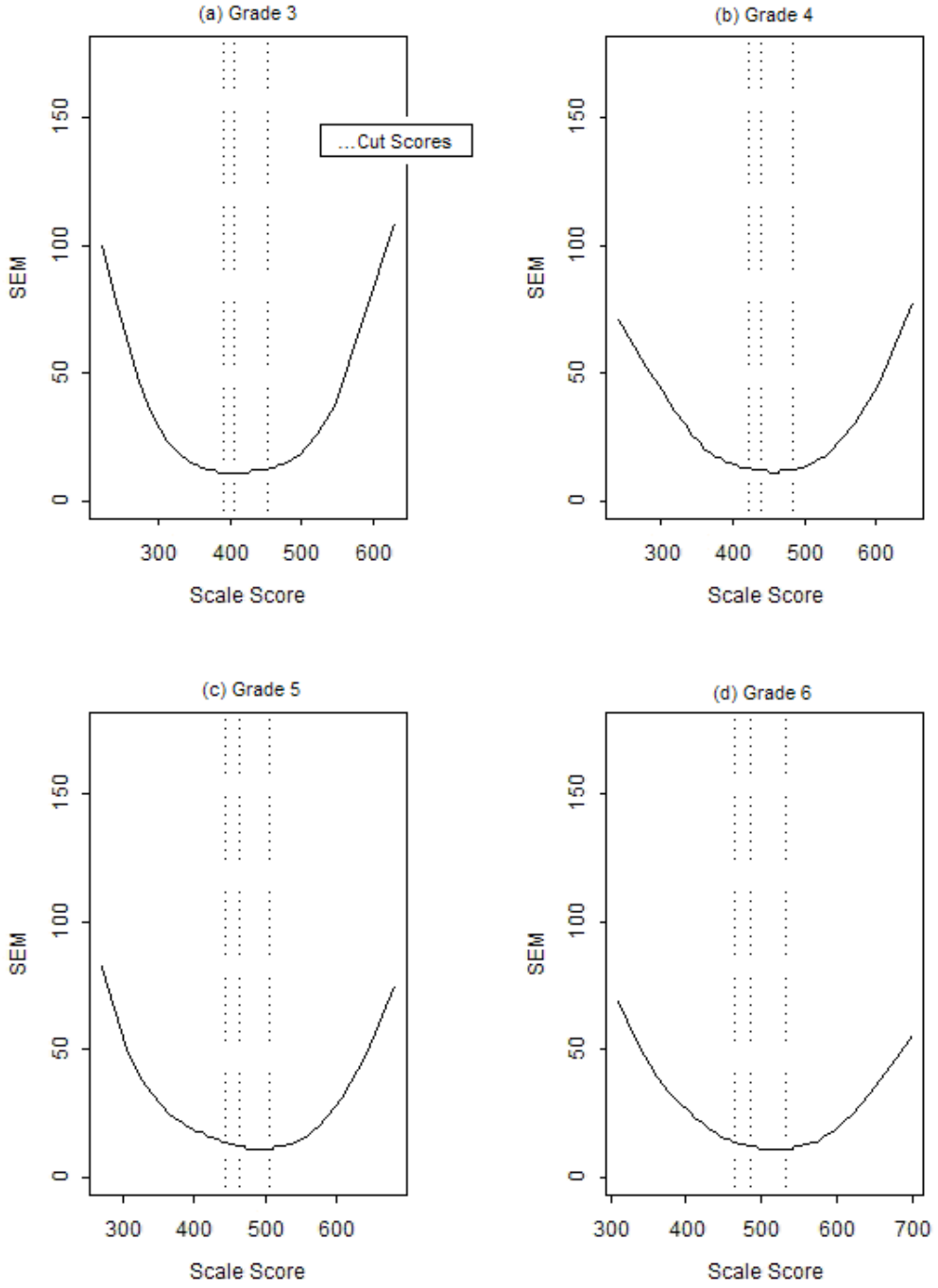
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-1 Cont'd
SEM Curves, Reading Grades 7, 8, 10



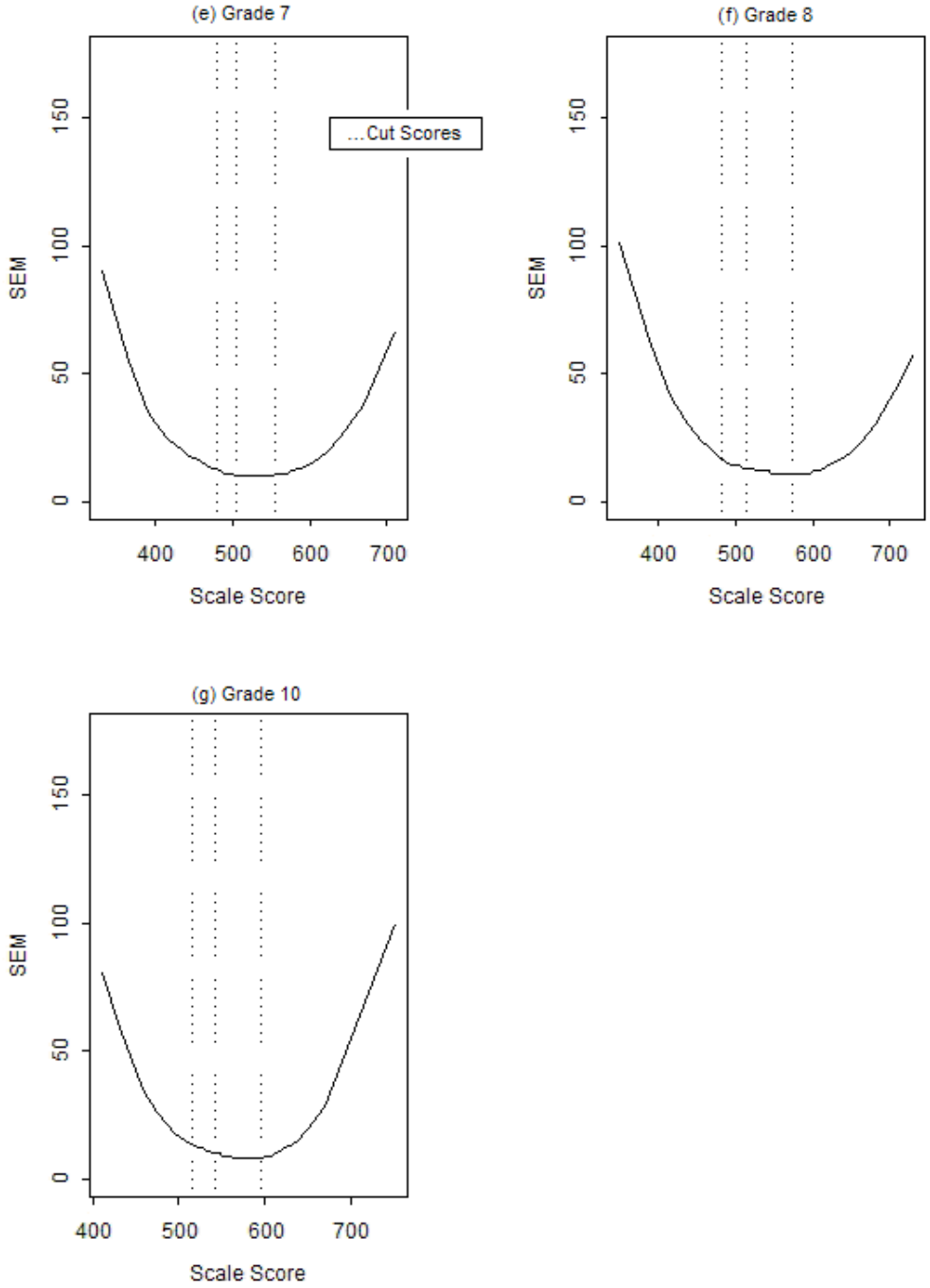
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-2
SEM Curves, Mathematics Grades 3-6



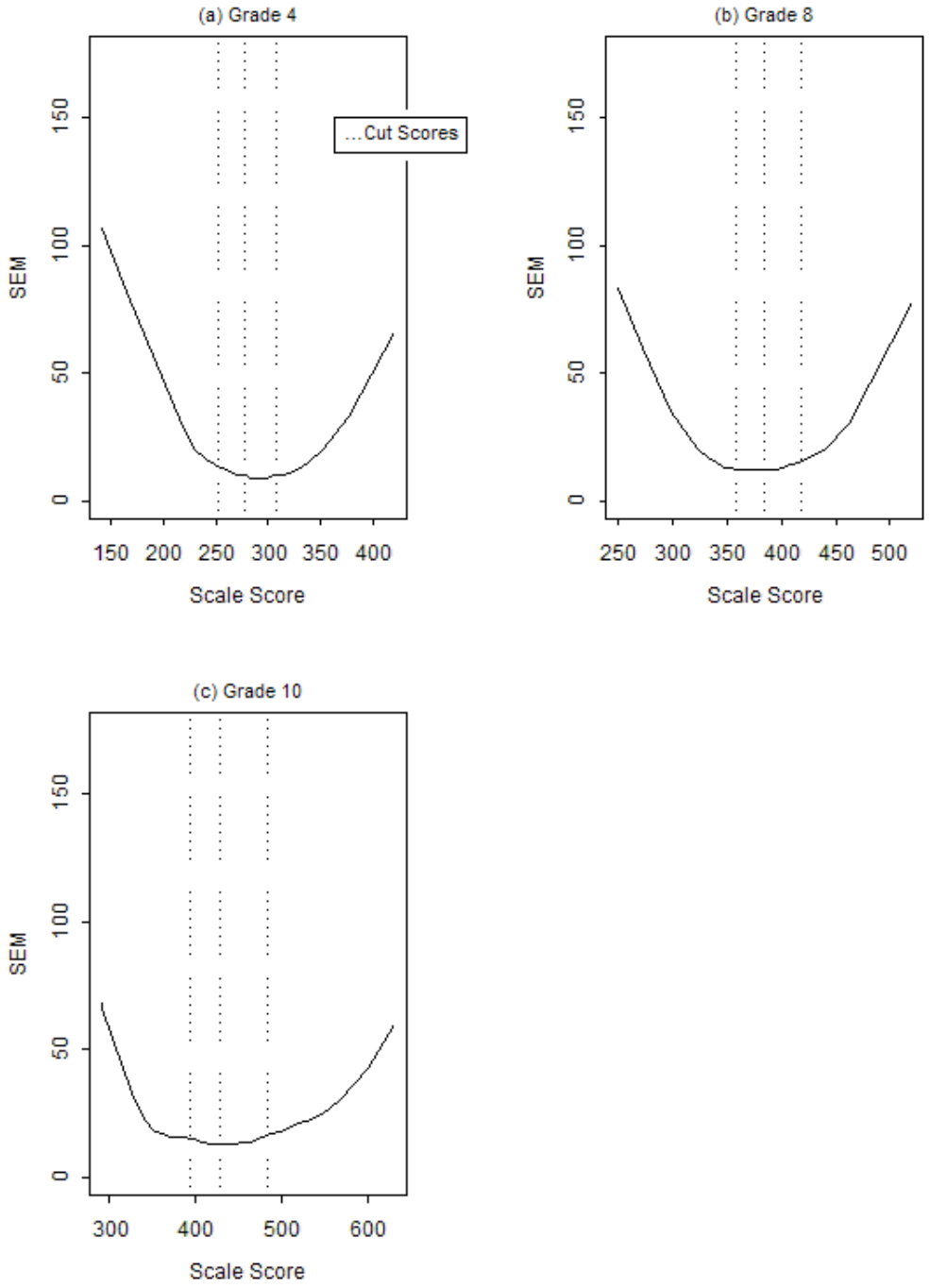
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-2 Cont'd
SEM Curves, Mathematics Grades 7, 8, 10



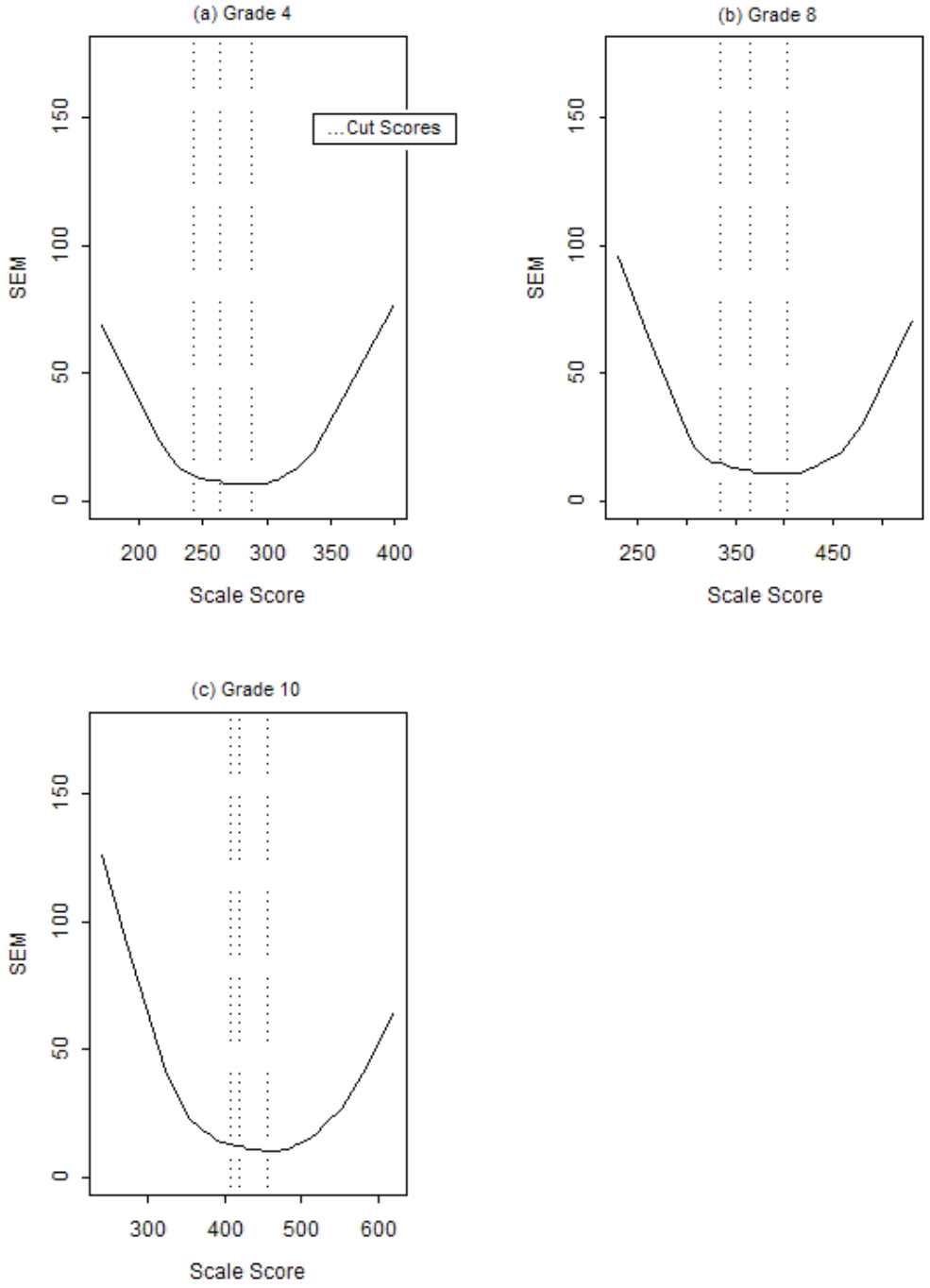
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-3
SEM Curves, Language Arts Grades 4, 8, 10



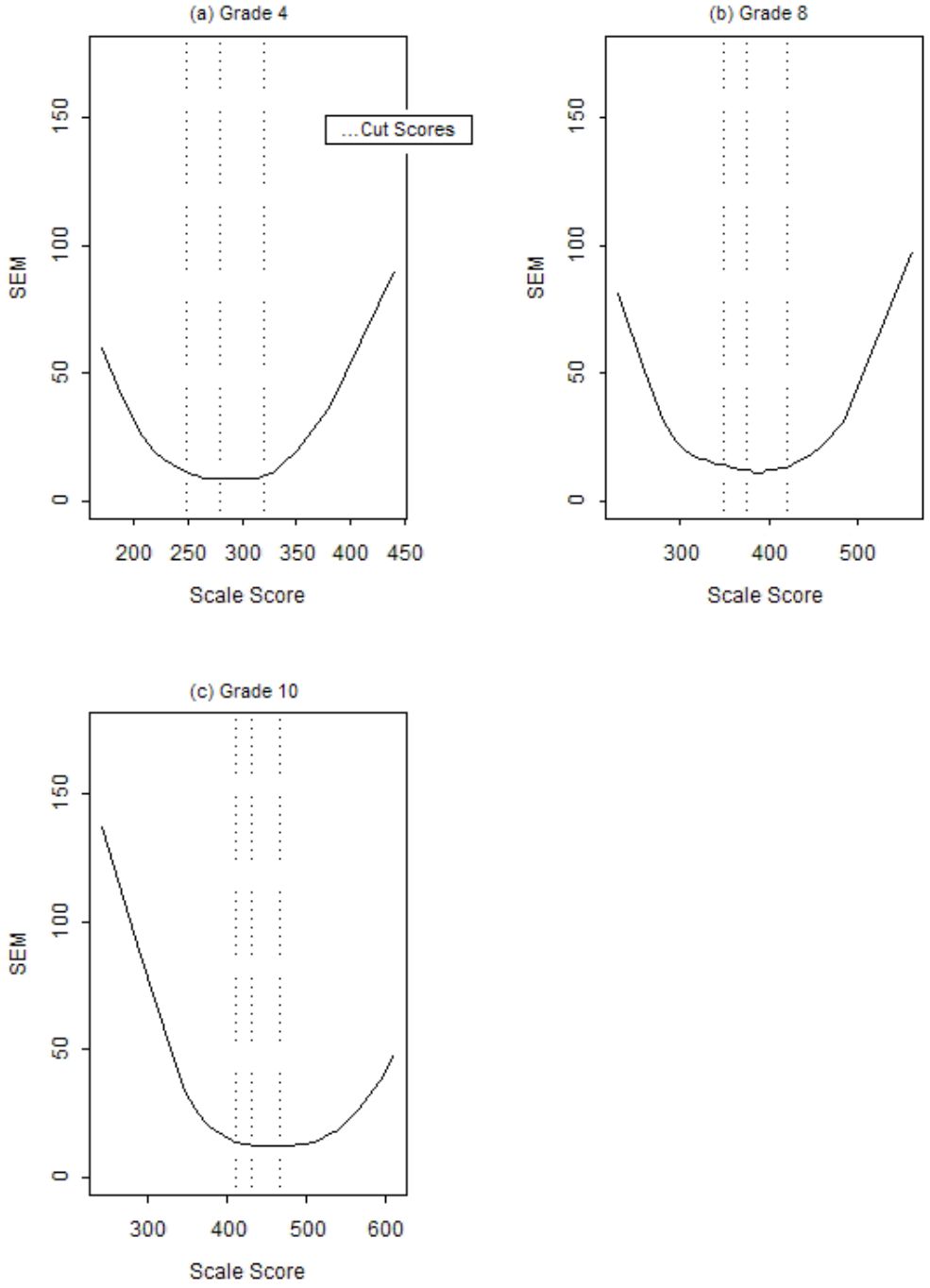
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-4
SEM Curves, Social Studies Grades 4, 8, 10



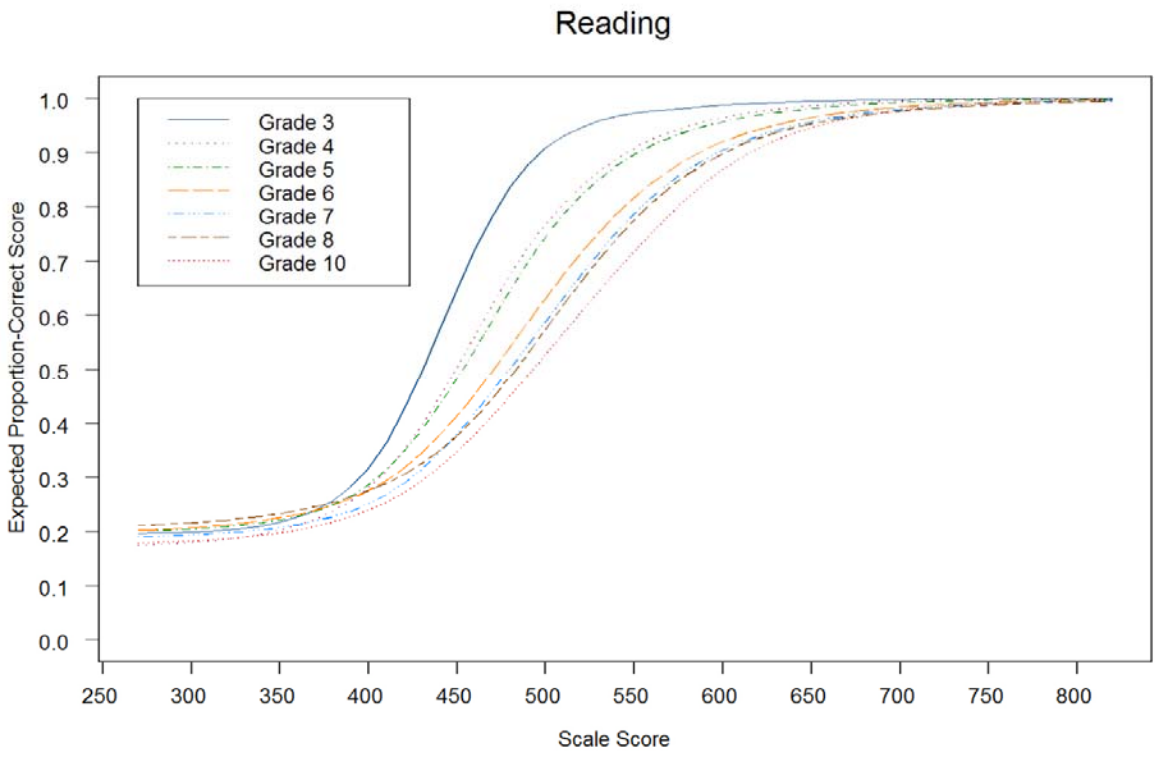
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-5
SEM Curves, Science Grades 4, 8, 10



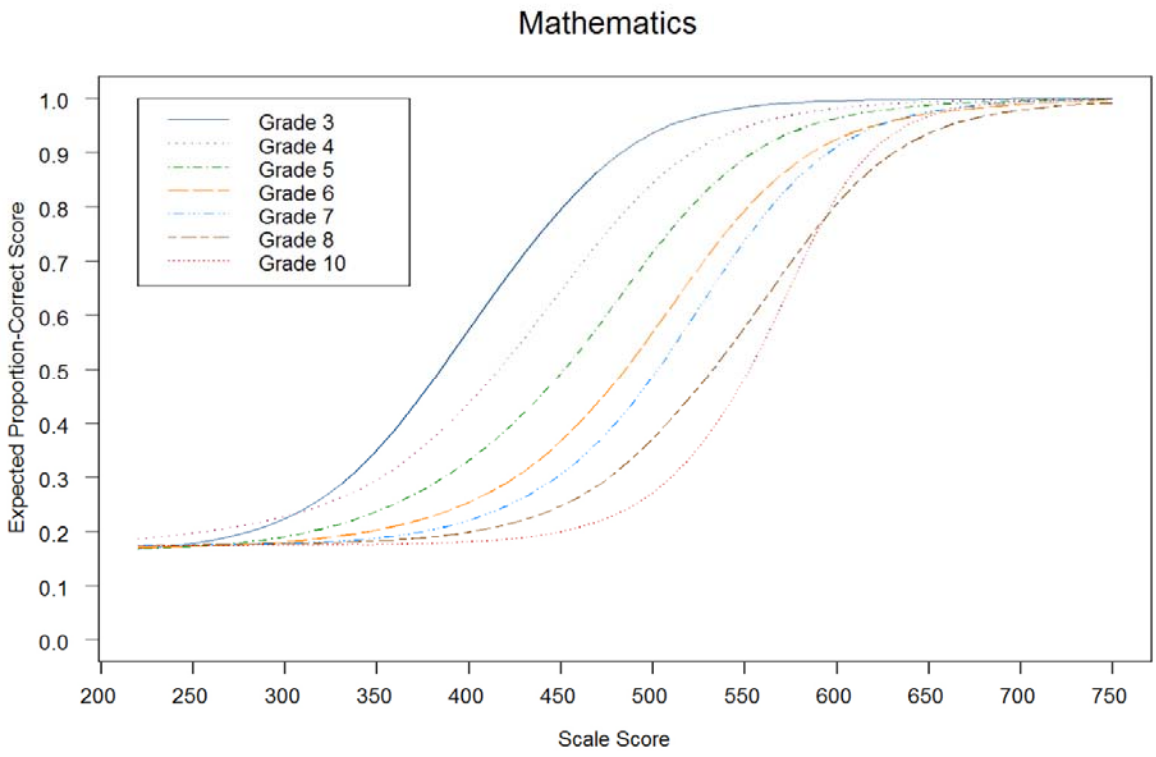
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-6
TCC Curve for Reading Grades 3-8, 10



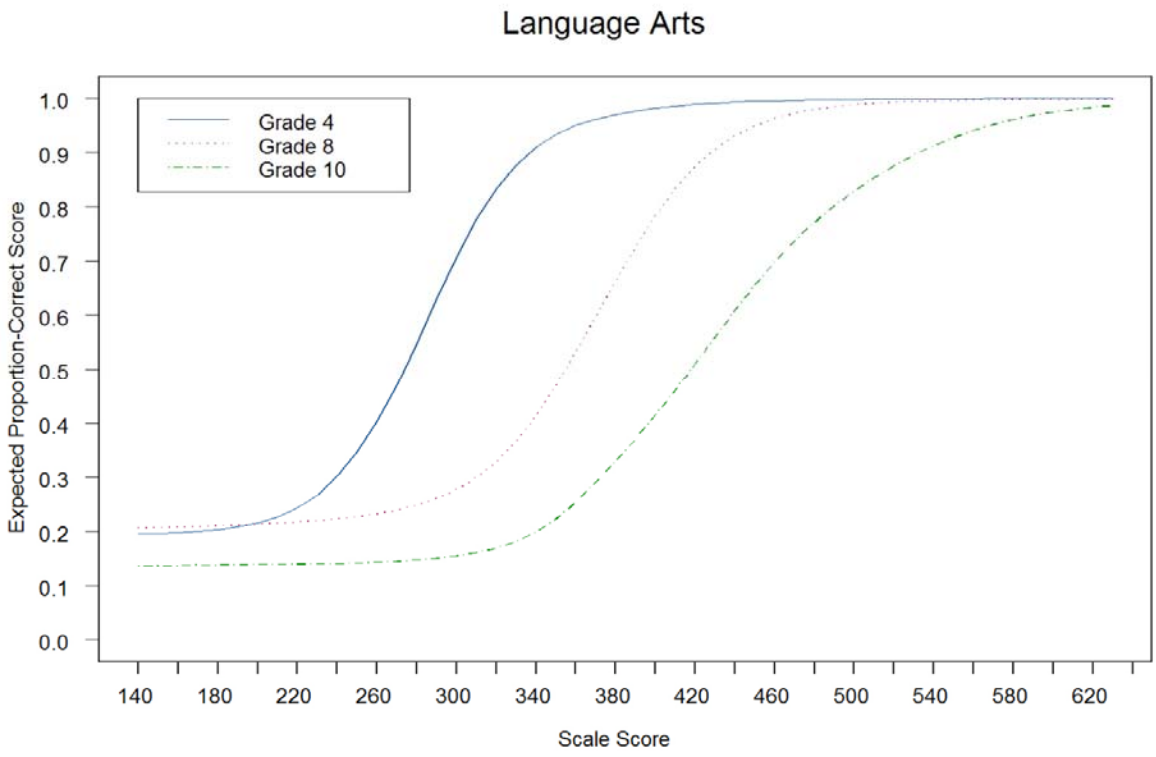
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-7
TCC Curve for Mathematics Grades 3-8, 10



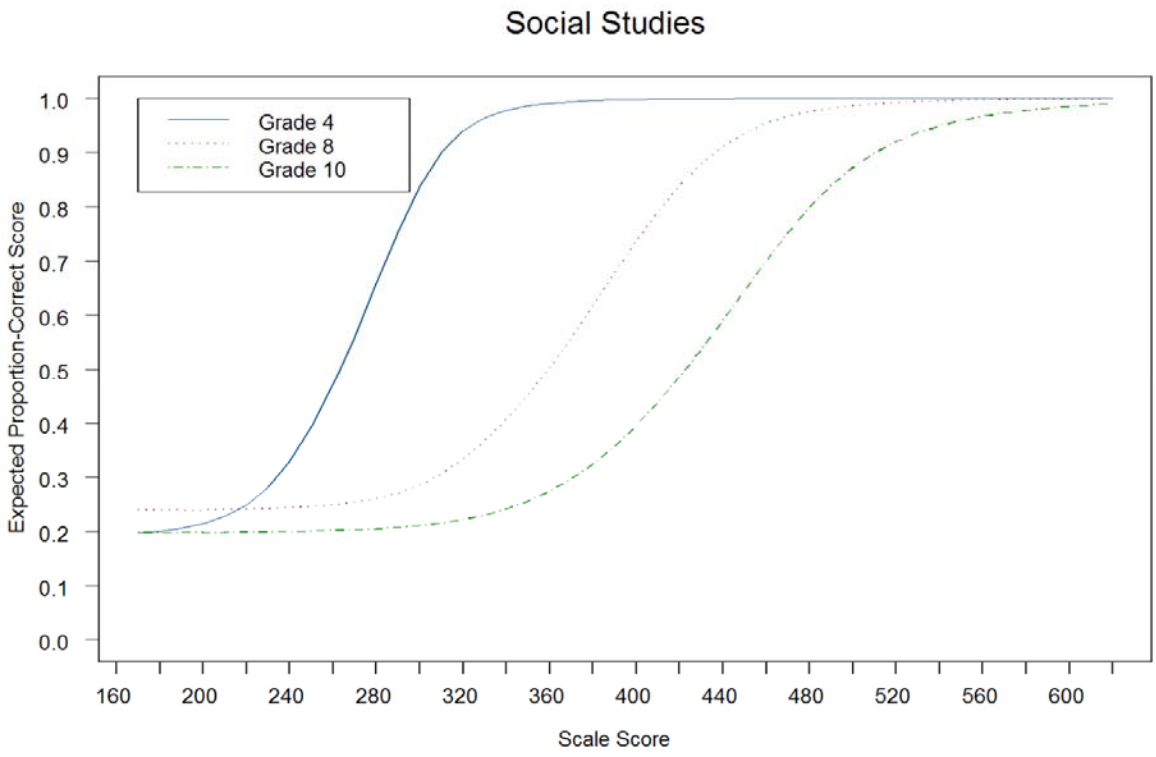
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-8
TCC Curve for Language Arts Grades 4, 8, 10



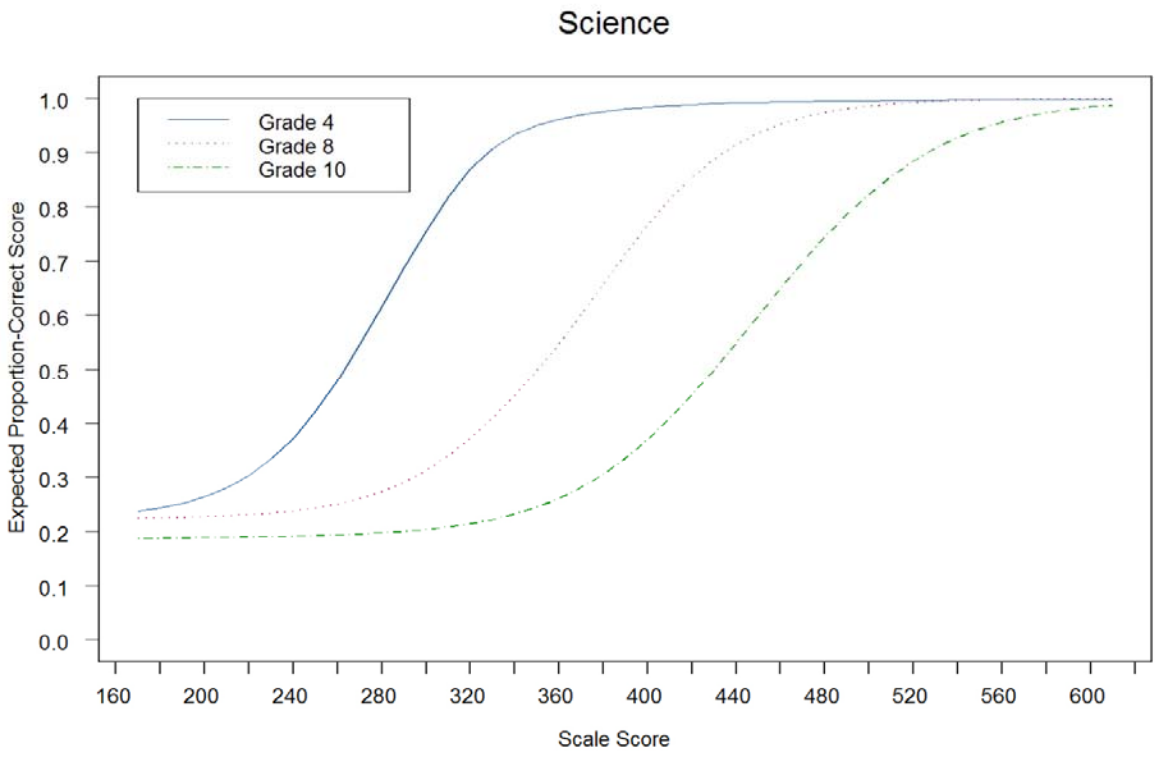
Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-9
TCC Curve for Social Studies Grades 4, 8, 10



Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 7-10
TCC Curve for Science Grades 4, 8, 10



Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 9-1
Reading Indices for Classification Consistency and Classification Accuracy

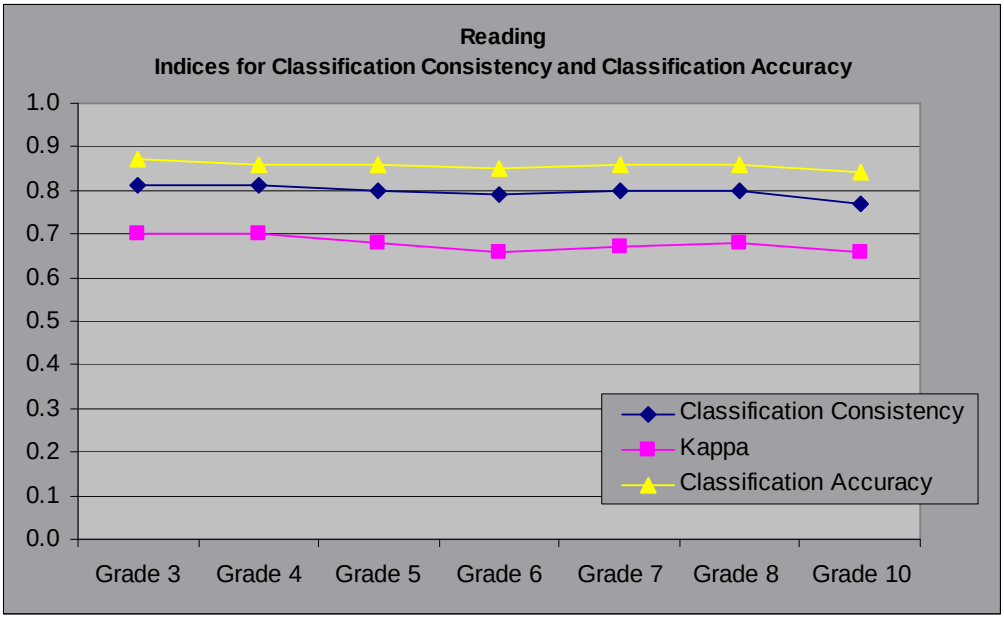
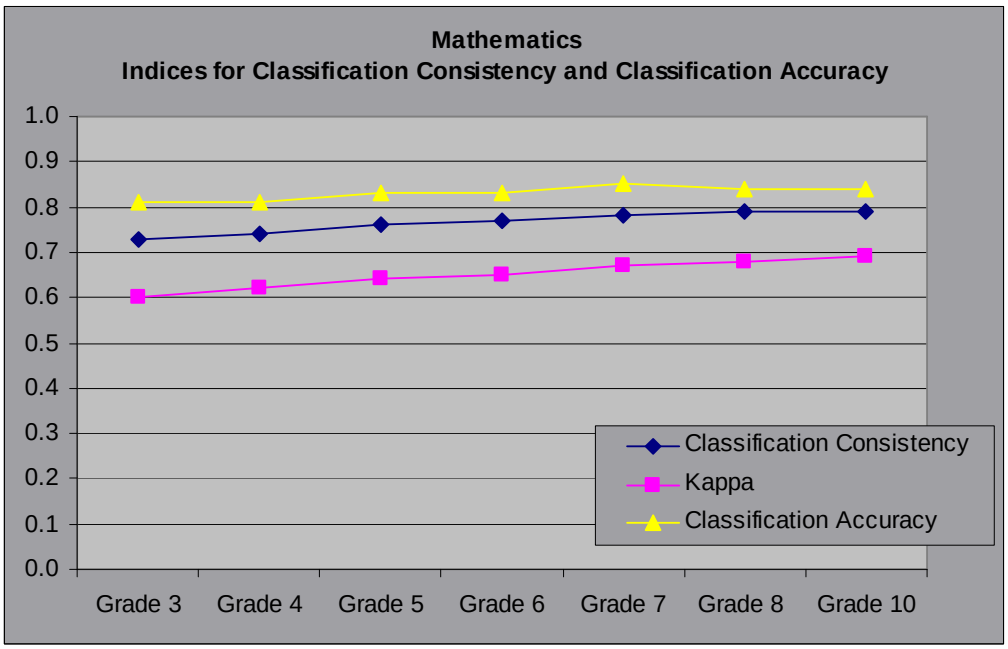


Figure 9-2
Mathematics Indices for Classification Consistency and Classification Accuracy



Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 9-3
Language Arts Indices for Classification Consistency and Classification Accuracy

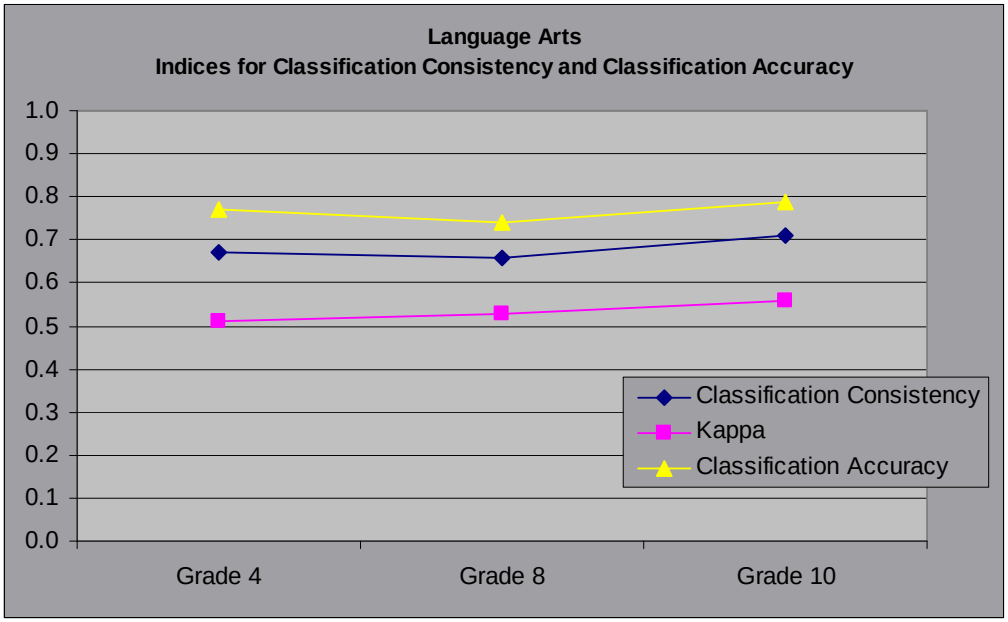
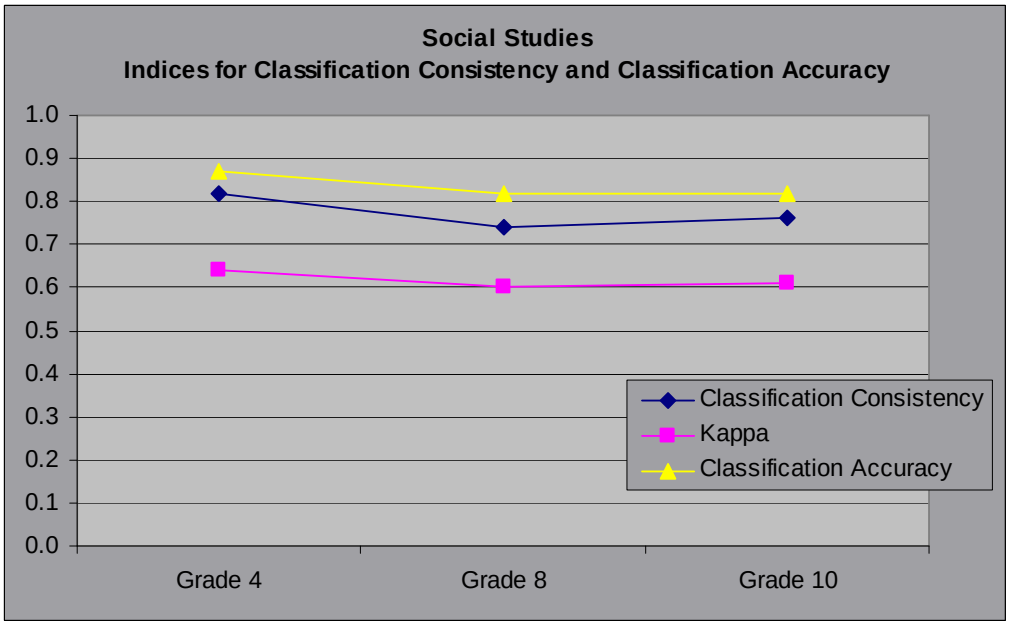
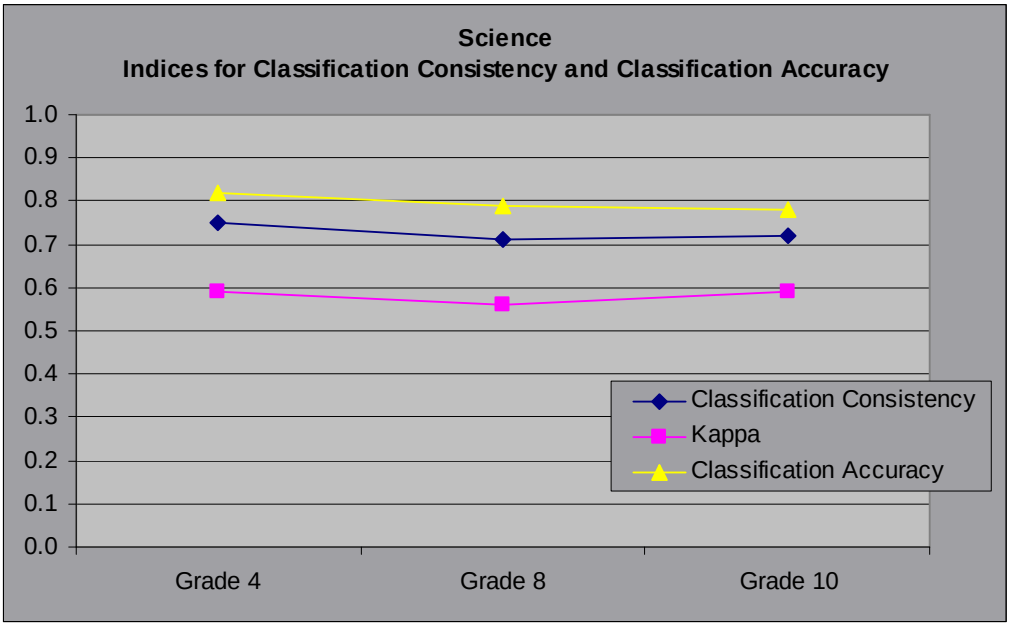


Figure 9-4
Social Studies Indices for Classification Consistency and Classification Accuracy



Copyright © 2011 by the Wisconsin Department of Public Instruction

Figure 9-5
Science Indices for Classification Consistency and Classification Accuracy



Copyright © 2011 by the Wisconsin Department of Public Instruction

Appendix 1: Fall 2007 Item Selection Check-Off Form

WKCE-CRT Form Selection Summary Document

Content Area:

Grade Level:

Test Description

	Form 2006				Anchor Items				Form 2007			
	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												
Total												

Blueprint Comparison (Number of items)

	2006 Blueprint Requirement		2007 Blueprint Requirement		Form 2006		Anchors		Form 2007	
Reporting Category	SR	CR	SR	CR	SR	CR	SR	CR	SR	CR
A										
B										
C										
D										
E										
F										

Copyright © 2008 by the Wisconsin Department of Public Instruction

WKCE-CRT Form Selection Summary Document, Cont'd

Blueprint Comparison (% Number of items)

	2006 Blueprint Requirement		**2007 Blueprint Requirement**		**Form 2006**		**Anchors 07**		**Form 2007**	
Reporting Category	SR	CR	SR	CR	SR	CR	SR	CR	SR	CR
A										
B										
C										
D										
E										
F										

Form 2007 Number of Items by DOK & Objective

Objective	Obj DoK	Dok Level 1	Dok Level 2	Dok Level 3	Dok Level 4
A					
B					
C					
D					
E					
F					

*Combine SR & CR items

Number of Items on DPI Watch List

	Anchor 07	Form 2007
Number of items		

Number of easy and difficult items for preventing ceiling and floor effect

	Form 2006		**Anchors 07 (SR only)**		**Form 2007**	
	SR	CR	SR	CR	SR	CR
Mean *p*-value						
No. of items: *p*-value < 0.30						
No. of items: 0.30 < *p*-value < 0.40						
No. of items: *p*-value > 0.80						
No. of items: 0.80 < *p*-value < 0.90						

Copyright © 2008 by the Wisconsin Department of Public Instruction

WKCE-CRT Form Selection Summary Document, Cont'd

TCCs overlay each other closely

	2006 & Anchor	**Anchor & 2007**	**2006 & 2007**
TCCs of Selected Form			

SE curves are smoothly bow-shaped without dips, bumps, and twists

SE curves of Selected Form	

Expected % Max. RS difference between any two Selected Forms < or = 0.05:

	2006 & Anchor
Max Raw Score Difference	

Copyright © 2008 by the Wisconsin Department of Public Instruction

WKCE-CRT Form Selection Summary Document, Cont'd

Number of Items with DIF

	DIF Code (Ag/Fav)	**2006** Against/favor	**Anchor** Against/favor	**2007** Against/favor
Gender	F			
	M			
Race/Ethnicity	White			
	African American			
	Hispanic			
	Asian			
	American Indian			
ELP	Proficient			
	Not Proficient			
SES	Disadvantaged			
	Not Disadvantaged			
Disability	Disabled			
	Not Disabled			

Number of Items with Poor Fit

	2006	**Anchor**	**2007**
Fit = 3			

Approval

Content Editor	Content Supervisor/Lead	Date	Project Manager	Date	Research Monitor	Date

Copyright © 2008 by the Wisconsin Department of Public Instruction

WKCE-CRT Form Selection Summary Document, Cont'd

Plots

- **2006 Form (Reference) vs. Anchor (Working) Set:**
  (TCC & SE curves & Summary Report)

- **Anchor (Reference) vs. 2007(Working) Set:**
  (TCC & SE curves & Summary Report)

- **2006 (Reference) vs. 2007 (Working) Set:**
  (TCC & SE curves & Summary Report)

- **2006 Form, Anchor Set, 2007 Form TCC & SE Curves:**

Copyright © 2008 by the Wisconsin Department of Public Instruction

Appendix 2:
Fall 2006 WKCE Technical Report, Parts 2, 3, 4, 8, and 12

Copyright © 2009 by the Wisconsin Department of Public Instruction

Part 2: Involvement of Wisconsin Educators

Standard 3.5 of the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) advises that relevant experts should review test specifications; that the nature, processes, and results of their reviews should be documented; and that relevant background information on the reviewers should also be documented. Part 2 of the Technical Report speaks to this standard by documenting the nature, processes, and results involvement of Wisconsin educators (and WDPI) in the test development process, as well as providing their background information, when available. As described below, Wisconsin educations were directly involved, for example, in the development of the test blueprints, item selection, evaluations of fairness and bias, reviewing the passages used for CR items, establishing cut scores and performance levels, and developing plain-language descriptions of the performance levels. The role of Wisconsin educators was an essential component of the development of the WKCE, because the professional expertise and judgment they provided plays a major role in providing content-related validity evidence in test development.

2.1 Establishing Test Content

According to the most recent edition of the *Standards for Educational and Psychological Testing* (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 tasks 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 some 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. While validity evidence is necessary to *support* inferences from test scores, responsibility for the validity of the actual *use* of the test scores lies with the person or agency using the test scores. Throughout the test development process, Wisconsin educators have provided the expert judgment to develop the Wisconsin Knowledge and Concepts Examinations–Criterion-Referenced Tests (WKCE) and interpretations of the test results that are consistent with each other.

Involvement of Wisconsin educators began in August 2003 with a four-day meeting to establish the content framework, eligible test content, and test blueprints. At this workshop facilitated by CTB and WDPI staff, approximately 80 educators (6 per content/grade group for Reading and Mathematics, grades 3–8 and 10) from throughout the state examined the Wisconsin Model Academic Standards at grades 4, 8, and 12 and considered what summative test information would be useful on a test report for each grade and what test reporting categories and subskills would be most informative. Because content standards exist only for grades 4, 8, and 12, the committees carefully considered what knowledge and skills students should have by

the fall of each school year by using expert judgment to extrapolate and interpolate the standards to grades 3, 5, 6, 7, and 10.

The agenda for the Assessment Limits Meeting is presented in Table 2-1. Following the meeting, CTB completed the editing of the test specifications documents drafted during the meeting, reviewed them for content clarity and articulation across grade levels, and submitted them to WDPI for review, comment, and then approval. The summary of the results and outcomes of the Assessment Limits meeting is provided in Part 3.

WDPI organized and sponsored a meeting of Wisconsin science educators for a science frameworks meeting January 18-20, 2005. The purpose of this meeting was to identify the eligible content based on the Wisconsin Model Academic Standards (WMAS) for grades 4, 8, and 12. Because the WKCE is administered in the fall and the WMAS are end-of-year standards, the educators needed to identify the specific assessment limits that would be appropriate for a fall test. For grade 10, the Science Frameworks Committee needed to determine what content knowledge students should have by the beginning of grade 10 by using professional judgment to interpolate the grade 8 and grade 12 Model Academic Standards. The Reading and Mathematics content frameworks had been established by 2005; only the Science frameworks for grades 4, 8, and 10 were developed during 2005.

2.2 Writing and Developing Assessment Materials

Wisconsin educators were involved in selecting Reading passages and reviewing test items prior to field testing. Reading passage review meetings were held in September 2003, December 2004, and August 2005. This section describes the participation of Wisconsin educators in the process of selecting reading passages. Additional information about the outcome of passage selection meetings is presented in Part 3.

The September 2003 meeting was the first passage review meeting and was for the purpose of selecting the passages necessary for the May 2004 and December 2004 administrations to generate the number of passages needed for three operational forms per grade level. The Reading Passage Review Committee consisted of approximately 35–40 Wisconsin educators from throughout the state. Members were selected by WDPI staff to achieve balance and representation from across the state. The committee was subdivided by grade level, with 5–7 educators per grade for grades 3–8. The meeting agenda is provided in Table 2-2. The passage review criteria were used for the first and subsequent passage review meetings.

The December 2004 passage review meeting was for the purpose of selecting passages to develop and field test in fall 2005, and the August 2005 meeting was to select passages to be developed in 2005 and field tested in fall 2006. For each of these meetings, 16–20 educators participated and were divided into two groups, a group for grades 3–5 and a group for grades 6–8, consisting of 8–10 educators in each group.

The Reading Passage Review Committee members discussed the passages presented for consideration for each grade level and recommended which passages should or should not be used for the development of new items. The committee's deliberations and discussions addressed

the interest level of the topic, grade appropriateness of vocabulary and graphics, and accessibility of the text to a diverse student population. Occasionally, the committee recommended that a passage be used at a different grade level than the grade for which it was submitted for review. CTB made final recommendations to WDPI regarding which two passages at each grade level should be developed for the fall 2005 field test, taking into consideration the number and type of passages already in the item pool and what types of passages were needed to build future operational forms (beyond the three forms required per the contract).

Another significant way in which Wisconsin educators have been involved in the development of the WKCE has been as item reviewers at item content review and selection meetings. Since the initiation of the contract with CTB, six meetings have occurred:

- December 2003: review of items to field test in May 2004
- April 2004: review items for field testing in December 2004
- May 2004: review and realign grade 10 reading and mathematics items
- March 2005: review items for embedded field testing in fall 2005
- November 2005: review items for embedded field testing in fall 2006
- January 2007: review items for embedded field testing in fall 2007

For each meeting, CTB staff provided orientation and training regarding the purpose of the review and the importance of the committee members' role in the item review meeting, with emphasis on how their professional judgment contributes to establishing the content validity of the assessments. The whole-group training took approximately one hour at the beginning of the first day. When participants reported to their content area and grade level group, the CTB facilitators provided additional content and grade specific information regarding the nature and scope of the review tasks. Participants were provided review guidelines, which were continually referenced as the items were reviewed and discussed.

The content review meetings served the purpose of substantiating that there is categorical concurrence between the content specified in the Assessment Frameworks document and the content represented by items in the pool of items developed for the assessment. As the educators reviewed the items, they considered that if a content topic, as defined by subskills and assessment limits within the Assessment Framework, is both broad and complex, then the items should also represent that depth and complexity and not just those aspects of the topic that are easiest to assess. Thus, great attention was given to content integrity and eschewing items that measure obscure facts, minor details, or ideas and skills that are of marginal importance to the overall acquisition and demonstration of content knowledge and skills. In both the development and review of the items, attention was given to ensuring that the items in the pool represented a range of cognitive skills and depth of cognitive complexity. That is, items should require analyzing, comparing, summarizing, concluding, inferring, evaluating, etc., and not solely recalling information.

CTB trained the Wisconsin educators to consider items for fairness and sensitivity issues in addition to content considerations. The PowerPoint presentation used for the whole-group training on the first day of each review meeting addressed the definition of test and item bias and possible factors that may contribute to systematic error of measurement on individual items for

groups of examinees sharing similar demographic characteristics, after controlling for overall performance on the test. CTB provided the review checklist (see below) and explained that the purpose of the fairness review is to identify any construct-irrelevant factors that might plausibly prevent members of a group of test takers from demonstrating that they possess the knowledge and skills being measured. As each item was reviewed for content, the committee participants also reviewed the item for fairness and sensitivity, using the *Checklist for the Sensitivity Reviewer*, and marking on the review form whether each item was acceptable or not acceptable. Ensuring that items do not have construct-irrelevant features also contributes to establishing content-relevant validity evidence. The review guidelines for both the content review and sensitivity review are presented in Tables 2-3 and 2-4.

The first item review meeting was held December 3–5, 2003 for the purpose of reviewing the items to be field tested in May 2004. New items for Reading and Mathematics for grades 3–8 were reviewed. Approximately 36 Reading educators and 36 Mathematics educators participated, with six educators assigned to each grade level group. The second review meeting was held April 28–30, 2004 to review items to be included as embedded field test items during the December 2004 forms calibration administration. Again, approximately 36 Reading educators and 36 Mathematics educators participated, with six educators assigned to each grade level group.

On May 25–27, 2004, a Reading and a Mathematics committee met to review the grade 10 items that were developed for the High School Graduation Test (HSGT) secondary item bank. The items in this bank had been developed in 2000 and field tested in February 2001 with grade 10 students. This item bank had been developed for the HSGT, which was to be administered in the spring of grade 11. The item bank was now being repurposed for use in the fall for the grade 10 test. Therefore, it was necessary to involve Wisconsin educators to review the item pool to determine which items would be appropriate for a grade 10 assessment. Approximately 8–10 educators for each content area reviewed the approximately 600 Mathematics items and 350 Reading items and made recommendations for which items should and should not be used based on opportunity to learn the content represented by the items by the end of grade 9. The items had been through previous content reviews and had been field tested; therefore, the committees were making recommendations for use of the items for the WKCE and were not recommending editorial revisions to the items. This meeting, and the recommendations regarding item use, contributed to establishing the content validity of the grade 10 test by assuring the consistency between the grade 10 content framework and the test items as appropriate and representative samples of the content domain.

An item review meeting was held March 8–10, 2005 for Reading, Mathematics, Language Arts, Science, and Social Studies. The committees for Reading and Mathematics reviewed newly-developed items for embedding as field test items on the fall 2005 operational test. The Reading and Mathematics committees consisted of approximately 24 educators for each content area. The committee for each content area was subdivided into two groups of 12 each: grades 3–5 and grades 6–8. The committees for Language Arts, Science, and Social Studies for grades 4 and 8 consisted of 8 educators per content and grade level. These committees reviewed *TerraNova* items and aligned them to the WMAS. The Language Arts, Science, and Social Studies content area committees reviewed and aligned items from *TerraNova* Complete Battery (Language Arts, Science, and Social Studies), Second Edition, levels 13, 14, 17, and 18. The Grade 10 committees for Language Arts, Science, and Social Studies consisted of 8 educators

per content area. The grade 10 committees reviewed custom items for these content areas that were previously developed for the Wisconsin High School Graduation Test and to identify which items would be appropriate for use on a fall test. The Science committees also finalized test blueprints. Wisconsin does not conduct a separate review for bias and sensitivity; however, as part of the training for the item content review, the Wisconsin educators are provided with guidelines for reviewing items for sensitivity issues.

An item selection review meeting was held November 30–December 2, 2005 to review draft items that, if approved, were eligible for field testing in fall 2006. The committees consisted of 12 Wisconsin educators for Reading, 12 for Mathematics, and 14 for Science. Table 2-5 identifies the CTB, WDPI, and Wisconsin educators who attended the November, 2005 meeting. The Reading and Mathematics committees were subdivided into two groups with six per group: grades 3–5 and grades 6–8. The Science committee was divided into two subgroups: one group for grade 4 and one group for grade 8 and 10. Draft items in Reading and Mathematics for grades 3–8 and Science for grades 4, 8, and 10 were reviewed. The results of the November 2005 item review selection are presented in Table 2-8.

In January, 2007, an item selection review meeting was held to review, revise, and approve items for field testing in fall 2007. Table 2-6 identifies the CTB, WDPI, and Wisconsin educators who attended this meeting. Table 2-9 presents the results of the review meeting. The processes, procedures, discussions, and results of these review meetings are described in greater detail in section 4.2.2 of this report.

Following the May 2004 field test administration, an Item Functioning Review meeting was held October 15, 2004 to review field test items that were flagged for differential item functioning (DIF). The committee consisted of approximately 20 participants, representing educators (K-12 and post-secondary) and community representatives.

CTB prepared materials for the meeting that included the 78 Reading and 58 Mathematics items across grades 3–8 from the May 2004 field test that were flagged for DIF. The review materials included the items and a spreadsheet listing each item, the content objective and subskill measured, the item's scale score location, and the item's DIF flag.

CTB Research staff conducted training using a PowerPoint presentation that included the following topics:

- item development processes aimed at minimizing the possibility of differential item functioning (i.e., adherence to guidelines for bias-free publishing)
- potential sources of bias in test items (i.e., stereotypes)
- definition of bias (i.e., systematic error of measurement)
- evaluative procedures used to minimize DIF
- statistical methods used to detect DIF (i.e., Linn-Harnisch and Mantel-Haenszel)
- number of students in Wisconsin at each grade included in the field test sample by ethnicity and gender subgroup
- gender and ethnicity DIF flags and how to interpret the flags

CTB subject matter experts then facilitated the review of items flagged for DIF and the committee discussion of the items. The committee's task was to review each item flagged for DIF and to propose hypotheses for why the item may have been flagged for DIF and make a recommendation to either keep the item in the item pool or avoid using the item on operational forms. The committee's recommendations regarding item use were considered along with the Technical Advisory Committee's guidance on use of items flagged for DIF. CTB provided WDPI with spreadsheets documenting the committee's comments and recommendation for each item.

Another way that Wisconsin educators were involved in the test development process was by participating in range finding activities for constructed response items. Range finding is the process of identifying samples of student responses that represent "solid" performance for each score point on the rubric as well as sample responses that represent the upper and lower boundaries of performance for each score point. A few Wisconsin educators for reading and mathematics traveled to Mather, California in February of 2005, 2006, and 2007 for range finding activities. These educators contributed their professional judgment to review sample student responses to determine which responses should be used as anchor papers or as training papers when training the scoring personnel.

2.3 Descriptor Writing

In February 2003, Wisconsin educators participated in setting standards for Reading, Language Arts, Science, and Social Studies for the Grade 4, 8, and 10 tests. CTB Research staff facilitated the meeting and implemented the Bookmark method of standard setting. Each grade/content committee consisted of approximately 24 educators from throughout the state. WDPI selected committee members to be representative of the student population in Wisconsin. Following the administration of the fall 2005 WKCE, CTB researchers used a linear interpolation process to set cut scores for Grades 3, 5, 6, and 7. Committees of Wisconsin educators were convened June 20–22, 2006 in order to develop performance level descriptors to accompany the performance standards for Reading and Mathematics for grades 3–8 and 10. Each grade/content area committee consisted of 5–7 Wisconsin educators. Descriptor writing provides plain-language description of the content students must know at each grade level to be *Proficient*. This information may be used by teachers and the public to fully understand the performance levels on the WKCE. The descriptor writing is described in detail in Part 11.

Part 3: Test Design

The degree to which valid inferences can be made from a test is closely linked to our confidence that the appropriate content is included on the test, and that the test tasks adequately sample the domain of content knowledge from which inferences about students' performance are made. Part 3 of the Technical Report describes how CTB, WDPI, and Wisconsin educators collaborated through a series of test development and design processes to in order to ensure that the appropriate content was included in the WKCE, and to ensure that the test tasks adequately sampled the domain of content knowledge pertinent to making legitimate inferences about student performance. A series of workshops, exercises, meetings, and design activities went into developing the WKCE. Part 3 documents the step-by-step workings of that entire design process. Part 3 reviews the instructions provided to the workshop participants, the principles that guided their work processes, the guidelines employed, and the results of their work. As described below, the *Wisconsin Model Academic Standards* were central to the entire test design process. Part 3 of the Technical Report demonstrates WCKE-CRT adherence to AERA/ APA /NCME standards 1.2, 1.6, 3.1, 3.2, 3.3, 3.11, 6.4, 6.15, 13.3, and 13.5.

3.1 Test Specifications

3.1.1 Content Framework and Assessment Limits

Appropriate identification of a test's content is critical, as the underlying content both defines and limits the inferences that can be derived from the test results. The degree to which valid inferences can be made from a test is closely linked to confidence that the appropriate content is included on the test and the test tasks (items) adequately sample the domain of content knowledge from which inferences about students' performance is made. Therefore, defining the assessment framework for Reading, Mathematics, Science, Language Arts, and Social Studies was a key task in the development of the WKCE.

Wisconsin has state standards at grades 4, 8, and 12. These standards, called the Model Academic Standards, are benchmark, end-of-year standards. Because the WKCE tests are administered in the fall at grades 3–8 and 10, it was first necessary for CTB and WDPI to collaborate to establish the grade-level Reading and Mathematics content to be assessed at each grade. For the grade 4, 8, and 10 Science assessments, the Model Academic Standards served as the foundation for the creation of the Science Assessment Frameworks. The Model Academic Standards for Language Arts and Social Studies provide the content framework for these content area tests at grades 4, 8, and 10.

The following principles guided the test development process to establish the assessment framework and assessment limits for WKCE tests:

- provide valid, equitable measurement of achievement;
- offer multiple ways of measuring student progress;
- give information useful for improving student's understanding of key concepts;
- engage and motivate students so they will perform their best work; and reflect current curricula and state standards.

The assessment framework specifies the broad categories within the content area at which test subscores may be reported, for example, Number Operations and Relations, Geometry, and Measurement for Mathematics and Understands Text and Analyzes Text for Reading. These broad categories are further delineated into subskills. For example, Number Operations and Relations is further defined as place value; reading, writing, and representing number, ordering and comparing numbers, and so forth. Assessment limits are even more granular and specify the specific content that is eligible for testing at each grade level and may clarify how the content may or may not 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.

The Assessment Framework documents created by WDPI provide information about the content measured at each grade level and explains the relationships among the Model Academic Standards, the Assessment Framework, and classroom instruction. The Framework documents are located on WDPI's website at http://www.WDPI.wi.gov/oea/wkce.html.

What is the framework?

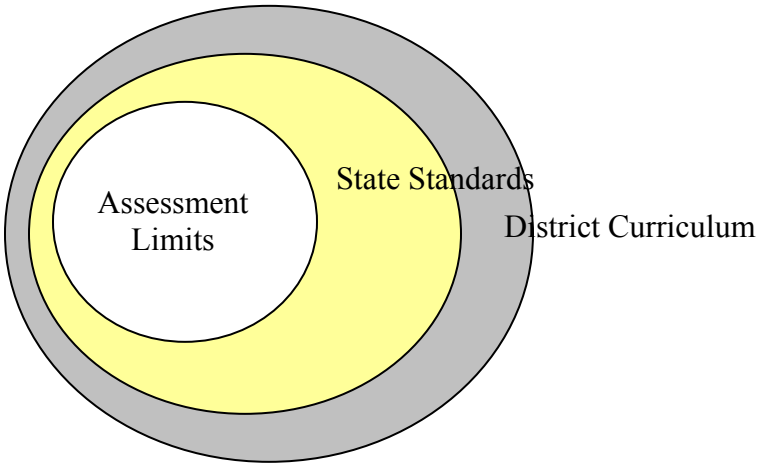
Establishing the content framework and eligible test content for Reading and Mathematics began in August 2003 with a workshop with Wisconsin Educators. This meeting is described in detail in section 2.1. At this workshop facilitated by CTB and WDPI staff, educators considered what summative test information would be useful on a test report and then designed the test reporting categories and subskills backward. Because the Wisconsin Model Academic Standards are only for grades 4, 8, and 12, the committees carefully considered what knowledge and skills students should have by the fall of each school year. They used professional judgment to extrapolate and interpolate the knowledge and skills needed from the standards for 4, 8, and 12. Committees then defined the eligible test content and assessment limits, ensuring the test framework they designed incorporated the content and performance standards enumerated in the Wisconsin Model Academic Standards. The Wisconsin Model Academic Standards for grades 4, 8, and 12 were used as the starting point and foundation for establishing the grade-specific content frameworks. Professional judgment was paramount in making decisions about what content knowledge and skills students at each grade level should have mastered at the beginning of the school year in order to be successful with the content taught at each grade. Throughout the process, the committee members referred to the Model Academic Standards to verify there was a clear connection between the content frameworks they were creating and the Model Academic Standards.

CTB provided the participants with instruction and guidelines for writing clear, precise assessment limits. These guidelines were:

Characteristics

- define the eligible test content and the upper limit of complexity and difficulty of assessable content
- address the content knowledge and skills that students need by the end of the grade level in order to be successful at the next grade level

- are a subset of the state standards, focusing on what can be assessed by a large-scale, paper-and-pencil test



- are **measurable** by multiple-choice and/or constructed-response test items
- are **comprehensive** (cover and sample the content domain) and represent a common core of high academic expectations for all students, no matter what school they attend are pedagogically definite, leaving no question as to what the boundaries of the testable content are
- address the range of content difficulty at the grade level—not just minimum competency, so that performance at basic, proficient, and advanced levels can be discerned from test results
- are of increasing intellectual difficulty at each higher grade and cover all important indices of learning in the content area
- are linguistically **unambiguous** and clearly state (without jargon) the specific content that can be addressed by test items

Sample Limits

Unclear:
Apply principles, concepts, and strategies from various strands of mathematics to solve problems that originate within the discipline of mathematics or in the real world.

Analyze the properties of plane geometric figures.

Identify literary elements in fiction passages.

Clear:
Given a pattern of whole numbers between 0 and 999 consisting of not more than 8 elements and not more than two operations (+, –, x), predict the next two numbers in the sequence.

Identify or describe parallel or perpendicular lines or line segments in plane geometric figures or pictures of shapes and figures.

Identify and describe major and minor characters, setting, problem, plot events, and solution in literary passages.

During the August 2003 workshop, the Wisconsin educators reviewed the assessment limits to determine which could be efficiently and effectively measured using multiple-choice items and which were best measured using constructed-response items, and made recommendations regarding how much emphasis should be given to each content standard on any given test form. The WKCE tests sample commonly taught processes, skills, and knowledge. They do not measure all of the skills that make up an educational domain. The outcomes of the workshops were the test framework for each grade and content area. Following the workshop, WDPI conducted follow-up meetings with educators to refine and articulate the content and subskills in the test framework across grade levels. The content frameworks established at the August 2003 meeting were then used to create the test blueprints for each content area and grade level.

Establishing the content framework and eligible content for Science began in January 2005 with a workshop hosted by WDPI with Wisconsin educators. At this workshop, facilitated jointly by WDPI and CTB staff, educators considered what content was intended by the Model Academic Standards and designed a framework to serve as a guide for educators throughout the state. At future meetings in March 2005 and May 2005, educators worked to finalize their recommendations for the framework and to determine appropriate assessment limits for each grade. A cross-grade comparison was completed to ensure appropriate scaffolding was present and that expectations of students increased appropriately from grade to grade. The frameworks and assessment limits documents were used to create the science test blueprints for grades 4, 8, and 10.

3.1.2 Test Blueprint

For the process of creating the test blueprints at the August 2003 meeting, CTB provided instructions and facilitated the process. CTB researchers provided guidelines regarding the number of items needed to achieve reliable tests. The result was a draft test blueprint that specified the amount of testing time required for each content area, how many score points for each test, how many score points for each content standard, and how many MC and CR test items would be on a form. CTB provided the following set of instructions for determining the test blueprint to participants, which essentially describes the process followed to develop the blueprints.

Test Blueprint Task

The test blueprint is created after the reporting categories, subskills, and assessment limits have been defined. The test blueprint is the recipe for constructing an operational test form. It identifies the total number of score points on the test and the distribution of score points across reporting categories and item formats.

1. In your small group, assign a weight to each reporting category for each grade you are working with. First, do this individually using the Test Blueprint Planner. Discuss your weights as a group and come to consensus (or majority) on the weight for each reporting category. The weight assigned should reflect the importance of the reporting category to the total test. The weight should also reflect the relative emphasis given to the reporting category content during instruction. Use the following weighting system:

3 = High Importance
2 = Moderate Importance
1 = Low Importance

2. The key entry person for the small group should enter the agreed-upon weight for the reporting categories on the Blueprint Planner. The spreadsheet will calculate the percentage of score points and the number of score points assigned. Repeat for each grade level. Print a copy for each person.

3. Based on the information derived by using the group weights, discuss whether the resulting percentage of score points reflects the relative importance of the content covered by the reporting category to the whole test and the emphasis given the content during instruction. Make adjustments as needed. Be sure each reporting category has the minimum number of score points (7). At this time, it is permissible to assign half-point weights if the resulting percentages reflect the reporting category's importance and emphasis.

4. As a large, cross-grade group, review the weights and percentages and discuss your group's rationale for the assignment of weights. The weight assigned to each reporting category may shift as the grade levels progress, reflecting shifts in importance and emphasis during instruction. Make adjustments to the weights as needed.

5. Return to your small group and transfer the weights to the detailed test blueprint form, which also shows the subskills and assessment limits. Next, complete the following steps:

    a. Review the subskills and assessment limits and determine which subskills are **best** assessed using constructed response items. Write "CR" in the column labeled "Allowable Item Formats." Discuss how many CR items are appropriate for each grade level (maximum = 8), given the content that should be assessed using CR items and the developmental level of students at the grade level.

    b. Based on the results of Step a., determine which reporting categories will have constructed response items and how many. Enter the number of CR items in the appropriate column in the row for the reporting category. The spreadsheet will then calculate the number of SR items needed for the reporting category based on the total score points for the reporting category.

    c. Assign a weight to each subskill within each reporting category. The weights reflect the relative importance of the subskill to the reporting category (not the entire test). The spreadsheet will calculate the number of SR items per subskill based on the recommended number of score points and BCR items assigned to the reporting category. Review the results and make adjustments to the weights as necessary.

    d. Repeat the process for each grade level.

    e. Complete the task of assigning Allowable Item Formats for all subskills and assessment limits.

CTB and WDPI then reviewed the draft blueprints with the committee to ensure the tests would provide a balanced measure of the eligible performance standards and yield highly reliable and valid scores and made modifications, as necessary, to achieve appropriate content coverage and balance. Together, the Wisconsin educators, WDPI, and CTB reviewed a variety of sample test items and discussed the characteristics of the types of items that would be best suited for inclusion on the WKCE field test. The test blueprints were then used to construct the May and December 2004 field tests and the fall 2005 operational test.

The following summary of the meeting was submitted to WDPI in September 2003 as documentation of the meeting outcomes.

Objective	Summary of Progress and Next Steps
Define reporting categories for WKCE, reading and mathematics	• Mathematics established six reporting categories, which are consistent with the WMAS content standards for mathematics. • Reading established four reporting categories: Determines meaning of words or phrases in context, Understands Text, Analyzes Text, Evaluates and Extends Text

Establish reporting category subskills for structuring assessment limits for cross-grade consistency	• Reading and Mathematics committees were successful in defining subskill categories that are consistent across grades.
	• A nomenclature for the four levels of the coding structure was determined that avoids confusion with the WMAS performance standard coding:
	  • Reporting Category (Math—A, B, C, D, E, F; Reading—A)
	  • Subskill (e.g., Aa, Ab, Ba, etc.)
	  • Subskill Indicator (e.g., Aa1, Aa2, etc.)
	  • Assessment Limit (2-digit code that identifies both grade level and assessment limit) (e.g., Aa13a, Aa25b)
	• The 5-digit code for the assessment limit may be grade-level specific.
	• The coding structure will fit CTB's PEID system.

Establish assessment limits for reading and mathematics, grades 2–9	• Initially, participants had difficulty understanding the concept of "assessment limits" and how they differ from the WMAS. CTB was able to provide examples to clarify the task. The handout materials included sample grade 3 limits for reading or mathematics, which were helpful in providing the participants with models to follow. CTB facilitators were able to prompt the groups to clarify and specify the limits. • Reading and mathematics committees defined assessment limits for the test at grades 3, 4, 5, 6, 7, 8. Assessment limits are very specific and identify the content that can be measured by test items. Where appropriate, the committees were specific in identifying content to be assessed at each grade, such as which affixes, inflectional endings, rhetorical devices, or the size of numbers to use in computation problems either with or without a context. • Grade 10 assessment limits for mathematics are incomplete, as there was insufficient time. This can be done later, but should be completed by summer 2004. • The reading committee did not review the existing assessment limits for high school, which were developed for the HSGT. • Participants for mathematics grades 7, 8, 10 were primarily high school teachers and were not familiar with instruction at grades 6 and 7. They expressed concern about being able to define appropriate assessment limits for the grade 7 and 8 tests.
Make recommendations for test blueprint by assigning weights, distributing score points and items across reporting categories and subskills, and identifying allowable item formats	• The committees worked first in their grade-level groups to identify the number of selected-response and constructed-response items for each grade level. • CTB provided blueprint templates in Excel that calculate recommended percentage and number of score points for each reporting category based on the number of SR and CR items and the weight assigned to each reporting category to reflect the relative importance. • The mathematics committee reviewed the weights and score point distributions as a whole group and made adjustments to smooth the distribution of score points for each reporting category across all grades. • Both the reading and mathematics committees identified which reporting categories should have CR items assigned, and how many CR items per reporting category. • There was not sufficient time to distribute score points across subskills within each reporting category. Diana Kasbaum completed this task for mathematics

	by 09/05/03. CTB Reading content supervisors are completing the blueprints and making recommendations to WDPI for distribution of score points across subskills with the reporting categories. • In order to complete the blueprints, CTB and WDPI will need to identify allowable item formats for each subskill or subskill indicator.
Review draft item specifications and make recommendations for refinements	• CTB brought draft reading and mathematics item specifications to the meeting. While the committee members may have reviewed them, there was not sufficient time to review and analyze them in detail. • The mathematics committee recommended that the constructed response items consist of two steps. Step A will be scored 0–1 and the score point will be reported to category B, C, D, E, or F. Step B will require the student to explain or justify their response to Step A and will be scored 0–2 using a generic rubric. • Decisions still need to be made regarding a Mathematics Formula Reference sheet. The secondary item pool was standardized with a reference sheet. *In order to proceed with item development, WDPI will need to determine whether a reference sheet will be provided at any other grade level and, if so, which formulas would be presented.* • The mathematics item specifications have been revised to reflect the CR format and are currently being reviewed by Teresa Hall and will be submitted to WDPI for review and approval. • CTB is working on revising the reading item specifications and will submit them to WDPI for review.
Review draft reading passage specifications and make recommendations for refinements	• The reading committee reviewed the draft reading passage specifications and made suggestions for revision. The passage specifications have been revised and are currently being reviewed by CTB content supervisors and will be submitted to WDPI for review and approval.
Review options for mathematics punch-out manipulatives and make recommendations for each grade	• The mathematics grade-level groups made recommendations for punch-out tools. • CTB provided WDPI with the list of recommended punch-out tools on 09/04/03 for review. Upon approval, CTB Manufacturing will work with the vendor. • Grade 4 recommended the inclusion of the L-shaped Pentomino piece. CTB does not currently have a die for this piece and is investigating cost. • The upper-level mathematics grades inquired about a protractor with a punch-out center. CTB is investigating this. They also inquired about having rulers and protractors on acetate stock. Margie asked CTB Manufacturing about this. CTB has explored this in the past but discovered that the acetate edges are sharp and can cause cuts.
Review CTB House Styles for art and identify WKCE exceptions	• The mathematics groups reviewed the CTB House Styles for art and marked comments on the art style specifications for grades 2–4, 5–7, and 8–12. Margie prepared a one-page summary of the art exceptions and modifications and submitted it to WDPI for review on 09/04/03.
Page Layout Specifications	• Copies of the HSGT General Page Layout Specifications, ELA page specifications, and Mathematics page specifications were left with WDPI staff to review (Sue Grady, Diana Kasbaum, Jacque Karbon. • WDPI and the CTB Art & Production manager will need to discuss whether there will be any modifications to the HSGT page layout specifications for use in the WKCE program. CTB would like to discuss the specifications with WDPI at the Passage Review Meeting Sept 30 – Oct 1, 2003.

Following the meeting, CTB finalized the creation of the following test specifications documents and submitted them to WDPI for approval. Sign-off on all documents was complete

by the end of November 2003. The specifications documents listed below exist as separate documents and are retained by both CTB and WDPI.

- Test Design for May 2004 and December 2004 Administration
- Test Blueprints
- Assessment Limits
- Passage Specifications
- Item Specifications
- Mathematics Manipulatives Specifications
- Art Specifications and Exceptions
- Page Layout Specifications
- Style Guide

3.1.3 Reading Passage Selection

In Part 2, the reading passage review meetings were identified and the role of Wisconsin educators was described. This section provides information about the outcomes of the passage review meetings. To date, three passage review meetings have occurred: September 2003, December 2004, and August 2005.

The first meeting in September 2003 was for the purpose of selecting passages to develop and field test in May 2004 and December 2004. In preparation for the meeting, CTB reading content editors selected 325 reading passages from a variety of literary sources. Reading passages were selected in accordance with the reading passage specifications, which identified the desired characteristics of reading passages for each grade level, such as genre, length, difficulty, and topics. CTB content supervisors reviewed and screened the passages to verify conformance to the passage specifications and that each passage could support test items measuring a broad range of content objectives and subskills. Of the 325 passages initially selected, 283 (approximately 40–50 per grade) were taken to the meeting to present to the reading passage selection committee. The following information summarizes the processes and outcomes of the passage review meeting held in September 2003; this information was submitted to WDPI following the meeting.

Objective	Summary of Progress and Next Steps
Review Reading Passage Specifications to become familiar with passage requirements	• As the first activity, participants reviewed and discussed the passage specifications. Participants asked questions about how passages might be paired on an operational form. Participants did not disagree with anything in the specifications.
Review assessment limits for reading, grades 3–8	• Participants reviewed the assessment limits documents. CTB had posed some queries about some aspects of the limits. Participants discussed the queries and recommended appropriate revisions in response to the queries. • CTB Reading content supervisors collected marked-up copies of the assessment limits from the table leaders and will revise the documents. • CTB will submit revised assessment limits documents to WDPI by 10/17/03.

Read and review reading passages and make recommendations for passages to develop for field testing in May 2004 and December 2004.	• Participants worked in grade-level groups under the direction of a WI educator table leader. CTB instructed the table leaders in how to document individual and group consensus on the passage review form and in how to move passages to another grade level. • CTB tracked the movement of reading passages from grade to grade in the passage database. • CTB will locate a few additional passages to make good passage pairings or to fulfill the requirement of field testing four passages of each type in 2004. CTB will send the additional passages to WDPI for their review.
Review passage recommendations with WDPI staff	• On Thursday, CTB staff reviewed the passage recommendations with WDPI staff. • CTB identified passages to be paired and passages that will appear together in May 2004 field test sessions. • CTB prepared and submitted to Maggie Burke a complete set of passages selected and recommended by each grade level group. • CTB selected 325 passages total and brought 283 to the meeting to present to the committees. • The committee identified 79 passages that were first priority and 29 that were second priority for developing. • Ultimately, 99 passages across grades 3–8 were identified as the passages to develop and field test.

In preparation for the subsequent reading passage review meetings, CTB staff examined the pool of passages and identified types of passages to target for the passage search. CTB is contracted to deliver three operational forms as a result of the May 2004 field testing and December 2004 forms calibration and field testing; on-going field testing occurs each year beginning with the fall 2005 administration, and these embedded field test items can contribute to making additional operational forms. Tables 3-1 and 3-2 present information about the number of passages presented for review at the December 2004 and August 2005 passage review meetings and the results of the committee recommendations. Detailed documentation of the committee comments for each passage and their recommendations were submitted to WDPI following the meetings and are not presented here, because the detailed information identifies the titles of passages in the operational pool, which would compromise test security. From among the passages approved at the August 2005 meeting, CTB identified which passages should be developed for field testing in fall 2006 and obtained WDPI's approval.

3.1.4 Development and Alignment of Items

A staff of professional item writers, many of them experienced teachers, wrote the WKCE test items that first appeared as operational items in fall 2005. Item development for the WKCE operational test forms began with selecting a variety of literary, informational, and everyday text reading passages. The emphasis was on selecting reading passages that are engaging to students and contain appropriate subject matter, but are not familiar to the students (which would create a potential source of bias). Materials were reviewed and approved by committees of Wisconsin educators. See Part 2 for additional information about the participation of Wisconsin educators in the test development process, and see Part 3.1.3 for the results of passage review meetings.

Every item that appears as an operational item on a test form was field tested, either in May 2004, December 2004, or as an embedded field test item in fall 2005. Embedded field test items appear in the test book with operational items that contribute to the students' scores, but they do contribute to the scores. Field test items appear in a separate session, following the sessions containing operational items. All assessment materials were carefully reviewed for content and editorial accuracy by test development specialists and the content specialists at the Wisconsin Department of Public Instruction and Wisconsin classroom teachers. The items that were included as embedded field test items in fall 2005 were written by the CTB content editors, as they have become very familiar with the content frameworks and the preferences of WDPI staff and Wisconsin Educators. All field test items were reviewed internally by CTB supervisors who are familiar with the Wisconsin content frameworks and item specifications. During all item reviews, careful attention was paid on verifying that each item measured the intended objective, subskill, and assessment limit. If there was any misalignment, the item was edited to achieve greater alignment or a different subskill or assessment limit was assigned. Following the internal editorial reviews, the new items were reviewed by committees of Wisconsin educators. The content review committee meetings are described in detail in Part 2.

3.2 Test Blueprints

Tables 3-1 through 3-5 show the blueprint for the operational portion of the fall 2006 tests. In order to report reliable subscores for a reporting category, a guideline of at least six score points per reporting category was used. In addition to the operational Reading and Mathematics items, there were embedded field test items. Section 3.3 provides greater detail about each test.

3.3 Description of the WKCE 2006 Tests

The 2006 test books contained Reading and Mathematics content in a single test booklet at each grade for grades 3, 5, 6 and 7. Test content for grades 4, 8, and 10 were included in two unique test booklets. Reading, Mathematics, and Science content for grades 4, 8, 10 were contained in Book 1, and Language Arts, Writing, and Social Studies content was in Book 2. Tables 3-6 through 3-10 provide the test design for the fall 2006 tests, including the number of operational and embedded field test (EFT) items and the amount of testing time allotted.

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 multiple-choice items were added for content standards not adequately covered by *TerraNova* items. The Grade 10 Language Arts, Science, and Social Studies tests consist of custom items previously developed for Wisconsin. Wisconsin educators reviewed the Grade 10 item pools in March 2005 and identified which items were appropriate for a test administered at the beginning of Grade 10; the items were originally developed for a test to be administered during the spring of grade 11. Only items that were vetted for use on the WKCE were included on the fall 2006 test.

3.3.1 Reading

Table 3-6 presents the Reading test structure. The Reading test for grades 3–8 had six operational reading passages, one each for six types of passages: short literary, long literary, short informational, long informational, poetry, everyday text. The embedded field test session had one or two passages, which could be any combination of the six types of passages. There were four test sessions: three containing operational items and the fourth containing the field test items. Each grade had at least one pair of paired reading passages with a few items that required analyzing or synthesizing ideas in both passages. Each of the three sessions with operational items had approximately 18 multiple-choice items. Two of the three operational sessions included a constructed response item. One of CR items was for the reporting category Analyzing Text, while the other was for the reporting category Evaluate and Extend Text. Each session was allotted 40 minutes of testing time. The field test session for each grade was allotted 30 minutes.

The Grade 10 test consisted of three sessions, with Sessions 1 and 2 having 35 minutes and Session 3 having 40 minutes. Session 1 had 18 MC items, session 2 had 15 MC items, and session 3 had 17 MC items and 1 CR item.

3.3.2 Mathematics

Table 3-7 shows the Mathematics test structure. The Mathematics test for grades 3, 4, and 5 had three sessions with operational items and one session for field test items. Grades 6, 7, and 8 had five sessions—four with operational items and one with field test items. The Grade 10 test had four operational sessions.

The first session at each grade and the first part of the field test session at grades 3–8 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 or the first part of the field test session.

For grades 3–8, there were four different forms. The operational items in all forms were the same, but the embedded field test items differed by form. Grade 10 had one form and did not contain any embedded field test items for Reading and Mathematics. There were a few embedded field test items in Science.

3.3.3 Language Arts

Table 3-8 presents the Language Arts test structure. The grade 4 and 8 Language Arts tests consisted of 24 *TerraNova* multiple-choice items and six custom multiple-choice items that measure Content Standard F, Research and Inquiry. The entire session was allotted 30 minutes of testing time. There was a writing session in grades 4 and 8 that presented an operational writing prompt that had been field tested in 2005. 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 the 30 MC items, and the second session contained the writing prompt.

3.3.4 Social Studies

Table 3-9 presents 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 test. There was one test session at these grades. The Grade 10 test consisted of 50 custom multiple-choice (MC) items developed for Wisconsin. The test was administered in two sessions. Sessions 1 and 2 contained 25 MC items in each session, and each session was timed at 30 minutes.

3.3.5 Science

Table 3-10 presents Science test structure. 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 test and 10 EFT items per form. There were two test sessions at these grades. Session 1 for grades 4 and 8 contained 40 MC items and was allotted 45 minutes. Session two contained 10 MC items and was allotted 15 minutes.

At the March 2005 meeting to review and align the *TerraNova* Science items to the Wisconsin Model Academic Standards, one outcome was that there were no Grade 8 items aligned to Standard A, Science Connections. New item development corrected this void for future years.

The Grade 10 test consisted entirely of custom items developed for Wisconsin. The test was administered in two sessions. Session 1 had 35 MC items and was allotted 40 minutes. Session 2 had 25 MC items and was allotted 25 minutes.

Part 4: Test Development

Part 4 of the Technical Report focuses on key development tasks and issues related to creating the fall 2006 operational test forms and the items that appeared as embedded field test items in the fall 2006 test books. The test development activities described in Parts 2 and 3 explain how specific development processes contributed to substantiating test validity, primarily through the use of expert professional judgment from Wisconsin educators and from CTB test development specialists. The foundation test specifications documents—Assessment Framework, 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 continued to serve as critical guides for the ongoing development and embedded field testing of items. These documents contribute to ensuring that each form of the test accurately measures the content that it should measure and measures that content in consistent and stable ways, thus providing judgmental evidence of the test's validity. Information is provided in Part 4 relating to the following topics:

- a general discussion of CTB's test book creation and editing process;
- a description of the item development process for embedded field test items;
- the process of selecting operational test items;
- the process of developing and selecting field test items;
- the resolution of style and formatting concerns; and
- the process of obtaining customer 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 4 provides details of how these processes were implemented in Wisconsin. This section of the Technical Report addresses the following AERA/APA/NCME 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.

4.1 Overall Test Book Creation and Editing Process

4.1.1 Solution Management

The first segment of test development is Solution Management. During this phase of the development process, the test design documents (item specifications, style guide, blueprints) created at the beginning of the contract are reviewed to determine if any adjustments are needed.

4.1.2 Documents and Materials Development

During the first year of the contract, the test specifications documents were developed through an extended, collaborative process with WDPI and based on the contributions of Wisconsin educators during the August 2003 frameworks meeting (see Part 2 and Part 3). Test

specifications include the test blueprint, passage specifications, item specifications, page specifications, and style guide. Prior to the development of the new field test items, CTB content editors reviewed the item specifications documents and added any additional details or sample item stems based on clarifying discussions with WDPI staff that occurred throughout the previous year. In May 2004, WDPI collapsed two subskills in Mathematics in two areas; this change resulted in a cosmetic change to the test blueprint documents to reflect the change to the content framework. Under Geometry, the two subskills of "Spatial Relationships" and "Transformations" were collapsed into a single subskill called "Spatial Relationships and Transformations." In Statistics and Probability, the subskills "Data Analysis" and "Statistical Models" were collapsed into a subskill called "Data Analysis and Statistics." Some of the multiple-choice score points were redistributed within and across reporting categories. However, because these changes were made prior to the first operational test administration, these changes did not impact the overall test or interpretation of scores. The blueprint, as revised in May 2004, became the test blueprint for the fall 2005 administration.

The project item specifications define the types of items that are used on the test and the particular attributes of each item format. The item specifications provide detailed information regarding the following:

- item type
- content strand, standard, objective, subskills to be measured
- clarification statement of the task students will perform when answering each item type
- assessment limits
- stimulus attributes (stems, graphics, narratives)
- response attributes (general, correct response, acceptable distractors, unacceptable distractors)
- scoring rubric attributes (general or item/task specific)
- sample items

All test material publications were created using Adobe InDesign software and routed through production and editorial reviews using CTB's electronic workflow management software, Monarch. CTB developed standardized conventions for electronic editorial markup in Adobe Acrobat PDF documents, which were applied during the editorial review of test materials. The editorial conventions used by all CTB Publishing staff were developed to ensure clarity and consistency of communicating edits within CTB and with external page production vendors. In 2006, CTB provided WDPI staff with training in using electronic editing markup tools and conventions.

4.1.3 Item Development and Editing

The development of quality test items requires content and assessment expertise and the ability to be creative while adhering to the test blueprint, detailed item specifications, and content limits. The test blueprint and item specifications provide clear direction as items for content framework assessment limits are written and edited. The test blueprint identifies how many multiple-choice and constructed-response items for each reporting category and subskill. The item specifications are detailed prescriptions for how items are to be written and include sample

stems or sample items in order to provide item writers with clear models for acceptable test items. During the first year of the contract, WDPI reviewed and approved the following test specifications documents, which are retained both by CTB and WDPI:

- Test Design for May 2004 and December 2004 Administrations
- Test Blueprint
- Assessment Limits
- Passage Specifications
- Item Specifications
- Mathematics Manipulatives Specifications
- Exceptions to the CTB House Art Specifications
- Page Layout Specifications
- Style Guide

Test items were developed using a template designed to capture all item attribute information and supporting information such as objective, subskill, assessment limit, score points, and content reference documentation. Test items were edited and revised by in-house content editors, content supervisors, style editors, and art specialists before being presented to teachers and state-level administrators for review and approval.

Item development and subsequent test material development were guided by a detailed, multi-module Publishing Process. The Publishing Process provides all publishing staff with a detailed, common set of strategies, procedures, and documentation that governs the production of all test materials. The publishing work modules address test specifications, item development, item reviews, manuscript creation and submission, page production cycles, quality assurance, release to manufacturing, and post-production tasks such as documentation of item attributes and hand scoring support. The result is that, regardless of content area or grade level, all materials are prepared in accordance with the same stringent and exacting standards. Table 4-1 shows the item development process flow used by CTB content development staff when developing test items.

Once items and associated artwork have been created, CTB editors review the items in relation to established criteria. Edits are made to each item as needed to ensure these criteria are met. Criteria for review include the following:

- the item validly measures the intended content standard;
- for selected-response items, distractors are plausible, parallel, and mutually exclusive;
- selected-response items have only one correct response;
- the item adheres to item specifications and content limits;
- the item stem, answer choices, and art are clear and concise;
- the item is of the appropriate level of difficulty;
- the item is bias free;
- the language and content are age appropriate; and
- the content is relevant.

Following content reviews of the items, style editors review the items for grammar, punctuation, and adherence to the customer-approved style sheet and guidelines for the assessments. Items are checked, for example, to ensure that language is clear and consistent within and across items; formatting is in accordance with the agreed-upon type fonts and sizes; words are hyphenated correctly; and techniques used to emphasize words are applied consistently.

4.1.4 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.

During a MIR, a proctor simulates the test administration experience by administering the test to two test takers for each grade and content area using the examiner's manual developed for the project. 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. A side benefit of this review is the possible revision of any unclear items prior to submission to Quality Assurance and the creation of camera copy, thus reducing the number of blue line changes required. The goal of this work module is to ensure all test components are precisely coordinated and free of errors and ambiguities. Clear and error-free materials ensure a smooth test administration and reflect the high professional quality of CTB products and staff.

The purpose of the QA review is to ensure all publishable products meet the high quality standards and expectations of CTB's customers. 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 clearly correct answer to each item. This QA review comes at the end of the process to augment the excellent work that takes place at each stage of the publishing cycle. It is QA's job to find any problems that may have been overlooked by the project team. This review is an important and irreplaceable step in the publishing process.

In addition to the MIR and QA review steps, the WKCE test books also go through Technology reviews to verify the scannable test books are built to meet CTB's exacting scanning and scoring specifications. With each round of page production, CTB production staff view the position of answer choice bubbles to confirm they are "on grid" and will be readable by CTB scanners. In addition, at the second pages stage, all test books are reviewed by Technology specialists to verify that bubbles are on grid, there is no "bubble back-up" that would interfere with accurate scanning, and other scannable page elements are properly placed. The 2006 WKCE test books were submitted to all of these reviews.

4.2 Item Pool

4.2.1 Item Writing

Items for the WKCE tests were written by trained, professional item writers familiar with the test blueprint and item specifications. The operational items on the fall 2005 and fall 2006 test were developed in the first year of the contract and field tested in May 2004, December 2004, or as embedded field test items in fall 2005. Items in the embedded field test sessions were developed in early 2005 and reviewed at the November 2005 content review meeting. Parts 2 and 3 describe the passage and item review processes in detail. Table 4-1 shows how many multiple-choice, constructed-response and total items have been written to date. Item development for the embedded field testing in fall 2006 was, for the most part, distributed across objectives and subskills proportionately per the test blueprints. Table 4-2 shows the number of items field tested through fall 2006. Tables 4-3 through 4-17 show the number of items developed and reviewed during 2005 for use as embedded field test items in fall 2006.

Item development for embedded field testing in fall 2007 occurred during 2006. CTB and WDPI staff reviewed the alignment reports independently and then held a series of conference calls to discuss the implications for item development during the 2006–2007 project year. The first conference call was held on September 30, 2006.

CTB staff examined how the alignment study participants had assigned DOK levels to items and looked for commonalities among items assigned DOK 3 versus 2 in order to determine if patterns emerged. For example, in Reading, it appeared that at the lower grades, DOK 3 was assigned to items requiring inferring a theme or main idea for part of a passage or summarizing a part of a passage. Similar items at the upper grades were assigned DOK 2 and DOK 3 was assigned to items that required drawing inferences over the entire passage.

In early October, CTB and WDPI staff held conference calls to further discuss the alignment results and observations gleaned from a close analysis of the DOK levels assigned to items. These discussions were helpful in determining the focus of item development for items to be field tested in fall 2007. Following these conference calls, CTB editors developed detailed item development plans. In addition to developing new items to meet DOK needs, other goals for item development included adding additional items for certain subskills to increase the item pool, providing flexibility in meeting the blueprint, increasing overall flexibility in selecting items for forms.

The 2007 item development effort focused on writing items at DOK level 3 for specific objectives in response to the Norm Webb alignment study recommendations. Mathematics focused on creating more DOK level 3 items for Statistics and Probability. In addition, CTB developed multiple-choice items for Mathematical Processes in order to implement a blueprint change to include multiple-choice items for this standard. The Reading item development focused on writing items for existing item sets to have more items for objective 3, Analyze Text, at DOK level 3. For both Reading and Mathematics, item development also addressed adding items for other, selected objectives in order to expand the item pool, improve overall content coverage in the item pool, and to increase flexibility when selecting operational forms. CTB editors prepared item development plans, discussed the plans with WDPI during conference

calls, and then submitted the plans to WDPI for approval in early November 2006. The item development plans included a total of 172 Reading items and 156 Mathematics items for a total of 328. Tables 4-18 through 4-21 summarize the item development plans that were submitted to WDPI.

4.2.2 Content/Bias Review

In May 2004, committees of reading and mathematics secondary educators were convened. They reviewed the existing secondary item pool and identified which items would be appropriate for use on the WKCE. Content and Bias Reviews of the new items to be included as embedded field test items were conducted in March 2005 by Wisconsin educators and facilitated by CTB content editors. In addition to committees for Reading and Mathematics, committees of educators were convened for Language Arts, Science, and Social Studies. The committees reviewed *TerraNova* items for grades 4 and 8 and the existing secondary item bank. The purpose of this review was to align *TerraNova* items to the Wisconsin Model Academic Standards and to identify items that would be appropriate for use on the WKCE.

An item selection review meeting was held November 30–December 2, 2005 to review draft items developed in 2005 that, if approved, were eligible for field testing in fall 2006. Draft items in Reading and Mathematics for grades 3–8 and Science for grades 4, 8, and 10 were reviewed. The results of the November 2005 item review selection for each content area and grade level are presented in Tables 4-3 through 4-17.

Each content area session began with an introductory presentation designed to prepare participants for their role in the item review. The general core of the presentation was customized for the Mathematics, Reading, and Science groups in order to provide the most meaningful information and examples for each content area. A list of topics presented during the training session is presented in the Item Selection Review Meeting Summary report presented to WDPI in January 2006. Additional description of the review meetings, the training provided, and the review guidelines provided to the participants is provided in Part 2 and Part 3.

Issues that were addressed during the Mathematics item review session included the desire of the committee and its WDPI representative to “raise the bar” in the lower grades by using higher level mathematical terminology adjacent to the soft math terminology often used in the classroom, e.g., corner (vertices). At the upper grades, it was the general sense of the committee that Strand A logic problems should apply more mathematical logic, rather than logical contexts. Despite discussion on how to measure mathematical logic independent of general logic skills, committee members were at a loss for being able to offer specific ideas for items. This topic needs further exploration and discussion between CTB and WDPI.

Of the 180 mathematics items presented, two were rejected. The review committee rejected a grade 4 brief constructed response item representing Category E Statistics and Probability, because the committee had concerns about students providing mathematical explanations regarding probability and spinners when the spinner is divided into more than four sections. A grade 8 selected response item representing Category A Mathematical Processes was

rejected, because the committee determined this item relied too heavily on the students' reading ability.

In Reading, both grade level groups identified one area of concern that was related to a particular subskill or assessment limit. The grade 3–5 group expressed dissatisfaction with several of the items written to measure *Aa2: Uses knowledge of word structure to determine meaning of words.* One example was a grade 3, Aa2 item that asked readers to identify the meaning of a contraction in a poem. Committee members rejected this item, because they feared items focusing on contractions in poetry would encourage "poor instruction." Subsequent to the review meeting, when the Reading blueprints were revised and reduced for the fall 2006 administration, this subskill was excluded from the blueprint.

In the grade 6–8 group, the discussion of items written to measure tone indicated a need to more clearly define how tone should be assessed. Wisconsin educators in the group seemed to be suggesting a clear preference for items focusing on the overall tone of a passage as opposed to items that focus on more subtle uses of tone within a passage. If items focusing on tone are restricted to passages with a pervasive overall tone, this could have an impact on passage selection.

A total of 17 items were rejected in Reading. The reasons for rejection are summarized according to reporting category.

Determines Meaning of Words and Phrases in Context: Four SR items that focused on determining the meaning of a word in context were rejected, because it was agreed the sentences did not provide sufficient context. One SR item was rejected, because of objections to asking about contractions in poetry.

Understands Text: One SR item was rejected, because it required the use of the word "not" in the stem.

Analyzes Text: A grade 3 SR item requiring students to identify which words rhymed in a poem was rejected, because it asked readers to identify rather than to analyze the author's use of a literary device. Previous to this review, grade 3 items that required students to identify rhyming words in poetry have been accepted. One SR item that focused on determining theme was rejected, because no consensus could be reached on what the theme was. One SR item that focused on analyzing the purpose of a rhetorical device was rejected due to disagreement regarding the rhetorical function of the phrase. Another SR item that focused on identifying the purpose of text features was rejected due to disagreement about the purpose of the convention. One BCR item was rejected, because it was deemed too sophisticated for the grade level; another BCR was rejected for not being rich enough to generate three levels of response.

Evaluates and Extends Text: One SR item was rejected, because of disagreement with the premise the item used to extend the theme of a literary passage to another situation. A similar SR item that focused on extending the scientific concept of an informational passage was rejected as being too difficult. One SR item that asked students to distinguish between important and unimportant details was rejected, because it was believed to emphasize a sensitive issue in the passage. One BCR item was rejected, because it was feared that prior knowledge would put

some students at a disadvantage; another BCR was rejected for not being rich enough to generate three levels of response.

The Science committees rejected four grade 4 items and four grade 8 items out of 314 total items submitted for review. None of the 26 grade 10 items was rejected. For grade 4, the review committee rejected one Standard A Science Connections selected response item, one Standard B Nature of Science selected response item, and two Standard C Scientific Inquiry extended constructed response items. The review committee felt all four of these items used content or scenarios that were above grade level. For grade 8, one Standard A Science Connections selected response item was rejected as the committee felt the scientific model used in the item was above grade level. Three grade 8 Standard D Physical Science selected response items were also rejected. The committee rejected one item that focused on predicting heat transfer, because they had concerns about the use of vocabulary in the answer choices. The other two items were written to objective D.8.10. The review committee decided the content in the assessment limit for the objective was above grade level, so they rejected both items written to that assessment limit.

For the items developed in 2006 for embedding as field test items in fall 2007, an item selection review meeting was held January 11–12, 2007. At the general meeting on Thursday, January 11, 2007, the CTB publishing project manager for the WKCE, provided a PowerPoint presentation to provide background information to the participants and orientation to the review criteria and procedures. CTB and WDPI collaborated in the identification of the training topics. CTB developed the PowerPoint presentation and submitted it to WDPI for review, suggestions for modification, and approval.

The CTB project manager expanded the PowerPoint presentation for the overall training to include more information about the difference between item difficulty and depth of knowledge. The presentation included sample released items illustrating items that were easy, medium, or difficult based on item statistics and low, medium, or high cognitive complexity. The presentation also included an expanded discussion of item statistics and how to interpret the data with sample released items and their data. The training focused on providing guidelines for reviewing items for content, cognitive complexity, linguistic accessibility, and fairness and sensitivity issues. The presentation included many sample Wisconsin public release items to illustrate the concepts. The summary report for the Item Selection Review Meeting presents a complete list of training topics and the PowerPoint slide deck, as well as a full description of the review process and results. Tables 4-22 through 4-33 present the results of the January 2006 Item Selection Review meeting.

In addition to the newly-developed items, CTB editors revised some items that had been flagged for revision and re-field testing and brought them to the review meeting. A total of 210 Reading items and 194 Mathematics items were presented at the review meeting.

For the 2007 review meeting, WDPI created an Access database to capture participants' discussions and judgments of items' content alignment and DOK assignment. A WDPI staff person used a laptop to record information in the database during the committee discussions. As participants reviewed each item prior to discussing the item, participants individually identified an objective and subskill to which they thought the item aligned and the DOK level. Participants

recorded the content objective and subskill and DOK level on their review form and then, in turn, verbally reported the information to the WDPI recorder. In this manner, WDPI collected data regarding the consensus alignment. As discussion or editing ensued for each item, participants could revise their judgment regarding the item's content and DOK alignment.

The concurrence between CTB editors and committee participants in DOK assignment was variable across content areas and grade levels. Based on the survey, participants had greater difficulty assigning a DOK level to items than assigning a content objective/subskill; 52% rated their response to question 6 with a "6" or "7" compared to 68% responding "6" or "7" to question 7. CTB recommends that CTB and WDPI collaborate to investigate ways to better document what types of items for each reporting category or subskill are DOK level 1, 2, or 3 in order to improve inter-rater reliability when assigning DOK. Improved standardization of assigning DOK level will help ensure consistency and stability in DOK assignments and that the WKCE test meets DOK alignment criteria. Consideration should also be given to the need for additional training information for committee participants to clarify what depth of knowledge is and is not. Although the group training information emphasized the distinction between DOK and item difficulty, participants confused the concepts at times.

Some participants reported feeling rushed. Consideration should be given to determining the optimal amount of time for review, given the number of items and the use of the Access database to record participants' DOK and content alignment discussions.

WDPI and CTB should review the information WDPI captured regarding alignment to objectives and DOK and the details of the discussions, so as to document lessons learned and to plan for future item review meetings.

4.2.3 Item Alignment

Throughout the item development and review process, the alignment between the item and the content standard/subskill/assessment limit was checked during each editing phase and again at the content and bias review. An alignment study, arranged by CTB, was conducted with Norm Webb of the University of Wisconsin, Madison using the fall 2005 test forms. The results of the alignment study were reviewed in the TAC Meeting in December 2005 and it was determined that the DOK level of the items should be reviewed for Reading, Math and Science. The TAC recommended, based on the Norm Webb study, that during item development the team should attempt to achieve 65% or more items at DOK 3 or better for a fully aligned test. CTB should utilize the same DOK rating system that Norm Webb employs. CTB and WDPI have validated all DOK levels in the above content areas and confirmed the DOK for all items for fall 2006 item development. In July 2006, a three-day Alignment Analysis Institute was conducted for WKCE math and reading content standards in Madison, Wisconsin. Eight reviewers, including content experts, district supervisors, and teachers. These committees analyzed the agreement between the state's standards and the 2006 assessments for grades 3-8 and 10.

The committee of mathematics educators determined that the alignment between the standards and the assessments was found to be reasonable for four of the seven grades and needs slight improvement for the other three grades (grades 5, 6, and 7). For all seven grades, the

assessments had a sufficient number of items that were adequately distributed among the objectives for each of the six mathematics standards. The main alignment issue was that not a high enough proportion of items had a DOK level, that was the same as or higher than the DOK level of the matching objective. This was the case primarily for one standard (Standard E, Statistics and Probability). Reviewers judged that items corresponding to Standard E had DOK levels of 1 or 2, whereas the DOK levels for the objectives under Standard E were judged to have DOK levels of 2 and 3. About nine or ten items would need to be replaced on each of the assessments for grades 5, 6, and 7 to attain full alignment. The alignment for the other grades was found to be reasonable. Overall, the alignment is reasonable across the grades with the exception of one standard. By replacing a few items with those at a DOK level of 2 or 3, full alignment would be attained.

The Language Arts committee concluded that the alignment between the reading standards from the assessment framework and the 2006 assessments was found to be reasonable for the seven grades. Reviewers judged that the complexity of the standards was high, with 50% of the 12 objectives under the four standards with a depth-of-knowledge level of 3 (drawing inferences, using information beyond the text, drawing conclusions, and analyzing author's purpose, etc.). For five of the seven grades, the alignment between the standards and assessment was found to be reasonable. Three or fewer items would need to be replaced to achieve full alignment at these grades. The alignment for grades 3 and 4 was found to need slight improvement. This is largely due to the relatively high depth-of-knowledge (DOK) levels expected for these grades, 50% at DOK Level 3. For grades 3 and 4, seven to 10 items need to be replaced by items primarily corresponding to Standard 3 (Analyze text) and with a DOK level of 3. Also, for grades 3 and 4, two or three items need to be added or replaced by items that correspond to Standard 4 (evaluate and extend text). A large proportion of the items for grades 3 and 4 were judged to have a DOK level of 2, fairly demanding for these grade levels. Overall, the alignment is considered to be reasonable and the assessment appropriately increasing in complexity over the grades.

A two-day Alignment Analysis Institute for science was conducted July 10-11, 2006. Nine reviewers analyzed the agreement between the Wisconsin Science Assessment Framework standards and the 2005 assessments for grades 4, 8, and 10. Two reviewers were from states other than Wisconsin, two were graduate students in science and science education from other states, and five were from around the state of Wisconsin. The reviewers included science content experts, district and state supervisors of science, and science teachers.

The results from the analyses indicate the alignment between the standards and assessments needs improvement for all three grades 4, 8, and 10. The main alignment issue for all grades is an insufficient number of items on the assessment to produce information on students' knowledge of each of the eight standards and the large number of objectives under these standards, as many as 145 for grade 10. The results from the analysis indicate the assessments of 40 or 64 items had less than six items for five standards at grades 4 and 8 and for four standards at grade 10. The assessments for the three grades also only had items targeting over half of the objectives under a standard for at most one standard. Thus, the assessments did not meet an acceptable level for Categorical Concurrence and Range-of-Knowledge Correspondence for the majority of the standards. The depth-of-knowledge of the items compared to the complexity of the standards for grades 4 and 8 was acceptable, but failed to meet an acceptable level for four of the grade 10 standards. To attain full alignment would

require adding 16 items and replacing nine items at grade 4; adding 30 items and replacing nine items at grade 8; and adding 24 items and replacing 13 items at grade 10. The number of science standards and objectives is large. The assessment could more easily be aligned to the standards if the structure of the standards was modified by consolidating the objectives and standards into a reduced number. The current level of specificity in the standards is not necessary for guiding an accountability system.

During the joint CTB/WDPI planning meeting in September 2006 and again at the December, 2006 TAC meeting, the science alignment results were discussed. It is significant to note that, because the science framework assessment limits were numbered, Norm Webb considered them objectives for the purpose of calculating Range-of-Knowledge alignment. The assessment limits in the Reading and Mathematics frameworks were not numbered and, therefore, not considered to be objectives for the purpose of the alignment study.

A primary purpose and emphasis of the item review meeting conducted in January 2006 was to verify the alignment of each item to an objective, subskill, and assessment limit of the Wisconsin Assessment Framework and to a Depth of Knowledge level. CTB developed the items to target specific objectives, subskills, and depth of knowledge level and documented the alignment of the items to the Framework. However, in order to simulate an external content alignment study, the participants were asked to identify the objective and subskill to which each item best aligned. Therefore, the content and DOK alignment information was not included on the hardcopy item cards in the review books, on the review forms, nor on the item templates projected on screen.

During all previous phases of item development, Depth of Knowledge levels were assigned to WKCE reading and mathematics items using the DOK Framework approved in the item specifications. This DOK framework was developed by CTB in collaboration with Norman Webb of the University of Wisconsin, Madison, but differs somewhat from the framework Dr. Webb uses when conducting alignment studies. At the December Technical Advisory Committee meeting, the TAC advised WDPI that if the alignment of the WKCE was to be judged using a particular framework, then they would be well advised to use the same framework. When CTB developed the items presented at the January 2007 review meeting, they were developed to target a specific DOK level using Norm Webb's framework.

4.3 Item Selection of 2005 WKCE

4.3.1 WKCE Item Selection

The original test design proposed administering the first of three forms calibrated during the December 2004 forms calibration administration. Because the contract was awarded late, the item development during the first year was, of necessity, divided into two phases. Half the items were developed and field tested in May 2004, whereas the May administration was originally planned to be for field testing all items developed during the first year of the contract. The December 2004 administration was to be a calibration of three operational forms. However, because the second half of the items developed during the first year needed to be field tested,

forms E04 and F04 contained field test items in addition to items that had been field tested in May 2004. The first of the three main forms administered in December 2004, D04, consisted of items field tested in May 2004. The other two forms, E04 and F04, consisted of both previously-field tested items and new field test items. The details of the modification to the test design and the results of the May and December 2004 administrations are documented in previous technical reports. When form D04 was assembled for December 2004, item statistics from the May 2004 administration were not yet available. Therefore, form D04 was selected using professional judgment.

For the fall 2006 administration, CTB Research recommended changing to a test design that uses a year-to-year anchor design and live calibration to ensure that each year's form can be equated to the fall 2005 form. The number of anchor items per content area are as follows:

- Reading: 18 MC, 0–1 CR
- Mathematics: 18 MC, 1 CR
- Language Arts: 15 MC
- Science: grades 4 and 8, 15 MC; grade 10, 18 MC
- Social Studies: grades 4 and 8, 15 MC; grade 10, 18 MC

CTB content experts used CTB's proprietary software ITEMWIN to select items for the fall 2006 operational test forms for all content areas and grade levels. The ITEMWIN software (Burket, 2000) allows the content editor to make informed decisions regarding an item selection. This software monitors the impact of each decision made during the item selection process and offers a variety of options for grouping, classifying, sorting, and ranking items to highlight key information as it is needed.

The ITEMWIN program has three parts. The first part is used 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, a bias rating indicating whether the item shows DIF to a particular population of students, the item parameters, and a fit rating indicating how well the item fits the expectations based on the IRT model used.

The second part of the ITEMWIN program uses the working item pool created in the first step to perform the actual test selection. Typically, the developer begins by specifying the number of items to be included in the test and a target number of items for each item format. The program can then be prompted to select automatically a test that represents the best possible statistical combination of items. These automatic selections can then be used as a reference set to which other selections are compared. Successive selections are plotted on a graphic display that shows the test characteristic curve for each set of selected items. In the case of the WKCE, the test characteristic curve for form D04 (administered in December 2004) for each grade and content area was generated, since form D04 was designed to become the first operational form in fall 2005.

In the third part of the program, a table shows both expected number correct and standard error of measurement as functions of scale score, as well as statistical and graphical summaries

on bias, 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 biased items, or does not adequately cover part of the range, becomes immediately apparent as the final statistics are generated. Content editors and research staff examined these statistics for each of the WKCE selections to confirm they each had an appropriate scale score range for the grade level and when the test characteristic curves for all grades were compared side-by-side, that there was an appropriate progression in difficulty.

CTB content editors submitted their selections to the content supervisor for review and then to the publishing project manager. The supervisor and manager may have requested changes to the selections in order to improve the test characteristic curve (TCC) or standard error (SE) curve. Form selections were then submitted to the Research scientist for review. Additional revisions may have been requested at this stage. For the Reading and Mathematics selections, it was especially important to ensure the test characteristic curves for all grade levels formed a progression. The grade 3, 4, and 5 selections and the grade 8 and 10 selections needed revisions to ensure that adjacent TCCs were non-overlapping. Upon approval of the selections by Research, the CTB editor submitted the selections to WDPI for review. For some selections, WDPI requested revisions for content, difficulty, or other reasons. Upon making the changes requested and submitting revised selection summary forms, all operational forms were approved by WDPI.

4.3.2 WKCE Field Test Item Selection

In addition to the operational items, a set of new field test items were included in the fall 2006 test books. Table 4-2 presents unique items field tested each year and total to date, through fall 2006.

In order to contribute to a bank of items that measure and support the curriculum and state content frameworks, development of the field test items for fall 2006 was guided by the test blueprints (See Tables 3-1 and 3-2.). The number of field test items developed for each objective or subskill was proportional to the number of items indicated on the blueprint. For future item development, consideration is given to the distribution of items that survive field testing and in the entire item pool across objectives and subskills. Following the fall 2005 administration, the test design used beginning fall 2006 and thereafter was changed; the number of items was reduced and a year-to-year anchor item design was used to ensure year-to-year equating. Other than anchor items, which are used in two successive years, the multiple-choice items should not be used more than once in two years and constructed-response items should be used once in four years.

4.4 Style and Format Decisions

A detailed Wisconsin Style Guide is used when style editing WKCE items and test book pages. The Style Guide includes capitalization and punctuation conventions, abbreviations, wording and formatting preferences, use of symbols, and other specific and general editing guidelines. This guide was initially developed for the Wisconsin High School Graduation Test

and was then augmented and revised to reflect WDPI's preferences for the WKCE. The Style Guide was developed during the first year of the contract prior to the development of test materials, and it continued to evolve as the project progressed and style issues were addressed. Additional updates were done based on editorial decisions made during the editing of the field test materials for May and December 2005. The Style Guide is a "living" document, and the revisions serve the purpose of bringing clarity and consistency to the test items and test materials.

The psychometric properties of the items need to remain stable across successive administrations. In order to achieve this stability, items should not be changed between successive administrations (e.g., field test and operational administration; operational and anchor administration). Furthermore, there should be no changes in the broader context in which the item is administered. Any editing or art change that may affect the statistical characteristics of an item should be avoided. Ideally, there should be no change in the wording of the stem or answer options, position of key, or formatting of answer choices. Any cosmetic changes to the items were reviewed and approved by CTB Research.

4.5 Customer Approvals

Approvals from WDPI staff were obtained during the phases of development:

- item content and bias review results
- item selections for the fall 2006 operational forms
- manuscripts
- second pages
- final pages (prior to release to manufacturing)

4.5.1 Item Content and Bias Review

Following the review of items each day, CTB and WDPI staff reviewed the edits recommended by the educator committees. WDPI staff initialed each item in the review books to indicate acceptance of the item accepted, accepted with revisions, or rejected. WDPI and CTB each kept a copy of the item review book with the edits marked.

4.5.2 Item Selection Approval

ITEMWIN selection summary reports were submitted to WDPI, which included graphics of the test characteristic and standard error curves, lists of items selected, summary test statistics. WDPI approval was obtained using a sign-off form.

4.5.3 Manuscript Approvals

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

4.5.4 Second Pages Approvals

The second pages represent WDPI's first review of the composited test book or test administration manual pages. By this point, all content issues had been resolved. That is, the focus of the approval was on format and presentation issues, rather than on content issues. WDPI approval was obtained using a sign-off form.

4.5.5 Final Pages Sign-off

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

Part 8: Calibration and Scaling

Part 8 of the Technical Report serves to describe the calibration and scaling processes, procedures, and results. The WKCE program primarily uses scores based item response theory (IRT), (scale scores) rather than raw scores. The scores reported to test users, such as students and teachers, are established through the processes of calibration and scaling described here. As described by item response theory, calibration is the process of determining the parameters of the response function for an item. Some references to introductory and advanced discussions of IRT are provided below. Scaling is the process of creating a scale score. Scaling may enhance test score interpretation by placing scores onto a common scale (AERA/APA/ NCME standards). The validity of all inferences derived from WKCE scores depends on proper calibration and scaling. The software applied, the vertical relationship across grades, successful estimation of parameters, fit, and the standard error of measurement are all key dimensions in the assessment of calibration and scaling procedures, and all are discussed here. Part 8 of this report addresses the following AERA/APA/NCME standards: 1.13, 2.1, 2.2, 2.14, 4.1, 4.2, 4.3, 4.11, 6.4, 6.5, and 13.6.

8.1 Calibration Methods

As indicated, the Fall 2006 WKCE was calibrated and scaled using item response theory (IRT). 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 Wisconsin Knowledge and Concept Exam (WKCE) (CTB/McGraw-Hill, 1997-2004).

Because the characteristics of MC and CR items are different, two different item response theory models were used. The three-parameter logistic model (Lord & Novick, 1968; Lord, 1980) was used to scale the MC items and the two-parameter partial credit model (Muraki, 1992; Yen, 1993) was used to scale the CR items. The three-parameter logistic model (3PL) defines a MC item in terms of three item parameters: the item difficulty (or its location on a scale of difficulty/ability), the item discrimination (or item differences on discrimination), and the level of guessing. The two-parameter partial credit model (2PPC) defines a CR item in terms of an item discrimination parameter and a location parameter for each score point. 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).

8.1.1 Calibration Models

As stated, the 3PL model (Lord & Novick, 1968; Lord, 1980) was used in the analysis of MC items. In this model, the probability that a student with scale score θ responds correctly to item i is:

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]}$$

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.

For analysis of the CR items in the 2006 WKCE, the 2PPC model (Muraki, 1992; Yen, 1993) was used. 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 that 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.

8.1.2 Calibration Software

The IRT models were implemented using CTB's PARDUX software (Burket, 1991). PARDUX estimates parameters simultaneously for MC and CR items using marginal maximum likelihood procedures implemented with the expected maximum (EM) algorithm (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 also shown that the IRT procedures used here produce accurate vertical scaling (Yen & Burket, 1997). The Stocking and Lord (1983) procedure was used to place the estimated parameters on the scale from which the anchor items (i.e., *TerraNova*) were drawn.

8.2 Scaling Procedures

As indicated, the Fall 2006 WKCE was calibrated and scaled using the calibration methods, models, and software discussed above in sections 8.1, 8.1.1, and 8.1.2. The reader should also know that the Fall 2005 scale was the first operational scale in the current assessment program. In 2005, the Wisconsin Department of Public Instruction augmented the existing program to create the WKCE. This change in 2005 moved the state assessment from a norm-referenced test to a criterion-referenced test. As a part of the change, a special linking study was conducted in order to relate scores from the previous assessment to scores under the new assessment. Also, as described in the Fall 2005 Technical Manual, the 2005 scales for Reading and Mathematics were established through a Form Standardization process in December 2004. Scales were later established for Language Arts, Social Studies, and Science in a separate, but related process for which the Form Standardization provided the basis. Like the 2005 scale, the Fall 2006 scale was calibrated and scaled based on the responses of WI students. For further information on the origin of the present scales, consult the Fall 2005 Technical Manual.

8.2.1 Reading and Mathematics

One important feature in the present scales is the vertical relationship across grades. Because the 2004 WKCE Form Standardization was on a vertical scale, the 2005 scale transformation to the 2004 scale preserved that vertical relationship. The vertical relationship is still preserved in the present scales because the 2006 scale was linked to the 2005 scale using the following two steps:

- Step 1: Fall 2006 WKCE items calibrated for each grade and content.
- Step 2: For each grade and content, the items which appeared in both the 2005 WKCE and the 2006 WKCE were treated as anchor items. Then, the Stocking and Lord (1983) formula was applied to estimate the transformation slope and intercept. The transformation slope and intercept were applied to 2006 WKCE item parameters.

As can be seen in Table 9-1 and Table 9-2, the means and standard deviations of both the 14 CD and WI census data show this vertical order across grades for Reading and Mathematics. The same vertical relationship can also be observed in Figures 8-4 to 8-5. Additional information on the planning and implementation of the WKCE vertical scales can be also obtained from WDPI.

8.2.2 Language Arts, Social Studies, and Science

Vertical scales were not set up for Language Arts, Social Studies, and Science during the 2004 Form Standardization. Without administering tests for all continuous grades in a given content area, it is difficult to build a vertical scale for the content area. Language Arts, Social Studies, and Science were administered only to grades 4, 8, and 10. However, these scales were artificially constructed in such a way so as to show a vertical relationship across grades.

There are psychometric and practical reasons for this arrangement. Without building in a vertical relationship across grades, all grades could show the same (or a very similar) mean scores, (assuming the same level of performance across grades). For example, in such an arrangement, one could see mean scale scores that were lower for 8th grade students than for 4th grade students. WDPI and CTB were concerned that this arrangement could confuse test users especially, for example, among parents with two children in two different grades.

To avoid this situation, an artificial vertical relationship was set up across grades for these three contents in the 2005 WKCE. The 2006 WKCE scale preserved this artificial vertical relationship because the 2006 scale was linked to the 2005 scale by applying the equating procedure described above for Reading and Mathematics.

The mean and standard deviation for all grades and contents can be found in Tables 9-1 and Table 9-2. There, the means and standard deviations from both the 14 CD and the WI census show this vertical order for Language Arts, Social Studies, and Science. The same vertical relationship can be observed in Figures 8-6 to 8-8.

8.3 Calibration and Scaling Results

As stated above, items that appeared in both 2005 and 2006 were treated as anchor items for calibrating and scaling. Field test items were calibrated together with operational items, and transformed to the scale of the 2006 WKCE using the item parameters of the 2006 WKCE operational items as anchor items. While all responses of field test MC items were included, about 2,000 of the responses of field test CR items and Writing prompts were used for both calibration and scaling. Note that about 2,000 responses were scored for each field test CR item and Writing prompt. The number of responses for CR items and Writing prompt can be found in Part 6.4.1 (Distribution of CR items).

8.3.1 IRT Item Parameters

All operational items converged, meaning parameters were successfully estimated for every item. There were some field test items that did not converge, or for which parameters could not be estimated during calibration. These items will not be used for any future testing without revising these items and re-field testing them. The items in Reading were: Grade 3 Form B item 62, Grade 6 Form B item 67, and Grade 7 all forms item 68. In Mathematics, one item did not converge: Grade 8 Form B item 61. Two FT items were also suppressed during range finding for Mathematics. These were Grade 4 Form C item 52, and Grade 7 Form B item 56. The Grade 8 Form A Writing Prompt, Part B, also did not converge.

The estimated item parameters from the 14 calibration districts were used for scoring. Although using item parameters from census data is ideal, parameters from the 14 calibration districts were used due to the time limitation. As can be seen in Part 7.1, the 14 CD seemed to represent the WI census well. The current Technical Report does not contain the item parameters

used for the Fall 2006 WKCE scoring, because of the large size of the data files. Separate excel files containing item parameters will be delivered to WDPI for a database.

8.3.2 IRT Item Fit

A statistical procedure was used to identify items that did not fit the IRT model. Item model fit information was obtained for each item using a *Z*-statistic. 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:

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

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*.

The *Z*-statistic is an index of the degree to which obtained proportions of students with each item score match the proportions that would be predicted by the estimated student ability and item parameters. These values, along with the associated chi-squares (Q_1), are computed for ten intervals corresponding to deciles of the ability distribution (Yen, 1984). 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*.

Table 8–1 presents items that were flagged based on the Z statistics above. For example, the Reading Grade 4 operational MC item #48 was flagged because its Z value of 22.04 is larger than the critical Z value of 17.07 based on a sample size of 6,402. The third column in the table, "Form," is blank for those items that appear on all forms. The table shows both CR and MC items as flagged. For CR items, there are, in general, a small number of students at the lower and higher score levels, and with these small sample sizes, misfit is thereby easily introduced between the observed ICC and the expected ICC. With a small sample size, it is not easy to get a stable expected ICC. In a similar manner, misfit for MC items often happens at the lower ability range or at the higher ability range, where there are fewer students. As shown in Table 8-1, more Mathematics items were flagged than Reading items. This is due to the fact that Mathematics contains more CR items than Reading.

The main issue in item fit is where the misfit happens. If the misfit happens around the lower or higher ability range, where there are not many students, this is a smaller issue and we need not worry much about it. If the misfit happens around the middle of ability range, where

there are many students, we are more concerned. Because the fit index itself does not show where the misfit happens on the ability range, the appropriate graphical information was produced for each item by PARDUX. The fit index for all items and the graphical information for items flagged are not included in this report, but will be separately delivered to WDPI.

The flagging of an item does not require that the item not be used. This item fit is just one of the criteria for selecting sound operational items. However, as with all items flagged, the list of items flagged based on the Z statistics and graphical information has been delivered to Development for future item selection.

8.3.3 Scoring and Standard Error of Measurement

As indicated, item-pattern scoring was applied to the 2006 WKCE. Within the broader context of large scale educational assessments, either raw score or item-pattern scoring could be chosen for the purposes of the WKCE. For groups of 25 or more students, the two methods produce tau equivalent results. However, item-pattern scoring is generally recommended because it produces more accurate scores for individual students. The increase in accuracy is equivalent, on the average, to approximately a 15 to 20% increase in test length (Yen, 1984; Yen & Candell, 1991). Item-pattern scoring utilizes more information about students' responses than number-correct scoring. Psychometrically speaking, the item pattern score is the most probable estimation of a student's true ability (it is the maximum likelihood estimate), produced within the context of known item parameter estimates.

With item pattern scoring, students with the same raw score can get different scale scores even if they correctly answered the same number of items. Unlike number correct scoring, the difficulty of items is a factor in the score with item-pattern scoring. So, for example, if student A and student B both correctly answered the same number of items, but student A correctly answered more of the difficult items than student B did, student A would have a higher scale score than student B.

Because of the nature of item pattern scoring, a scoring table showing a simple, direct relationship between raw score and scale score cannot be applied to the 2006 WKCE. However, scoring tables showing a rough relationship among raw score, scale score, and standard error of measurement (SEM) can be produced, and that data is presented in tables 8-2 through 8-24.

The standard error of measurement is used to obtain a range within which a student's true score is likely to fall, that is, with a certain degree of probability. An obtained 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. It is expected that 68% of the time a student's score obtained from a single testing would fall within one SEM of that student's true score and that 95% of the time the obtained score would fall within two standard errors of true score.

Standard errors of measurement (SEM) for the 2006 WKCE scale scores, obtained from item-pattern scoring, are displayed graphically for each of the test configurations in Figures 8-5

through 8-10. Each figure includes a SEM curve of a given grade and content area. Each curve is plotted as a function of the scale scores. Note that for convenience, the highest and lowest obtainable scale score (HOSS and LOSS) of the 2006 WKCE were used as the starting scale score and the last scale score.

These figures show the scale score range within which measurement is most accurate. The figures also show that extreme scale scores have more measurement error than moderate scores. The forms lose accuracy of measurement for scale scores near the high or low extremes because there are fewer students at these score ranges.

Part 12: Cut Scores and Descriptor Writing

When a test is used to make categorical decisions, such as pass/fail, the scores used for the decision are called "cut scores." The primary purpose of Chapter 12 is to describe the origin and meaning of the cut scores applied in the WKCE. As described below, the cut scores were originally established through a process called *standard setting*, and then adapted to the current scale. As detailed below, cut scores derive their meaning through *descriptors*. "Descriptor writing" is the process of establishing a plain-language description of what students must know in order to fall into each of the performance levels established though cut scores. Descriptors thereby firmly root the cut scores and performance levels in the content that students are supposed to learn. Descriptors and cut scores together define, in qualitative and quantitative terms, the difference between a student who is Proficient, and a student who is not.

In addition to describing the origin and meaning of the cut scores, the present chapter serves to describe the descriptor writing process in detail. Chapter 12 of thus functions to address standard 4.19 of the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999): "When proposed score interpretation involves one or more cut scores, the rationale and procedures used for establishing cut scores should be clearly documented." In terms of the validity of the WKCE, it is essential to understand that descriptors and cut scores are established in a collaborative, participatory process, largely driven by the input of Wisconsin teachers and educators. In addition, as cited in the *Standards,* validity extends to the interpretation of test scores. The descriptors clearly establish, in plain language, the proper frame of reference for understanding how to interpret test scores, and cut scores in particular.

12.1 Cut Scores

Student performance on the Wisconsin Knowledge and Concepts Examinations (WKCE) is reported in terms of four *proficiency categories*—Advanced, Proficient, Basic, and Minimal Performance. Each proficiency category is defined by a range of scale scores. The cut scores divide the range of possible scale scores into the four proficiency categories and define the minimal scale score needed to be classified into each category. Descriptor writing, described in section 12.2, is the process of creating the plain-language description of the content that students at each grade level actually demonstrate for their performance to be classified as Minimal Performance, Basic, Proficient, and Advanced. The performance category descriptors link test performance back to the assessment framework and the content knowledge and skills measured and the degree of proficiency with the content that students at each performance category demonstrate. The descriptors contribute to validity evidence that the test actually measures the content that it purports to measure.

In 2002, cut scores were established for the WKCE, and these cut scores were used through the Fall 2004 WKCE. In 2005, the Wisconsin Department of Public Instruction moved the state assessment from a norm-referenced test to a criterion-reference test and it also expanded the number of grades assessed in Reading and Mathematics. The new assessment was placed on a new scale and a special linking study was conducted in order to relate scores from the previous

assessment to scores under the new assessment. Results of that study, including crosswalk tables, which linked the two scores, were presented in the Fall 2005 WKCE Technical Report.

For present purposes, the reader should understand that the cut scores in the current assessment were established during the changeover to the criterion-referenced WKCE. Cut scores were established with reference to their impact data. "Impact data" refers to the percentage of students classified in each achievement level. In other words, cut scores were established with reference to how the cuts would impact the distribution of students across performance levels.

The cut scores established for the 2005 criterion-referenced WKCE yielded the same impact data (or minimized the difference in impact data) as the 2004 WKCE while reporting scores on a new scale. Cut scores for grades previously not assessed were established by extrapolating and interpolating based on the 2004 cut scores that existed for grades 4, 8 and 10. Readers interested in more information can consult the Fall 2005 WKCE Technical Report, Part 11.

Table 12-2 shows the cut scores for Grades 3–8 and 10 for Reading, along with the impact data associated with these cut scores. Table 12-3 shows the cut scores and associated impact data for Grades 3–8 and 10 for Mathematics. Tables 12-4, 12-5, and 12-6 present the cut scores and associated impact data for Grades 4, 8, and 10 for Language Arts, Social Studies, and Science. Figures 12-1 through 12-10 present the cut scores and percentages for all performance levels based on impact data across all grades and content areas.

12.2 Descriptor Writing

As has been discussed, the assessment in Wisconsin is criterion-referenced. Simply speaking, that means students are assessed with reference to specific knowledge and skills. Accordingly, the performance levels established through cut scores can also be understood with reference to specific knowledge and skills demonstrated, as well as the *degree* of proficiency with the content knowledge and skills demonstrated.

In June of 2006, approximately 100 Wisconsin educators were convened for a three-day meeting to develop performance level "descriptors" to accompany the performance levels established through cut scores. The process of "descriptor writing" establishes a plain-language description of the content that students must know at each grade level in order to be Basic, Proficient, and Advanced. The descriptors can be used by teachers and the public to understand the performance levels on the WKCE. The final descriptors are based on the collective input of the approximately 100 educators convened. The agenda for the meeting is in Table 12-1.

In the descriptor writing workshop, participants were asked to define the knowledge, skills, and abilities that are required of students in each grade to be Basic, Proficient, and Advanced. To inform their descriptions, participants reviewed ordered item booklets and item maps and the assessment framework documents, and identified the specific knowledge and skills required to answer each item correctly and identified the reasons why each item is more difficult

than the preceding item. Participants were shown the cut scores and then wrote descriptors for each grade/content area. The image below is an example of an item map.

Wisconsin Knowledge and Concepts Examinations Criterion-Referenced Test
Grade 4 Mathematics Item Map

Order of Difficulty	Loca-tion	Test / Session	Item No.	Part	Item Type	Score Key	Sub-skill	What does this item measure? That is what do you know about a student who can respond successfully to this item/score point?	Why is this item more difficult than the preceding items?
1	424	MA G4 E S3	70		MC	4	E.a		**NA**
2	425	MA G4 F S1	12		MC	4	B.b		
3	426	MA G4 F S4B	86		MC	3	C.a		
4	430	MA G4 E S3	59		MC	4	B.a		
5	436	MA G4 F S1	21		MC	2	F.a		
6	436	MA G4 F S4B	83		MC	1	D.a		

Prior to the workshop, CTB and WDPI discussed the final format of the descriptors. WDPI requested three formats:

Brief Narrative Description

These one-paragraph descriptions of each proficiency level may be most useful for those who simply want an overview of the knowledge and skills students typically demonstrate at each level.

Detailed Narrative Description

These descriptions contain more detail but are still structured in a way that makes the information easy to grasp.

Elements of Proficiency Levels

The elements are descriptions of discrete knowledge and skills students typically demonstrate at each proficiency level. They complement the narratives by enumerating specific examples of knowledge and skills described in the narratives.

The morning of the first day of the descriptor writing workshop, CTB presented a PowerPoint presentation which reviewed the purpose of the descriptor writing workshop, how the cut scores for each performance category were established, the specific tasks to be completed, the characteristics of well-written descriptors, and how the descriptors should reflect the progression of abilities within and across grade levels.

The educators were assigned to content and grade level groups with 4–6 participants per grade. Two CTB facilitators were assigned to each content area group. The CTB facilitators guided the committees through a series of tasks designed to build familiarity with the test and the content frameworks and then to draft and revise descriptors. CTB provided the participants with a handout that listed hints for writing clear and precise descriptors. Participants were also

provided the descriptors for grades 4, 8, and 10 from the 2003 standard setting and descriptor writing.

CTB provided participants in each content and grade-level group with hardcopy templates for recording their draft descriptors for each assessment framework reporting category and performance category. An example is shown below.

		Minimal Performance	Basic	Proficient	Advanced
		Students at this level are able to:	Students at this level are able to:	Students at this level are able to:	Students at this level are able to:
1	**Determines meaning of words and phrases in context**				
1.1	**Uses context clues to determine meaning of words or phrases** (List items for this objective from the Ordered Item Book in order) ↓				

Participants first identified which items in the ordered item book addressed each content standard at each performance level by writing the number of the item in the ordered item book in the appropriate cell on the template. Next, they wrote descriptors based on the individual items. Participants were instructed to describe the knowledge, skill, and cognitive demand measured by the item without describing the specific content of the item. A sample of a portion of the completed template for Reading grade 4 is shown below.

		Minimal Performance	Basic	Proficient	Advanced
		Students at this level are able to:	Students at this level are able to:	Students at this level are able to:	Students at this level are able to:
1	**Determines meaning of words and phrases in context**				
1.1	**Uses context clues to determine meaning of words or phrases** (List items for this objective from the Ordered Item Book in order) ↓	OIB: none Using clear context clues, illustrations and graphics to determine meanings of words.	OIB: 1S Identifying the meaning of simple idiomatic expressions and common sayings. (1S)	OIB: 5, 8, 9, 10, 11 Using context clues and prior knowledge to determine the meaning of familiar and multiple-meaning words. (5, 8, 9, 10, 11) Identifying synonyms of familiar words.	OIB: 32, 37, 47 Using context clues and prior knowledge to determine the meaning of unfamiliar words, multiple-meaning words, or words and phrases used figuratively. (32, 37, 47)

Because items did not exist for every performance level and every content standard, especially for the Minimal Performance and Basic categories, participants applied their professional judgment to augment the information provided by the test items in order to develop a more complete set of descriptors. Specifically, the sequence of tasks was:

- take the Fall 2005 test
- review the ordered item book and describe each item using the item map
- review the cut scores and identify the cut score location in the ordered item book
- review the existing performance level descriptors for grades 4, 8, 10 (established in 2003)
- organize ordered items by content objective and performance level
- draft descriptors by content objective and performance level
- review descriptors for each content objective within the grade level group
- review descriptors by content objective in cross-grade level groups

- revise descriptors by content objective to reflect level to level and grade-to-grade progression
- draft multi-paragraph narrative descriptors from the bulleted list of descriptors by objective
- review and revise narratives across performance levels within a grade

Throughout the process, CTB facilitators answered questions about the process, provided guidance on how to state the descriptors clearly, and encouraged them to carefully consider the content represented by the individual items and how the content difficulty and cognitive complexity changed as the scale score level of the item increased.

Following the meeting, CTB content specialists reviewed the draft descriptors, checking the accuracy of the description written for each item by checking it against the item in the ordered item book. The content specialists also verified the accuracy of the descriptors in terms of consistency with the assessment frameworks and content terminology. The CTB content specialist also edited the descriptors for consistency in style and to ensure that the descriptors appropriately described the increasing level of knowledge and skills across performance levels within a grade and across the grades. The revised descriptors were submitted to WDPI for review. WDPI distributed the draft descriptors to the table leaders for their review, and a conference call was conducted with WDPI, CTB, and the table leaders in attendance. The conference calls were helpful for providing feedback on both general and specific issues. The CTB content specialists then revised the bulleted descriptors and the multi-paragraph narratives based on the feedback and submitted them to WDPI for a second review. WDPI reviewed the descriptors and provided feedback, which focused primarily on the narrative descriptors. CTB did a final edit of the bulleted and multi-paragraph narrative and then wrote the single-paragraph, condensed narrative. WDPI then completed the formatting of the descriptors to prepare them for presentation to the superintendent's cabinet prior to release to the public. The final descriptor documents are available on the WDPI website at http://www.WDPI.wi.gov/oea/profdesc.html.

Table 2-1
Wisconsin WKCE-CRT Assessment Limits Meeting, August 25–28, 2003

Wisconsin WKCE-CRT
Assessment Limits Meeting
August 25–28, 2003

Meeting Objectives

1. Define reporting categories for WKCE-CRT, reading and mathematics
2. Establish reporting category subskills for structuring assessment limits for cross-grade consistency
3. Establish assessment limits for reading and mathematics, grades 2–9
4. Make recommendations for test blueprint by assigning weights, distributing score points and items across reporting categories and subskills, and identifying allowable item formats
5. Review draft item specifications and make recommendations for refinements
6. Review draft reading passage specifications and make recommendations for refinements

Monday, August 25

10:00 – 11:00	Welcome Overview of Test Development Processes Purpose of Establishing Assessment Limits • What assessment limits are • Criteria for good assessment limits • Examples of clear and unclear limits
11:00 – 12:00	Defining Reporting Categories for WKCE-CRT • Content area small groups (2–4, 5–6, 7–9) • Content area large group Subskill Framework • Content area small groups
12:00 – 12:30	Lunch
12:30 – 2:00	Subskill Framework • Content area small groups (cont.) • Content area large group
2:00 – 2: 15	Break

2:15 – 4:00	Assessment Limits
	• Review samples and models
	• First reporting category & subskills
	  o Small groups
	  ▪ define grades 4 and 8 first
	  o Large group cross-grade articulation

4:00 – 4:30	Debriefing and Questions

Tuesday, August 26

8:00 – 8:30	Continental Breakfast

Table 2-1 Cont'd

8:30 – 10:00	Assessment Limits (small groups)
10:00 – 10:15	Break
10:15 – 11:15	Assessment Limits (large group, cross-grade articulation)
11:15 – 12:00	Assessment Limits (small groups)
12:00 – 12:30	Lunch
12:30 – 2:00	Assessment Limits (small groups)
2:00 – 2:15	Break
2:15 – 3:15	Assessment Limits (small groups)
3:15 – 4:00	Assessment Limits (large group, cross-grade articulation)

Wednesday, August 27

8:00 – 8:30	Continental Breakfast
8:30 –10:00	Complete Assessment Limits • Small groups • Large group, cross-grade articulation
10:00 – 10: 15	Break
11:00 – 12:00	Cross Content Area Articulation Small Groups (reading & math 2–4, 5–6, 7–9)
12:00 – 12:30	Lunch
12:30 – 1:00	Introduction to Test Blueprints
1:00 – 2:00	Test Blueprint Design • Overview of Process by CTB • Small Groups to assign weight to each reporting category and subskill to determine percentage of score points • Large Group Articulation to refine score point distribution
2:00 – 2:15	Break
2:15 – 4:00	Assignment of Allowable Item Formats Distribution of Items Across Subskills within Reporting Categories

Thursday, August 28

8:00 – 8:30	Continental Breakfast
9:00 – 10:00	Review Item Specifications and Reading Passage Specifications
10:00 – 10:15	Break
10:15 – 12:00	Review, Revisit Revise: Assessment Limits, Blueprint, Item Specifications
12:00	Dismissal

Table 2-2
Wisconsin WKCE-CRT Reading Passage Review Meeting, September 29 – October 1, 2003

Wisconsin WKCE-CRT
Reading Passage Review Meeting
September 29 – October 1, 2003

Meeting Objectives

1. Become familiar with reading passage specifications and review criteria
2. Review grade level assessment limits
3. Read and evaluate reading passages according to specifications and criteria
4. Document passage recommendations

Tuesday, September 29

8:30 – 9:00	Registration and Refreshments
9:00 – 9:30	Welcome, Introductions, Paperwork • Participant Agreement Forms • Travel Reimbursement Forms Task Orientation • Passage Materials • Passage Cover Sheets • Review Criteria • Documentation Forms • Routing Slips
9:30 – 12:00	Review passages in grade-level groups
12:00 – 12:45	Lunch
12:45 – 1:00	Questions and Recentering
1:00 – 4:30	Review passages in grade-level groups (afternoon refreshment break provided)
4:30 – 4:45	Debriefing and Questions

Wednesday, October 1

8:00 – 8:30	Registration and Refreshments
8:30 – 12:00	Review passages in grade-level groups
12:00 – 12:45	Lunch
12:45 – 4:30*	Review passages in grade-level groups (afternoon refreshment break provided) *Participants will be dismissed earlier if the work is completed.

Table 2-2 Cont'd
Wisconsin WKCE-CRT Reading Passage Review Meeting, September 29 – October 1, 2003

Reading Passage Review Criteria

- Does the passage meet the specifications with respect to:
  - ❑ passage type
  - ❑ avoiding inappropriate topics
  - ❑ passage length
  - ❑ text features

- Is the passage appropriate for the grade level in terms of:
  - ❑ readability?
  - ❑ interest level?
  - ❑ being accessible to students with a range of ability levels and backgrounds?

- Will the passage be able to support a wide range of test items measuring the grade-level assessment limits?

Procedures to send a passage to another grade level:

1. Each person removes the passage from the binder, including cover sheet and any copyright information, and clips together with a paper clip.
2. The table leader gathers all copies of the passage (6), clips all copies together with a large paper clip, fills out and attaches a routing slip.

The table leader makes a notation on the review form that the passage was sent to another grade (indicate grade level), and gives the set of passages to the CTB facilitator.

Table 2-3
Checklist for the Content Reviewer

Checklist for the Content Reviewer

For All Items:

Check to ensure the content of each item:

- ❑ is targeted to assess only one objective or skill (unless specifications indicate otherwise)
- ❑ deals with material that is important in testing the targeted objective or skill
- ❑ uses grade-appropriate content and thinking skills
- ❑ is presented at a reading level suitable for the grade level being tested
- ❑ is accurate and documented against reliable, up-to-date sources

For Multiple-Choice Items:

Check to ensure that the content of each item:

- ❑ has a stem that facilitates answering the question or completing the statement without looking at the answer choices
- ❑ has a stem that does <u>not</u> present clues to the correct answer choice
- ❑ has answer choices that are plausible and attractive to the student who has not mastered the objective or skill
- ❑ is conceptually, grammatically, and syntactically consistent—between the stem and answer choices, and among the answer choices
- ❑ has mutually exclusive distractors
- ❑ has one and only one correct answer choice

For Constructed-Response Items:

Check to ensure that the content of each item:

- ❑ is written so that a student possessing the knowledge or skill being tested can construct a response that is scorable with the specified rubric or scoring tool; that is, the range of possible correct responses must be wide enough to allow for diversity of responses,
  but narrow enough so that students who do not clearly show their grasp of the objective or skill being assessed cannot obtain the maximum score
- ❑ is presented without clue to the correct response
- ❑ has precise and unambiguous directions for the desired response
- ❑ is free of extraneous words or expressions
- ❑ is appropriate for the question being asked and the intended response (For example, the item does not ask students to draw pictures of abstract ideas.)
- ❑ is conceptually, grammatically, and syntactically consistent

Table 2-4
Checklist for the Sensitivity Reviewer

Checklist for the Sensitivity Reviewer

To have confidence in test results, it is important to ensure that students are given a reasonable chance to do their best on the test. Test items must be accessible to a diverse student population with respect to gender, race, ethnicity, geographic region, socioeconomic status, and other factors.

Check to ensure that the content of each item is free of explicit references to or descriptions of:

- ❑ events involving extreme sadness or adversity
- ❑ acts of physical or psychological violence
- ❑ alcohol or drug abuse
- ❑ vulgar language
- ❑ sex

Check to ensure that if any religious, political, social, or philosophical issues are addressed:

- ❑ more than one point of view is expressed
- ❑ beliefs or biases do not interfere with factual accuracy
- ❑ contemporary issues that have already been proven to be controversial are absent
- ❑ stereotypic descriptions of beliefs or customs are absent

Test items must:

- ❑ be free of offensive, disturbing, or inappropriate language or content
- ❑ be free of stereotyping based on:
    - gender
    - race
    - ethnicity
    - religion
    - socioeconomic status
    - age
    - regional or geographic area
    - disability
    - occupation
- ❑ demonstrate sensitivity to historical representation of groups
- ❑ be free of differential familiarity for any group based on:
    - language
    - socioeconomic status
    - regional or geographic area
    - prior knowledge or experiences unrelated to the subject matter being tested

Table 2-5
November 2005 Item Selection Review Participants

CTB: Dennis Allion, Margie Tully, Jannette McMunn, Judy Staten, Kristina Summers, Heather Farina, Andrew Jones, Joshua Pierce, Deedra Pell, Andrina Ortiz

DPI: Lynette Russell, Shelley Lee, Philip Olsen, Jacque Karbon, Sandra Berndt, Phil Cranley, Laura Morancheck, Diana Kasbaum, Jim Marti, Visalakshi Somasundaram, Robert Kohl

Wisconsin Educators	Content	District	Attendance
Brahan, Larry	Mathematics	Sheboygan	✓
Carrington, Tanzi	Mathematics	Milwaukee	✓
Hollinger, Rosann	Mathematics	Milwaukee	✓
Richards, Mary	Mathematics	Waupaca	✓
Weber, Harlan	Mathematics	Sheboygan	✓
Hilgart, Faye	Mathematics	Madison	✓
Jensen, Margaret	Mathematics	Madison	✓
Schefelker, Beth	Mathematics	Milwaukee	✓
Truszynski, James	Mathematics	Waukesha	✓
Valentine, Carrie	Mathematics	Madison	✓
Viegut, Deb	Mathematics	Arrowhead	✓
Womack, Lois	Mathematics	Milwaukee	✓
Bennett, Judy	Reading	Mineral Point	✓
Castro, Mariana	Reading	Madison	Did not attend
Diaz, Linda	Reading	Racine	✓
Dvorak, Steve	Reading	Hayward	✓
Holloway, Reola	Reading	Milwaukee	✓
Schoen, Monica	Reading	Neenah	✓
Schumann, Susan	Reading	La Crosse	Did not attend
Van Hoof, Chris	Reading	Clintonville	✓
Washington, Doncella	Reading	Milwaukee	Did not attend
Wepking, Bart	Reading	Wheatland Center	✓
Wiedmann, Lisa	Reading	Rhinelander	Did not attend
Zarling, Debra	Reading	Oshkosh	✓
Baeseman, Pam	Science	Mercer	✓
Bergerson, David	Science	Wisconsin Rapids	✓
Boone, Bob	Science	Menomonie	✓
Bukosky, Richard	Science	Milwaukee	✓
Hanhan, Tolga	Science	Superior	✓
Johnson, Sue	Science	Madison	✓
Patrick, Judy	Science	Wauwatosa	✓
Schoenemann, Anne	Science	Madison	✓
Staten, Mary	Science	Milwaukee	Did not Attend
VanPay, Ellen	Science	Green Bay	✓
Wachtel, Lisa	Science	Madison	✓
Whitman, Kathryn	Science	Manitowoc	✓
Whitsett, Sue	Science	Fond du Lac	✓
Foss, Margaret	Science	Ladysmith Hawkins	✓

Table 2-6
January 2007 Item Selection Review Participants

CTB: Leslie Dodge, Margie Tully, Becky Fisher, Cathy Brown, Andrea Jachman, Teresa Kuntz, Jannette McMunn, Jana McCarty, Chris Williams

DPI: Sandra Berndt, Jason Bierbrauer, Brad Carl, Phil Cranley, Jason Engle, Jeremiah Holiday, Dacia Hopfensperger, Jacqueline Iribanen, Brian Johnson, Jackie Karbon, Diana Kasbaum, Lynette Russell, Viji Somasundaram, Michael St. Pierre, Jennifer Teasdale, Tami Lanier, Barb Ebben, Phil Olsen, Margaret Planner

Wisconsin Educators	Content	District	Attendance
Becher, Paul	Mathematics	Waukesha School District	✓
Burge, Cathy	Mathematics	Holmen District	✓
Cuellar, Cynthia	Mathematics	Milwaukee Public Schools	✓
George, Kimberly	Mathematics	Kenosha Unified School District	✓
Hilgart, Faye	Mathematics	Madison Metropolitan School District	✓
Hintz, Lori	Mathematics	Menominee Indian School District	✓
Jenkins, Mazie	Mathematics	MMSD (Dane County)	✓
Jensen, Margaret	Mathematics	Madison Metropolitan School District	✓
Santos, Jesus	Mathematics	Milwaukee	✓
Schefelker, Beth	Mathematics	Milwaukee	✓
Schewe, Bev	Mathematics	Mannette District	✓
Truszynski, Jim	Mathematics	Waukesha School District	✓
Valentine, Carrie	Mathematics	Dane County	✓
Alvara, Patricia	Reading	Kenosha Unified	✓
Bangert, Linda	Reading	Menominee Indian School District	✓
Bennett, Judy	Reading	Mineral Point Unified School District	✓
Dvorak, Steve	Reading	Hayward	✓
Edwards, Bonnie	Reading	Milwaukee Public Schools	✓
Eggert, Joan	Reading	McFarland School District	✓
Haertel, Sue	Reading	Hamilton District	✓
Holloway, Reola	Reading	Milwaukee	✓
Powell, Mary Lee	Reading	Appleton Area School District	✓
Schumann, Susan	Reading	LaCrosse District	✓
Van Hoof, Chris	Reading	Clintonville Public School	✓
Wepking, Bart	Reading	Wheatland District	✓

Table 2-7
Item Content Review Results, December 2003

Grade	Accepted As Is	Accepted w/Edits	Rejected	Total Items Reviewed
Reading				
3	*	*	*	285
4	*	*	*	221
5	*	*	*	253
6	*	*	*	292
7	*	*	*	213
8	*	*	*	209
Reading Total				

*****Note: breakdown of items not available.

Mathematics				
3	24 (15%)	132 (80%)	9 (5%)	165
4	34 (21%)	129 (79%)	0	163
5	36 (23%)	110 (71%)	9 (6%)	155
6	35 (23%)	105 (70%)	10 (7%)	150
7	70 (46%)	80 (53%)	2 (1%)	152
8	26 (17%)	117 (77%)	9 (6%)	152
Mathematics Total	225 (24%)	673 (72%)	39 (4%)	937
Grand Total				

Table 2-8
Item Content Review Results, April 2004

Grade	Accepted As Is	Accepted w/Edits	Rejected	Total Items Reviewed
Reading				
3	68 (40%)	91 (54%)	11 (6%)	170
4	128 (70%)	41 (23%)	13 (7%)	182
5	161 (83%)	29 (15%)	3 (2%)	193
6	109 (65%)	51 (30%)	9 (5%)	169
7	101 (71%)	33 (23%)	8 (6%)	142
8	49 (30%)	82 (51%)	31 (19%)	162
Reading Total	617 (60%)	330 (32%)	74 (7%)	1,021
Mathematics				
3	68 (40%)	91 (54%)	11 (6%)	170
4	128 (70%)	41 (23%)	13 (7%)	182
5	161 (83%)	29 (15%)	3 (2%)	193
6	109 (65%)	51 (30%)	9 (5%)	169
7	101 (71%)	33 (23%)	8 (6%)	142
8	49 (30%)	82 (51%)	31(19%)	162
Mathematics Total	617 (60%)	330 (32%)	74 (7%)	1,021
Grand Total	68 (40%)	91 (54%)	11 (6%)	170

Table 2-9
Item Content Review Results, March 2005

Grade	Accepted As Is	Accepted w/Edits	Rejected	Total Items Reviewed
Reading				
3	15 (60%)	7 (28%)	3 (12%)	25
4	20 (57%)	15 (43%)	0	35
5	23 (55%)	15 (36%)	4 (9%)	42
6	75 (86%)	10 (11%)	2 (2%)	87
7	21 (54%)	17 (43%)	1 (3%)	39
8	15 (44%)	18 (53%)	1 (3%)	34
Reading Total	169 (65%)	82 (31%)	11 (4%)	262
Mathematics				
3	9 (17%)	40 (75%)	4 (7%)	53
4	18 (31%)	37 (65%)	1 (5%)	56
5	8 (15%)	37 (72%)	1 (2%)	51
6	24 (32%)	46 (62%)	4 (5%)	74
7	6 (9%)	45 (70%)	12 (19%)	64
8	14 (18%)	49 (64%)	8 (10%)	77
Mathematics Total	79 (21%)	254 (68%)	30 (8%)	375
Grand Total	**248 (39%)**	**336 (53%)**	**41 (6%)**	**637**

Table 2-10
Item Content Review Results, November 2005

Grade	Accepted As Is	Accepted w/Edits	Rejected	Total Items Reviewed
Reading				
3	13 (38%)	15 (44%)	6 (18%)	34
4	22 (60%)	13 (35%)	2 (5%)	37
5	9 (27%)	21 (64%)	3 (9%)	33
6	22 (60%)	13 (35%)	2 (5%)	37
7	26 (65%)	10 (25%)	4 (10%)	40
8	33 (87%)	4 (10%)	1 (3%)	38
Reading Total	125 (57%)	76 (35%)	18 (8%)	219
Mathematics				
3	1 (6%)	17 (94%)	0	18
4	4 (13.3%)	25 (83.3%)	1 (3.3%)	30
5	3 (9%)	29 (91%)	0 (0%)	32
6	17 (42.5%)	23 (57.5%)	0 (0%)	40
7	13 (52%)	12 (48%)	0 (0%)	25
8	7 (20%)	27 (77%)	1 (3%)	35
Mathematics Total	45 (25%)	133 (74%)	2 (1%)	180
Science				
4	40 (25%)	113 (72%)	4 (3%)	157
8	53 (34%)	100 (64%)	4 (3%)	157
10	9 (35%)	17 (65%)	0 (0%)	26
Science Total	102 (30%)	230 (68%)	8 (2%)	340
Grand Total	**272 (37%)**	**439 (59%)**	**28 (4%)**	**739**

Table 2-11
Item Content Review Results, January 2007

Grade	Accepted As Is	Accepted w/Edits	Rejected	Total Items Reviewed
Reading				
3	13 (30%)	24 (56%)	6 (14%)	43 (100%)
4	22 (76%)	6 (21%)	1 (3%)	29 (100%)
5	10 (27%)	25 (68%)	2 (5%)	37 (100%)
6	15 (35%)	26 (60%)	2 (5%)	43 (100%)
7	21 (70%)	9 (30%)	0 (0%)	30 (100%)
8	21 (75%)	7 (25%)	0 (0%)	28 (100%)
Reading Total	102 (48%)	97 (47%)	11 (5%)	210
Mathematics				
3	11 (34%)	19 (57%)	3 (9%)	33 (100%)
4	16 (53%)	14 (47%)	0 (0%)	30 (100%)
5	12 (34%)	23 (66%)	0 (0%)	35 (100%)
6	11 (34%)	18 (57%)	3 (9%)	32 (100%)
7	7 (21%)	18 (57%)	7 (21%)	32 (99%)
8	17 (53%)	14 (44%)	1 (3%)	32 (100%)
Mathematics Total	74 (38%)	106 (55%)	14 (7%)	194
Grand Total	**177 (44%)**	**203 (50%)**	**25 (6%)**	**404**

Table 3-1
Reading Passage Review Results, September, 2003

Grade	Passages Reviewed	Develop & Field Test
3	53	19
4	49	13
5	53	20
6	41	15
7	48	22
8	39	10
Total	**283**	**99**

Table 3-2
Reading Passage Review Results, December 2004

Grade	Passages Reviewed	Use as Is	Use with Edits	Do not Use
3	11	7	1	3
4	16	8	1	7
5	10	3	3	4
6	15	8	4	3
7	17	9	2	6
8	9	4	2	3
Total	**78**	**39**	**13**	**26**

Table 3-3
Reading Passage Review Results, August 2005

Grade	Passages Reviewed	Use as Is	Use with Edits	Do not Use
3	9	0	7	2
4	7	3	1	3
5	8	3	0	5
6	7	2	3	2
7	8	0	4	4
8	8	3	2	3
Total	**47**	**11**	**17**	**19**

Table 3-4
Reading Test Blueprint: Grades 3–8, 10
*Note: Number of score points at the subskill indicator level (e.g., 1.1, 1.2, etc.) are for MC items only; CR items provide the balance of score points.

Reporting Category	Category Title	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
		% of Pts	# of Pts	% of Pts	# of Pts	% of Pts	# of Pts	% of Pts	# of Pts	% of Pts	# of Pts	% of Pts	# of Pts	% of Pts	# of Pts
1	**Determines meaning of words or phrases in context**	18%	11	17%	10	20%	12	18%	11	18%	11	18%	11	13%	7
1.1	Uses context clues to determine meaning of words or phrases		8		7		7		8		6		6		
1.2	Uses knowledge of word structure to determine meaning of words		2		2		4		1		3		2		
1.3	Uses word reference materials to determine meaning of words and phrases		1		1		1		2		2		3		
2	**Understands Text**	28%	17	30%	18	27%	16	23%	14	25%	15	22%	13	11%	6
2.1	Demonstrates understanding of literal meaning by identifying stated information in literary text		8		9		8		4		5		3		
2.2	Demonstrates understanding of literal meaning by identifying stated information in informational text		8		7		6		5		6		7		
2.3	Demonstrates understanding of explicitly stated sequence of events in literary and informational text		1		2		4		5		4		3		
3	**Analyzes Text**	42%	25	38%	23	33%	20	35%	21	35%	21	34%	20	46%	26
3.1	Analyzes literary text		11		8		8		8		7		8		
3.2	Analyzes informational text.		7		7		6		5		8		6		
33	Analyzes author's use of language in literary and informational text.		4		5		4		6		4		4		
4	**Evaluates and Extends Text**	12%	7	15%	9	20%	12	24%	14	22%	13	26%	16	30%	17
4.1	Evaluates and extends literary text		3		1		5		6		3		4		

Table 3-4
Reading Test Blueprint: Grades 3–8, 10 Cont'd
*Note: Number of score points at the subskill indicator level (e.g., 1.1, 1.2, etc.) are for MC items only; CR items provide the balance of score points.

Reporting Category	Category Title	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
		% of Pts	# of Pts	% of Pts	# of Pts	% of Pts	# of Pts	% of Pts	# of Pts	% of Pts	# of Pts	% of Pts	# of Pts	% of Pts	# of Pts
4.2	Evaluates and extends informational text		1		3		4		4		4		10		
4.3	Evaluates and extends author's use of language in literary and informational text		1		3		1		2		6		2		
	Number of MC Items (max = 54	54		53		54		54		54		54		54	
	Number of CR Items (max = 6	2		2		2		2		2		2		2	
	Total Score Points for Test	**60**		**59**		**60**		**60**		**60**		**60**		**56**	

Table 3-5
Mathematics Test Blueprint: Grades 3–8, 10
Note: Subskill score points represent MC score points only. CR item points make up the difference between total subskill points and reporting category points.

Reporting Category	Category Title	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
		% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill
A	**Mathematical Processes**	**14%**	**8**	**14%**	**8**	**13%**	**8**	**13%**	**8**	**13%**	**8**	**13%**	**8**	**15.5%**	**9**
Aa	Reasoning														
Ab	Communication														
Ac	Connections														
Ad	Representation														
Ae	Problem Solving														
B	**Number Operations and Relationships**	**21%**	**12**	**19%**	**11**	**19%**	**12**	**19%**	**12**	**19%**	**12**	**14%**	**9**	**12%**	**7**
Ba	**Number Concepts**		**6**		**5**		**6**		**6**		**6**		**6**		**4**
Ba1	Place Value														
Ba2	Reading, Writing, Representing Number														
Ba3	Ordering/Comparing														
Ba4	Number Theory														
Ba5	Counting/Set Concepts														
Ba6	Proportionality														
Ba7	Fraction/Decimal/Percent Equivalency														
Bb	**Number Computation**		**5**		**6**		**6**		**6**		**6**		**3**		**2**
Bb1	Whole Numbers														
Bb2	Fractions														
Bb3	Decimals														
Bb4	Percents														
Bb5	Irrational														
Bb6	Estimation														

Table 3-5 Cont'd
Mathematics Test Blueprint: Grades 3–8, 10
Note: Subskill score points represent MC score points only. CR item points make up the difference between total subskill points and reporting category points.

Reporting Category	Category Title	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
		% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill	% of Pts	Pts/obj, MC pts/subskill
Bb7	Integers														
C	**Geometry**	**17%**	**10**	**17%**	**10**	**16%**	**10**	**16%**	**10**	**19%**	**12**	**15%**	**9**	**17%**	**10**
Ca	Describing Figures		4		3		2		2		3		2		
Cb	Spatial Relationships and Transformations		4		5		5		4		4		4		
Cc	Coordinate System		1		1		2		3		3		2		
D	**Measurement**	**16%**	**9**	**16%**	**9**	**16%**	**10**	**16%**	**10**	**16%**	**10**	**19%**	**12**	**19%**	**11**
Da	Measurable Attributes		3		3		4		2		3		2		
Db	Direct Measurement		5		4		3		3		3		3		
Dc	Indirect Measurement		1		1		2		4		4		6		
E	**Statistics and Probability**	**16%**	**9**	**16%**	**9**	**18%**	**11**	**18%**	**11**	**15%**	**9**	**15%**	**9**	**15.5%**	**9**
Ea	Data Analysis and Statistics		4		4		7		7		5		5		
Eb	Probability		4		4		3		3		3		3		
F	**Algebraic Relationships**	**16%**	**9**	**18%**	**10**	**18%**	**11**	**18%**	**11**	**18%**	**11**	**24%**	**15**	**21%**	**12**
Fa	Patterns, Relations, and Functions		4		5		5		5		2		6		
Fb	Expressions, Equations, and Inequalities		2		2		3		2		4		6		
Fc	Properties		2		2		2		3		4		2		
	Number of MC Items	45		45		50		50		50		50		50	
	Number of CR Items	4		4		4		4		4		4		4	
	Total Score Points for Test	**57**		**57**		**62**		**62**		**62**		**62**		**58**	
	Minutes (item time	79		79		85		97		97		97		85	
	CR Score Pts as % of Total	21%		21%		19%		19%		19%		19%		14%	

Table 3-6
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	16	1	15	1
D	Language	5		8		9	
F	Research and Inquiry	6		6		6	
	Total Number of Items	**30**	**1**	**30**	**1**	**30**	**1**
	Total Number of Points	**30**	**9**	**30**	**9**	**30**	**9**

Table 3-7
Science Test Blueprint: Grades 4, 8, 10

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

Table 3-8
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 3-9
Reading Test Structure

Grade 3	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	54	1	1	54	54
CR items	2	3	5	6	10
EFT/FT MC items	10	1	1	10	10
EFT/FT CR items	1	3	5	3	5
Reading Time					60
TOTALS	**67**			**60**	**139**

Grade 4	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	53	1	1	53	53
CR items	2	3	5	6	10
EFT/FT MC items	10	1	1	10	10
EFT/FT CR items	1	3	5	3	5
Reading Time					60
TOTALS	**66**			**59**	**138**

Grade 5	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	54	1	1	54	54
CR items	2	3	5	6	10
EFT/FT MC items	14	1	1	14	14
EFT/FT CR items	1	3	5	3	5
Reading Time					60
TOTALS	**71**			**60**	**143**

Grade 6	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	54	1	1	54	54
CR items	2	3	5	6	10
EFT/FT MC items	10	1	1	10	10
EFT/FT CR items	1	3	5	3	5
Reading Time					60
TOTALS	**67**			**60**	**139**

Grade 7	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	54	1	1	54	54
CR items	2	3	5	6	10
EFT/FT MC items	15	1	1	15	15
EFT/FT CR items	1	3	5	3	5
Reading Time					60
TOTALS	**72**			**60**	**144**

Table 3-10
Reading Test Structure Cont'd

Grade 8	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	54	1	1	54	54
CR items	2	3	5	6	10
EFT/FT MC items	10	1	1	10	10
EFT/FT CR items	1	3	5	3	5
Reading Time					60
TOTALS	**67**			**60**	**139**

Grade 10	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	50	1	1	50	50
CR items	2	3	5	6	10
EFT/FT MC items					
EFT/FT CR items					
Reading Time					45
TOTALS	**52**			**56**	**105**

Table 3-11
Mathematics Test Structure

Grade 3	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	45	1	1.3	45	59
CR items	4	3	5	12	20
EFT/FT MC items	8	1	1.3	8	11
EFT/FT CR items	1	3	5	3	5
TOTALS	**58**			**57**	**95**

Grade 4	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	45	1	1.3	45	59
CR items	4	3	5	12	20
EFT/FT MC items	8	1	1.3	8	11
EFT/FT CR items	1	3	5	3	5
TOTALS	**58**			**57**	**95**

Table 3-12
Mathematics Test Structure Cont'd

Grade 5	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	50	1	1.3	50	65
CR items	4	3	5	12	20
EFT/FT MC items	8	1	1.3	8	11
EFT/FT CR items	1	3	5	3	5
TOTALS	**63**			**62**	**101**

Grade 6	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	50	1	1.3	50	65
CR items	4	3	8	12	32
EFT/FT MC items	8	1	1.3	8	11
EFT/FT CR items	1	3	8	3	8
TOTALS	**63**			**62**	**116**

*Note: A suppressed item in Grade 6 mathematics reduces the effective total OP score points from 62 to 61.

Grade 7	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	50	1	1.3	50	65
CR items	4	3	8	12	32
EFT/FT MC items	8	1	1.3	8	11
EFT/FT CR items	1	3	8	3	8
TOTALS	**63**			**62**	**116**

Grade 8	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	50	1	1.3	50	65
CR items	4	3	8	12	32
EFT/FT MC items	**10**	1	1.3	**10**	**13**
EFT/FT CR items	1	3	8	3	8
TOTALS	**65**			**62**	**118**

Grade 10	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	50	1	1.3	50	65
CR items	4	2	5	8	20
ECR items	0	4	10	0	0
EFT/FT MC items		1	1.3	0	0
EFT/FT CR items		2	5	0	0
EFT/FT ECR items		4	10	0	0
TOTALS	**54**			**58**	**85**

Table 3-13
Language Arts Test Structure

Grade 4	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	30	1	1.14	30	35
CR items	1	9	30	9	30
EFT/FT MC items					
EFT/FT CR items					
TOTALS	**31**			**39**	**65**

Grade 8	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	**30**	**1**	**1.14**	**30**	**35**
CR items	**1**	**9**	**30**	**9**	**30**
EFT/FT MC items					
EFT/FT CR items					
TOTALS	**31**			**39**	**65**

Grade 10	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	**30**	**1**	**1**	**30**	**30**
CR items	**1**	**9**	**30**	**9**	**30**
EFT/FT MC items					
EFT/FT CR items					
TOTALS	**31**			**39**	**60**

Table 3-14
Social Studies Test Structure

Grade 4	**No. of Items**	**Pts per Item**	**Minutes per Item**	**Total OP Points**	**Total Minutes**
MC items	38	1	~1	38	40
CR items					
EFT/FT MC items					
EFT/FT CR items					
TOTALS	**38**			**38**	**40**

Grade 8	**No. of Items**	**Pts per Item**	**Minutes per Item**	**Total OP Points**	**Total Minutes**
MC items	40	1	1	40	40
CR items					
EFT/FT MC items					
EFT/FT CR items					
TOTALS	**40**			**40**	**40**

Grade 10	**No. of Items**	**Pts per Item**	**Minutes per Item**	**Total OP Points**	**Total Minutes**
MC items	50	1	1	50	50
CR items					
EFT/FT MC items					
EFT/FT CR items					
TOTALS	**50**			**50**	**50**

Table 3-15
Science Test Structure

Grade 4	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	40	1	1	40	40
CR items					
EFT/FT MC items	**10**	**1**	**1**	**0**	**10**
EFT/FT CR items					
TOTALS	**50**			**40**	**50**

Grade 8	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	40	1	1	40	40
CR items					
EFT/FT MC items	10	1	**1**	**0**	**10**
EFT/FT CR items					
TOTALS	**50**			**40**	**50**

Grade 10	No. of Items	Pts per Item	Minutes per Item	Total OP Points	Total Minutes
MC items	50	1	1	50	50
CR items					
EFT/FT MC items	**10**	1	1	**0**	10
EFT/FT CR items					
TOTALS	**60**			**50**	**60**

Table 4-1
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	Total MC to date	Total CR to date
Grade 3										
Reading	411	52	23	2	30	4	40	3	504	61
Math	317	36	33	14	18	2	30	4	398	56
Total	**728**	**88**	**56**	**16**	**48**	**6**	**70**	**7**	**902**	**117**
Grade 4										
Reading	380	56	32	3	34	3	25	4	471	66
Math	265	35	45	9	29	1	26	4	365	49
Language Arts	0	0	0	10	0	0	0	0	0	10
Science	0	0	0	0	123	34	0	0	123	34
Total	**645**	**91**	**77**	**22**	**186**	**38**	**51**	**8**	**959**	**159**
Grade 5										
Reading	433	59	36	6	29	5	29	7	527	77
Math	305	49	38	11	26	3	30	5	399	68
Total	**738**	**108**	**74**	**17**	**55**	**8**	**59**	**12**	**926**	**145**
Grade 6										
Reading	511	56	32	5	42	5	37	6	622	72
Math	310	41	53	16	7	2	28	4	398	63
Total	**821**	**97**	**85**	**21**	**49**	**7**	**65**	**10**	**1020**	**135**
Grade 7										
Reading	359	44	35	4	38	4	25	5	457	57
Math	305	34	32	23	20	0	28	4	385	61
Total	**664**	**78**	**67**	**27**	**58**	**4**	**53**	**9**	**842**	**118**
Grade 8										
Reading	365	44	30	4	34	4	25	4	454	56
Math	289	51	47	25	20	2	28	4	384	82
Language Arts	0	0	0	10	0	0	0	0	0	10
Science	0	0	0	0	125	34	0	0	125	34
Total	**654**	**95**	**77**	**39**	**179**	**40**	**53**	**8**	**963**	**182**
Grade 10										
Reading	0	0	0	0	0	0	0	0	0	0
Math	0	0	0	0	0	0	0	0	0	0
Language Arts	0	0	0	0	0	0	0	0	0	0
Science	0	0	0	0	18	8	0	0	18	8
Total	**0**	**0**	**0**	**0**	**18**	**8**	**0**	**0**	**18**	**8**
TOTALS										
Reading	2,459	311	188	24	207	25	181	29	3,035	389
Mathematics	1,791	246	248	98	120	10	170	25	2,329	379
Language Arts	0	0	0	20	0	0	0	0	0	20
Science	0	0	0	0	266	76	0	0	266	76
Grand Total	**4,250**	**557**	**436**	**142**	**593**	**111**	**351**	**54**	**5,630**	**864**

Table 4-2
Unique Items Field Tested Each Year and Total to Date

	MC Items Field Tested in 2004	CR Items Field Tested in 2004	MC Items Field Tested in 2005	CR Items Field Tested in 2005	MC Items Field Tested in 2006	CR Items Field Tested in 2006	Total MC Field Tested to Date	Total CR Field Tested to Date
Grade 3								
Reading	242	12	24	2	27	2	293	16
Math	252	24	15	2	32	4	299	30
Total	**494**	**36**	**39**	**4**	**59**	**6**	**592**	**46**
Grade 4								
Reading	294	12	24	2	32	3	350	17
Math	231	29	15	2	32	4	278	35
Language Arts	0	0	0	6	0	0	0	6
Science	0	0	0	0	40	0	40	0
Social Studies	0	0	0	0	0	0	0	0
Total	**525**	**41**	**39**	**10**	**104**	**7**	**668**	**58**
Grade 5								
Reading	235	14	24	2	28	2	287	18
Math	257	34	15	2	32	4	304	40
Total	**492**	**48**	**39**	**4**	**60**	**6**	**591**	**58**
Grade 6								
Reading	259	14	24	1	33	3	316	18
Math	252	33	15	2	32	4	299	39
Total	**511**	**47**	**39**	**3**	**65**	**7**	**615**	**57**
Grade 7								
Reading	259	14	24	1	17	2	300	17
Math	243	33	15	2	32	4	290	39
Total	**502**	**47**	**39**	**3**	**49**	**6**	**590**	**56**
Grade 8								
Reading	274	14	24	1	33	4	331	19
Math	234	33	15	2	40	4	289	39
Language Arts	0	0	0	6	0	0	0	6
Science	0	0	0	0	40	0	40	0
Social Studies	0	0	0	0	0	0	0	0
Total	**508**	**47**	**39**	**9**	**113**	**8**	**660**	**64**
Grade 10								
Reading	0	0	0	0	0	0	0	0
Math	0	0	0	0	0	0	0	0
Language Arts	0	0	0	0	0	0	0	0
Science	0	0	0	0	10	0	10	0
Social Studies	0	0	0	0	0	0	0	0
Total	**0**	**0**	**0**	**0**	**10**	**0**	**10**	**0**
TOTALS								
Grand Totals	**3,032**	**266**	**234**	**33**	**460**	**40**	**3,726**	**339**

Table 4-3
Reading Grade 3 2005 Item Development and November 2005 Item Rev. Results

Reporting Category Subskill	Items Written		Items Accepted		Items Revised		Items Rejected	
	SR	CR	SR	CR	SR	CR	SR	CR
Aa								
Aa1	4	0	3	0	1	0	0	0
Aa2	3	0	1	0	0	0	2	0
Aa3	1	0	1	0	0	0	0	0
Total	8	0	5	0	1	0	2	0
Ab								
Ab1	1	0	1	0	0	0	0	0
Ab2	7	0	3	0	3	0	1	0
Ab3	1	0	1	0	0	0	0	0
Total	9	0	5	0	3	0	1	0
Ac								
Ac1	1	1	0	0	1	1	0	0
Ac2	7	2	2	0	5	1	0	1
Ac3	3	0	1	0	1	0	1	0
Total	11	3	3	0	7	2	1	1
Ad								
Ad1	1	0	0	0	1	0	0	0
Ad2	1	1	0	0	1	0	0	1
Ad3	0	0	0	0	0	0	0	0
Total	2	1	0	0	2	0	0	1
Grand Total	30	4	13	0	13	2	4	2

Table 4-4
Reading Grade 4 2005 Item Development and November 2005 Item Rev. Results

Reporting Category Subskill	Items Written		Items Accepted		Items Revised		Items Rejected	
	SR	CR	SR	CR	SR	CR	SR	CR
Aa								
Aa1	4	0	2	0	2	0	0	0
Aa2	2	0	2	0	0	0	0	0
Aa3	1	0	0	0	1	0	0	0
Total	7	0	4	0	3	0	0	0
Ab								
Ab1	7	0	5	0	2	0	0	0
Ab2	0	0	0	0	0	0	0	0
Ab3	2	0	0	0	2	0	0	0
Total	9	0	5	0	4	0	0	0
Ac								
Ac1	9	1	4	1	4	0	1	0
Ac2	0	0	0	0	0	0	0	0
Ac3	4	1	3	1	1	0	0	0
Total	13	2	7	2	5	0	1	0
Ad								
Ad1	4	1	3	0	0	1	1	0
Ad2	0	0	0	0	0	0	0	0
Ad3	1	0	1	0	0	0	0	0
Total	5	1	4	0	0	1	1	0
Grand Total	34	3	20	2	12	1	2	0

Table 4-5
Reading Grade 5 2005 Item Development and November 2005 Item Rev. Results

Reporting Category Subskill	Items Written		Items Accepted		Items Revised		Items Rejected	
	SR	CR	SR	CR	SR	CR	SR	CR
Aa								
Aa1	3	0	0	0	2	0	1	0
Aa2	2	0	0	0	2	0	0	0
Aa3	2	0	2	0	0	0	0	0
Total	7	0	2	0	4	0	1	0
Ab								
Ab1	3	0	2	0	1	0	0	0
Ab2	4	0	2	0	2	0	0	0
Ab3	1	0	0	0	1	0	0	0
Total	8	0	4	0	4	0	0	0
Ac								
Ac1	6	1	0	0	6	1	0	0
Ac2	4	1	3	0	1	1	0	0
Ac3	1	0	0	0	1	0	0	0
Total	11	2	3	0	8	2	0	0
Ad								
Ad1	0	1	0	0	0	0	0	1
Ad2	2	0	0	0	2	0	0	0
Ad3	2	0	0	0	1	0	1	0
Total	4	1	0	0	3	0	1	1
Grand Total	30	3	9	0	19	2	2	1

Table 4-6
Reading Grade 6 2005 Item Development and November 2005 Item Rev. Results

Reporting Category Subskill	Items Written		Items Accepted		Items Revised		Items Rejected	
	SR	CR	SR	CR	SR	CR	SR	CR
Aa								
Aa1	5	0	3	0	1	0	1	0
Aa2	1	0	1	0	0	0	0	0
Aa3	1	0	1	0	0	0	0	0
Total	7	0	5	0	1	0	1	0
Ab								
Ab1	6	0	5	0	1	0	0	0
Ab2	0	0	0	0	0	0	0	0
Ab3	2	0	2	0	0	0	0	0
Total	8	0	7	0	1	0	0	0
Ac								
Ac1	9	2	5	1	4	0	0	1
Ac2	0	0	0	0	0	0	0	0
Ac3	3	0	2	0	1	0	0	0
Total	12	2	7	1	5	0	0	1
Ad								
Ad1	6	2	2	1	4	1	0	0
Ad2		0	0	0	0	0	0	0
Ad3	2	0	1	0	1	0	0	0
Total	8	2	3	1	5	1	0	0
Grand Total	35	4	22	2	12	1	1	1

Table 4-7
Reading Grade 7 2005 Item Development and November 2005 Item Rev. Results

Reporting Category Subskill	Items Written		Items Accepted		Items Revised		Items Rejected	
	SR	CR	SR	CR	SR	CR	SR	CR
Aa								
Aa1	5	0	1	0	3	0	1	0
Aa2	2	0	1	0	1	0	0	0
Aa3	0	0	0	0	0	0	0	0
Total	7	0	2	0	4	0	1	0
Ab								
Ab1	3	0	3	0	0	0	0	0
Ab2	4	0	4	0	0	0	0	0
Ab3	2	0	1	0	1	0	0	0
Total	9	0	8	0	1	0	0	0
Ac								
Ac1	6	1	5	0	1	1	0	0
Ac2	5	00	2	0	2	0	1	0
Ac3	2	1	2	1	0	0	0	0
Total	13	2	9	1	3	1	1	0
Ad								
Ad1	3	1	1	1	1	0	1	0
Ad2	4	1	3	1	0	0	1	0
Ad3	2	0	2	0	0	0	0	0
Total	9	2	6	2	1	0	2	0
Grand Total	38	4	26	2	9	1	3	1

Table 4-8
Reading Grade 8 2005 Item Development and November 2005 Item Rev. Results

Reporting Category Subskill	Items Written		Items Accepted		Items Revised		Items Rejected	
	SR	CR	SR	CR	SR	CR	SR	CR
Aa								
Aa1	3	0	2	0	1	0	0	0
Aa2	2	0	2	0	0	0	0	0
Aa3	3	0	3	0	0	0	0	0
Total	8	0	7	0	1	0	0	0
Ab								
Ab1	6	0	6	0	0	0	0	0
Ab2	0	0	0	0	0	0	0	0
Ab3	1	0	1	0	0	0	0	0
Total	7	0	7	0	0	0	0	0
Ac								
Ac1	7	1	7	1	0	0	0	0
Ac2	0	0	0	0	0	0	0	0
Ac3	5	2	4	1	0	1	1	0
Total	12	3	11	2	0	1	1	0
Ad								
Ad1	5	1	4	0	1	1	0	0
Ad2	0	0	0	0	0	0	0	0
Ad3	2	0	2	0	0	0	0	0
Total	7	1	6	0	1	1	0	0
Grand Total	34	4	31	2	2	2	1	0

Table 4-9
Mathematics Grade 3 2005 Item Development and November 2005 Item Rev. Results

WKCE-CRT 2006 Item Development
Mathematics Grade 3

Total Items Developed for 2006							Items Accepted without Revision						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total				2		2	A Total						
B	Ba				5	5	B	Ba				1	1
	Bb				1	1		Bb					
B Total					6	6	B Total					1	1
C	Ca				2	2	C	Ca					
	Cb							Cb					
	Cc				3	3		Cc					
C Total					5	5	C Total						
D	Da				2	2	D	Da					
	Db				2	2		Db					
	Dc							Dc					
D Total					4	4	D Total						
E	Ea				1	1	E	Ea					
	Eb				2	2		Eb					
E Total					3	3	E Total						
F	Fa		1			1	F	Fa					
	Fb							Fb					
	Fc		1			1		Fc					
F Total			2			2	F Total						
Total			**2**	**2**	**18**	**22**	**Total**					**1**	**1**

Table 4-9, continued

WKCE-CRT 2006 Item Development
Mathematics Grade 3

Items Accepted with Revisions							Items Rejected						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total				2		2	A Total						
B	Ba				4*	4	B	Ba					
	Bb				1	1		Bb					
B Total					5	5	B Total						
C	Ca				2	2	C	Ca					
	Cb							Cb					
	Cc				3	3		Cc					
C Total					5	5	C Total						
D	Da				2	2	D	Da					
	Db				2*	2		Db					
	Dc							Dc					
D Total					4	4	D Total						
E	Ea				1	1	E	Ea					
	Eb				2	2		Eb					
E Total					3	3	E Total						
F	Fa		1			1	F	Fa					
	Fb							Fb					
	Fc		1			1		Fc					
F Total			2			2	F Total						
Total			**2**	**2**	**17**	**21**	**Total**						

***includes one Ba and one Db item that the teachers wrote at Rev.**

Table 4-10
Mathematics Grade 4 2005 Item Development and November 2005 Item Rev. Results

WKCE-CRT 2006 Item Development
Mathematics Grade 4

Total Items Developed for 2006							Items Accepted without Revision						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total				1		1	A Total						
B	Ba				3	3	B	Ba				1	1
	Bb				5	5		Bb				2	2
B Total					8	8	B Total					3	3
C	Ca				3	3	C	Ca				1	1
	Cb				6	6		Cb					
	Cc							Cc					
C Total					9	9	C Total					1	1
D	Da				3	3	D	Da					
	Db				3	3		Db					
	Dc							Dc					
D Total					6	6	D Total						
E	Ea				1	1	E	Ea					
	Eb		1			1		Eb					
E Total			1		1	2	E Total						
F	Fa				3	3	F	Fa				1	1
	Fb				2	2		Fb					
	Fc							Fc					
F Total					5	5	F Total					1	1
Total			**1**	**1**	**29**	**31**	**Total**					**5**	**5**

Table 4-10, continued

WKCE-CRT 2006 Item Development
Mathematics Grade 4

Items Accepted with Revisions							Items Rejected						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total							A Total				1		1
B	Ba				2	2	B	Ba					
	Bb				3	3		Bb					
B Total					5	5	B Total						
C	Ca				2	2	C	Ca					
	Cb				6	6		Cb					
	Cc							Cc					
C Total					8	8	C Total						
D	Da				3	3	D	Da					
	Db				3	3		Db					
	Dc							Dc					
D Total					6	6	D Total						
E	Ea				1	1	E	Ea					
	Eb							Eb		1			1
E Total					1	1	E Total			1			1
F	Fa				2	2	F	Fa					
	Fb				2	2		Fb					
	Fc							Fc					
F Total					4	4	F Total						
Total					**24**	**24**	**Total**			**1**	**1**		**2**

Table 4-11
Mathematics Grade 5 2005 Item Development and November 2005 Item Rev. Results

WKCE-CRT 2006 Item Development
Mathematics Grade 5

Total Items Developed for 2006							Items Accepted without Revision						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total				3		3	A Total						
B	Ba				6	6	B	Ba				1	
	Bb				1	1		Bb					
B Total					7	7	B Total					1	1
C	Ca				1	1	C	Ca					
	Cb				4	4		Cb					
	Cc							Cc					
C Total					5	5	C Total						
D	Da		1		4	5	D	Da					
	Db				2	2		Db					
	Dc				1	1		Dc					
D Total			1		7	8	D Total						
E	Ea				2	2	E	Ea				1	
	Eb				3	3		Eb					
E Total					5	5	E Total					1	1
F	Fa				1	1	F	Fa					
	Fb				4	4		Fb				1	
	Fc		2			2		Fc					
F Total			2		5	7	F Total					1	1
Total			**3**	**3**	**29**	**35**	**Total**					**3**	**3**

Table 4-11, continued

WKCE-CRT 2006 Item Development
Mathematics Grade 5

Items Accepted with Revisions							Items Rejected						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total				3		3	A Total						
B	Ba				5	5	B	Ba					
	Bb				1	1		Bb					
B Total					6	6	B Total						
C	Ca				1	1	C	Ca					
	Cb				4	4		Cb					
	Cc							Cc					
C Total					5	5	C Total						
D	Da		1		3	4	D	Da					
	Db				2	2		Db					
	Dc				1	1		Dc					
D Total			1		6	7	D Total						
E	Ea				1	1	E	Ea					
	Eb				3	3		Eb					
E Total					4	4	E Total						
F	Fa				1	1	F	Fa					
	Fb				3	3		Fb					
	Fc		2			2		Fc					
F Total			2		4	6	F Total						
Total			**3**	**3**	**25**	**31**	**Total**						

Table 4-12
Mathematics Grade 6 2005 Item Development and November 2005 Item Rev. Results

WKCE-CRT 2006 Item Development
Mathematics Grade 6

Total Items Developed for 2006							Items Accepted without Revision						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total				3		3	A Total				1		1
B	Ba				4	4	B	Ba				3	
	Bb		1		3	4		Bb				3	
B Total			1		7	8	B Total					6	6
C	Ca				4	4	C	Ca					
	Cb		1		5	6		Cb				1	
	Cc							Cc					
C Total			1		9	10	C Total					1	1
D	Da						D	Da					
	Db				3	3		Db				1	
	Dc				1	1		Dc				1	
D Total					4	4	D Total					2	2
E	Ea				6	6	E	Ea				1	
	Eb				5	5		Eb				4	
E Total					11	11	E Total					5	5
F	Fa		1		4	5	F	Fa				1	1
	Fb				1	1		Fb		1		1	2
	Fc				1	1		Fc					
F Total			1		6	7	F Total			1		2	3
Total			**3**	**3**	**37**	**43**	**Total**			**1**	**1**	**16**	**18**

Table 4-12, continued

WKCE-CRT 2006 Item Development
Mathematics Grade 6

Items Accepted with Revisions							Items Rejected						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total				2			A Total						
B	Ba				1	1	B	Ba					
	Bb		1			1		Bb					
B Total			1		1	2	B Total						
C	Ca				4	4	C	Ca					
	Cb		1		3	4		Cb					
	Cc							Cc					
C Total			1		7	8	C Total						
D	Da						D	Da					
	Db				2	2		Db					
	Dc				1	1		Dc					
D Total					3	3	D Total						
E	Ea				5	5	E	Ea					
	Eb				1	1		Eb					
E Total					6	6	E Total						
F	Fa				3	3	F	Fa					
	Fb							Fb					
	Fc				1	1		Fc					
F Total					4	4	F Total						
Total			**2**	**2**	**21**	**23**	**Total**						

Table 4-13
Mathematics Grade 7 2005 Item Development and November 2005 Item Rev. Results

WKCE-CRT 2006 Item Development
Mathematics Grade 7

Total Items Developed for 2006							Items Accepted without Revision						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total				3	2	5	A Total				1		1
B	Ba		1		3	4	B	Ba				3	3
	Bb				4	4		Bb				3	3
B Total			1		7	8	B Total					6	6
C	Ca						C	Ca					
	Cb		2		3	5		Cb		1		2	3
	Cc							Cc					
C Total			2		3	5	C Total			1		2	3
D	Da				4	4	D	Da				1	1
	Db							Db				1	1
	Dc				1	1		Dc					
D Total					5	5	D Total					2	2
E	Ea				2	2	E	Ea				1	1
	Eb							Eb					
E Total					2	2	E Total					1	1
F	Fa						F	Fa					
	Fb							Fb					
	Fc				3	3		Fc				1	1
F Total					3	3	F Total					1	1
Total			**3**	**3**	**22**	**28**	**Total**			**1**	**1**	**12**	**14**

Table 4-13, continued

WKCE-CRT 2006 Item Development
Mathematics Grade 7

Items Accepted with Revisions							Items Rejected						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total				2	2	4	A Total						
B	Ba		1			1	B	Ba					
	Bb				1	1		Bb					
B Total			1		1	2	B Total						
C	Ca						C	Ca					
	Cb		1		1	2		Cb					
	Cc							Cc					
C Total			1		1	2	C Total						
D	Da				3	3	D	Da					
	Db							Db					
	Dc							Dc					
D Total					3	3	D Total						
E	Ea				1	1	E	Ea					
	Eb							Eb					
E Total					1	1	E Total						
F	Fa						F	Fa					
	Fb							Fb					
	Fc				2	2		Fc					
F Total					2	2	F Total						
Total			**2**	**2**	**10**	**14**	**Total**						

Table 4-14
Mathematics Grade 8 2005 Item Development and November 2005 Item Rev. Results

WKCE-CRT 2006 Item Development Mathematics Grade 8														
Total Items Developed for 2006								Items Accepted without Revision						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa							A	Aa					
	Ab								Ab					
	Ac								Ac					
	Ad								Ad					
	Ae								Ae					
A Total				2	4	6		A Total					2	2
B	Ba				4	4		B	Ba				1	1
	Bb				4	4			Bb					
B Total					8	8		B Total					1	1
C	Ca				1	1		C	Ca				1	1
	Cb				2	2			Cb				2	2
	Cc				1	1			Cc					
C Total					4	4		C Total					3	3
D	Da		1		2	3		D	Da					
	Db				1	1			Db					
	Dc				1	1			Dc				1	1
D Total			1		4	5		D Total					1	1
E	Ea				1	1		E	Ea					
	Eb		1		3	4			Eb					
E Total			1		4	5		E Total						
F	Fa				3	3		F	Fa					
	Fb				4	4			Fb					
	Fc				2	2			Fc					
F Total					9	9		F Total						
Total			**2**	**2**	**33**	**37**		**Total**					**7**	**7**

Table 4-14, continued

WKCE-CRT 2006 Item Development
Mathematics Grade 8

Items Accepted with Revisions							Items Rejected						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total	Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	Aa						A	Aa					
	Ab							Ab					
	Ac							Ac					
	Ad							Ad					
	Ae							Ae					
A Total				2	1	3	A Total					1	1
B	Ba				3	3	B	Ba					
	Bb				4	4		Bb					
B Total					7	7	B Total						
C	Ca						C	Ca					
	Cb							Cb					
	Cc				1	1		Cc					
C Total					1	1	C Total						
D	Da		1		2	3	D	Da					
	Db				1	1		Db					
	Dc							Dc					
D Total			1		3	4	D Total						
E	Ea				1	1	E	Ea					
	Eb		1		3	4		Eb					
E Total			1		4	5	E Total						
F	Fa				3	3	F	Fa					
	Fb				4	4		Fb					
	Fc				2	2		Fc					
F Total					9	9	F Total						
Total			**2**	**2**	**25**	**29**	**Total**					**1**	**1**

Table 4-15
Science Grade 4 2005 Item Development and November 2005 Item Rev. Results

WKCE-CRT 2006 Item Development Science Grade 4											
Total Items Developed for 2006							Items Accepted without Revision				
Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total		Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total
A	A.4.2			13	13		A	A.4.2				
	A.4.3			12	12			A.4.3			2	2
	A.4.4			13	13			A.4.4			2	2
A Total				38	38		A Total				4	4
B	B.4.1			14	14		B	B.4.1			8	8
	B.4.2			14	14			B.4.2			6	6
	B.4.3			15	15			B.4.3			2	2
B Total				43	43		B Total				16	16
C	C.4.3			3	3		C	C.4.3				
	C.4.5		2		2			C.4.5		1		1
	C.4.6		2	6	8			C.4.6		1	5	6
	C.4.7		2	4	6			C.4.7			1	1
	C.4.8		2	5	7			C.4.8			1	1
C Total			8	18	26		C Total			2	7	9
D	D.4.1	1			1		D	D.4.1	1			1
	D.4.2	1			1			D.4.2				
	D.4.4	1			1			D.4.4				
	D.4.5	2			2			D.4.5	2			2
	D.4.6	1			1			D.4.6				
	D.4.7	1			1			D.4.7				
D Total		7			7		D Total		3			3
E	E.4.1	1			1		E	E.4.1				
	E.4.2	1			1			E.4.2				
	E.4.3	1			1			E.4.3				
	E.4.4	1			1			E.4.4				
	E.4.5	1			1			E.4.5				
	E.4.6	1			1			E.4.6				
	E.4.7	1			1			E.4.7				
E Total		7			7		E Total					
F	F.4.2	3			3		F	F.4.2				
	F.4.3	4			4			F.4.3				
F Total		7			7		F Total					
G	G.4.1			6	6		G	G.4.1			2	2
	G.4.2			5	5			G.4.2				
	G.4.3			5	5			G.4.3			1	1
	G.4.4			4	4			G.4.4			3	3
	G.4.5			4	4			G.4.5			2	2
G Total				24	24		G Total				8	8
H	H.4.3	2			2		H	H.4.3				
	H.4.4	3			3			H.4.4				

Table 4-15 Cont'd
Science Grade 4 2005 Item Development and November 2005 Item Rev. Results

WKCE-CRT 2006 Item Development Science Grade 4												
Total Items Developed for 2006							Items Accepted without Revision					
Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total		Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total
H Total		5			5		H Total					
Total		**26**	**8**	**123**	**157**		**Total**		**3**	**2**	**35**	**40**

Table 4-15, continued
Science Grade 4 2005 Item Development and November 2005 Item Rev. Results

Items Accepted with Revisions						Items Rejected					
Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total	Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total
A	A.4.2			13	13	A	A.4.2				
	A.4.3			10	10		A.4.3				
	A.4.4			10	10		A.4.4			1	1
A Total				33	33	A Total				1	1
B	B.4.1			5	5	B	B.4.1			1	1
	B.4.2			8	8		B.4.2				
	B.4.3			13	13		B.4.3				
B Total				26	26	B Total				1	1
C	C.4.3			3	3	C	C.4.3				
	C.4.5		1		1		C.4.5				
	C.4.6			1	1		C.4.6		1		1
	C.4.7		2	3	5		C.4.7				
	C.4.8		1	4	5		C.4.8		1		1
C Total			4	11	15	C Total			2		2
D	D.4.1					D	D.4.1				
	D.4.2	1			1		D.4.2				
	D.4.4	1			1		D.4.4				
	D.4.5						D.4.5				
	D.4.6	1			1		D.4.6				
	D.4.7	1			1		D.4.7				
D Total		4			4	D Total					
E	E.4.1	1			1	E	E.4.1				
	E.4.2	1			1		E.4.2				
	E.4.3	1			1		E.4.3				
	E.4.4	1			1		E.4.4				
	E.4.5	1			1		E.4.5				
	E.4.6	1			1		E.4.6				
	E.4.7	1			1		E.4.7				
E Total		7			7	E Total					
F	F.4.2	3			3	F	F.4.2				
	F.4.3	4			4		F.4.3				
F Total		7			7	F Total					
G	G.4.1			4	4	G	G.4.1				
	G.4.2			5	5		G.4.2				
	G.4.3			4	4		G.4.3				
	G.4.4			1	1		G.4.4				
	G.4.5			2	2		G.4.5				
G Total				16	16	G Total					
H	H.4.3	2			2	H	H.4.3				
	H.4.4	3			3		H.4.4				
H Total		5			5	H Total					
Total		**23**	**4**	**86**	**113**	**Total**		**0**	**2**	**2**	**4**

Table 4-16
Science Grade 8 2005 Item Development and November 2005 Item Rev. Results

Total Items Developed for 2006						Items Accepted without Revision					
Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total	Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total
A	A.8.3	4		5	9	A	A.8.3			1	1
	A.8.5	1		10	11		A.8.5			1	1
	A.8.6	1		13	14		A.8.6			3	3
A Total		6		28	34	A Total				5	5
B	B.8.1	1		4	5	B	B.8.1			2	2
	B.8.3			8	8		B.8.3			5	5
	B.8.4	2		4	6		B.8.4			2	2
	B.8.5	1		4	5		B.8.5			1	1
	B.8.6	2		4	6		B.8.6			3	3
B Total		6		24	30	B Total				13	13
C	C.8.1		2	2	4	C	C.8.1			1	1
	C.8.2			2	2		C.8.2			1	1
	C.8.3			3	3		C.8.3			2	2
	C.8.4			2	2		C.8.4				
	C.8.5		2	2	4		C.8.5				
	C.8.6		2	2	4		C.8.6			1	1
	C.8.7		1	1	2		C.8.7				
	C.8.9		1	2	3		C.8.9				
	C.8.10			2	2		C.8.10				
	C.8.11			3	3		C.8.11			1	1
C Total			8	21	29	C Total				6	6
D	D.8.5			2	2	D	D.8.5			1	1
	D.8.7			4	4		D.8.7			3	3
	D.8.8			2	2		D.8.8				
	D.8.9			2	2		D.8.9			2	2
	D.8.10			2	2		D.8.10				
D Total				12	12	D Total				6	6
E	E.8.2			2	2	E	E.8.2				
	E.8.4			1	1		E.8.4			1	1
	E.8.5			1	1		E.8.5				
	E.8.6			2	2		E.8.6			2	2
	E.8.7			1	1		E.8.7			1	1
E Total				7	7	E Total				4	4
F	F.8.2			1	1	F	F.8.2			1	1
	F.8.3			2	2		F.8.3			1	1
	F.8.6			3	3		F.8.6				
	F.8.7			1	1		F.8.7				
	F.8.9			1	1		F.8.9			1	1
F Total				8	8	F Total				3	3
G	G.8.1	1		1	2	G	G.8.1			1	1
	G.8.2	1		1	2		G.8.2	1		1	2
	G.8.3	3		1	4		G.8.3	3			3
	G.8.6	1		2	3		G.8.6			1	1
	G.8.7	1		1	2		G.8.7			1	1
G Total		7		6	13	G Total		4		4	8

Table 4-16 Cont'd
Science Grade 8 2005 Item Development and November 2005 Item Rev. Results

Total Items Developed for 2006							**Items Accepted without Revision**					
Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total		Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total
H	H.8.1	2		5	7		H	H.8.1			1	1
	H.8.2	3		4	7			H.8.2			1	1
	H.8.3	2		8	10			H.8.3			6	6
H Total		7		17	24		H Total				8	8
Total		**26**	**8**	**123**	**157**		**Total**		**4**	**0**	**49**	**53**

Table 4-16, continued

Items Accepted with Revisions						Items Rejected					
Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total	Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total
A	A.8.3	4		4	8	A	A.8.3				
	A.8.5	1		8	9		A.8.5			1	1
	A.8.6	1		10	11		A.8.6				
A Total		6		22	28	A Total				1	1
B	B.8.1	1		2	3	B	B.8.1				
	B.8.3			3	3		B.8.3				
	B.8.4	2		2	4		B.8.4				
	B.8.5	1		3	4		B.8.5				
	B.8.6	2		1	3		B.8.6				
B Total		6		11	17	B Total					
C	C.8.1		2	1	3	C	C.8.1				
	C.8.2			1	1		C.8.2				
	C.8.3			1	1		C.8.3				
	C.8.4			2	2		C.8.4				
	C.8.5		2	2	4		C.8.5				
	C.8.6		2	1	3		C.8.6				
	C.8.7		1	1	2		C.8.7				
	C.8.9		1	2	3		C.8.9				
	C.8.10			2	2		C.8.10				
	C.8.11			2	2		C.8.11				
C Total			8	15	23	C Total					
D	D.8.5			1	1	D	D.8.5				
	D.8.7			1	1		D.8.7				
	D.8.8			1	1		D.8.8			1	1
	D.8.9						D.8.9				
	D.8.10						D.8.10			2	2
D Total				3	3	D Total				3	3
E	E.8.2			2	2	E	E.8.2				
	E.8.4						E.8.4				
	E.8.5			1	1		E.8.5				
	E.8.6						E.8.6				
	E.8.7						E.8.7				
E Total				3	3	E Total					
F	F.8.2					F	F.8.2				
	F.8.3			1	1		F.8.3				
	F.8.6			3	3		F.8.6				
	F.8.7			1	1		F.8.7				
	F.8.9						F.8.9				
F Total				5	5	F Total					
G	G.8.1	1			1	G	G.8.1				
	G.8.2						G.8.2				
	G.8.3			1	1		G.8.3				
	G.8.6	1		1	2		G.8.6				
	G.8.7	1			1		G.8.7				
G Total		3		2	5	G Total					
H	H.8.1	2		4	6	H	H.8.1				
	H.8.2	3		3	6		H.8.2				
	H.8.3	2		2	4		H.8.3				
H Total		7		9	16	H Total					
Total		**22**	**8**	**70**	**100**	**Total**		**0**	**0**	**4**	**4**

Table 4-17
Science Grade 10 2005 Item Development and November 2005 Item Rev. Results

Total Items Developed for 2006							**Items Accepted without Revision**					
Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total		Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total
B	B.12.1			1	1		B	B.12.1				
	B.12.5			1	1			B.12.5				
B Total				2	2		B Total					
C	C.12.1		4		4		C	C.12.1		3		3
	C.12.2		1		1			C.12.2				
	C.12.3		2		2			C.12.3				
	C.12.4		1		1			C.12.4				
C Total			8		8		C Total			3		3
H	H.12.1			4	4		H	H.12.1				
	H.12.2			3	3			H.12.2			1	1
	H.12.5			3	3			H.12.5			3	3
	H.12.6			6	6			H.12.6			2	2
H Total				16	16		H Total				6	6
Total			**8**	**18**	**26**					**3**	**6**	**9**

Items Accepted with Revisions							Items Rejected					
Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total		Standard	Subskill	2pt-BCR	3-4 pt-ECR	SR	Grand Total
B	B.12.1			1	1		B	B.12.1				
	B.12.5			1	1			B.12.5				
B Total				2	2		B Total					
C	C.12.1		1		1		C	C.12.1				
	C.12.2		1		1			C.12.2				
	C.12.3		2		2			C.12.3				
	C.12.4		1		1			C.12.4				
C Total			5		5		C Total					
H	H.12.1			4	4		H	H.12.1				
	H.12.2			2	2			H.12.2				
	H.12.5							H.12.5				
	H.12.6			4	4			H.12.6				
H Total				10	10		H Total					
Total			**5**	**12**	**17**							

Table 4-18
Reading: 2006 Item Development Plan by Grade, Reporting Category, and Item Type

Grade	Reporting Category	CR	Paired CR	SR	Grand Total
3	1 Determines Meaning of Words			6	6
	2 Understands Text			6	6
	3 Analyzes Text	1		9	10
	4 Evaluate and Extends Text	1	1	3	5
3 Total		2	1	24	27
4	1 Determines Meaning of Words			2	2
	2 Understands Text			2	2
	3 Analyzes Text	3		8	11
	4 Evaluate and Extends Text	1		6	7
4 Total		4		18	22
5	1 Determines Meaning of Words			4	4
	2 Understands Text			2	2
	3 Analyzes Text	3		14	17
	4 Evaluate and Extends Text	3		7	10
5 Total		6		27	33
6	1 Determines Meaning of Words			6	6
	2 Understands Text			4	4
	3 Analyzes Text	3		18	21
	4 Evaluate and Extends Text	4		5	9
6 Total		7		33	40
7	1 Determines Meaning of Words			1	1
	2 Understands Text			2	2
	3 Analyzes Text	2		14	16
	4 Evaluate and Extends Text		3	6	9
7 Total		2	3	23	28
8	1 Determines Meaning of Words			2	2
	2 Understands Text			4	4
	3 Analyzes Text	2		10	12
	4 Evaluate and Extends Text		2	2	4
8 Total		2	2	18	22
Grand Total		**23**	**6**	**143**	**172**

Table 4-19
Reading: 2006 Item Development Plan by Rationale and Item Type

Grade	Rationale	CR	Paired CR	SR	Grand Total
3	Increase flexibility			9	9
	Meet DoK needs			9	9
	Provide greater flexibility in meeting blueprint	2	1	6	9
3 Total		2	1	24	27
4	Increase flexibility	1		3	4
	Increase pool			2	2
	Meet DoK needs	1		6	7
	Provide greater flexibility in meeting blueprint	2		7	9
4 Total		4		18	22
5	Increase pool			6	6
	Meet DoK needs			15	15
	Provide greater flexibility in meeting blueprint	6		6	12
5 Total		6		27	33
6	Increase pool			16	16
	Meet DoK needs			9	9
	Provide greater flexibility in meeting blueprint	7		8	15
6 Total		7		33	40
7	Increase flexibility	2	2	7	11
	Increase pool			3	3
	Meet DoK needs			10	10
	Provide greater flexibility in meeting blueprint		1	2	3
	Not specified			1	1
7 Total		2	3	23	28
8	Increase flexibility	2	2		4
	Increase pool			5	5
	Meet DoK needs			9	9
	Provide greater flexibility in meeting blueprint			2	2
	Not specified			2	2
8 Total		2	2	18	22
Grand Total		**23**	**6**	**143**	**172**

Table 4-20
Mathematics: 2006 Item Development Plan by Grade, Reporting Category, and Item Type

Grade	Reporting Category-Subskill	2 pt CR	3 pt CR	SR	Grand Total
3	Ba		1		1
	Ca			3	3
	Cb		1	6	7
	Cc			3	3
	Da			2	2
	Db			2	2
	Ea		1	1	2
	Eb			3	3
	Fa	1			1
3 Total		1	3	20	24
4	Ba			1	1
	Bb			1	1
	Ca		1	2	3
	Cb			4	4
	Cc			1	1
	Da			2	2
	Db		1	1	2
	Dc			1	1
	Ea		1	2	3
	Eb			3	3
	Fa	1		1	2
	Fb			1	1
4 Total		1	3	20	24
5	A			2	2
	Ba			1	1
	Ca	1	1	1	3
	Cb			3	3
	Cc			1	1
	Da			1	1
	Db			1	1
	Dc		1		1
	Ea		1	5	6
	Eb			2	2
	Fb			3	3
5 Total		1	3	20	24
6	Ca		1	3	4
	Cb			3	3
	Cc			1	1
	Da		1		1
	Dc			1	1
	Ea			8	8
	Eb			6	6
	Fa		1	2	3
	Fb		1		1

Table 4-20, continued

6 Total			4	24	28
7	A			2	2
	Ba			1	1
	Ca			3	3
	Cb		1	2	3
	Cc			2	2
	Dc	1			1
	Ea			7	7
	Eb		1	7	8
	Fa	1			1
7 Total		2	2	24	28
8	A			3	3
	Ca			2	2
	Cb			1	1
	Cc			1	1
	Da			2	2
	Db			1	1
	Dc			1	1
	Ea		1	6	7
	Eb	2	1	7	10
8 Total		2	2	24	28
Grand Total		**7**	**17**	**132**	**156**

Table 4-21
Mathematics: 2006 Item Development Plan by Rationale and Item Type

Grade	Rationale	2pt CR	BCR	SR	Grand Total
3	Increase flexibility	1	3	18	22
	Meet DoK Needs			2	2
3 Total		1	3	20	24
4	Increase flexibility		3	13	16
	Meet DoK Needs			3	3
	Provide greater flexibility in meeting blueprint	1		4	5
4 Total		1	3	20	24
5	Increase flexibility	1			1
	Increase Pool		1	9	10
	Meet DoK Needs		1	9	10
	Provide greater flexibility in meeting blueprint		1	2	3
5 Total		1	3	20	24
6	Increase flexibility			2	2
	Increase Pool		4	10	14
	Meet DoK Needs			12	12
6 Total			4	24	28
7	Increase flexibility			6	6
	Meet DoK Needs	2	2	18	22
7 Total		2	2	24	28
8	Increase flexibility			8	8
	Increase Pool			2	2
	Meet DoK Needs	2	2	14	18
8 Total		2	2	24	28
Grand Total		**7**	**17**	**132**	**156**

Table 4-22
Mathematics Grade 3 2006 Item Development and January 2007 Item Rev. Results

WKCE-CRT 2007 Item Development
Mathematics Grade 3

Total Items Developed for 2007								**Items Accepted without Revision**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	A				2			A	A					
A Total						2		A Total						
B	Ba		1	1	0			B	Ba					
	Bb				0				Bb					
B Total					0	1		B Total						
C	Ca		1	1	2			C	Ca		1	1		
	Cb				7				Cb				3	
	Cc				3				Cc					
C Total						13		C Total						4
D	Da				2			D	Da				1	
	Db				2				Db					
	Dc				0				Dc					
D Total						4		D Total						1
E	Ea				7			E	Ea				4	
	Eb		1	1	4				Eb				1	
E Total						12		E Total						5
F	Fa	1						F	Fa	1				
	Fb								Fb					
	Fc								Fc					
F Total						1		F Total						1
Total		**1**	**3**	**3**	**30**	**33**		**Total**		**1**	**1**	**1**	**9**	**11**

Table 4-22, continued

WKCE-CRT 2007 Item Development Mathematics Grade 3														
Items Accepted with Revisions								**Items Rejected**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A								A					2	
A Total								A Total						2
B	Ba				1			B	Ba					
	Bb								Bb					
B Total						1		B Total						
C	Ca				2			C	Ca					
	Cb				3				Cb				1	
	Cc				3				Cc					
C Total						8		C Total						1
D	Da				2			D	Da					
	Db				1				Db					
	Dc				0				Dc					
D Total						3		D Total						
E	Ea				3			E	Ea					
	Eb		1	1	3				Eb					
E Total						7		E Total						
F	Fa							F	Fa					
	Fb								Fb					
	Fc								Fc					
F Total								F Total						
Total		**0**	**1**	**1**	**18**	**19**		**Total**		**0**	**0**	**0**	**3**	**3**

Table 4-23
Mathematics Grade 4 2006 Item Development and January 2007 Item Rev. Results

WKCE-CRT 2007 Item Development Mathematics Grade 4														
Total Items Developed for 2007								**Items Accepted without Revision**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	A				5			A	A				3	
A Total						5		A Total						3
B	Ba							B	Ba					
	Bb				1				Bb				1	
B Total						1		B Total						1
C	Ca				1			C	Ca				1	
	Cb		1	1	4				Cb				1	
	Cc				1				Cc				1	
C Total						7		C Total						3
D	Da				2			D	Da				2	
	Db		1	1	2				Db					
	Dc								Dc					
D Total						5		D Total						2
E	Ea		1	1	5			E	Ea				2	
	Eb				3				Eb				3	
E Total						9		E Total						5
F	Fa	1			1			F	Fa	1				
	Fb				1				Fb				1	
	Fc								Fc					
F Total						3		F Total						2
Total		**1**	**3**	**3**	**26**	**30**		**Total**		**1**	**0**	**0**	**15**	**16**

Table 4-23, continued

WKCE-CRT 2007 Item Development Mathematics Grade 4														
Items Accepted with Revisions								**Items Rejected**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A					1			A						
A Total						1		A Total						
B	Ba							B	Ba					
	Bb								Bb					
B Total								B Total						
C	Ca				1			C	Ca					
	Cb		1	1	3				Cb					
	Cc								Cc					
C Total						5		C Total						
D	Da							D	Da					
	Db		1	1	2				Db					
	Dc								Dc					
D Total						3		D Total						
E	Ea		1	1	3			E	Ea					
	Eb								Eb					
E Total						4		E Total						
F	Fa				1			F	Fa					
	Fb								Fb					
	Fc								Fc					
F Total						1		F Total						
Total		**0**	**3**	**3**	**11**	**14**		**Total**		**0**	**0**	**0**	**0**	**0**

Table 4-24
Mathematics Grade 5 2006 Item Development and January 2007 Item Rev. Results

WKCE-CRT 2007 Item Development
Mathematics Grade 5

Total Items Developed for 2007								**Items Accepted without Revision**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	A				5			A	A				1	
A Total						5		A Total						1
B	Ba				1			B	Ba				1	
	Bb								Bb					
B Total						1		B Total						1
C	Ca	1	1	1	2			C	Ca					
	Cb				2				Cb					
	Cc				1				Cc				1	
C Total						7		C Total						1
D	Da				1			D	Da					
	Db				1				Db					
	Dc		1	1					Dc					
D Total						3		D Total						
E	Ea		1	1	10			E	Ea				6	
	Eb	1			4				Eb				1	
E Total						16		E Total						7
F	Fa							F	Fa					
	Fb				3				Fb				2	
	Fc								Fc					
F Total						3		F Total						2
Total		**2**	**3**	**3**	**30**	**35**		**Total**					**12**	**12**

Table 4-24, continued

WKCE-CRT 2007 Item Development Mathematics Grade 5														
Items Accepted with Revisions								**Items Rejected**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A					4			A						
A Total						4		A Total						
B	Ba							B	Ba					
	Bb								Bb					
B Total						0		B Total						
C	Ca	1	1	1	2			C	Ca					
	Cb				2				Cb					
	Cc								Cc					
C Total						6		C Total						
D	Da				1			D	Da					
	Db				1				Db					
	Dc		1	1					Dc					
D Total						3		D Total						
E	Ea		1	1	4			E	Ea					
	Eb	1			3				Eb					
E Total						9		E Total						
F	Fa							F	Fa					
	Fb				1				Fb					
	Fc								Fc					
F Total						1		F Total						
Total		**2**	**3**	**3**	**18**	**23**		**Total**		**0**	**0**	**0**	**0**	**0**

Table 4-25
Mathematics Grade 6 2006 Item Development and January 2007 Item Rev. Results

WKCE-CRT 2007 Item Development
Mathematics Grade 6

Total Items Developed for 2007								**Items Accepted without Revision**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	A				4			A	A				1	
A Total						4		A Total						1
B	Ba							B	Ba					
	Bb								Bb					
B Total						0		B Total						0
C	Ca				3			C	Ca				2	
	Cb		1	1	3				Cb				2	
	Cc				1				Cc				0	
C Total						8		C Total						4
D	Da							D	Da					
	Db		1	1					Db					
	Dc				1				Dc					
D Total						2		D Total						
E	Ea				8			E	Ea				3	
	Eb				5				Eb				3	
E Total						13		E Total						6
F	Fa		2	2	2			F	Fa					
	Fb								Fb					
	Fc				1				Fc					
F Total						5		F Total						
Total		**0**	**4**	**4**	**28**	**32**		**Total**		**0**	**0**	**0**	**11**	**11**

Table 4-25, continued

WKCE-CRT 2007 Item Development Mathematics Grade 6														
Items Accepted with Revisions								**Items Rejected**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A					2			A					1	
A Total						2		A Total						1
B	Ba							B	Ba					
	Bb								Bb					
B Total						0		B Total						0
C	Ca							C	Ca				1	
	Cb		1	1	1				Cb					
	Cc				1				Cc					
C Total						3		C Total						1
D	Da							D	Da					
	Db		1	1					Db					
	Dc				1				Dc					
D Total						2		D Total						0
E	Ea				4			E	Ea				1	
	Eb				2				Eb					
E Total						6		E Total						1
F	Fa		2	2	2			F	Fa					
	Fb								Fb					
	Fc				1				Fc					
F Total						5		F Total						0
Total		**0**	**4**	**4**	**14**	**18**		**Total**		**0**	**0**	**0**	**3**	**3**

Table 4-26
Mathematics Grade 7 2006 Item Development and January 2007 Item Rev. Results

WKCE-CRT 2007 Item Development
Mathematics Grade 7

Total Items Developed for 2007								**Items Accepted without Revision**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	A	1			5			A	A	1				
A Total						6		A Total						1
B	Ba				2			B	Ba					
	Bb								Bb					
B Total						2		B Total						0
C	Ca				3			C	Ca				1	
	Cb		1	1	1				Cb		1	1	1	
	Cc				2				Cc				1	
C Total						7		C Total						4
D	Da							D	Da					
	Db								Db					
	Dc								Dc					
D Total						0		D Total						0
E	Ea				9			E	Ea				2	
	Eb				6				Eb					
E Total						15		E Total						2
F	Fa							F	Fa					
	Fb		2	2					Fb					
	Fc								Fc					
F Total						2		F Total						0
Total		**1**	**3**	**3**	**28**	**32**		**Total**		**1**	**1**	**1**	**5**	**7**

Table 4-26, continued

WKCE-CRT 2007 Item Development Mathematics Grade 7														
Items Accepted with Revisions								**Items Rejected**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A					3			A					2	
A Total						3		A Total						2
B	Ba				1			B	Ba				1	
	Bb								Bb					
B Total						1		B Total						1
C	Ca				1			C	Ca				1	
	Cb								Cb					
	Cc				1				Cc					
C Total						2		C Total						1
D	Da							D	Da					
	Db								Db					
	Dc								Dc					
D Total						0		D Total						0
E	Ea				6			E	Ea				1	
	Eb				4				Eb				2	
E Total						10		E Total						3
F	Fa							F	Fa					
	Fb		2	2					Fb					
	Fc								Fc					
F Total						2		F Total						0
Total		**0**	**2**	**2**	**16**	**18**		**Total**		**0**	**0**	**0**	**7**	**7**

Table 4-27
Mathematics Grade 8 2006 Item Development and January 2007 Item Rev. Results

WKCE-CRT 2007 Item Development
Mathematics Grade 8

Total Items Developed for 2007								**Items Accepted without Revision**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A	A				7			A	A				2	
A Total								A Total						2
B	Ba							B	Ba					
	Bb								Bb					
B Total								B Total						0
C	Ca				2			C	Ca				1	
	Cb				1				Cb				1	
	Cc				1				Cc					
C Total								C Total						2
D	Da				2			D	Da				1	
	Db								Db					
	Dc				2				Dc				1	
D Total								D Total						2
E	Ea		1	1	6			E	Ea		1	1	4	
	Eb	2	1	1	7				Eb	1	1	1	4	
E Total								E Total						11
F	Fa							F	Fa					
	Fb								Fb					
	Fc								Fc					
F Total								F Total						
Total		**2**	**2**	**2**	**28**	**32**		**Total**		**1**	**2**	**2**	**14**	**17**

Table 4-27, continued

WKCE-CRT 2007 Item Development Mathematics Grade 8														
Items Accepted with Revisions								**Items Rejected**						
Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total		Reporting Category	Subskill	2pt-CR	A-BCR	B-BCR	SR	Grand Total
A					4			A					1	
A Total						4		A Total						1
B	Ba							B	Ba					
	Bb								Bb					
B Total						0		B Total						0
C	Ca				1			C	Ca					
	Cb								Cb					
	Cc				1				Cc					
C Total						2		C Total						0
D	Da				1			D	Da					
	Db								Db					
	Dc				1				Dc					
D Total						2		D Total						0
E	Ea				2			E	Ea					
	Eb	1			3				Eb					
E Total						6		E Total						0
F	Fa							F	Fa					
	Fb								Fb					
	Fc								Fc					
F Total						0		F Total						0
Total		**1**	**0**	**0**	**13**	**14**		**Total**		**0**	**0**	**0**	**1**	**1**

Table 4-28
Reading Grade 3 2006 Item Development and January 2007 Item Rev. Results

Reporting Category Subskill	**Items Written**		**Items Accepted**		**Items Revised**		**Items Rejected**	
	SR	**CR**	**SR**	**CR**	**SR**	**CR**	**SR**	**CR**
Aa								
Aa1	8	0	5	0	1	0	2	0
Aa2	1	0	0	0	0	0	1	0
Aa3	0	0	0	0	0	0	0	0
Total	**9**	**0**	**5**	**0**	**1**	**0**	**3**	**0**
Ab								
Ab1	7	0	1	0	6	0	0	0
Ab2	3	0	2	0	1	0	0	0
Ab3	1	0	0	0	1	0	0	0
Total	**11**	**0**	**3**	**0**	**8**	**0**	**0**	**0**
Ac								
Ac1	10	0	3	0	7	0	0	0
Ac2	1	1	0	1	1	0	0	0
Ac3	3	0	0	0	1	0	2	0
Total	**14**	**1**	**3**	**1**	**9**	**0**	**2**	**0**
Ad								
Ad1	3	1	0	0	3	1	0	0
Ad2	3	1	1	0	1	1	1	0
Ad3	0	0	0	0	0	0	0	0
Total	**6**	**2**	**1**	**0**	**4**	**2**	**1**	**0**
Grand Total	**40**	**3**	**12**	**1**	**22**	**2**	**6**	**0**

Table 4-29
Reading Grade 4 2006 Item Development and January 2007 Item Rev. Results

Reporting Category Subskill	**Items Written**		**Items Accepted**		**Items Revised**		**Items Rejected**	
	SR	**CR**	**SR**	**CR**	**SR**	**CR**	**SR**	**CR**
Aa								
Aa1	2	0	1	0	0	0	1	0
Aa2	1	0	1	0	0	0	0	0
Aa3	0	0	0	0	0	0	0	0
Total	**3**	**0**	**2**	**0**	**0**	**0**	**1**	**0**
Ab								
Ab1	0	0	0	0	0	0	0	0
Ab2	2	0	2	0	0	0	0	0
Ab3	0	0	0	0	0	0	0	0
Total	**2**	**0**	**2**	**0**	**0**	**0**	**0**	**0**
Ac								
Ac1	5	1	3	1	2	0	0	0
Ac2	5	2	3	2	2	0	0	0
Ac3	4	0	3	0	1	0	0	0
Total	**14**	**3**	**9**	**3**	**5**	**0**	**0**	**0**
Ad								
Ad1	2	0	2	0	0	0	0	0
Ad2	2	1	1	1	1	0	0	0
Ad3	2	0	2	0	0	0	0	0
Total	**6**	**1**	**5**	**1**	**1**	**0**	**0**	**0**
Grand Total	**25**	**4**	**18**	**4**	**6**	**0**	**1**	**0**

Table 4-30
Reading Grade 5 2006 Item Development and January 2007 Item Rev. Results

Reporting Category Subskill	**Items Written**		**Items Accepted**		**Items Revised**		**Items Rejected**	
	SR	**CR**	**SR**	**CR**	**SR**	**CR**	**SR**	**CR**
Aa								
Aa1	4	0	2	0	2	0	0	0
Aa2	0	0	0	0	0	0	0	0
Aa3	1	0	1	0	1	0	0	0
Total	**5**	**0**	**3**	**0**	**3**	**0**	**0**	**0**
Ab								
Ab1	0	0	0	0	0	0	0	0
Ab2	1	0	1	0	0	0	0	0
Ab3	2	0	0	0	2	0	0	0
Total	**3**	**0**	**1**	**0**	**2**	**0**	**0**	**0**
Ac								
Ac1	9	1	1	0	8	1	0	0
Ac2	7	4	1	3	6	1	0	0
Ac3	0	0	0	0	0	0	0	0
Total	**16**	**5**	**2**	**3**	**14**	**2**	**0**	**0**
Ad								
Ad1	4	1	1	0	3	0	0	1
Ad2	1	1	0	0	1	0	0	1
Ad3	0	0	0	0	0	0	0	0
Total	**5**	**2**	**1**	**0**	**4**	**0**	**0**	**2**
Grand Total	**29**	**7**	**7**	**3**	**23**	**2**	**0**	**2**

Table 4-31
Reading Grade 6 2006 Item Development and January 2007 Item Rev. Results

Reporting Category Subskill	**Items Written**		**Items Accepted**		**Items Revised**		**Items Rejected**	
	SR	**CR**	**SR**	**CR**	**SR**	**CR**	**SR**	**CR**
Aa								
Aa1	6	0	1	0	4	0	1	0
Aa2	1	0	0	0	1	0	0	0
Aa3	0	0	0	0	0	0	0	0
Total	**7**	**0**	**1**	**0**	**5**	**0**	**1**	**0**
Ab								
Ab1	1	0	1	0	0	0	0	0
Ab2	1	0	1	0	0	0	0	0
Ab3	1	0	0	0	1	0	0	0
Total	**3**	**0**	**2**	**0**	**1**	**0**	**0**	**0**
Ac								
Ac1	10	1	3	0	6	1	1	0
Ac2	9	1	6	1	3	0	0	0
Ac3	5	1	2	0	3	1	0	0
Total	**24**	**3**	**11**	**1**	**12**	**2**	**1**	**0**
Ad								
Ad1	1	1	0	0	1	1	0	0
Ad2	1	1	0	0	1	1	0	0
Ad3	1	1	0	0	1	1	0	0
Total	**3**	**3**	**0**	**0**	**3**	**3**	**0**	**0**
Grand Total	**37**	**6**	**14**	**1**	**21**	**5**	**2**	**0**

Table 4-32
Reading Grade 7 2006 Item Development and January 2007 Item Rev. Results

Reporting Category Subskill	**Items Written**		**Items Accepted**		**Items Revised**		**Items Rejected**	
	SR	**CR**	**SR**	**CR**	**SR**	**CR**	**SR**	**CR**
Aa								
Aa1	1	0	0	0	1	0	0	0
Aa2	0	0	0	0	0	0	0	0
Aa3	0	0	0	0	0	0	0	0
Total	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**
Ab								
Ab1	0	0	0	0	0	0	0	0
Ab2	3	0	3	0	0	0	0	0
Ab3	0	0	0	0	0	0	0	0
Total	**3**	**0**	**3**	**0**	**0**	**0**	**0**	**0**
Ac								
Ac1	6	1	4	0	2	1	0	0
Ac2	8	2	7	1	1	1	0	0
Ac3	3	0	1	0	2	0	0	0
Total	**17**	**3**	**12**	**1**	**5**	**2**	**0**	**0**
Ad								
Ad1	4	1	4	0	0	1	0	0
Ad2	0	1	0	1	0	0	0	0
Ad3	0	0	0	0	0	0	0	0
Total	**4**	**2**	**4**	**1**	**0**	**1**	**0**	**0**
Grand Total	**25**	**5**	**19**	**2**	**6**	**3**	**0**	**0**

Table 4-33
Reading Grade 8 2006 Item Development and January 2007 Item Rev. Results

Reporting Category Subskill	**Items Written**		**Items Accepted**		**Items Revised**		**Items Rejected**	
	SR	**CR**	**SR**	**CR**	**SR**	**CR**	**SR**	**CR**
Aa								
Aa1	1	0	0	0	1	0	0	0
Aa2	0	0	0	0	0	0	0	0
Aa3	1	0	1	0	0	0	0	0
Total	**2**	**0**	**1**	**0**	**1**	**0**	**0**	**0**
Ab								
Ab1	2	0	2	0	0	0	0	0
Ab2	1	0	1	0	0	0	0	0
Ab3	1	0	1	0	0	0	0	0
Total	**4**	**0**	**4**	**0**	**0**	**0**	**0**	**0**
Ac								
Ac1	7	1	5	0	2	1	0	0
Ac2	7	1	7	1	0	0	0	0
Ac3	1	0	0	0	1	0	0	0
Total	**15**	**2**	**12**	**1**	**3**	**1**	**0**	**0**
Ad								
Ad1	3	0	2	0	1	0	0	0
Ad2	0	2	0	1	0	1	0	0
Ad3	0	0	0	0	0	0	0	0
Total	**3**	**2**	**2**	**1**	**1**	**1**	**0**	**0**
Grand Total	**24**	**4**	**19**	**2**	**5**	**2**	**0**	**0**

Table 8-1
Item Flagged Based on Yen's Q_1

Content	Grade	Form	Test Book_ID	CR Part	Status	Type	N	Z	Critical Z
RD	4		48		OP	MC	6402	22.04	17.07
	4	B	62		FT	CR	3013	8.19	8.03
	6		17		OP	MC	6557	19.52	17.49
	6		28		OP	MC	6493	18.97	17.31
	7		11		OP	MC	6654	56.69	17.74
	8		45		OP	MC	6843	27.98	18.25
MA	3		19	B	OP	CR	6251	18.01	16.67
	4		12	B	OP	CR	6274	18.54	16.73
	5		29	B	OP	CR	6408	88.09	17.09
	5		41	B	OP	CR	6458	34.94	17.22
	6		33	B	OP	CR	6448	28.90	17.19
	6	B	56	B	FT	CR	2991	11.81	7.98
	7		9		OP	MC	6783	20.71	18.09
	7		31	B	OP	CR	6483	49.42	17.29
	7		51		OP	MC	6709	21.24	17.89
	7	A	63	A	FT	CR	2987	8.82	7.97
	7	C	56	B	FT	CR	3061	8.51	8.16
	8	A	57	B	FT	CR	3547	14.05	9.46
	10		15		OP	MC	7619	25.16	20.32
	10		30		OP	CR	6838	29.85	18.23
	10		36		OP	CR	6989	26.95	18.64
	10		44		OP	MC	7581	23.07	20.22
	10		47		OP	CR	7125	25.65	19.00
LA	4		18		OP	MC	6328	20.88	16.87
	8		25		OP	MC	6635	22.78	17.69
	8		28		OP	MC	6599	31.54	17.60
	10		17		OP	MC	7407	21.01	19.75
SS	10		13		OP	MC	7370	21.14	19.65
	10		33		OP	MC	7352	19.88	19.61
SC	4		23		OP	MC	6317	17.11	16.85
	8		20		OP	MC	6624	18.06	17.66

Table 8-2
Scoring Table for Reading Grade 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	270	112	42	457	8
1	270	112	43	460	8
2	270	112	44	462	8
3	270	112	45	465	8
4	270	112	**46**	**468**	**8**
5	270	112	47	471	8
6	270	112	48	475	9
7	270	112	49	479	9
8	270	112	50	483	10
9	270	112	51	487	10
10	270	112	52	492	11
11	274	108	53	497	11
12	342	40	54	504	13
13	360	27	55	511	14
14	371	21	56	521	17
15	379	18	57	534	20
16	386	15	58	553	26
17	391	13	59	584	38
18	**396**	**12**	60	640	74
19	400	11			
20	403	10			
21	407	9			
22	410	9			
23	412	9			
24	415	8			
25	418	8			
26	420	8			
27	422	7			
28	425	7			
29	427	7			
30	429	7			
31	**431**	**7**			
32	434	7			
33	436	7			
34	438	7			
35	440	7			
36	442	7			
37	445	7			
38	447	7			
39	449	7			
40	452	7			
41	454	7			

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

Table 8-3
Scoring Table for Reading Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	280	109	**42**	**490**	**11**
1	280	109	43	494	12
2	280	109	44	498	12
3	280	109	45	503	12
4	280	109	46	508	13
5	280	109	47	512	13
6	280	109	48	518	13
7	280	109	49	523	14
8	280	109	50	529	14
9	280	109	51	536	15
10	280	109	52	543	16
11	332	57	53	551	17
12	355	40	54	560	19
13	370	31	55	570	21
14	381	25	56	584	24
15	389	21	57	603	30
16	**396**	**19**	58	634	43
17	403	16	59	650	52
18	408	15			
19	413	14			
20	417	13			
21	421	12			
22	425	11			
23	428	11			
24	432	10			
25	435	10			
26	438	10			
27	**441**	**10**			
28	444	9			
29	447	9			
30	450	9			
31	453	9			
32	456	9			
33	459	9			
34	462	9			
35	465	9			
36	469	10			
37	472	10			
38	475	10			
39	479	10			
40	482	11			
41	486	11			

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

Table 8-4
Scoring Table for Reading Grade 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	290	92	42	494	11
1	290	92	**43**	**497**	**11**
2	290	92	44	501	11
3	290	92	45	505	11
4	290	92	46	509	11
5	290	92	47	513	12
6	290	92	48	518	12
7	290	92	49	523	12
8	290	92	50	528	13
9	290	92	51	533	13
10	290	92	52	539	14
11	290	92	53	546	15
12	290	92	54	553	16
13	328	56	55	562	18
14	353	37	56	574	21
15	369	30	57	588	25
16	381	26	58	609	32
17	390	23	59	646	49
18	398	21	60	690	78
19	**405**	**19**			
20	411	18			
21	417	16			
22	422	15			
23	427	14			
24	431	14			
25	435	13			
26	439	12			
27	443	12			
28	**446**	**12**			
29	450	11			
30	453	11			
31	457	11			
32	460	11			
33	463	10			
34	466	10			
35	470	10			
36	473	10			
37	476	10			
38	480	10			
39	483	10			
40	486	11			
41	490	11			

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

Table 8-5
Scoring Table for Reading Grade 6

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	300	93	42	511	13
1	300	93	**43**	**515**	**13**
2	300	93	44	520	14
3	300	93	45	525	14
4	300	93	46	530	14
5	300	93	47	536	15
6	300	93	48	542	15
7	300	93	49	548	16
8	300	93	50	554	16
9	300	93	51	561	17
10	300	93	52	569	18
11	300	93	53	578	19
12	300	93	54	588	21
13	332	64	55	600	24
14	355	48	56	615	27
15	372	37	57	634	33
16	384	30	58	661	43
17	394	25	59	709	65
18	403	21	60	730	77
19	410	19			
20	416	18			
21	**422**	**17**			
22	428	16			
23	433	16			
24	438	15			
25	443	14			
26	447	14			
27	451	14			
28	456	13			
29	**460**	**13**			
30	463	12			
31	467	12			
32	471	12			
33	475	12			
34	479	12			
35	482	12			
36	486	12			
37	490	12			
38	494	12			
39	498	12			
40	502	12			
41	506	13			

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

Table 8-6
Scoring Table for Reading Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	310	114	42	528	12
1	310	114	43	532	12
2	310	114	44	536	12
3	310	114	45	541	13
4	310	114	46	545	13
5	310	114	47	550	13
6	310	114	48	555	13
7	310	114	49	560	14
8	310	114	50	565	14
9	310	114	51	571	14
10	310	114	52	577	15
11	310	114	53	584	16
12	354	70	54	591	17
13	380	44	55	600	18
14	396	31	56	611	21
15	407	25	57	624	24
16	416	21	58	642	29
17	424	19	59	673	42
18	431	17	60	780	149
19	**437**	**16**			
20	442	15			
21	447	14			
22	452	14			
23	456	13			
24	460	13			
25	464	12			
26	**468**	**12**			
27	472	12			
28	476	12			
29	480	12			
30	483	12			
31	487	12			
32	490	12			
33	494	12			
34	498	12			
35	501	12			
36	505	12			
37	509	12			
38	513	12			
39	516	12			
40	520	12			
41	**524**	**12**			

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

Table 8-7
Scoring Table for Reading Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	330	106	42	533	12
1	330	106	43	537	12
2	330	106	**44**	**542**	**12**
3	330	106	45	546	13
4	330	106	46	550	13
5	330	106	47	555	13
6	330	106	48	560	14
7	330	106	49	565	14
8	330	106	50	571	15
9	330	106	51	577	15
10	330	106	52	584	16
11	330	106	53	592	18
12	330	106	54	601	20
13	362	74	55	612	22
14	389	52	56	625	26
15	406	39	57	644	32
16	419	31	58	673	44
17	428	26	59	730	75
18	436	22	60	790	121
19	443	19			
20	**449**	**17**			
21	454	16			
22	459	15			
23	464	14			
24	468	13			
25	472	13			
26	476	12			
27	**480**	**12**			
28	483	12			
29	487	12			
30	491	12			
31	494	12			
32	498	12			
33	501	11			
34	504	11			
35	508	11			
36	511	11			
37	515	12			
38	519	12			
39	522	12			
40	526	12			
41	530	12			

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

Table 8-8
Scoring Table for Reading Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	350	55	42	566	15
1	350	55	43	572	15
2	350	55	44	577	16
3	350	55	45	584	16
4	350	55	46	590	17
5	350	55	47	597	17
6	350	55	48	605	19
7	350	55	49	614	20
8	350	55	50	624	22
9	350	55	51	637	25
10	350	55	52	652	29
11	350	55	53	672	35
12	356	52	54	703	46
13	379	40	55	764	78
14	395	33	56	820	123
15	408	28			
16	419	25			
17	428	23			
18	437	21			
19	444	20			
20	452	19			
21	**458**	**19**			
22	465	18			
23	471	18			
24	477	17			
25	482	17			
26	488	16			
27	493	16			
28	498	16			
29	**503**	**16**			
30	508	15			
31	513	15			
32	518	15			
33	522	15			
34	527	15			
35	532	14			
36	536	14			
37	541	14			
38	546	14			
39	551	14			
40	**556**	**15**			
41	561	15			

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

Table 8-9
Scoring Table for Mathematics Grade 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	220	87	42	437	12
1	220	87	43	442	12
2	220	87	44	446	12
3	220	87	45	451	13
4	220	87	**46**	**455**	**13**
5	220	87	47	460	13
6	220	87	48	466	14
7	220	87	49	472	14
8	220	87	50	478	15
9	220	87	51	485	16
10	252	57	52	493	18
11	274	41	53	503	20
12	290	32	54	514	22
13	302	26	55	531	27
14	311	23	56	558	39
15	319	20	57	630	99
16	326	18			
17	333	17			
18	339	16			
19	344	15			
20	349	14			
21	354	14			
22	359	13			
23	363	13			
24	367	13			
25	371	13			
26	376	12			
27	380	12			
28	384	12			
29	387	12			
30	391	12			
31	**395**	**12**			
32	399	12			
33	403	11			
34	406	11			
35	**410**	**11**			
36	414	11			
37	418	11			
38	421	12			
39	425	12			
40	429	12			
41	433	12			

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

Table 8-10
Scoring Table for Mathematics Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	240	72	42	472	11
1	240	72	43	476	11
2	240	72	44	480	11
3	240	72	**45**	**484**	**12**
4	240	72	46	489	12
5	240	72	47	494	12
6	240	72	48	499	13
7	240	72	49	505	14
8	240	72	50	511	14
9	240	72	51	518	16
10	249	65	52	526	17
11	281	47	53	536	20
12	302	38	54	549	23
13	317	32	55	566	29
14	330	28	56	593	42
15	340	24	57	650	84
16	350	21			
17	358	19			
18	365	18			
19	371	17			
20	377	16			
21	383	15			
22	389	15			
23	394	14			
24	399	14			
25	404	14			
26	408	13			
27	413	13			
28	417	13			
29	**421**	**12**			
30	425	12			
31	429	12			
32	433	11			
33	437	11			
34	**441**	**11**			
35	445	11			
36	449	11			
37	452	11			
38	456	11			
39	460	11			
40	464	11			
41	468	11			

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

Table 8-11
Scoring Table for Mathematics Grade 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	270	84	42	491	11
1	270	84	43	495	11
2	270	84	44	498	11
3	270	84	45	502	11
4	270	84	**46**	**506**	**11**
5	270	84	47	510	11
6	270	84	48	514	12
7	270	84	49	518	12
8	270	84	50	522	12
9	270	84	51	527	12
10	270	84	52	531	13
11	280	74	53	537	13
12	312	47	54	542	14
13	331	35	55	548	15
14	346	28	56	555	16
15	357	24	57	564	18
16	367	22	58	573	19
17	375	20	59	586	23
18	383	19	60	603	28
19	390	18	61	633	41
20	396	17	62	680	75
21	402	17			
22	408	16			
23	414	16			
24	419	15			
25	424	15			
26	429	14			
27	434	14			
28	438	13			
29	442	13			
30	**447**	**13**			
31	451	13			
32	455	12			
33	458	12			
34	462	12			
35	**466**	**12**			
36	470	12			
37	473	11			
38	477	11			
39	480	11			
40	484	11			
41	487	11			

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

Table 8-12
Scoring Table for Mathematics Grade 6*

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	310	89	42	520	9
1	310	89	43	523	10
2	310	89	44	526	10
3	310	89	45	530	10
4	310	89	**46**	**533**	**10**
5	310	89	47	536	10
6	310	89	48	540	10
7	310	89	49	544	11
8	310	89	50	548	11
9	311	88	51	552	12
10	354	50	52	557	12
11	376	36	53	562	13
12	391	28	54	568	14
13	402	24	55	575	15
14	412	21	56	583	17
15	419	19	57	592	19
16	426	18	58	605	22
17	433	16	59	622	28
18	438	15	60	653	42
19	443	14	61	700	77
20	448	14			
21	453	13			
22	457	12			
23	461	12			
24	**465**	**12**			
25	468	11			
26	472	11			
27	475	11			
28	478	11			
29	482	10			
30	**485**	**10**			
31	488	10			
32	491	10			
33	494	10			
34	497	10			
35	500	10			
36	503	10			
37	506	9			
38	508	9			
39	511	9			
40	514	9			
41	517	9			

*A suppressed item here reduces maximum raw score from 62 to 61.
** **Bold** represents SEM around cut score.

Table 8-13
Scoring Table for Mathematics Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	330	90	42	545	10
1	330	90	43	549	10
2	330	90	44	552	10
3	330	90	**45**	**555**	**10**
4	330	90	46	558	10
5	330	90	47	562	10
6	330	90	48	565	10
7	330	90	49	569	10
8	330	90	50	573	10
9	330	90	51	576	10
10	330	90	52	581	11
11	364	59	53	585	11
12	391	40	54	590	12
13	408	29	55	596	13
14	420	24	56	602	14
15	430	21	57	610	16
16	438	19	58	619	18
17	445	17	59	631	21
18	452	16	60	648	27
19	458	16	61	678	40
20	464	15	62	710	60
21	469	15			
22	474	14			
23	479	14			
24	**483**	**13**			
25	487	13			
26	491	12			
27	495	12			
28	499	12			
29	503	11			
30	**507**	**11**			
31	510	11			
32	513	11			
33	517	11			
34	520	11			
35	523	10			
36	527	10			
37	530	10			
38	533	10			
39	536	10			
40	539	10			
41	542	10			

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

Table 8-14
Scoring Table for Mathematics Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	350	77	42	564	10
1	350	77	43	567	10
2	350	77	44	570	10
3	350	77	**45**	**573**	**10**
4	350	77	46	577	10
5	350	77	47	580	10
6	350	77	48	584	10
7	350	77	49	587	10
8	350	77	50	591	11
9	350	77	51	595	11
10	350	77	52	599	11
11	350	77	53	604	11
12	350	77	54	608	11
13	387	47	55	613	12
14	411	35	56	619	13
15	427	30	57	626	14
16	440	26	58	634	15
17	450	23	59	644	18
18	459	21	60	660	24
19	467	20	61	687	37
20	474	18	62	730	71
21	481	17			
22	**487**	**16**			
23	492	16			
24	497	15			
25	502	15			
26	507	14			
27	511	14			
28	**516**	**13**			
29	520	13			
30	524	12			
31	527	12			
32	531	12			
33	535	11			
34	538	11			
35	541	11			
36	545	11			
37	548	10			
38	551	10			
39	554	10			
40	557	10			
41	561	10			

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

Table 8-15
Scoring Table for Mathematics Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	410	63	42	589	9
1	410	63	43	592	9
2	410	63	**44**	**596**	**10**
3	410	63	45	599	10
4	410	63	46	603	10
5	410	63	47	606	10
6	410	63	48	611	11
7	410	63	49	615	11
8	410	63	50	620	12
9	410	63	51	625	13
10	410	63	52	632	14
11	410	63	53	640	16
12	422	52	54	649	18
13	447	36	55	662	22
14	463	29	56	682	30
15	475	24	57	719	47
16	485	21	58	750	65
17	493	20			
18	500	18			
19	507	17			
20	513	16			
21	**518**	**15**			
22	523	14			
23	527	14			
24	532	13			
25	536	12			
26	539	12			
27	**543**	**12**			
28	547	11			
29	550	11			
30	553	11			
31	557	10			
32	560	10			
33	563	10			
34	566	10			
35	569	10			
36	572	9			
37	574	9			
38	577	9			
39	580	9			
40	583	9			
41	586	9			

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

Table 8-16
Scoring Table for Language Arts Grade 4

Raw Score	Scale Score	SEM
0	140	108
1	140	108
2	140	108
3	140	108
4	140	108
5	140	108
6	186	62
7	220	28
8	232	18
9	241	13
10	247	12
11	**253**	**11**
12	258	10
13	263	10
14	267	10
15	272	10
16	276	9
17	**280**	**9**
18	284	9
19	289	9
20	293	9
21	297	9
22	302	9
23	307	10
24	**312**	**10**
25	318	11
26	324	11
27	332	12
28	343	15
29	361	23
30	420	82

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

Table 8-17
Scoring Table for Language Arts Grade 8

Raw Score	Scale Score	SEM
0	250	85
1	250	85
2	250	85
3	250	85
4	250	85
5	250	85
6	250	85
7	265	70
8	300	35
9	315	23
10	326	18
11	334	15
12	341	14
13	347	13
14	352	12
15	357	12
16	**362**	**11**
17	367	11
18	372	11
19	377	11
20	382	11
21	**387**	**11**
22	392	11
23	398	11
24	403	11
25	410	12
26	**418**	**13**
27	428	15
28	443	20
29	468	31
30	520	59

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

Table 8-18
Scoring Table for Language Arts Grade 10

Raw Score	Scale Score	SEM
0	290	66
1	290	66
2	290	66
3	290	66
4	290	66
5	290	66
6	290	66
7	323	33
8	340	23
9	351	19
10	360	17
11	368	16
12	375	16
13	382	16
14	388	16
15	**395**	**15**
16	401	15
17	407	15
18	412	14
19	418	14
20	423	14
21	**429**	**13**
22	434	13
23	439	13
24	444	13
25	450	13
26	455	13
27	461	13
28	467	13
29	473	14
30	480	15
31	**488**	**15**
32	496	16
33	505	17
34	515	18
35	526	19
36	540	21
37	557	26
38	586	37
39	630	68

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

Table 8-19
Scoring Table for Social Studies Grade 4

Raw Score	Scale Score	SEM
0	170	71
1	170	71
2	170	71
3	170	71
4	170	71
5	170	71
6	170	71
7	170	71
8	194	47
9	216	25
10	226	16
11	233	13
12	238	11
13	**242**	**10**
14	246	9
15	250	8
16	253	8
17	256	8
18	259	7
19	262	7
20	**265**	**7**
21	267	7
22	270	7
23	273	7
24	275	7
25	278	7
26	281	7
27	284	7
28	287	7
29	**290**	**7**
30	293	7
31	296	7
32	300	8
33	304	8
34	310	10
35	317	12
36	327	15
37	344	22
38	400	78

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

Table 8-20
Scoring Table for Social Studies Grade 8

Raw Score	Scale Score	SEM
0	230	99
1	230	99
2	230	99
3	230	99
4	230	99
5	230	99
6	230	99
7	230	99
8	230	99
9	254	75
10	291	38
11	306	23
12	316	17
13	324	15
14	330	13
15	**336**	**13**
16	341	12
17	346	12
18	351	11
19	355	11
20	359	10
21	363	10
22	**367**	**10**
23	370	10
24	374	10
25	378	9
26	381	9
27	385	9
28	389	9
29	393	9
30	397	9
31	401	10
32	**405**	**10**
33	410	10
34	415	11
35	421	12
36	428	13
37	437	15
38	450	19
39	470	26
40	530	83

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

Table 8-21
Scoring Table for Social Studies Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	240	141	42	482	11
1	240	141	43	487	11
2	240	141	44	493	12
3	240	141	45	499	13
4	240	141	46	507	14
5	240	141	47	517	16
6	240	141	48	530	18
7	240	141	49	549	24
8	240	141	50	620	95
9	240	141			
10	240	141			
11	279	102			
12	328	53			
13	348	33			
14	361	24			
15	370	20			
16	378	18			
17	385	16			
18	391	15			
19	396	14			
20	401	13			
21	405	13			
22	**409**	**12**			
23	413	12			
24	417	11			
25	**421**	**11**			
26	424	11			
27	428	10			
28	431	10			
29	434	10			
30	438	10			
31	441	9			
32	444	9			
33	447	9			
34	451	9			
35	454	9			
36	**457**	**10**			
37	461	10			
38	465	10			
39	469	10			
40	473	10			
41	477	11			

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

Table 8-22
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	185	45
10	204	27
11	215	21
12	224	18
13	231	16
14	238	14
15	243	13
16	248	12
17	**253**	**12**
18	257	11
19	261	11
20	265	10
21	269	10
22	272	10
23	276	9
24	**280**	**9**
25	283	9
26	286	9
27	290	9
28	293	9
29	297	9
30	300	9
31	304	9
32	308	9
33	312	9
34	317	10
35	**322**	**10**
36	329	12
37	337	14
38	350	19
39	372	30
40	440	96

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

Table 8-23
Scoring Table for Science Grade 8

Raw Score	Scale Score	SEM
0	230	103
1	230	103
2	230	103
3	230	103
4	230	103
5	230	103
6	230	103
7	230	103
8	230	103
9	230	103
10	259	74
11	292	41
12	308	28
13	320	22
14	329	19
15	336	17
16	342	15
17	348	14
18	**353**	**13**
19	358	13
20	363	12
21	367	12
22	372	11
23	**376**	**11**
24	380	11
25	384	11
26	389	11
27	393	11
28	397	11
29	402	11
30	406	11
31	411	11
32	417	12
33	**422**	**12**
34	428	13
35	435	14
36	443	15
37	453	17
38	467	21
39	492	32
40	560	98

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

Table 8-24
Scoring Table for Science Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	240	134	42	498	13
1	240	134	43	504	14
2	240	134	44	511	14
3	240	134	45	519	16
4	240	134	46	528	18
5	240	134	47	540	21
6	240	134	48	557	26
7	240	134	49	587	39
8	240	134	50	610	54
9	240	134			
10	255	119			
11	318	56			
12	341	36			
13	356	26			
14	367	21			
15	375	18			
16	382	16			
17	389	15			
18	394	15			
19	400	14			
20	405	13			
21	410	13			
22	**414**	**12**			
23	418	12			
24	422	12			
25	427	11			
26	**430**	**11**			
27	434	11			
28	438	11			
29	442	11			
30	446	11			
31	450	11			
32	453	11			
33	457	11			
34	461	11			
35	465	11			
36	**469**	**11**			
37	473	11			
38	478	11			
39	482	11			
40	487	12			
41	492	12			

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

Table 12-1
WKCE-CRT Descriptor Writing Agenda, June 20–22, 2006

WKCE-CRT Descriptor Writing Agenda
Reading 3–8, 10
Math 3–8, 10
June 20–22, 2006

Madison Concourse Hotel
1 West Dayton Street, Madison, WI 53702
608-257-6000, *Website:* *http://www.concoursehotel.com/*

Purpose

Description Writing provides plain-language description of the content that students must know at each grade level to be Proficient. This information may be used by teachers and the public to fully understand the performance levels on the WKCE-CRT. Description Writing allows for teacher input regarding performance-level descriptors.

In the Description Writing Workshop, participants will be asked to record the knowledge, skills, and abilities that are required of students in each grade to be Basic, Proficient, and Advanced. To inform their descriptions, participants will review ordered item booklets and item maps and identify the knowledge and skills required to answer each item correctly and why each item is more difficult than the preceding item. Participants will be shown the statistically set cut scores and will write descriptors for each grade/content area.

Tuesday, June 20

7:30—9:00	Registration, Continental Breakfast
9:00—10:30	Large Group Orientation • Introductions • Overview of Descriptor Writing Task • Review of Cut Score Procedures • Descriptor Writing Hints
10:30–10:45	Break, reconvene in content area rooms
10:45–11:30	Test Taking Activity
11:30–12:00	Ordered Item Book and Item Map • Training & modeling by group leader
12:00–12:30	Lunch
12:30—2:30	Review of Ordered Item Book, cont.
2:30–4:00	Descriptor Writing: Reporting Category • Presentation of cut scores • Presentation of 2003 descriptors for grades 4, 8, 10 • Organization of ordered items by content objective

Wednesday, June 21

Table 12-1
WKCE-CRT Descriptor Writing Agenda, June 20–22, 2006

8:00–8:30	Continental Breakfast
8:30–10:30	Descriptor Writing: Reporting Category • Draft descriptors by content objective and performance level
10:30–10:45	Break
10:45–12:00	Descriptor Writing • Completion of descriptors by content objective
12:00–12:30	Lunch
12:30–4:00	Review of Descriptors & Revision • Cross-grade review of descriptors by content objective • Revision of descriptors by content objective

Thursday, June 22

8:30–9:00	Continental Breakfast
9:00–10:30	Synthesis of Descriptors • Synthesis of descriptors by objective into grade-level descriptors by performance category • Cross-grade review to verify articulation
10:30–10:45	Break
10:45–12:00	Review of Synthesized Descriptors • Cross-grade review to verify articulation
12:00–12:30	Lunch
12:30–3:00	Final Polishing and Review of Descriptors • Final revision of grade-level descriptors
Final cross-grade review	

Table 12-2
Cut scores and Associated Impact Data for WKCE-CRT Reading

Score Range					Impact Data				
Grade	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
3	270-393	394-429	430-465	466-640	4.54%	13.57%	36.59%	45.30%	81.89%
4	280-395	396-439	440-488	489-650	4.05%	12.88%	40.71%	42.36%	83.07%
5	290-400	401-443	444-496	497-690	4.59%	10.45%	42.22%	42.74%	84.96%
6	300-417	418-456	457-513	514-730	4.79%	9.28%	41.58%	44.35%	85.93%
7	310-433	434-466	467-522	523-780	5.17%	9.48%	40.45%	44.90%	85.35%
8	330-444	445-479	480-538	539-790	5.36%	9.45%	42.07%	43.12%	85.19%
10	350-455	456-502	503-554	555-820	9.54%	14.15%	33.15%	43.16%	76.31%

Table 12-3
Cut scores and Associated Impact Data for WKCE-CRT Mathematics

Score Range					Impact Data				
Grade	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
3	220-391	392-406	407-451	452-630	16.57%	9.26%	38.74%	35.43%	74.17%
4	240-420	421-437	438-483	484-650	13.12%	9.19%	42.49%	35.20%	77.69%
5	270-444	445-462	463-504	505-680	13.93%	10.90%	38.55%	36.63%	75.18%
6	310-463	464-484	485-531	532-700	11.96%	11.41%	43.69%	32.95%	76.64%
7	330-479	480-503	504-554	555-710	9.39%	11.01%	46.89%	32.66%	79.55%
8	350-482	483-512	513-572	573-730	10.06%	14.39%	48.70%	26.85%	75.55%
10	410-515	516-540	541-594	595-750	14.77%	13.56%	45.74%	25.93%	71.67%

Table 12-4
Cut scores and Associated Impact Data for WKCE-CRT Language Arts

Score Range					Impact Data				
Grade	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
4	140-251	252-276	277-307	308-420	4.63%	17.28%	45.46%	32.64%	78.10%
8	250-357	358-384	385-417	418-520	13.28%	23.94%	38.87%	23.91%	62.78%
10	290-392	393-427	428-483	484-630	7.86%	18.99%	55.84%	17.31%	73.15%

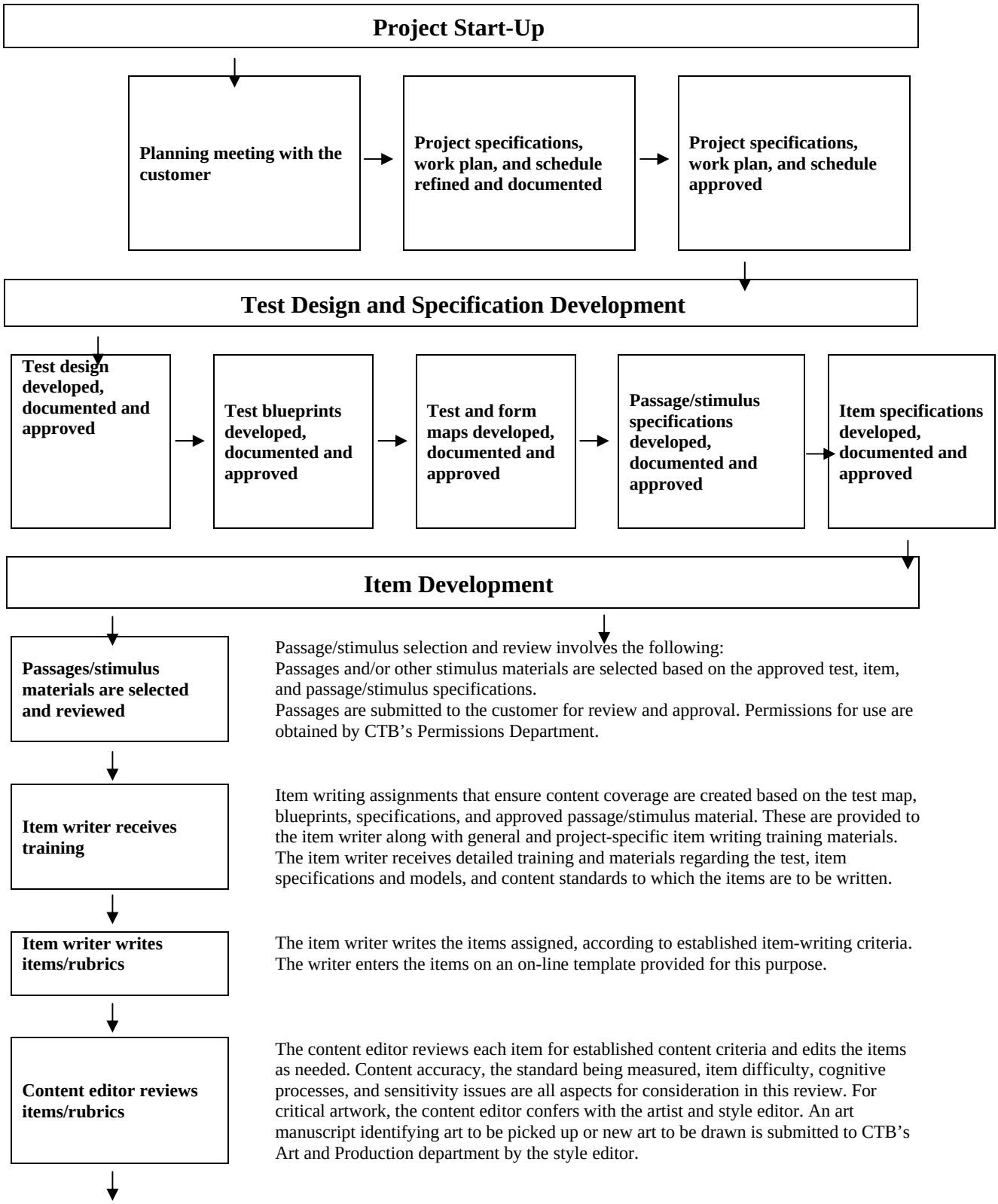
Table 12-5
Cut scores and Associated Impact Data for WKCE-CRT Social Studies

Score Range					Impact Data				
Grade	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
4	170-241	242-262	263-287	288-400	1.55%	5.34%	25.59%	67.53%	93.12%
8	230-333	334-363	364-402	403-530	3.49%	12.26%	40.30%	43.96%	84.26%
10	240-407	408-419	420-454	455-620	14.95%	6.71%	31.73%	46.61%	78.34%

Table 12-6
Cut scores and Associated Impact Data for WKCE-CRT Science

Score Range					Impact Data				
Grade	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
4	170-248	249-278	279-319	320-440	4.86%	16.46%	56.22%	22.46%	78.68%
8	230-348	349-374	375-418	419-560	8.97%	15.17%	47.02%	28.85%	75.87%
10	240-410	411-428	429-465	466-610	14.50%	11.79%	35.84%	37.86%	73.70%

Figure 4-1
CTB's Item Development Process



Style editor reviews items/rubrics	The style editor reviews and edits the items for grammar, punctuation, and adherence to the customer-approved style sheet for the HSGT. Items are checked to ensure that language, punctuation, and formatting are clear and consistent within and across items; that words are hyphenated correctly; and that techniques used to emphasize words (such as underlining) are consistently applied. Style edits and queries are presented to the content editor for review and resolution.
↓	
Supervisor/manager review	The development supervisor or manager reviews each item to verify that established criteria have been met. Edits or queries are presented to the content editor for review and resolution.
↓	
Interdepartmental reviews and local pilot	The items are provided to the Research and Hand-Scoring departments for review. These departments verify that items are acceptable from a research or hand-scoring point of view. Edits or queries are presented to the content editor for review and resolution.
↓	
Accuracy check	All updated items are checked for accuracy by the style editor or proofreader before presentation to the CUSTOMER and its item review committees. Critical art is developed, edited, corrected, and placed on the template forms prior to this review.
↓	
Agency reviews	The items are provided to the customer for review with art in place. Items will be reviewed for both content and bias by the customer's content specialists. The customer's edits or queries are marked on the hard copy as needed. CTB will then conduct the review meetings. Pending customer approval of the committees' suggested edits, the content editor updates item templates with any revisions.
↓	
Final reviews	Following the field test, CTB reviews items and their field-test statistics. Based on the item statistics, items may be deleted from the item pool or have minor revisions made and be re-field tested.

Figure 4-2
CTB's Test Material Development Process

Selecting test items	All final edits are made to the items. The content editor selects test items for each test form, according to established item selection procedures, and orders the items as they are to appear in the tests.
Formatting booklets	The electronic item templates and a hard copy of the selected items and art showing in what sequence they are to appear in the test(s) are submitted to CTB's Art and Production Staff, who will assemble the items in proper order and electronically format the test books.
First-page reviews	The "first pages" provided by the Art and Production Department are proofread against the material originally submitted. The style editor and content editor, review the first pages for established criteria. Criteria for this review include clueing, juxtapositions, workability, and navigability. There should be no individual item edits at this point.
Production of second pages	The marked-up first pages are returned to Art and Production for revisions. The revisions are called "second pages."
Second page reviews	The style editor reviews the second pages to verify that revisions are accurate. The content editor checks the manuscript and provides it to the customer for review. The customer reviews the manuscript, noting all requested corrections in writing.
Production of camera copy	The content editor reviews the customer's requested edits. These are submitted to Art and Production for production of camera-ready copy.
Camera copy sign-offs	Once the camera-ready copy is produced and approved by the style editor and CTB's Quality Assurance Department, mock-ups are made of the booklets. These are submitted for approval by the content editor, the development supervisor/manager, CTB's Technology Department, and the customer. Following sign-offs, the materials are released for printing.

Figure 8-1
TCC Curve for Reading Grades 3-8, 10

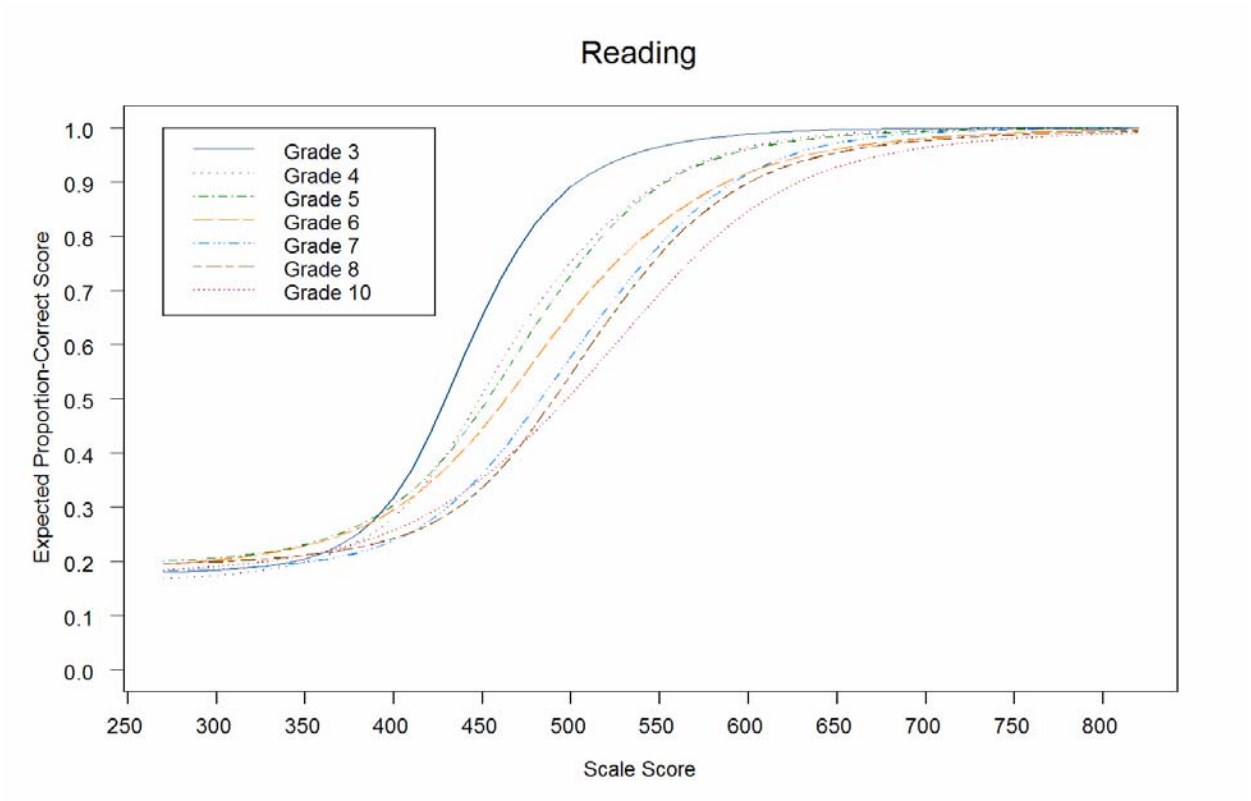


Figure 8-2
TCC Curve for Mathematics Grades 3-8, 10

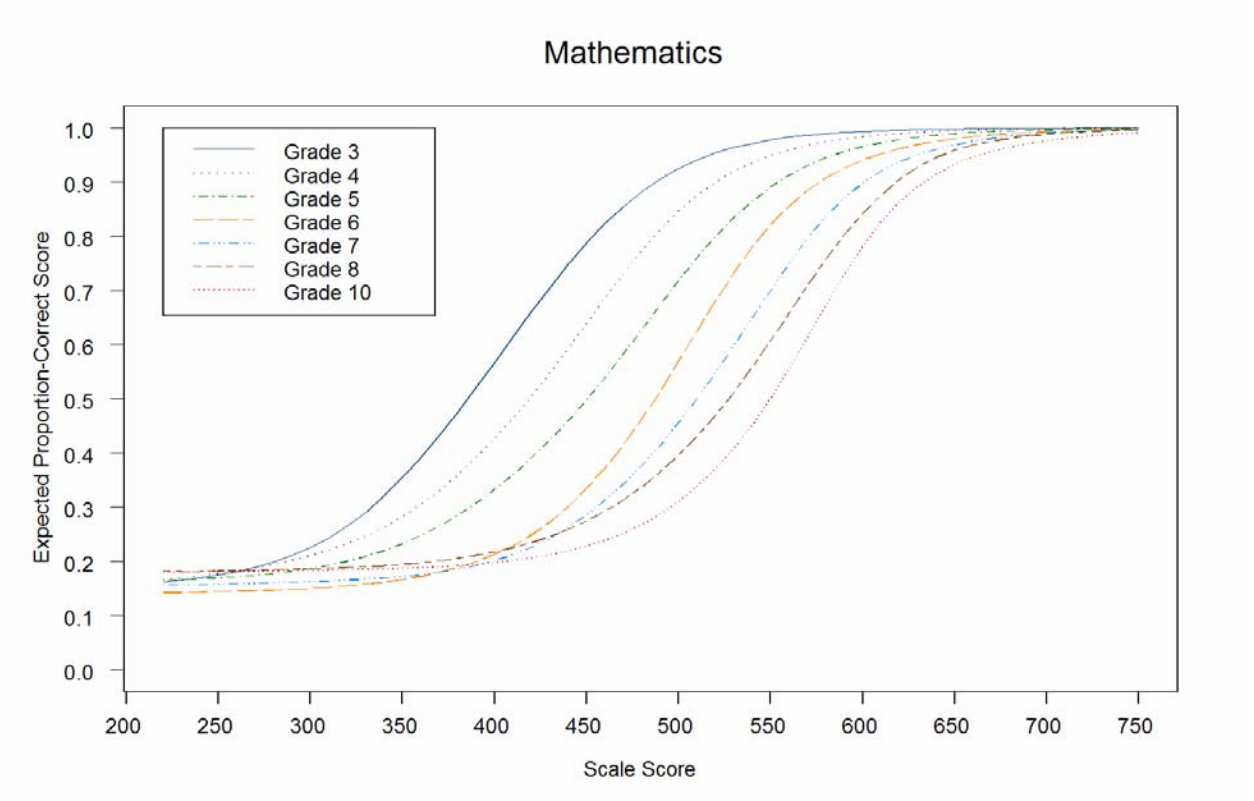


Figure 8-3
TCC Curve for Language Arts Grades 4, 8, 10

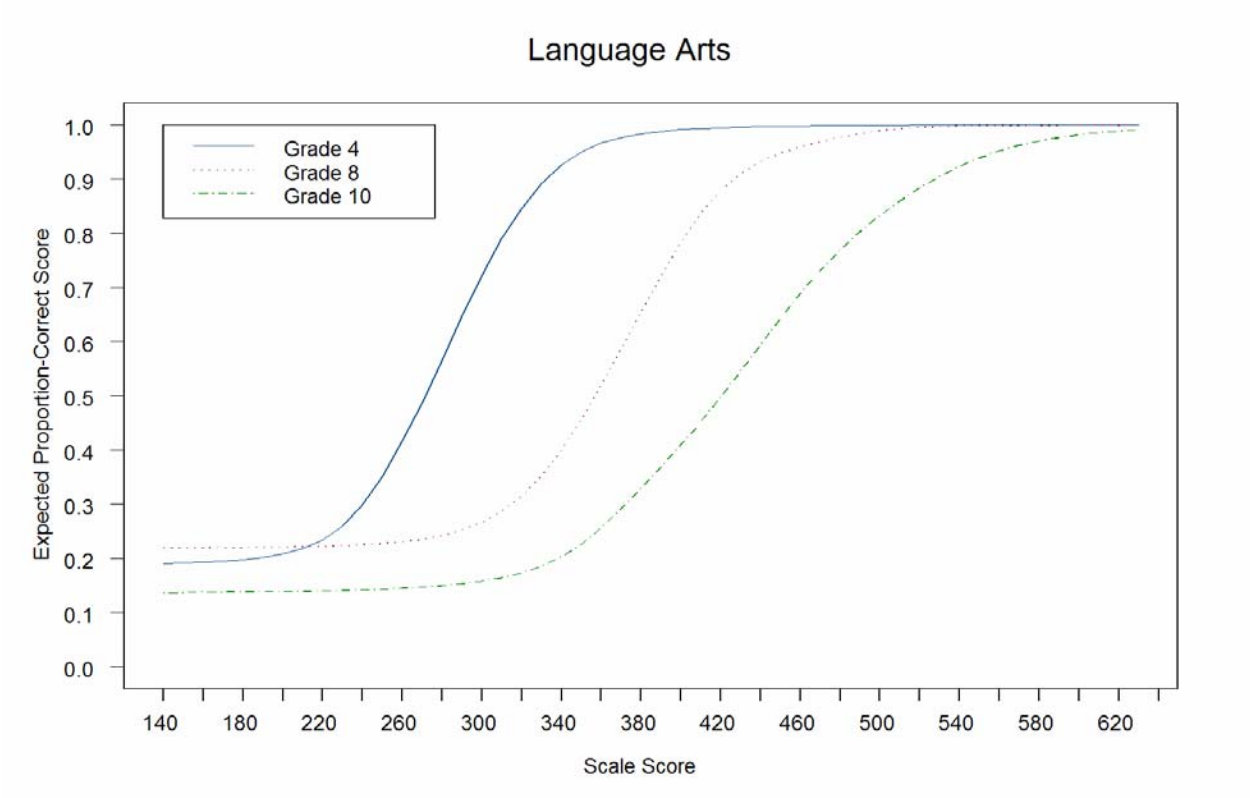


Figure 8-4
TCC Curve for Social Studies Grades 4, 8, 10

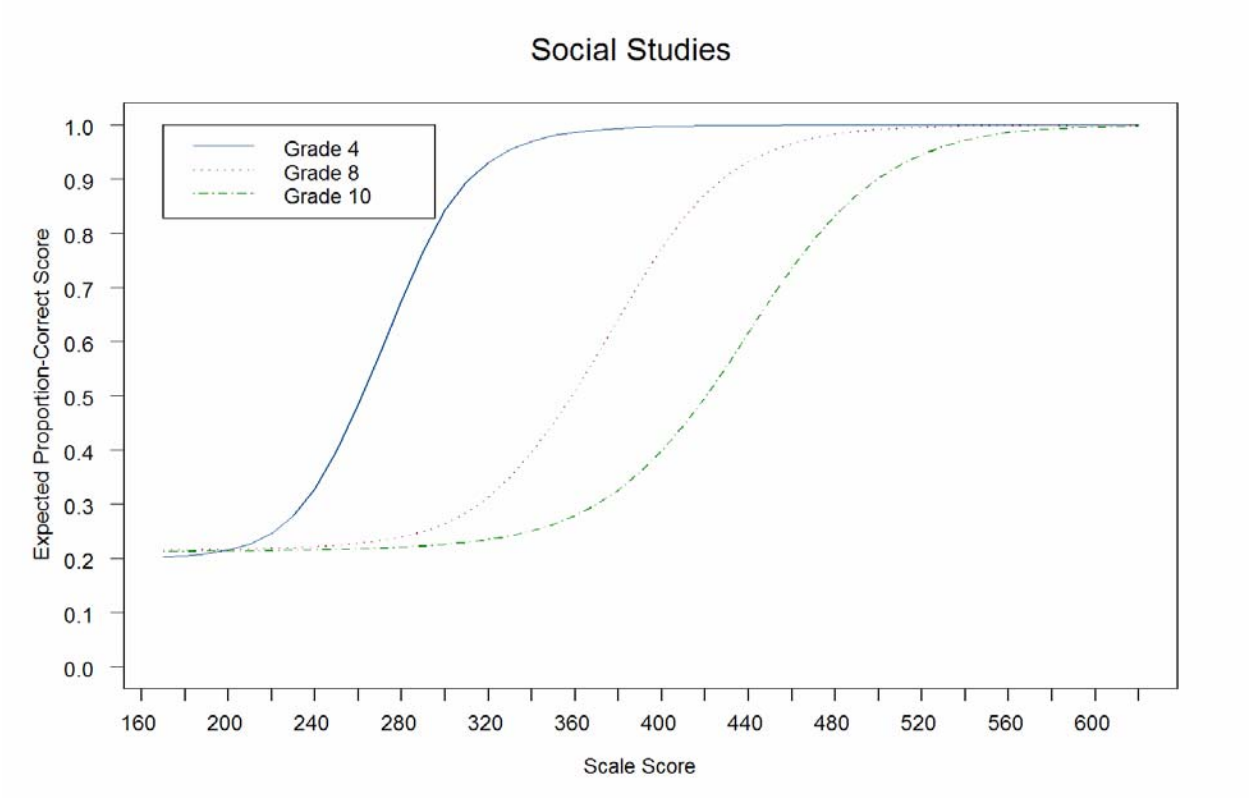


Figure 8-5
TCC Curve for Science Grades 4, 8, 10

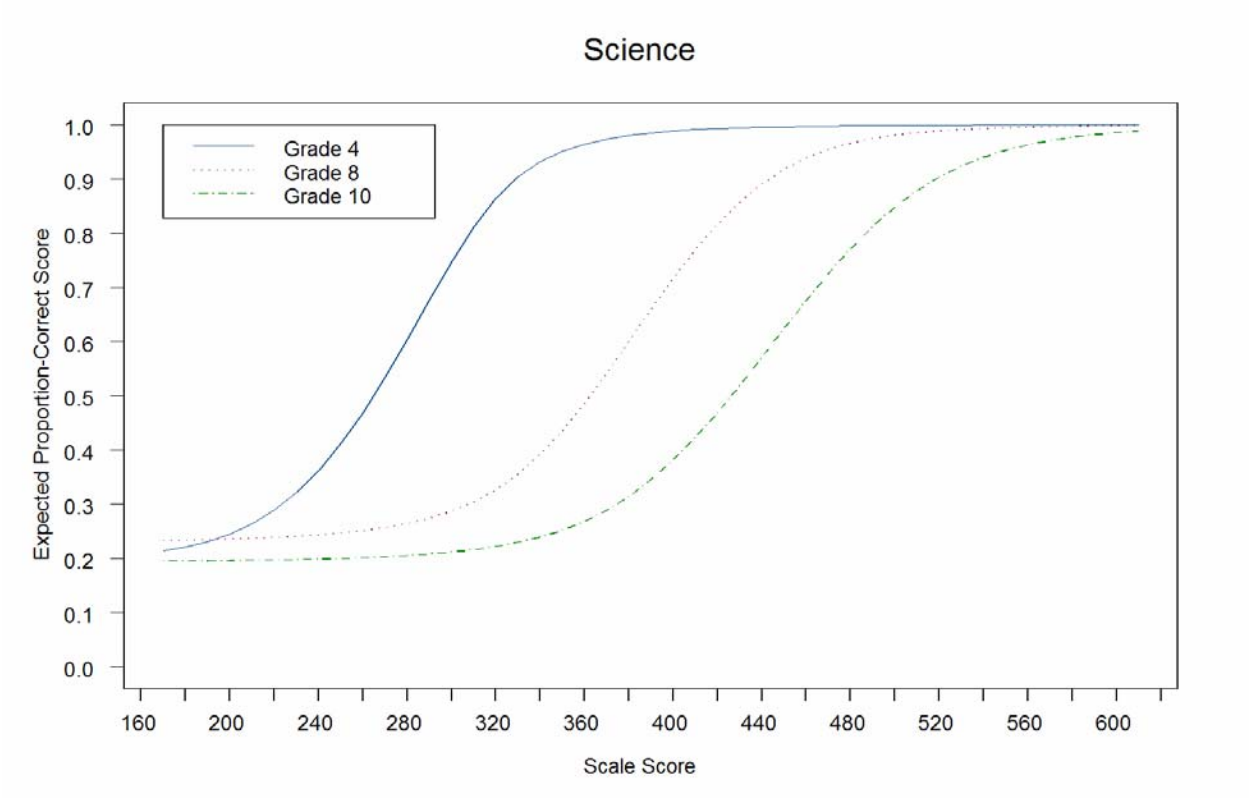


Figure 8-6
SEM Curves, Reading Grades 3-6

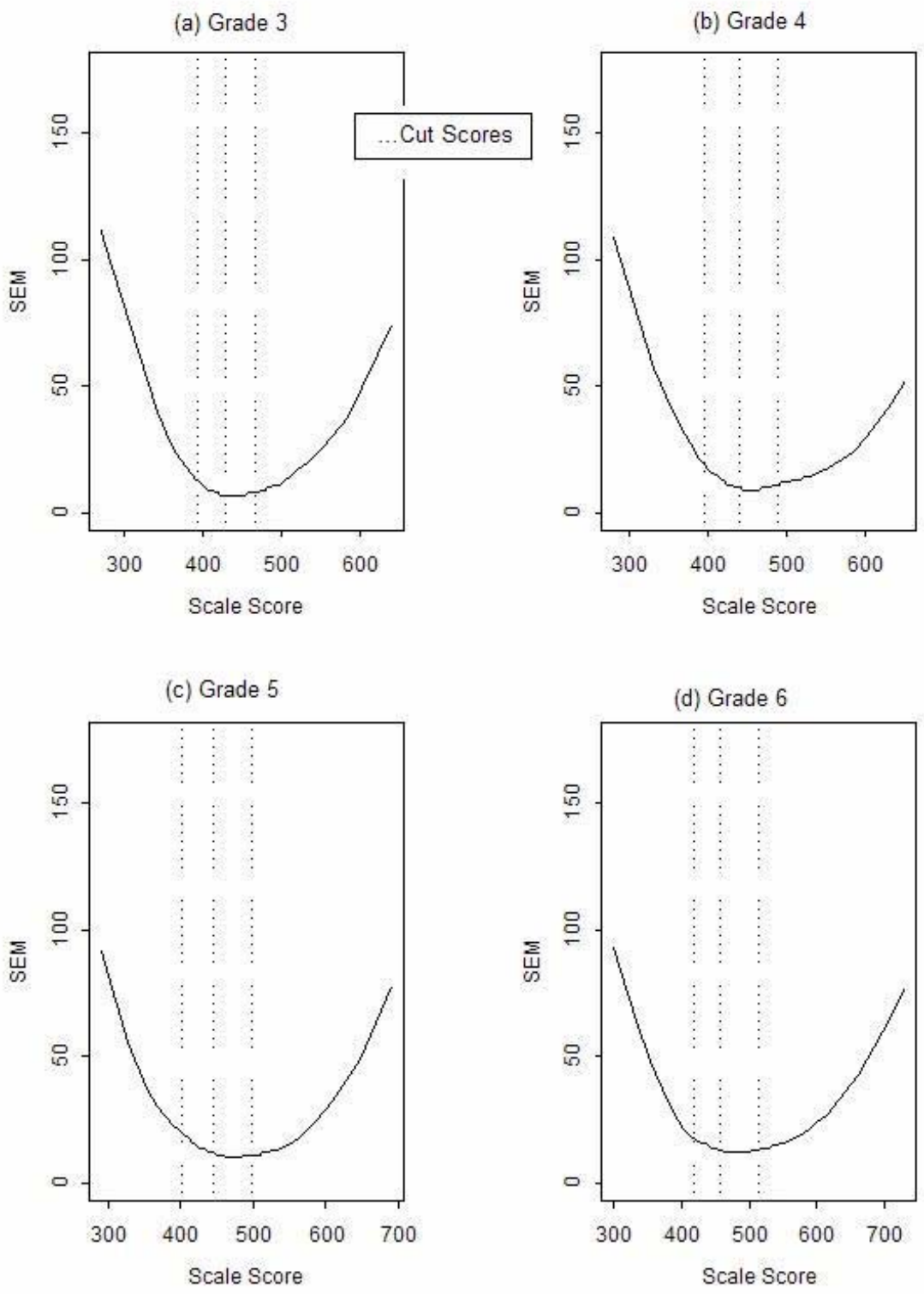
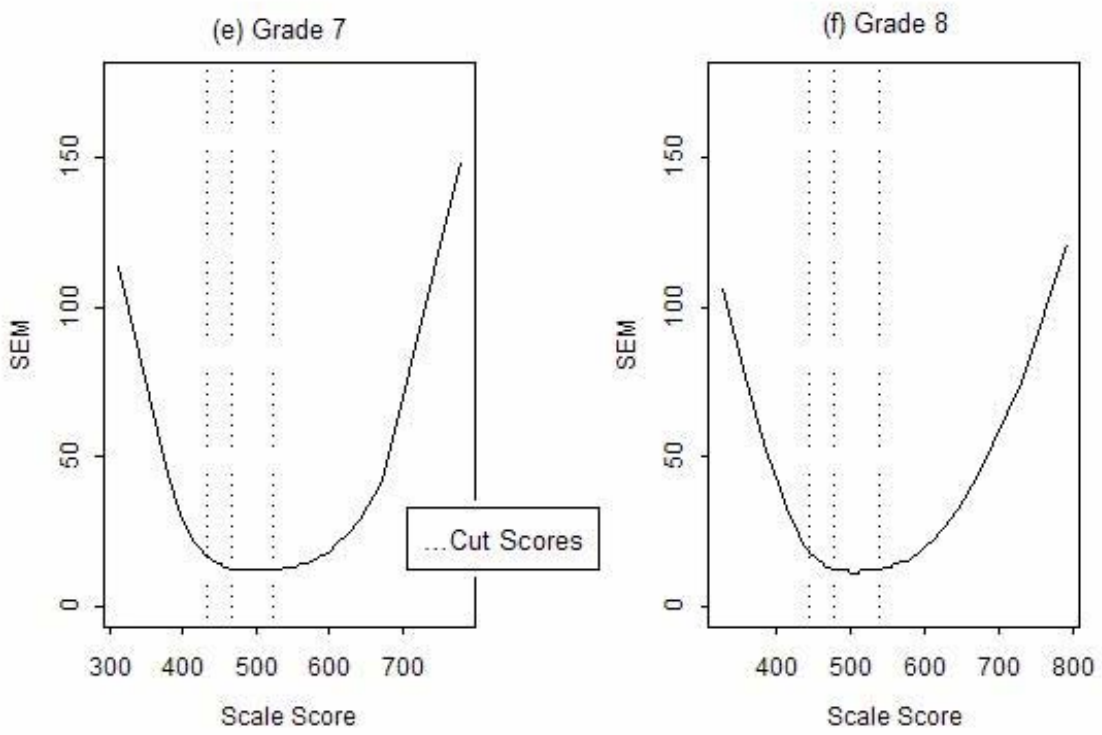


Figure 8-6 Cont'd
SEM Curves, Reading Grades 7, 8, 10



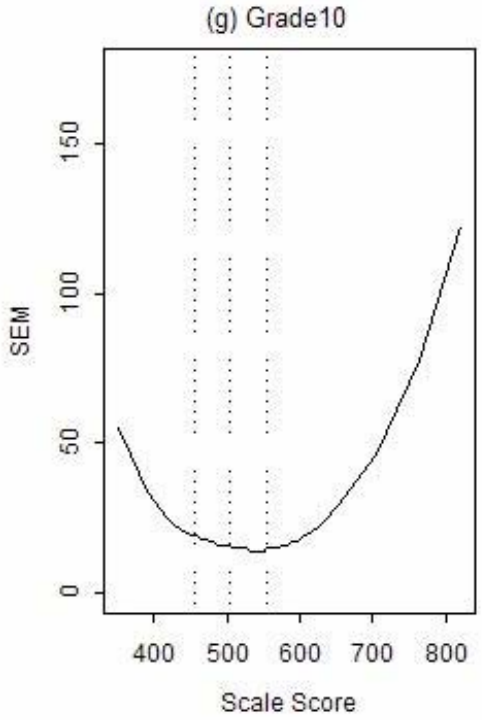


Figure 8-7
SEM Curves, Mathematics Grades 3-6

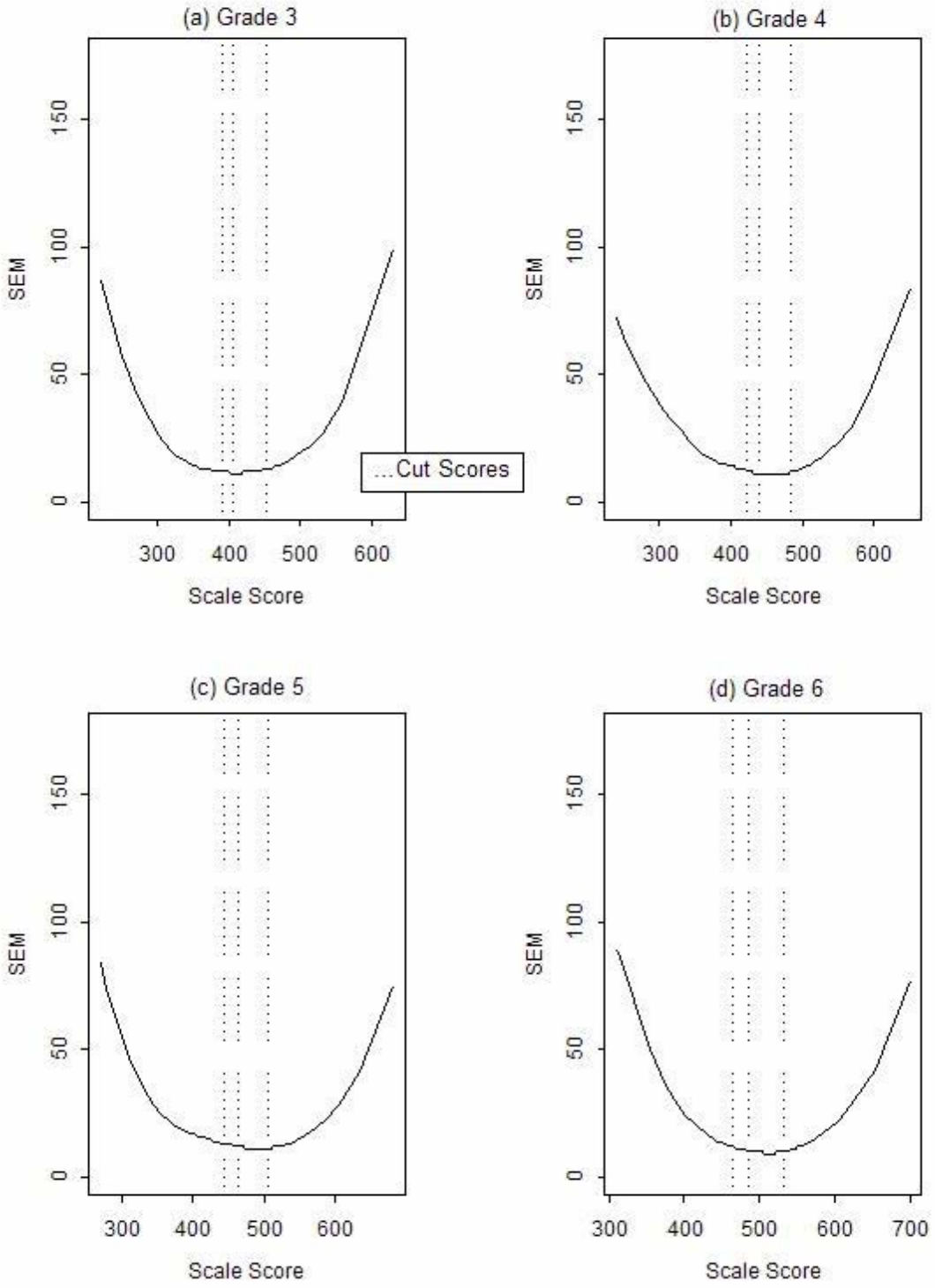


Figure 8-7 Cont'd
SEM Curves, Mathematics Grades 7, 8, 10

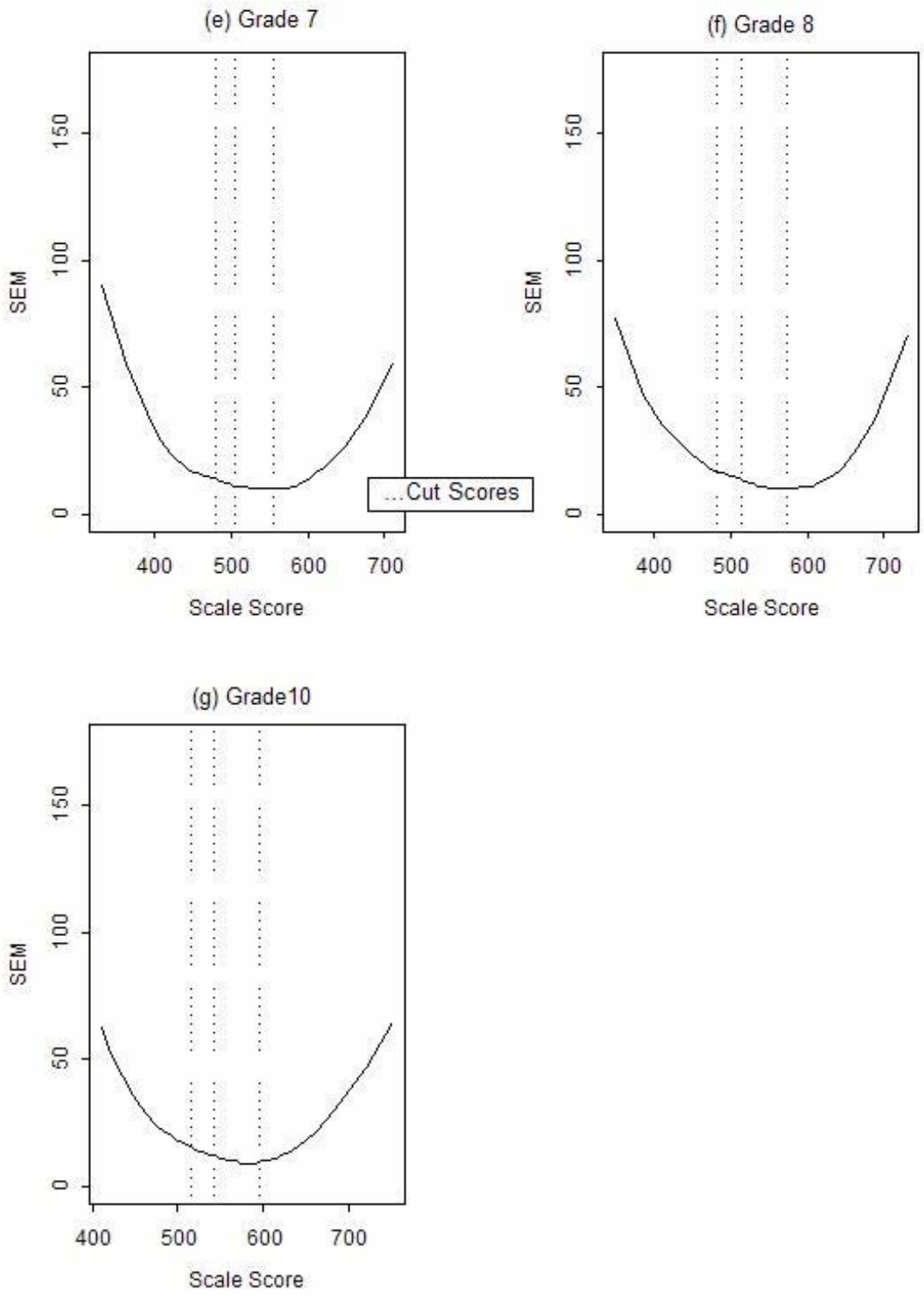


Figure 8-8
SEM Curves, Language Arts Grades 4, 8, 10

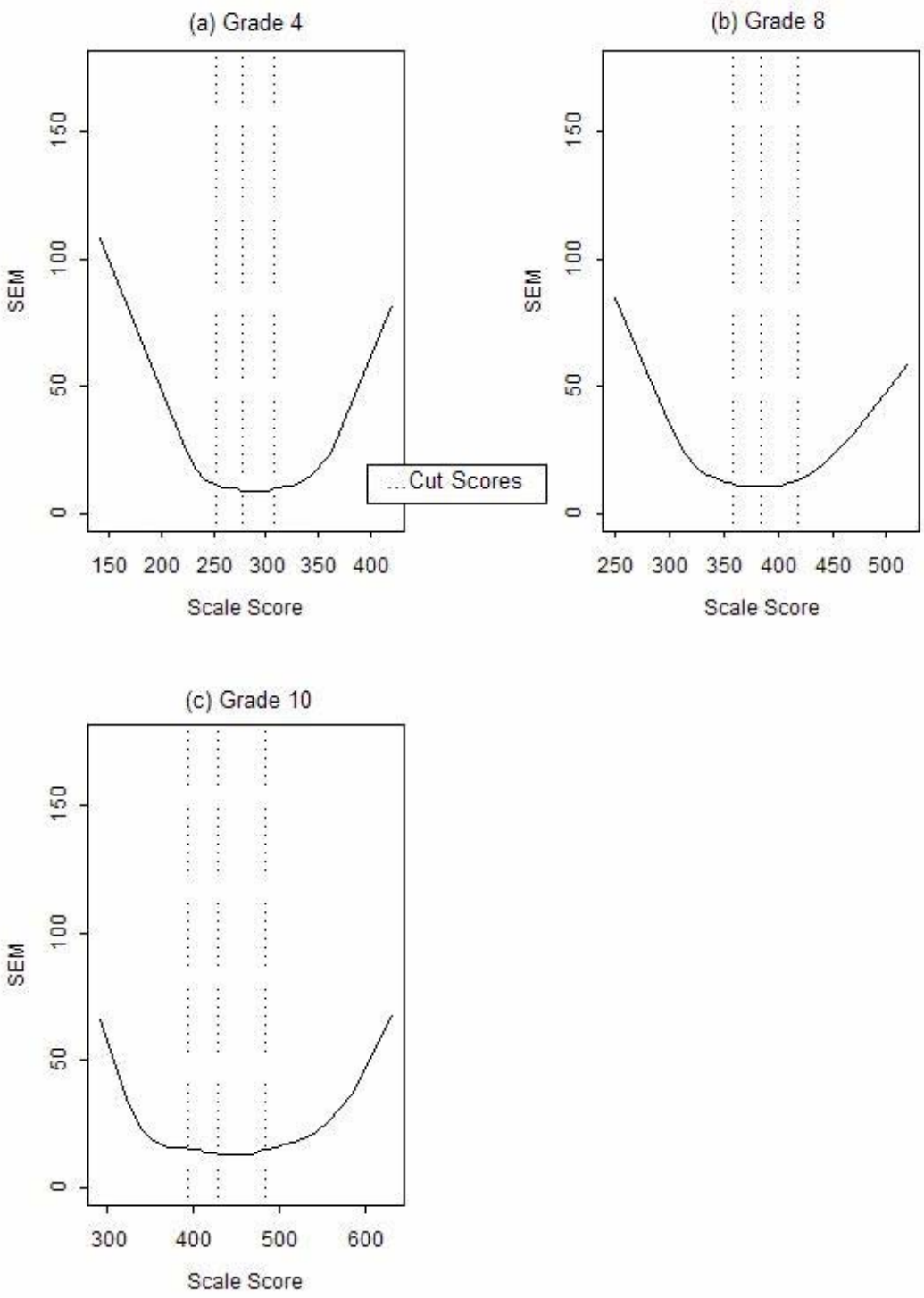


Figure 8-9
SEM Curves, Social Studies Grades 4, 8, 10

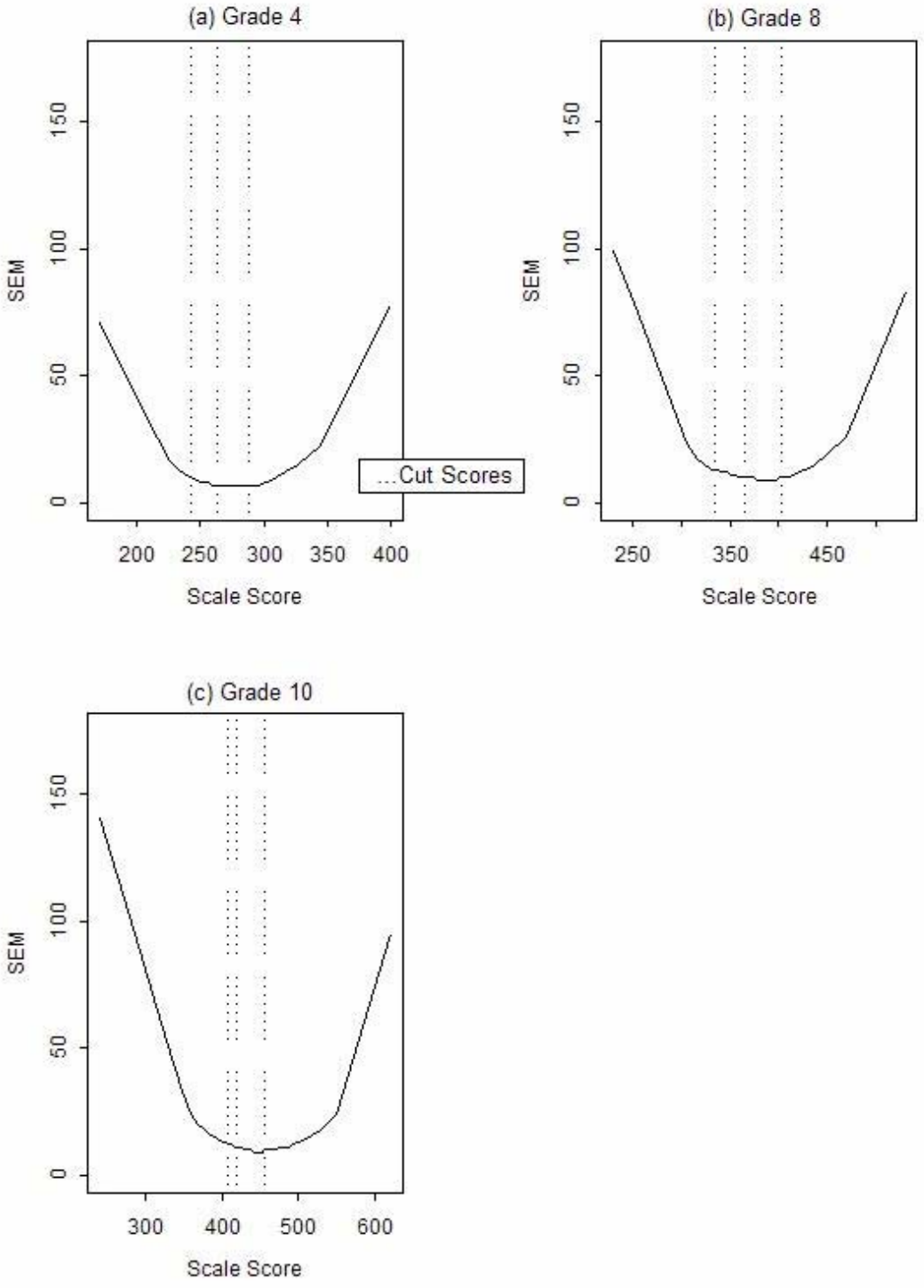


Figure 8-10
SEM Curves, Science Grades 4, 8, 10

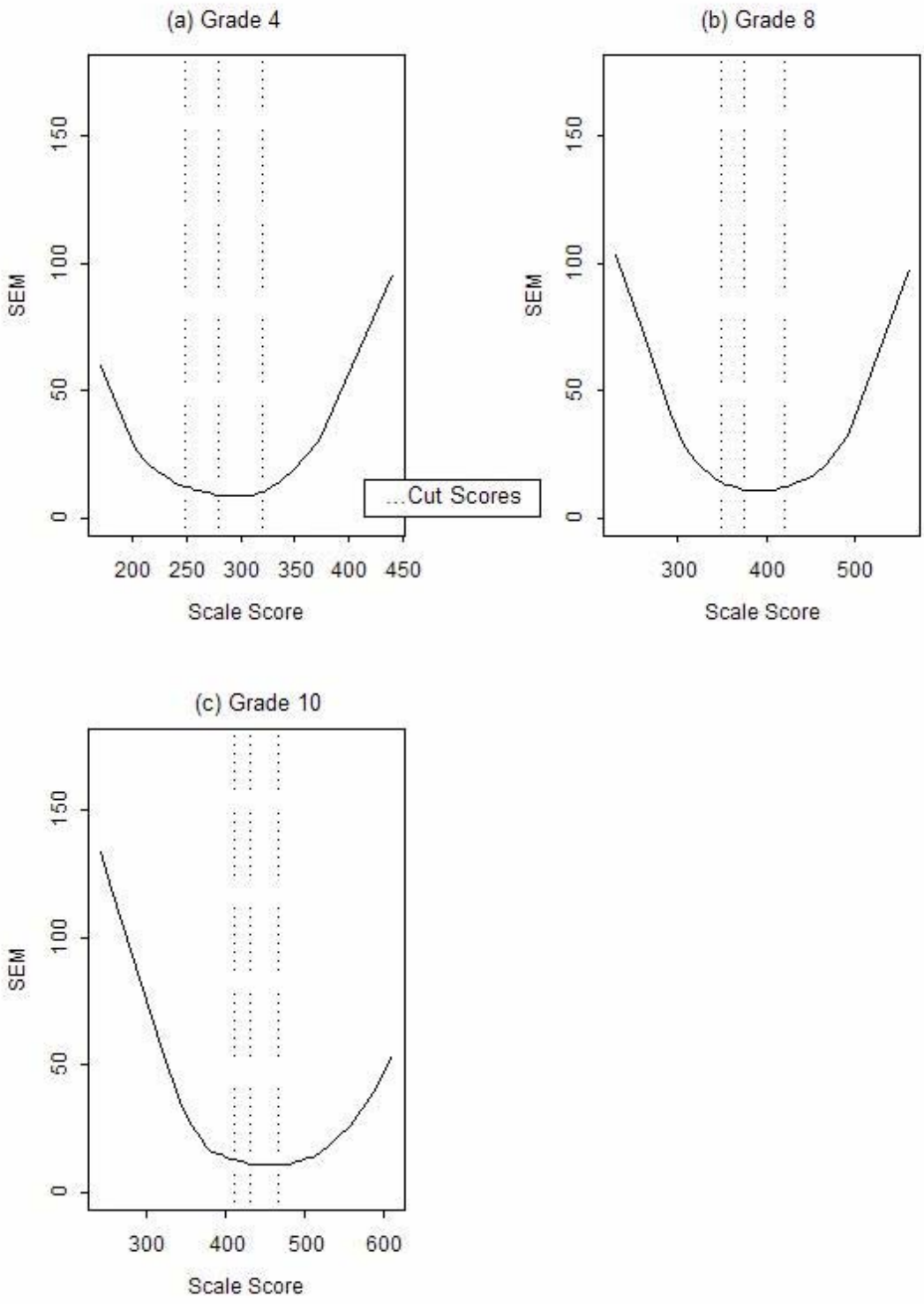


Figure 12-1
Cut Scores for Reading

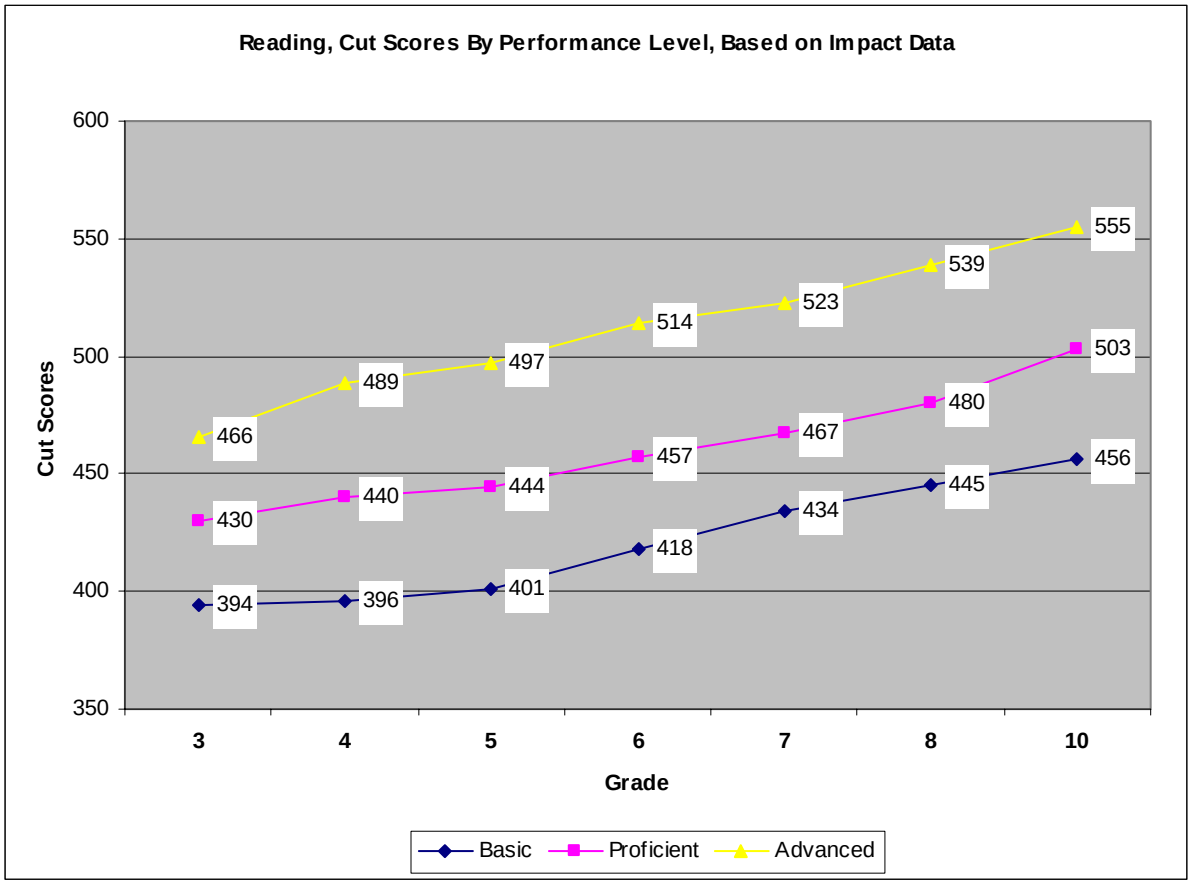


Figure 12-2
Cut Scores for Mathematics

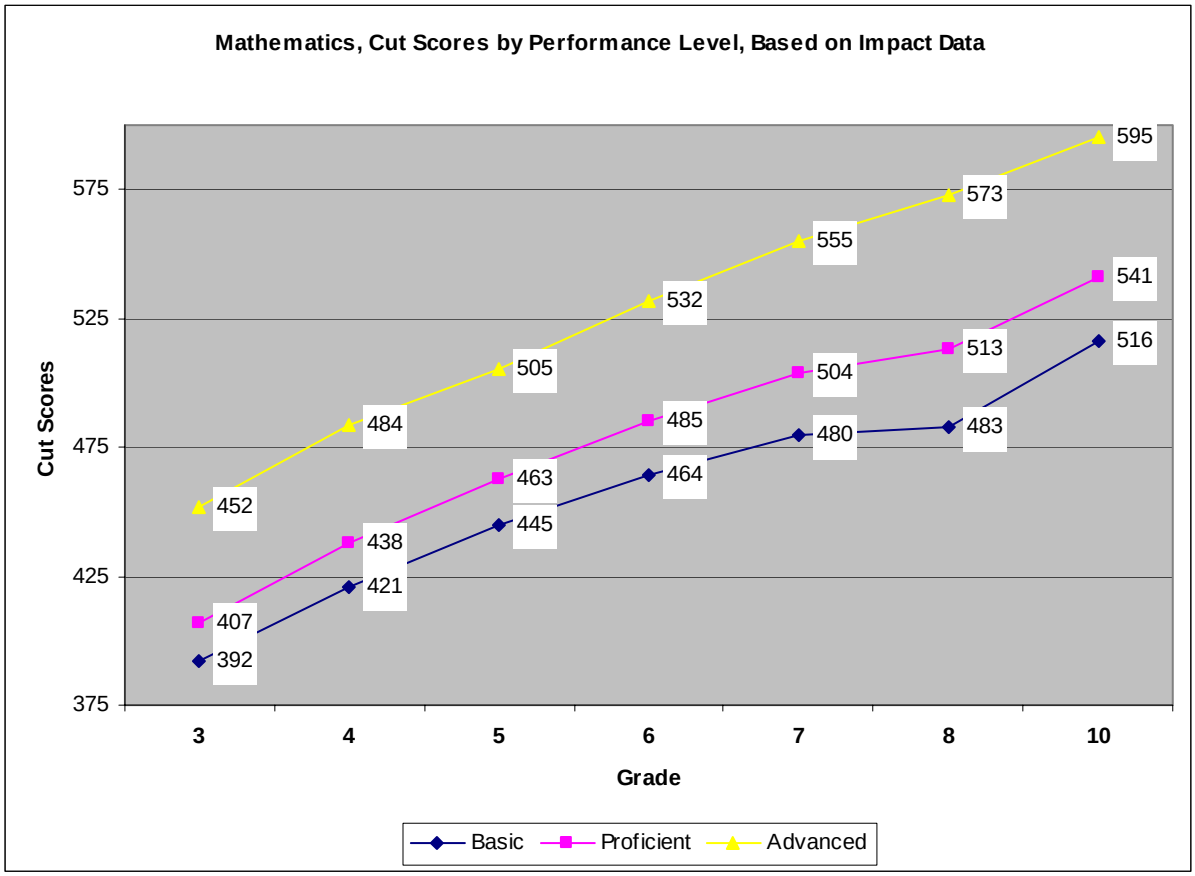


Figure 12-3
Cut Scores for Language Arts

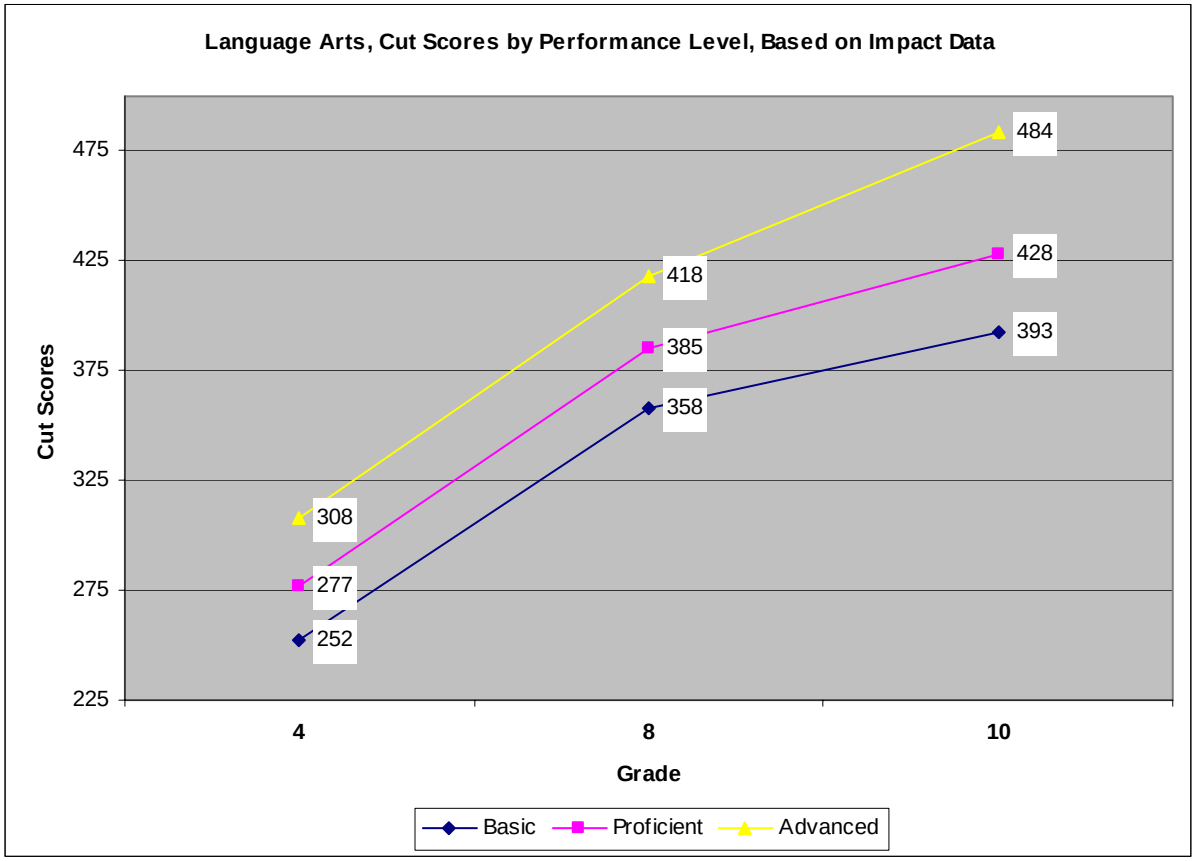


Figure 12-4
Cut Scores for Social Studies

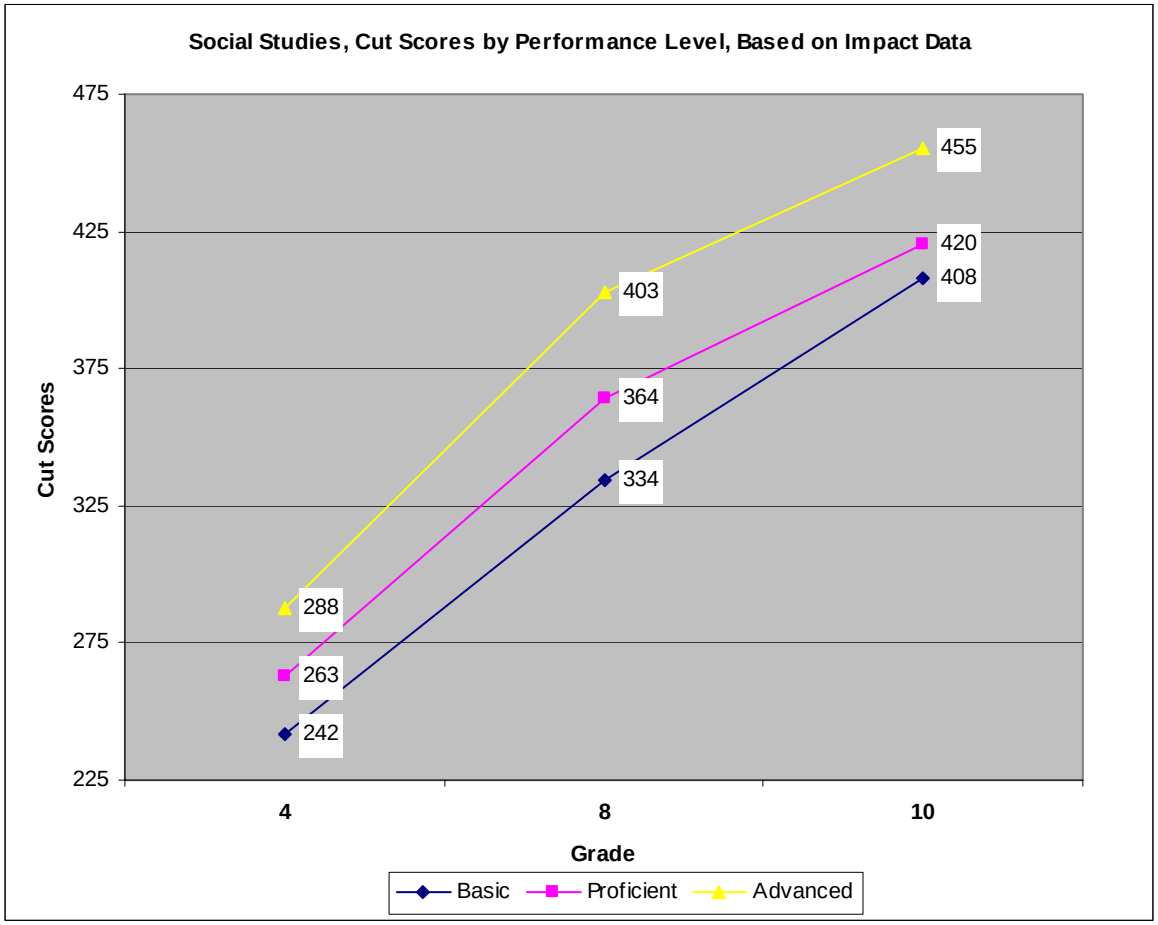


Figure 12-5
Cut Scores for Science

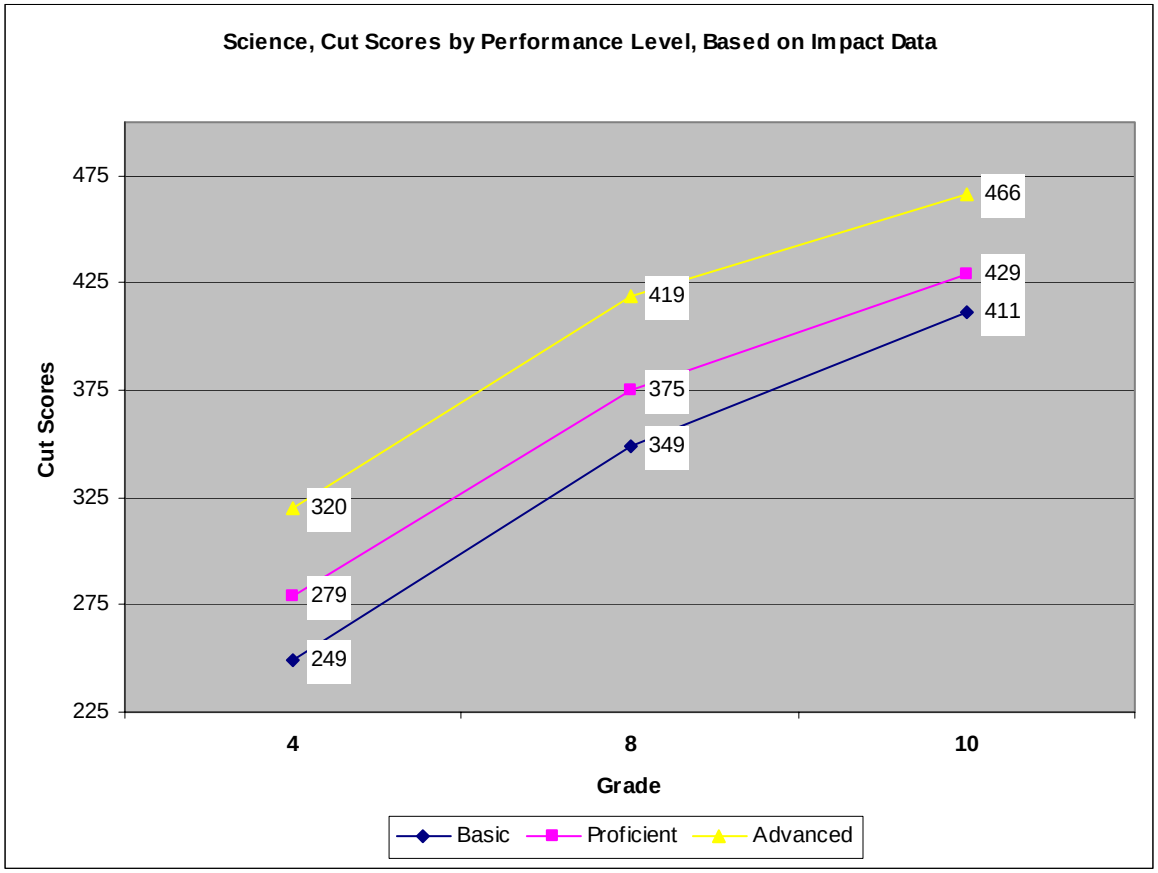


Figure 12-6
Percent of Students for Reading

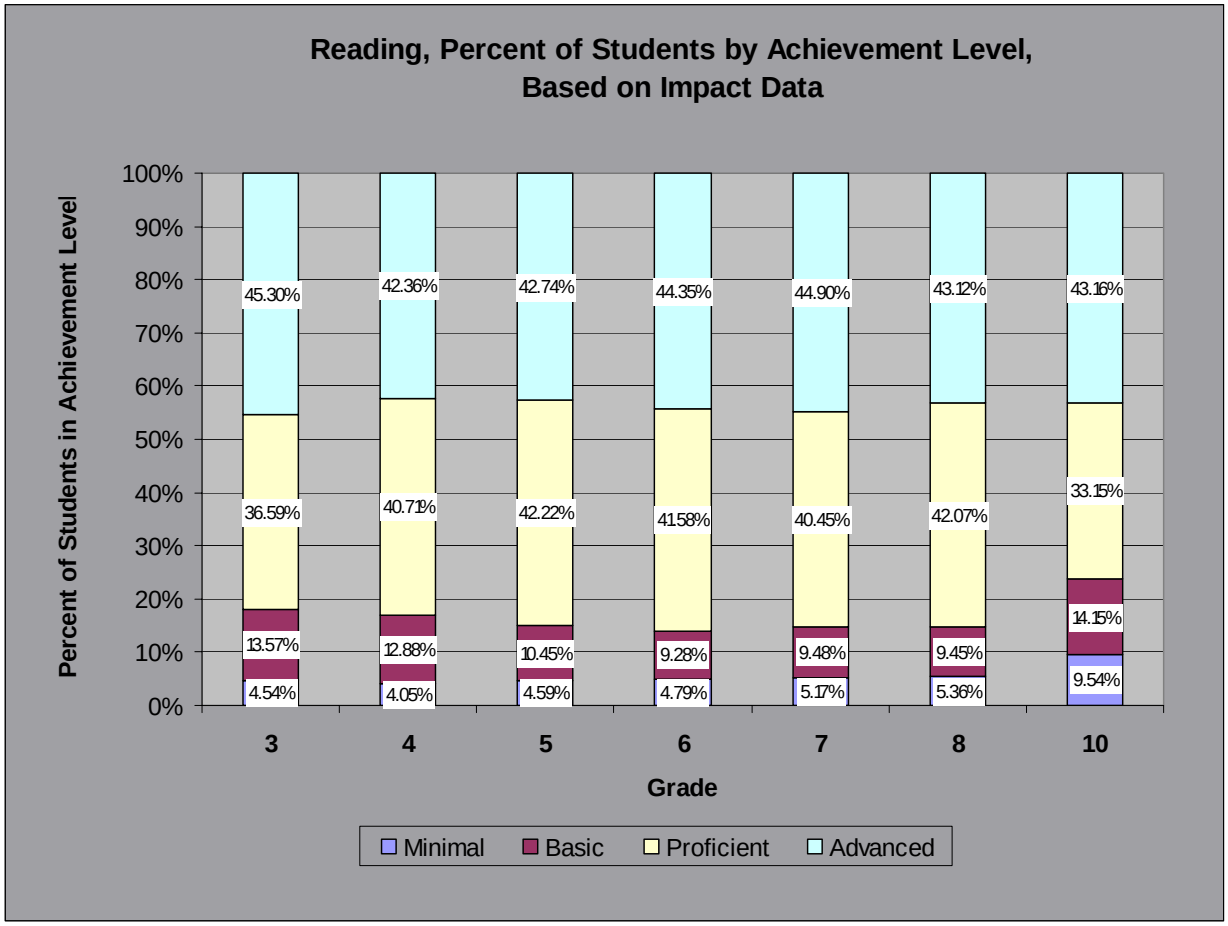


Figure 12-7
Percent of Students for Mathematics

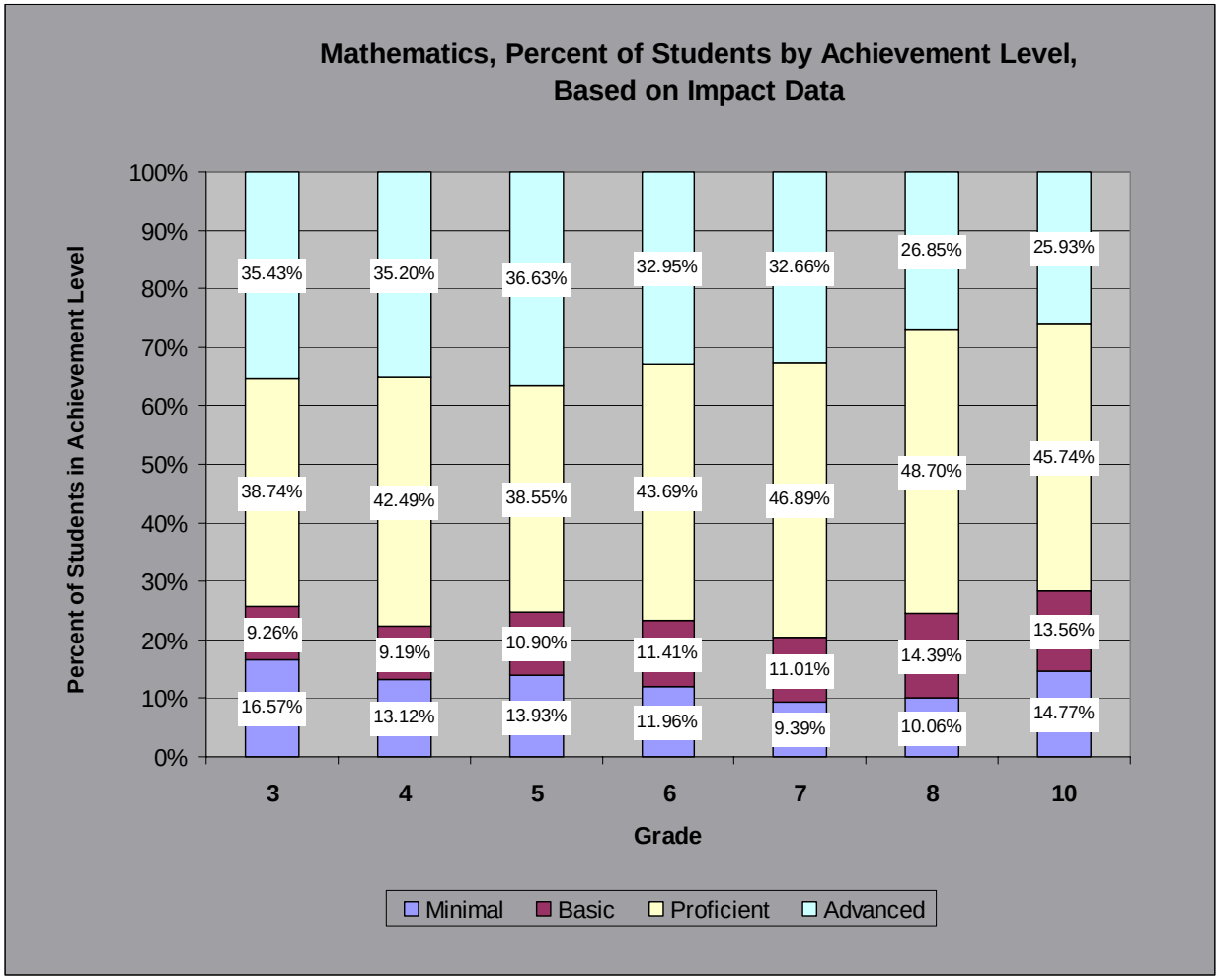


Figure 12-8
Percent of Students for Language Arts

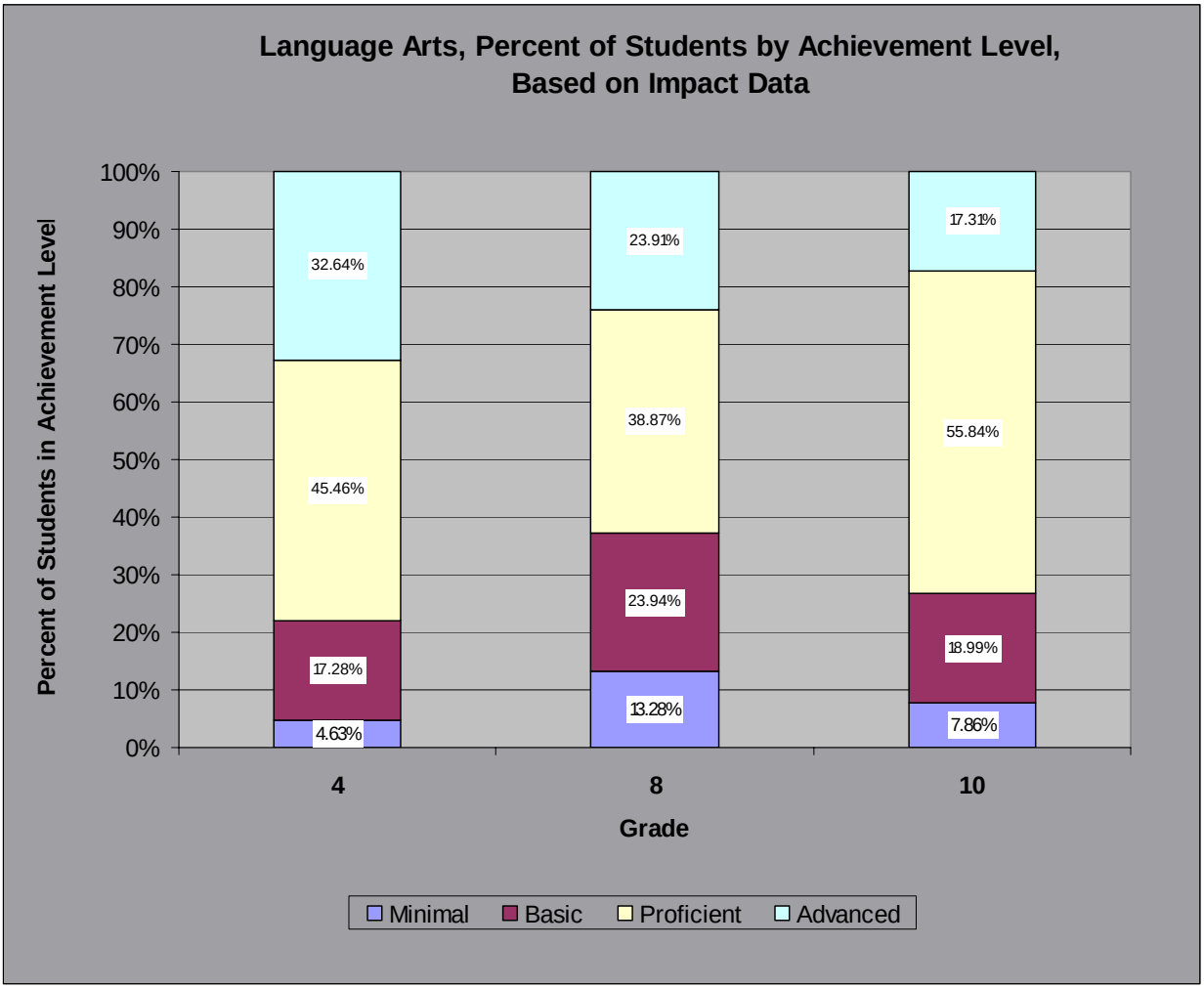


Figure 12-9
Percent of Students for Social Studies

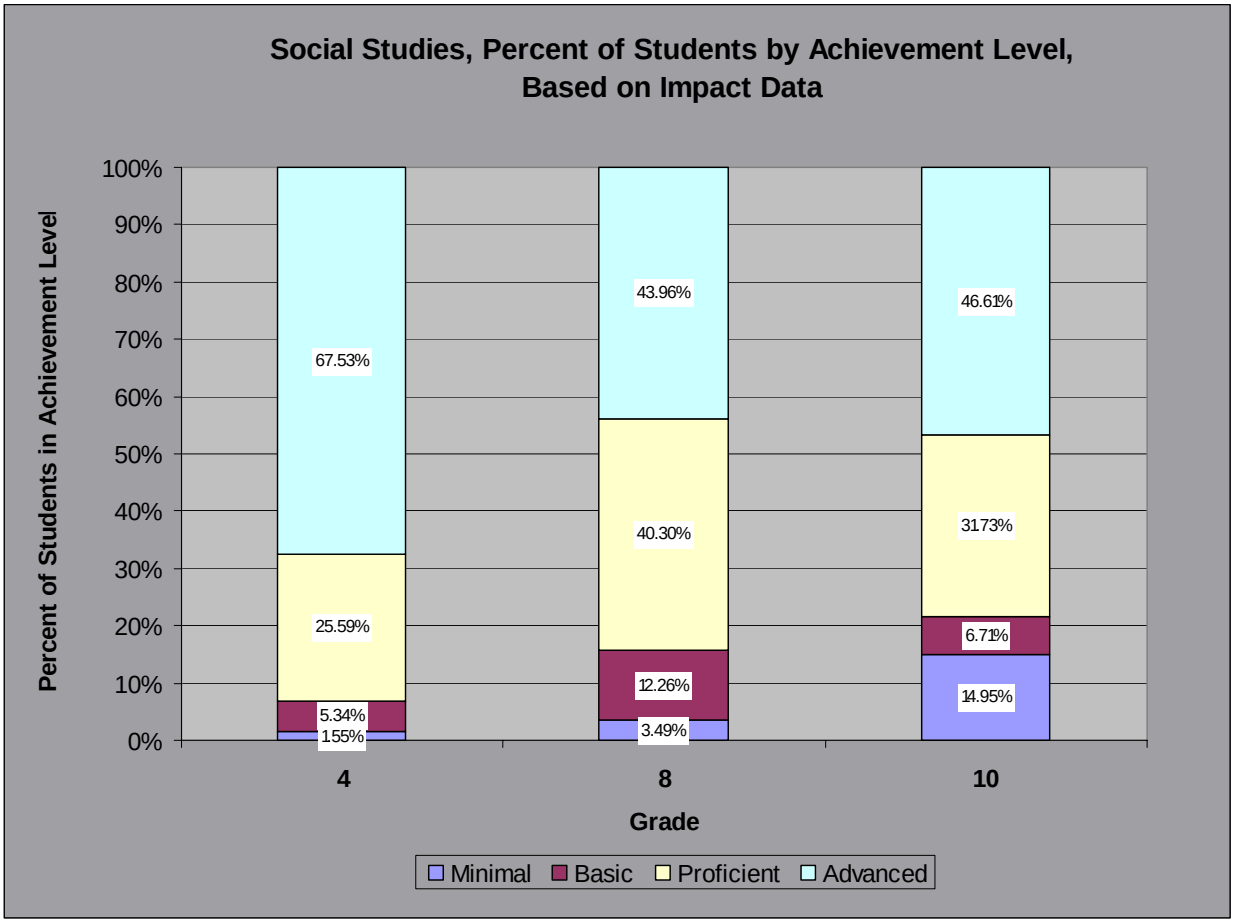
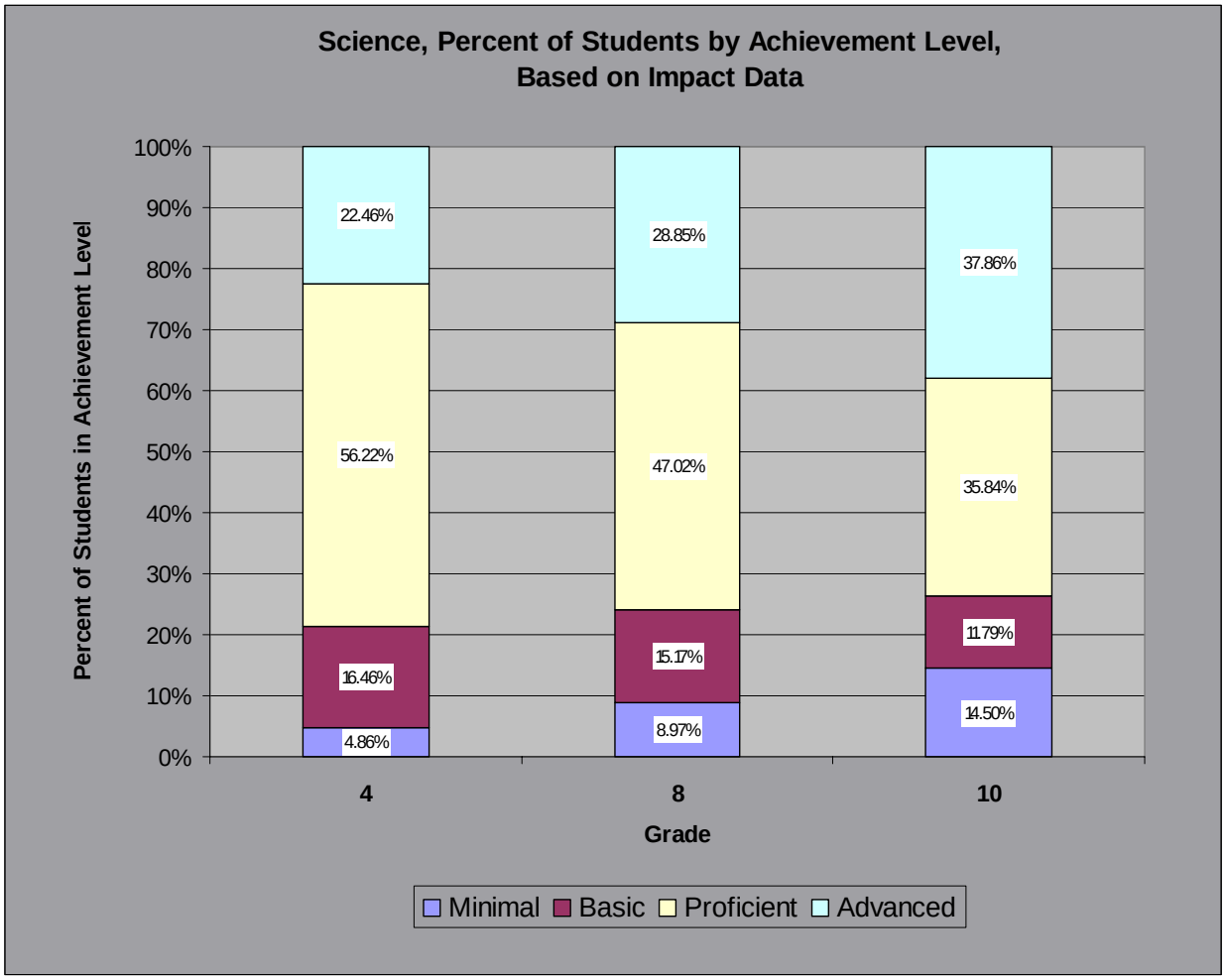


Figure 12-10
Percent of Students for Science



Appendix 3:
Fall 2005 WKCE Technical Report, Parts 8 and 11

Copyright © 2009 by the Wisconsin Department of Public Instruction

Part 8: Calibration and Scaling

Part 8 describes the calibration and scaling procedures applied to the 2005 WKCE-CRT. There were two main differences between the 2004 WKCE and the 2005 WKCE-CRT. First, most of the Fall 2004 WKCE items were directly from *TerraNova*, while for the Fall 2005 WKCE-CRT, all Reading and Mathematics items were customized to Wisconsin standards. Second, the *TerraNova* item parameters estimated in 2002 *TerraNova* standardization, using a national sample, were applied to score the 2004 WKCE, whereas the 2005 WKCE-CRT item parameters were estimated using 14 calibration districts (CD) from Wisconsin. Note that the scale of the 2005 WKCE-CRT is the first operational scale. That is, there is no scale connection between the 2004 WKCE and 2005 WKCE-CRT. The scale scores of the 2004 WKCE and WKCE-CRT can not be directly compared. The score conversion tables should be used for this purpose. The relationship between the 2004 WKCE and the 2005 WKCE-CRT is described in detail in Part 11.

8.1 Calibration Methods

The 2005 Fall WKCE-CRT was calibrated and scaled using item response theory (IRT) procedures similar to those followed in the development of the *TerraNova* test (CTB/McGraw-Hill, 1997), *TerraNova* 2nd Edition (CTB/McGraw-Hill, 2000), and the Wisconsin Knowledge and Concept Exam (WKCE) (CTB/McGraw-Hill, 1997-2004).

Because the characteristics of MC and CR items are different, two different item response theory models were used in the analysis of the data. The three-parameter logistic model (Lord & Novick, 1968; Lord, 1980) was used to scale the MC items and the two-parameter partial credit model (Muraki, 1992; Yen, 1993) was used to scale the CR items. The three-parameter logistic model (3PL) defines a MC item in terms of three item parameters: the item difficulty (or its location on a scale of difficulty/ability), the item discrimination (or item differences on discrimination), and the level of guessing. The two-parameter partial credit model (2PPC) defines a CR item in terms of an item discrimination parameter and a location parameter for each score point. 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).

8.1.1 Calibration Models

The 3PL model (Lord & Novick, 1968; Lord, 1980) was used in the analysis of MC items. In this model, the probability that a student with scale score θ responds correctly to item *i* is:

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]}$$

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.

For analysis of the CR items in 2005 WKCE-CRT, the 2PPC model (Muraki, 1992; Yen, 1993) was used. 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

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

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 that 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.

8.1.2 Calibration Software

The IRT models were implemented using CTB's PARDUX software (Burket, 1991). PARDUX estimates parameters simultaneously for MC and CR items using marginal maximum likelihood procedures implemented with the expected maximum (EM) algorithm (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 also shown that the IRT procedures used here produce accurate vertical scaling (Yen & Burket, 1997). The Stocking and Lord (1983) procedure was used to place the estimated parameters on the scale from which the anchor items (i.e., *TerraNova*) were drawn.

8.2 Scaling Procedures

The scaling procedure for Reading and Mathematics was different from that for the remaining contents because the vertical scale was applied to Reading and Mathematics only.

8.2.1 Reading and Mathematics

Scales for Reading and Mathematics were based on the scale established during the December 2004 Form Standardization. For the December 2004 administration, three forms, D, E, and F, were constructed and administered. Using Form D, the vertical relationship for Reading and Mathematics grades 3 through 10, except for Grade 9, were constructed. In the 2005 Fall WKCE-CRT, an almost intact Form D was administered except for Reading Grade 4, where a combination of Form D and Form E was administered. The following two steps were used to place the 2005 WKCE-CRT scale on the 2004 Form Standardization scale:

- Step 1: 2005 WKCE-CRT items were calibrated for each grade and content.

- Step 2: For each grade and content, the items which appeared in both 2004 Form Standardization and the 2005 WKCE-CRT were treated as anchor items. Using the anchor items, item parameters for the 2005 WKCE-CRT were transformed.

Then, the Stocking and Lord (1983) formula was applied to estimate the transformation slope and intercept. The transformation slope and intercept was applied to 2005 WKCE item parameters. Because the 2004 WKCE Form Standardization was on a vertical scale across grades, the 2005 scale transformation to the 2004 scale means that the vertical relationship across grades still exists for the 2005 WKCE-CRT. The mean and standard deviation for Reading and Mathematics can be found Part 9.1 Summary Statistics for Scale Score.

Figure 8-3 and 8-4 show the vertical relationships of Reading and Mathematics scales across grades. Although some test characteristics curves for Reading were overlapped in some ability ranges, this overlapping was not a major concern because this type of vertical relationship pattern for Reading has been found in many large scale State assessments, and the vertical order of the state mean and standard deviation was considered more important. As can be seen Table 9-11 and Table 9-12, the means and standard deviations of the 14 CD and WI census show this vertical order for Reading and Mathematics.

8.2.2 Language Arts, Social Studies, and Science

Vertical scaling was planned for and applied to Reading and Mathematics, but because Language Arts, Social Studies, and Science are not given, those scales are grade specific. In the 2005 WKCE-CRT, Reading and Mathematics were administered to grades 3–8 and 10, while the remaining three contents were administered to grades 4, 8, and 10. Without administering tests for all continuous grades for a given content area, it is difficult to build a vertical scale for the content area. Although the vertical relationship across grades was not set up for these grades in

the 2004 Form Standardization testing, the scales for grades 4, 8, 10 were artificially constructed in such a way so as to show a vertical relationship across grades. DPI and CTB were concerned that, had this not been done, test users could wrongly interpret the scales because different grades would show similar means and standard deviations. The typical scales without the vertical relationship across grades are set up to use the same mean and standard deviation for all grades. In that arrangement, two students from two different grades considered to have similar performances in their respective grades, could actually see a scale score for the higher grade student lower than the scale score of the lower grade student. To avoid this situation, an artificial vertical relationship was set up across grades for these three contents. The mean and standard deviation for these three grades can be found Part 9.1 Summary Statistics for Scale Score. Figure 8-5 and 8-7 show the vertical relationships of Language Arts, Social Studies, and Science across grades. Although the three TCCs for the three grades show the vertical relationship across grades, this relationship was artificially built, as mentioned.

8.3 Calibration and Scaling Results

As described, the items that appeared in both the 2004 Form Standardization and 2005 WKCE-CRT were treated as anchor items for calibrating and scaling the operational items. For some contents and grades, the 2005 WKCE-CRT contains field tested items together with operational items. Part 7.4 Classical Item Analysis shows information for these field test items. These field test items were calibrated together with 2005 WKCE-CRT operational items, and transformed to the scale of the 2005 WKCE-CRT using the item parameters of the 2005 WKCE-CRT operational items. While all responses of field test MC items were included, about 2,000 responses of field test CR items and Writing prompts were used for both calibration and scaling. Note that about 2,000 responses were scored for each field test CR item and Writing prompt. The number of responses for CR items and Writing prompt can be found in Part 6.4.1 (Distribution of CR items).

8.3.1 IRT Item Parameters

All operational items were converged, meaning parameters were successfully estimated for each item, but there were three field test items not converged, or for which parameters could not be estimated during calibration: Reading Grade 7 Form A/B/C #82, Mathematics Grade 7 Form B #73, and Language Arts Grade 8 item #32. Figure 8-1 shows the item characteristic curve (ICC) of Reading item #82. The horizontal axis represents the range of student ability (or performance trait) from -4.0 to 4.1. The vertical axis presents the proportion of students correct on the item. The figure clearly shows that the expected ICC based on the IRT theory did not fit to the observed ICC across all ranges. If the two ICCs fit well, the two lines would be almost overlapped across all ability ranges. While the expected ICC always expects the monotonic increase of performance as students' ability increases, the observed ICC for the item did not show that monotonic pattern. The observed ICC shows that low-ability students did better for this item than high-performing students did for this item. Figure 8-2 shows the ICC for Mathematics Grade 7 item #73. As can be seen in the figure, this item did not discriminate students across all ability levels. This item was relatively easy for all students. These two items

will not be used as operational items without re-field testing after the items are revised. Figure 8-3 shows the ICC for Language Arts Grade 8 item #32. Like the mathematics item, this item does not discriminate between students across all ability levels. These three items will not be used for any future testing without re-field testing after the items are revised.

Whenever item parameters were used, as when used for scoring, the estimated item parameters from the 14 calibration districts were used in the 2005 WKCE-CRT. Although using item parameters from census data is ideal, the item parameters from the 14 calibration districts were used due to the time limitation. As can be seen in Part 7.1, the 14 CD seemed to represent the WI census well.

The current technical report does not contain item parameters used for the 2005 Fall WKCE-CRT scoring because of the large size of the data files. Separate excel files containing item parameters will be delivered to DPI for a database.

8.3.2 IRT Item Fit

A statistical procedure was used to identify items that did not fit the IRT model. Item model fit information was obtained for each item using a *Z*-statistic. 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:

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

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*.

The *Z*-statistic is an index of the degree to which obtained proportions of students with each item score are close to the proportions that would be predicted by the estimated student ability and item parameters. These values, along with the associated chi-squares (Q_1), are computed for ten intervals corresponding to deciles of the ability distribution (Yen, 1984). 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*.

Table 8–1 presents items that were flagged based on the Z statistics above. For example, the second row shows that Reading Grade 3 operational MC item #40 was flagged because its Z value of 43.04 is larger than the critical Z value of 16.10 based on the sample size of 6,306. The

third column does not show form numbers for items that appear on all forms. Many CR operational or field test items were flagged, though the ratio of CR items to MC items on a test form is small. In general, there are a small number of students at the lower score level or higher score level for CR items, and these small sample sizes easily introduce the misfit between the observed ICC and expected ICC. With a small sample size, it is not easy to get a stable expected ICC. In a similar manner, the misfit for MC items often happens at the lower ability range or higher ability range, where there are a small number of students. As shown in Table 1, more Mathematics items were flagged than Reading items because Mathematics contains more CR items than Reading. Because the index itself does not show where the misfit happens on the ability range, graphical information was produced for each item by PARDUX. The main concern for the item fit is where the misfit happens. If the misfit happens around the lower or higher ability range, where there are not many students, we do not worry as much about the misfit. If the misfit happens around the middle of ability range, where there are many students, we are more concerned. The flagging of an item does not require that the item not be used. This item fit is just one of the criteria for selecting sound operational items. The fit index for all items and the graphical information for items flagged are not included in this report, but will be separately delivered to DPI. As with all items flagged, the list of items flagged based on the Z statistics and graphical information was delivered to Development for future item selection.

8.3.3 Scoring and Standard Error of Measurement

Item-pattern scoring utilizes more information about students' responses than number-correct scoring. The item-pattern score is the maximum likelihood estimate for students with a given response pattern and known item parameter estimates. Either raw score or item-pattern scoring can be chosen. For groups of 25 or more students, the two methods produce tau equivalent results. Item-pattern scoring is generally recommended because it produces more accurate scores for individual students. This increase in accuracy is equivalent, on the average, to approximately a 15 to 20% increase in test length (Yen, 1984; Yen & Candell, 1991). This item-pattern score was applied to the 2004 WKCE and the 2005 WKCE-CRT. Note that the pattern score means that students with the same raw score can get different scale scores. Students with the same raw score can have different scale scores even if they correctly answered the same number of items. If a student A correctly answered more difficult items than student B, with the same raw score for the same test, the scale score of the student A would be higher than that of student B. Students who correctly answered difficult items will have higher scale scores than the students who correctly answered easy items. Therefore, a scoring table, which shows the relationship between raw score and scale score, can not be applied to the 2005 WKCE-CRT. However, to show the rough relationship among raw score, scale score, and standard error of measurement (SEM), scoring tables were included. Tables 8-2 through 8-25 show these scoring tables.

Standard error of measurement is used to obtain a range within which a student's true score is likely to fall. An obtained 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. It is expected that 68% of the time a student's score obtained from a single testing would fall within

one SEM of that student's true score and that 95% of the time the obtained score would fall within two standard errors of true score.

Standard errors of measurement (SEM) for the 2005 WKCE-CRT scale scores, obtained from item-pattern scoring, are displayed graphically for each of the test configurations in Figures 8-9 through 8-13. Each figure includes a SEM curve of a given grade level. The curve for each form is plotted as a function of the scale scores. Note that for convenience, the highest and lowest obtainable scale score (HOSS and LOSS) of 2005 WKCE-CRT were used as the starting scale score and the last scale score.

These figures show the scale score range within which measurement is most accurate and that extreme scale scores have more measurement error than moderate scores. The forms lose accuracy of measurement for scale scores near the high or low extremes because there are fewer students at these score ranges.

Part 11: Linking Study and Descriptor Writing

11.1 Linking 2005 WKCE-CRT to 2004 WKCE

Cut scores were established for the Wisconsin Knowledge and Concepts Examination (WKCE) in 2002. These cut scores were used until the Fall 2004 WKCE. In 2005, the Wisconsin Department of Public Instruction augmented the testing program to create the WKCE-CRT, assessing students in Grades 3–8 and 10. Because the Fall 2005 WKCE-CRT is a criterion-reference test and the 2004 WKCE is a norm-referenced test, the 2005 WKCE-CRT scale was different than the 2004 WKCE scale, for all contents. Therefore, a linking study, which links the 2005 WKCE-CRT to the 2004 WKCE, was necessary. Note that the 2004 WKCE was administered only on grades 4, 8, and 10, for all contents: Reading, Mathematics, Language Arts, Social Studies, and Science. For 2005, the WKCE-CRT expanded the Reading and Mathematics assessments to include grades 3, 5, 6, and 7.

A study comparing three different linking procedures to link the 2004 WKCE and the 2005 WKCE-CRT was performed. The results and implications were delivered to DPI (see the two papers, "A Report for Linking 2004 WKCE Operational Test to 2005 WKCE-CRT Operational Test" and "Discussion for Implication of Three Linking Studies"). Two Technical Advisory Council (TAC) members, DPI, and CTB discussed the results of the study, and the linking procedure based on the assumption that there is flat growth between the 2004 WKCE and the 2005 WKCE-CRT for grades 4, 8, and 10 was accepted.

After reviewing different options for interpolating and extrapolating the cut scores for the 2005 WKCE-CRT assessments in Reading and Mathematics, the Wisconsin Department of Public Instruction (DPI) opted to use a method of linear interpolation based on impact data. In the option chosen, Reading and Mathematics cut scores for grades 3, 5, 6, and 7 are to be interpolated/extrapolated using the cuts scores for grades 4, 8, and 10.

To find cut scores for Grades 3, 5, 6, and 7, the *impact data* for Grades 4 and 8 were first calculated. Impact data indicate the percentage of students classified in each achievement level. The cut scores for Grades 4, 8, and 10 were derived from the no-growth model, which preserved the existing Grade 4, 8, and 10 cut scores. Linear interpolation was then used to find the desired impact data for Grades 5, 6, and 7 based on the previously calculated impact data for Grades 4 and 8. Linear extrapolation was used to find the desired impact data for Grade 3 by extending the trend. The cut scores which most closely gave the desired impact data were then found, as described below.

There exists no one preferred method by which to identify cut scores using impact data, as previously described. Rather, the method used represents a policy decision by DPI. To find cut scores for a given grade and content area, the desired percent of students in an achievement level was first found through either linear interpolation or extrapolation. For each achievement level, if a cut score existed which yielded exactly this impact data, the cut score was adopted. If no cut score gave exactly this percentage, then the highest cut score which yielded the desired or next greater percentage point was found.

For example, assume that 30.0% of students should be classified as *Proficient* or above in a sample grade and content area. If a cut score was available which classified exactly 30.0% of students as *Proficient* or above, then it was adopted. However, if exactly 30.0% was not possible because of slight variations in the scoring table, then the lowest cut score which yielded at least 30.0% was adopted. This method ensures that the percent of students classified as *Proficient* and above will not decline simply as an artifact of the interpolation process.

The WKCE-CRT assessments for Reading and Mathematics are on a vertical scale, and it is important that the cut scores for a given achievement level rise from grade to grade. To promote this type of *vertical moderation*, the cut score for *Advanced* in Grade 10 Reading was raised from 538 to 555, and the cut score for *Basic* in Grade 3 was lowered from 396 to 394.

Table 11-1 shows the cut scores for Grades 3–8 and 10 for Reading, along with the impact data associated with these cut scores. Table 11-2 shows the cut scores and associated impact data for Grades 3–8 and 10 for Mathematics. Tables 11-3, 11-4, and 11-5 present the cut scores and associated impact data for Grades 4, 8, and 10 for Language Arts, Social Studies, and Science. Figures 11-1 through 11-10 present the cut scores and percentages for all performance levels based on impact data across all grades and content areas.

Crosswalk tables, which show the relationship between the 2004 WKCE scale score and the 2005 WKCE-CRT scale score for each percentile, were also generated for all contents and grades. These are tables 11-6 to 11-20. The first column, "Fall 2004 WKCE," and the third column, "Fall 2005 WKCE-CRT," provide the scale scores corresponding to each percentile in the second column.

11.2 Descriptor Writing

Committees of Wisconsin educators were convened June 20–22, 2006 in order to develop performance level descriptors to accompany the performance standards. Description writing provides plain-language description of the content that students must know at each grade level to be Proficient. This information may be used by teachers and the public to fully understand the performance levels on the WKCE-CRT. Description Writing allows for teacher input regarding performance-level descriptors.

In the description writing workshop, participants were asked to record the knowledge, skills, and abilities that are required of students in each grade to be Basic, Proficient, and Advanced. To inform their descriptions, participants reviewed ordered item booklets and item maps and identified the knowledge and skills required to answer each item correctly and why each item is more difficult than the preceding item. Participants were shown the statistically set cut scores and then wrote descriptors for each grade/content area.

Prior to the workshop, CTB and DPI discussed the final format of the descriptors. DPI requested three formats:

Brief Narrative Description

These one-paragraph descriptions of each proficiency level may be most useful for those who simply want an overview of the knowledge and skills students typically demonstrate at each level.

Detailed Narrative Description

These descriptions contain more detail but are still structured in a way that makes the information easy to grasp.

Elements of Proficiency Levels

The elements are descriptions of discrete knowledge and skills students typically demonstrate at each proficiency level. They complement the narratives by enumerating specific examples of knowledge and skills described in the narratives.

The morning of the first day, CTB presented a PowerPoint presentation which reviewed the purpose of the descriptor writing workshop, how the cut scores for each performance category were established, an overview of the specific tasks to be completed, the characteristics of well-written descriptors, and how the descriptors should reflect the progression of abilities within and across grade levels.

The educators were assigned to content and grade level groups with 4–6 participants per grade. Two CTB facilitators were assigned to each content area group. The CTB facilitators guided the committees through a series of tasks designed to build familiarity with the test and the content frameworks and then to draft and revise descriptors. Because there were not items for every performance level for each content standard, especially for the Minimal Performance and Basic categories, participants were instructed to use professional judgment to augment the information provided by the test items in order to develop a more complete set of descriptors. Specifically, the sequence of tasks was:

- take the Fall 2005 test
- review the ordered item book and describe each item using the item map
- review the cut scores and identify the cut score location in the ordered item book
- review the existing performance level descriptors for grades 4, 8, 10 (established in 2003)
- organize ordered items by content objective and performance level
- draft descriptors by content objective and performance level
- review descriptors for each content objective within the grade level group
- review descriptors by content objective in cross-grade level groups
- revise descriptors by content objective to reflect level to level and grade-to-grade progression
- draft multi-paragraph narrative descriptors from the bulleted list of descriptors by objective
- review and revise narratives across performance levels within a grade

Following the meeting, CTB content specialists reviewed the draft descriptors, checking the accuracy of the description written for each item by checking it against the item in the ordered item book. The CTB content specialist also edited the descriptors for consistency in style and to

ensure that the descriptors appropriately described the increasing level of knowledge and skills across performance levels within a grade and across the grades. The revised descriptors were submitted to DPI for review. DPI distributed the draft descriptors to the table leaders for their review, and a conference call was conducted with DPI, CTB, and the table leaders in attendance. The conference calls were helpful for providing feedback on both general and specific issues. The CTB content specialists then revised the bulleted descriptors and the multi-paragraph narratives based on the feedback and submitted them to DPI for a second review. DPI reviewed the descriptors and provided feedback, which focused primarily on the narrative descriptors. CTB did a final edit of the bulleted and multi-paragraph narrative and then wrote the single-paragraph, condensed narrative. DPI then completed the formatting of the descriptors to prepare them for presentation to the superintendent's cabinet prior to release to the public.

Table 8-1
Item Flagged Based on Yen's Q_1

Content	Grade	Form	Test Book_ID	CR Part	Status	Type	N	Z	Critical Z
RD	3		40		OP	MC	6036	43.04	16.10
	3	A	87		FT	CR	2712	10.41	7.23
	4		12		OP	MC	6248	18.2	16.66
	4	A	79		FT	CR	2902	14.38	7.74
	6		44		OP	MC	6454	28.47	17.21
	7		46		OP	MC	6659	33.9	17.76
	7		62		OP	MC	6642	30.54	17.71
	8		57		OP	MC	6876	18.44	18.34
	10		18		OP	MC	7490	20.04	19.97
	10		39		OP	MC	7448	38.11	19.86
	10		42		OP	MC	7466	30.29	19.91
MA	3	A	60		FT	CR	1905	5.54	5.08
	3	C	60		FT	CR	1927	9.69	5.14
	3	C	67		FT	CR	1937	8.26	5.17
	4		20	A	OP	CR	6251	27.56	16.67
	4		37	B	OP	CR	6105	29.11	16.28
	5		1		OP	MC	6297	38.09	16.79
	5		10	B	OP	CR	6243	18.9	16.65
	5		32	B	OP	CR	6109	99.74	16.29
	5		59		OP	MC	6284	17.45	16.76
	5	C	67	A	FT	CR	1925	6.63	5.13
	5	C	67	B	FT	CR	1911	13.01	5.10
	6		53	A	OP	CR	6349	20.22	16.93
	6		53	B	OP	CR	6284	16.96	16.76
	6		57	B	OP	CR	6141	33.18	16.38
	6	A	67		FT	CR	1904	9.21	5.08
	7		29	B	OP	CR	6293	24.04	16.78
	7		58		OP	MC	6510	18.99	17.36
	7		60	A	OP	CR	6292	20	16.78
	7	B	68		FT	CR	1889	33.27	5.04
	7	B	79		FT	CR	1915	12.01	5.11
	8		2		OP	MC	6848	21.05	18.26
	8		4	B	OP	CR	6791	23.87	18.11
	8		8		OP	MC	6855	31.54	18.28
	8		22	A	OP	CR	6582	17.65	17.55
	8		33	B	OP	CR	6436	31.23	17.16

Table 8-1 (Cont.)
Item Flagged Based on Yen's Q_1

Content	Grade	Form	Test Book_ID	CR Part	Status	Type	N	Z	Critical Z
MA	8		35		OP	MC	6796	18.53	18.12
	8		51		OP	MC	6790	19.27	18.11
	8	A	68		FT	MC	21586	74.48	57.56
	8	C	66		FT	MC	20680	216.75	55.15
	10		8		OP	MC	7449	25.77	19.86
	10		15		OP	CR	7095	35.9	18.92
	10		53		OP	MC	7454	20.58	19.88
SS	4		13		OP	MC	5978	30.16	15.94
	4		35		OP	MC	5962	20.78	15.90
	8		42		OP	MC	6677	24.88	17.81
	10		30		OP	MC	7361	39.45	19.63
SC	8		19		OP	MC	6812	22.08	18.17
	8		20		OP	MC	6812	20.57	18.17
	10		4		OP	MC	7393	24.56	19.71
LA	4	B	34		FT	MC	9713	25.96	25.90
	4	F	1	A	FT	CR	1907	5.37	5.09
	8		15		OP	MC	6245	24.44	16.65
	8		26		OP	MC	6673	27.94	17.79
	8		1	A	OP	CR	65069	196.08	173.52
	8	D	1	B	FT	CR	1932	6.95	5.15
	10		18		OP	MC	7435	35.32	19.83

Table 8-2
Scoring Table for Reading Grade 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	270	126	33	435	7
1	270	126	34	437	7
2	270	126	35	439	7
3	270	126	36	441	7
4	270	126	37	443	7
5	270	126	38	445	7
6	270	126	39	447	7
7	270	126	40	449	7
8	270	126	41	452	7
9	270	126	42	454	7
10	270	126	43	456	7
11	270	126	44	459	8
12	270	126	45	461	8
13	342	54	46	464	8
14	364	32	**47**	**467**	**8**
15	375	22	48	470	8
16	383	17	49	473	9
17	389	15	50	476	9
18	**394**	**13**	51	479	9
19	398	12	52	482	9
20	402	11	53	486	10
21	406	10	54	490	10
22	409	9	55	494	10
23	412	9	56	499	11
24	414	8	57	503	11
25	417	8	58	509	12
26	419	8	59	515	13
27	422	7	60	522	14
28	424	7	61	531	16
29	426	7	62	543	19
30	428	7	63	560	25
31	**431**	**7**	64	587	36
32	433	7	65	635	57
			66	640	60

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

Table 8-3
Scoring Table for Reading Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	280	126	39	473	9
1	280	126	40	475	9
2	280	126	41	478	9
3	280	126	42	481	9
4	280	126	43	484	9
5	280	126	44	487	9
6	280	126	**45**	**489**	**9**
7	280	126	46	492	9
8	280	126	47	496	9
9	280	126	48	499	9
10	280	126	49	502	10
11	280	126	50	505	10
12	280	126	51	509	10
13	336	70	52	512	10
14	362	44	53	516	11
15	378	32	54	520	11
16	389	26	55	525	11
17	**397**	**22**	56	529	12
18	404	19	57	534	12
19	410	17	58	540	13
20	415	15	59	546	14
21	419	14	60	553	15
22	424	13	61	561	17
23	427	12	62	571	19
24	431	11	63	583	21
25	434	11	64	600	26
26	437	10	65	627	37
27	**440**	**10**	66	650	51
28	443	10			
29	446	9			
30	449	9			
31	452	9			
32	454	9			
33	457	9			
34	460	9			
35	462	9			
36	465	9			
37	467	9			
38	470	9			

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

Table 8-4
Scoring Table for Reading Grade 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	290	80	39	462	11
1	290	80	40	465	11
2	290	80	41	469	11
3	290	80	42	472	11
4	290	80	43	475	11
5	290	80	44	478	11
6	290	80	45	481	11
7	290	80	46	485	11
8	290	80	47	488	11
9	290	80	48	491	11
10	290	80	49	495	11
11	290	80	**50**	**498**	**11**
12	290	80	51	502	11
13	290	80	52	506	11
14	290	80	53	510	12
15	322	51	54	514	12
16	344	35	55	518	12
17	358	28	56	523	12
18	369	24	57	528	13
19	377	21	58	533	13
20	385	19	59	539	14
21	391	18	60	545	15
22	397	16	61	552	16
23	**402**	**15**	62	560	17
24	407	15	63	570	19
25	412	14	64	581	21
26	416	13	65	594	24
27	420	13	66	612	29
28	424	13	67	637	37
29	428	12	68	679	55
30	432	12	69	690	61
31	436	12			
32	439	12			
33	**442**	**11**			
34	446	11			
35	449	11			
36	452	11			
37	456	11			
38	459	11			

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

Table 8-5
Scoring Table for Reading Grade 6

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	300	84	39	482	12
1	300	84	40	485	12
2	300	84	41	489	11
3	300	84	42	492	11
4	300	84	43	496	11
5	300	84	44	499	11
6	300	84	45	503	11
7	300	84	46	506	11
8	300	84	47	510	12
9	300	84	**48**	**514**	**12**
10	300	84	49	518	12
11	300	84	50	522	12
12	300	84	51	526	12
13	300	84	52	530	12
14	328	59	53	535	13
15	350	44	54	540	13
16	365	35	55	545	14
17	377	29	56	550	14
18	386	24	57	556	15
19	394	21	58	562	15
20	401	19	59	568	16
21	407	17	60	576	17
22	413	16	61	583	17
23	**418**	**15**	62	592	18
24	423	15	63	601	19
25	428	14	64	612	20
26	433	14	65	625	23
27	437	14	66	642	28
28	441	14	67	667	38
29	445	13	68	711	59
30	449	13	69	730	69
31	453	13			
32	**457**	**13**			
33	460	13			
34	464	12			
35	468	12			
36	471	12			
37	475	12			
38	478	12			

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

Table 8-6
Scoring Table for Reading Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	310	103	39	499	10
1	310	103	40	502	10
2	310	103	41	506	10
3	310	103	42	509	10
4	310	103	43	512	10
5	310	103	44	515	10
6	310	103	45	518	11
7	310	103	**46**	**522**	**11**
8	310	103	47	525	11
9	310	103	48	529	11
10	310	103	49	532	11
11	310	103	50	536	11
12	310	103	51	540	12
13	341	72	52	544	12
14	368	45	53	548	12
15	385	32	54	552	13
16	396	26	55	557	13
17	406	21	56	562	14
18	413	19	57	567	14
19	420	17	58	573	15
20	426	16	59	579	16
21	**431**	**15**	60	586	17
22	437	14	61	594	18
23	441	14	62	604	20
24	446	13	63	615	23
25	450	13	64	628	26
26	454	13	65	644	30
27	458	12	66	667	36
28	462	12	67	700	48
29	**466**	**12**	68	774	89
30	470	12	69	780	93
31	473	12			
32	477	11			
33	480	11			
34	483	11			
35	487	11			
36	490	11			
37	493	11			
38	496	11			

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

Table 8-7
Scoring Table for Reading Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	330	83	39	512	12
1	330	83	40	516	12
2	330	83	41	519	12
3	330	83	42	523	12
4	330	83	43	526	12
5	330	83	44	530	12
6	330	83	45	534	12
7	330	83	**46**	**537**	**12**
8	330	83	47	541	12
9	330	83	48	545	12
10	330	83	49	549	12
11	330	83	50	553	12
12	330	83	51	557	13
13	330	83	52	562	13
14	330	83	53	566	13
15	359	60	54	571	14
16	382	43	55	576	14
17	398	34	56	582	15
18	410	28	57	587	15
19	420	24	58	593	15
20	428	22	59	600	16
21	435	20	60	606	16
22	442	19	61	614	17
23	**447**	**18**	62	622	18
24	453	17	63	631	19
25	458	16	64	642	21
26	463	15	65	655	24
27	467	15	66	673	30
28	472	14	67	699	40
29	476	14	68	750	65
30	**480**	**13**	69	790	90
31	484	13			
32	487	13			
33	491	12			
34	495	12			
35	498	12			
36	502	12			
37	505	12			
38	509	12			

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

Table 8-8
Scoring Table for Reading Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	350	56	39	523	14
1	350	56	40	527	14
2	350	56	41	531	14
3	350	56	42	535	14
4	350	56	43	539	14
5	350	56	44	544	14
6	350	56	45	548	14
7	350	56	**46**	**552**	**14**
8	350	56	47	557	14
9	350	56	48	561	15
10	350	56	49	566	15
11	350	56	50	571	15
12	350	56	51	576	16
13	350	56	52	582	16
14	354	54	53	587	16
15	375	42	54	593	17
16	391	35	55	600	18
17	404	30	56	606	18
18	414	26	57	614	19
19	423	24	58	622	20
20	431	22	59	631	21
21	438	21	60	641	22
22	444	20	61	652	24
23	451	19	62	665	26
24	**456**	**18**	63	681	29
25	462	17	64	702	35
26	467	17	65	735	48
27	472	16	66	803	89
28	477	16	67	820	103
29	481	16			
30	486	15			
31	490	15			
32	494	15			
33	499	15			
34	**503**	**14**			
35	507	14			
36	511	14			
37	515	14			
38	519	14			

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

Table 8-9
Scoring Table for Mathematics Grade 3

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	220	75	39	412	12
1	220	75	40	415	12
2	220	75	41	419	12
3	220	75	42	422	12
4	220	75	43	426	12
5	220	75	44	430	12
6	220	75	45	433	12
7	220	75	46	437	12
8	220	75	47	441	12
9	220	75	48	445	12
10	237	60	49	449	12
11	264	40	**50**	**454**	**12**
12	280	31	51	458	12
13	292	26	52	463	13
14	302	22	53	467	13
15	310	20	54	473	13
16	318	18	55	478	13
17	324	17	56	484	14
18	330	16	57	490	15
19	335	15	58	498	16
20	340	14	59	507	18
21	345	14	60	517	20
22	349	13	61	530	24
23	354	13	62	548	30
24	358	13	63	573	39
25	362	12	64	617	58
26	366	12	65	630	65
27	369	12			
28	373	12			
29	377	12			
30	380	12			
31	384	12			
32	387	12			
33	**391**	**11**			
34	394	11			
35	398	11			
36	401	11			
37	405	11			
38	**408**	**11**			

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

Table 8-10
Scoring Table for Mathematics Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	240	68	**39**	**438**	**11**
1	240	68	40	441	11
2	240	68	41	444	11
3	240	68	42	447	11
4	240	68	43	451	11
5	240	68	44	454	11
6	240	68	45	457	11
7	240	68	46	461	11
8	240	68	47	464	11
9	240	68	48	467	11
10	240	68	49	471	11
11	260	54	50	474	11
12	285	40	51	478	11
13	303	33	52	482	11
14	316	28	**53**	**485**	**12**
15	327	25	54	489	12
16	336	23	55	493	12
17	345	20	56	498	12
18	352	19	57	502	13
19	358	18	58	507	13
20	364	16	59	512	14
21	370	15	60	518	15
22	375	14	61	524	15
23	379	14	62	531	16
24	384	13	63	539	18
25	388	13	64	549	20
26	392	12	65	560	22
27	396	12	66	577	27
28	400	12	67	605	40
29	404	12	68	650	73
30	407	11			
31	411	11			
32	414	11			
33	418	11			
34	**421**	**11**			
35	425	11			
36	428	11			
37	431	11			
38	434	11			

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

Table 8-11
Scoring Table for Mathematics Grade 5

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	270	60	39	460	11
1	270	60	**40**	**463**	**10**
2	270	60	41	466	10
3	270	60	42	469	10
4	270	60	43	472	10
5	270	60	44	475	10
6	270	60	45	477	10
7	270	60	46	480	10
8	270	60	47	483	10
9	270	60	48	486	10
10	270	60	49	489	10
11	270	60	50	492	10
12	279	55	51	494	10
13	304	43	52	497	10
14	322	35	53	500	10
15	336	30	54	503	10
16	348	27	**55**	**506**	**10**
17	358	24	56	509	10
18	366	23	57	512	10
19	374	21	58	516	10
20	381	20	59	519	10
21	388	18	60	522	11
22	393	17	61	526	11
23	399	17	62	530	11
24	404	16	63	534	12
25	409	15	64	538	12
26	414	14	65	543	12
27	418	14	66	548	13
28	422	14	67	553	14
29	426	13	68	559	15
30	430	13	69	566	16
31	434	12	70	574	16
32	437	12	71	582	17
33	441	12	72	592	17
34	**444**	**12**	73	605	20
35	447	11	74	624	29
36	451	11	75	657	43
37	454	11	76	680	53
38	457	11			

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

Table 8-12
Scoring Table for Mathematics Grade 6

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	310	75	39	496	10
1	310	75	40	499	10
2	310	75	41	501	10
3	310	75	42	504	10
4	310	75	43	506	9
5	310	75	44	509	9
6	310	75	45	512	9
7	310	75	46	514	9
8	310	75	47	517	9
9	310	75	48	519	9
10	310	75	49	521	9
11	310	75	50	524	9
12	319	67	51	526	9
13	351	42	52	529	9
14	370	32	**53**	**531**	**9**
15	384	26	54	534	9
16	395	23	55	536	9
17	404	21	56	539	9
18	412	19	57	541	9
19	419	18	58	544	9
20	425	17	59	547	9
21	431	16	60	550	9
22	436	15	61	553	9
23	441	15	62	556	9
24	445	14	63	559	10
25	450	14	64	562	10
26	454	13	65	566	11
27	458	13	66	570	11
28	462	12	67	575	12
29	**465**	**12**	68	579	13
30	469	12	69	585	14
31	472	12	70	591	15
32	475	11	71	599	17
33	478	11	72	609	19
34	482	11	73	622	24
35	**485**	**11**	74	641	31
36	487	11	75	675	48
37	490	10	76	700	65
38	493	10			

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

Table 8-13
Scoring Table for Mathematics Grade 7

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	330	86	39	521	10
1	330	86	40	523	10
2	330	86	41	526	9
3	330	86	42	529	9
4	330	86	43	532	9
5	330	86	44	534	9
6	330	86	45	537	9
7	330	86	46	539	9
8	330	86	47	542	9
9	330	86	48	545	9
10	330	86	49	547	9
11	334	82	50	550	9
12	373	46	51	553	9
13	393	34	**52**	**555**	**9**
14	407	28	53	558	9
15	418	24	54	561	9
16	427	21	55	564	9
17	434	19	56	566	9
18	441	18	57	569	9
19	447	16	58	572	9
20	453	16	59	575	9
21	458	15	60	578	9
22	463	14	61	581	9
23	467	14	62	585	10
24	471	13	63	588	10
25	476	13	64	592	10
26	**479**	**13**	65	596	11
27	483	12	66	600	11
28	487	12	67	605	12
29	490	12	68	610	13
30	494	11	69	616	13
31	497	11	70	623	14
32	500	11	71	630	15
33	**503**	**11**	72	639	17
34	506	11	73	651	19
35	509	10	74	667	25
36	512	10	75	695	37
37	515	10	76	710	46
38	518	10			

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

Table 8-14
Scoring Table for Mathematics Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	350	71	39	554	10
1	350	71	40	557	9
2	350	71	41	560	9
3	350	71	42	562	9
4	350	71	43	565	9
5	350	71	44	568	9
6	350	71	45	570	9
7	350	71	**46**	**573**	**9**
8	350	71	47	576	9
9	350	71	48	578	9
10	350	71	49	581	9
11	350	71	50	583	9
12	373	55	51	586	9
13	402	40	52	588	9
14	421	33	53	591	9
15	436	28	54	594	9
16	448	25	55	596	9
17	458	23	56	599	9
18	467	20	57	602	9
19	475	19	58	605	9
20	**481**	**17**	59	608	9
21	488	16	60	611	9
22	493	15	61	614	10
23	498	14	62	618	10
24	503	14	63	621	10
25	508	13	64	625	10
26	**512**	**12**	65	629	11
27	516	12	66	634	11
28	520	12	67	639	12
29	523	11	68	644	12
30	527	11	69	650	13
31	530	11	70	658	15
32	533	11	71	667	17
33	536	10	72	680	21
34	540	10	73	704	32
35	543	10	74	730	51
36	546	10			
37	549	10			
38	551	10			

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

Table 8-15
Scoring Table for Mathematics Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	410	82	39	578	8
1	410	82	40	580	8
2	410	82	41	582	8
3	410	82	42	585	8
4	410	82	43	587	8
5	410	82	44	589	8
6	410	82	45	592	8
7	410	82	**46**	**594**	**8**
8	410	82	47	597	8
9	410	82	48	599	8
10	410	82	49	602	8
11	410	82	50	604	8
12	427	65	51	607	9
13	459	40	52	609	9
14	477	29	53	612	9
15	488	24	54	615	9
16	497	20	55	618	9
17	504	18	56	621	9
18	511	16	57	625	10
19	**516**	**15**	58	628	10
20	521	14	59	632	10
21	526	13	60	637	11
22	530	13	61	641	11
23	534	12	62	647	12
24	537	11	63	653	13
25	**541**	**11**	64	660	15
26	544	11	65	669	17
27	547	10	66	680	20
28	550	10	67	697	25
29	553	10	68	726	39
30	555	9	69	750	53
31	558	9			
32	561	9			
33	563	9			
34	566	9			
35	568	8			
36	570	8			
37	573	8			
38	575	8			

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

Table 8-16
Scoring Table for Language Arts Grade 4

Raw Score	Scale Score	SEM
1	140	116
2	140	116
3	140	116
4	140	116
5	140	116
6	140	116
7	188	68
8	224	32
9	239	21
10	248	16
11	**255**	**14**
12	261	12
13	266	10
14	270	9
15	275	9
16	**278**	**9**
17	282	8
18	286	8
19	290	9
20	294	9
21	298	9
22	302	9
23	**307**	**9**
24	312	9
25	317	10
26	323	11
27	330	12
28	341	16
29	362	27
30	420	85

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

Table 8-17
Scoring Table for Language Arts Grade 8

Raw Score	Scale Score	SEM
1	250	84
2	250	84
3	250	84
4	250	84
5	250	84
6	250	84
7	280	54
8	304	30
9	317	22
10	327	17
11	334	15
12	341	13
13	347	12
14	352	12
15	**357**	**11**
16	362	11
17	366	11
18	371	11
19	376	11
20	381	11
21	**386**	**11**
22	392	12
23	398	12
24	404	13
25	412	13
26	**420**	**13**
27	429	14
28	441	17
29	461	25
30	520	79

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

Table 8-18
Scoring Table for Language Arts Grade 10

Raw Score	Scale Score	SEM
0	290	72
1	290	72
2	290	72
3	290	72
4	290	72
5	290	72
6	290	72
7	321	41
8	344	24
9	357	19
10	367	17
11	375	16
12	382	16
13	389	15
14	**395**	**15**
15	401	14
16	407	14
17	413	13
18	418	13
19	423	13
20	**429**	**13**
21	434	12
22	439	12
23	444	13
24	450	13
25	455	13
26	461	13
27	467	14
28	473	14
29	480	14
30	**487**	**15**
31	495	15
32	503	16
33	512	16
34	522	17
35	533	18
36	548	21
37	568	27
38	602	41
39	630	56

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

Table 8-19
Scoring Table for Social Studies Grade 4

Raw Score	Scale Score	SEM
0	170	71
1	170	71
2	170	71
3	170	71
4	170	71
5	170	71
6	170	71
7	170	71
8	170	71
9	190	51
10	214	27
11	225	17
12	232	13
13	237	11
14	**242**	**10**
15	246	9
16	249	8
17	252	8
18	255	7
19	258	7
20	260	7
21	**263**	**7**
22	265	6
23	268	6
24	270	6
25	273	6
26	275	6
27	278	6
28	280	6
29	283	6
30	286	7
31	**289**	**7**
32	293	7
33	297	8
34	302	9
35	309	11
36	320	15
37	341	27
38	400	86

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

Table 8-20
Scoring Table for Social Studies Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	230	108	39	419	11
1	230	108	40	425	11
2	230	108	41	431	12
3	230	108	42	440	14
4	230	108	43	452	18
5	230	108	44	474	29
6	230	108	45	530	79
7	230	108			
8	230	108			
9	230	108			
10	230	108			
11	232	106			
12	295	43			
13	312	26			
14	321	19			
15	329	15			
16	**334**	**13**			
17	339	12			
18	344	11			
19	347	10			
20	351	10			
21	355	9			
22	358	9			
23	361	9			
24	**365**	**9**			
25	368	9			
26	371	9			
27	374	9			
28	377	9			
29	380	9			
30	383	9			
31	387	9			
32	390	9			
33	394	9			
34	397	9			
35	401	9			
36	**405**	**9**			
37	409	10			
38	414	10			

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

Table 8-21
Scoring Table for Social Studies Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	240	136	39	443	9
1	240	136	40	445	9
2	240	136	41	447	8
3	240	136	42	450	8
4	240	136	43	452	8
5	240	136	**44**	**455**	**8**
6	240	136	45	457	8
7	240	136	46	459	8
8	240	136	47	462	9
9	240	136	48	464	9
10	240	136	49	467	9
11	240	136	50	469	9
12	240	136	51	472	9
13	240	136	52	475	9
14	289	87	53	478	9
15	332	44	54	480	9
16	350	30	55	483	9
17	361	24	56	487	10
18	370	20	57	490	10
19	377	18	58	493	10
20	383	16	59	497	10
21	388	15	60	501	11
22	393	14	61	505	11
23	397	13	62	510	12
24	401	12	63	516	13
25	405	12	64	522	14
26	**408**	**11**	65	530	16
27	411	11	66	539	18
28	415	10	67	551	22
29	417	10	68	569	29
30	**420**	**10**	69	600	44
31	423	10	70	620	56
32	426	9			
33	428	9			
34	431	9			
35	433	9			
36	436	9			
37	438	9			
38	440	9			

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

.

Table 8-22
Scoring Table for Science Grade 4

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	170	47	39	416	54
1	170	47	40	440	72
2	170	47			
3	170	47			
4	170	47			
5	170	47			
6	170	47			
7	170	47			
8	170	47			
9	170	47			
10	187	35			
11	205	27			
12	217	22			
13	227	18			
14	235	16			
15	241	15			
16	**247**	**13**			
17	252	13			
18	257	12			
19	262	11			
20	266	11			
21	270	10			
22	274	10			
23	**278**	**10**			
24	281	9			
25	285	9			
26	288	9			
27	292	9			
28	296	9			
29	299	9			
30	303	9			
31	307	9			
32	312	10			
33	317	10			
34	**322**	**11**			
35	329	12			
36	338	15			
37	350	19			
38	371	28			

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

Table 8-23
Scoring Table for Science Grade 8

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	230	95	39	520	48
1	230	95	40	560	82
2	230	95			
3	230	95			
4	230	95			
5	230	95			
6	230	95			
7	230	95			
8	230	95			
9	230	95			
10	278	47			
11	299	26			
12	311	20			
13	320	17			
14	328	16			
15	335	15			
16	342	14			
17	**348**	**13**			
18	353	13			
19	358	13			
20	363	12			
21	368	12			
22	373	12			
23	**377**	**11**			
24	382	11			
25	386	11			
26	391	11			
27	395	11			
28	400	11			
29	404	11			
30	409	11			
31	415	12			
32	**420**	**12**			
33	426	13			
34	433	13			
35	441	15			
36	451	17			
37	463	20			
38	482	27			

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

Table 8-24
Scoring Table for Science Grade 10

Raw Score	Scale Score	SEM	Raw Score	Scale Score	SEM
0	240	133	39	450	9
1	240	133	40	452	9
2	240	133	41	455	9
3	240	133	42	457	9
4	240	133	43	460	9
5	240	133	44	462	9
6	240	133	**45**	**465**	**9**
7	240	133	46	467	9
8	240	133	47	470	9
9	240	133	48	472	9
10	240	133	49	475	9
11	240	133	50	478	9
12	240	133	51	481	9
13	304	69	52	484	9
14	334	41	53	487	10
15	350	31	54	490	10
16	361	25	55	493	10
17	370	22	56	497	10
18	377	19	57	501	11
19	384	17	58	505	11
20	389	16	59	509	12
21	394	15	60	514	12
22	399	14	61	519	13
23	403	13	62	526	14
24	407	13	63	533	16
25	**410**	**12**	64	542	18
26	414	12	65	554	21
27	417	11	66	571	27
28	420	11	67	602	42
29	423	11	68	610	47
30	426	11			
31	**429**	**10**			
32	432	10			
33	434	10			
34	437	10			
35	440	10			
36	442	9			
37	445	9			
38	447	9			

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

Table 11-1
Cut scores and Associated Impact Data for WKCE-CRT Reading

	Score Range				Impact Data				
Grade	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
3	270-393	394-429	430-465	466-640	3.9%	13.8%	38.9%	43.4%	*82.3%*
4	280-395	396-439	440-488	489-650	4.5%	12.1%	40.4%	43.0%	*83.4%*
5	290-400	401-443	444-496	497-690	4.8%	11.3%	40.8%	43.0%	*83.4%*
6	300-417	418-456	457-513	514-730	5.2%	10.4%	41.4%	42.9%	*84.3%*
7	310-433	434-466	467-522	523-780	5.4%	9.8%	42.0%	42.8%	*84.8%*
8	330-444	445-479	480-538	539-790	5.6%	8.8%	43.4%	42.3%	*85.7%*
10	350-455	456-502	503-554	555-820	9.2%	14.1%	33.0%	43.6%	*76.6%*

Table 11-2
Cut scores and Associated Impact Data for WKCE-CRT Mathematics

	Score Range				Impact Data				
Grade	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
3	220-391	392-406	407-451	452-630	17.5%	9.5%	40.1%	32.9%	*73.0%*
4	240-420	421-437	438-483	484-650	16.3%	10.4%	40.9%	32.5%	*73.3%*
5	270-444	445-462	463-504	505-680	15.1%	11.6%	42.6%	30.7%	*74.0%*
6	310-463	464-484	485-531	532-700	13.9%	12.3%	44.9%	28.9%	*74.5%*
7	330-479	480-503	504-554	555-710	12.7%	12.7%	47.0%	27.6%	*74.6%*
8	350-482	483-512	513-572	573-730	11.6%	13.4%	49.5%	25.5%	*75.0%*
10	410-515	516-540	541-594	595-750	14.2%	12.7%	46.7%	26.4%	*73.1%*

Table 11-3
Cut scores and Associated Impact Data for WKCE-CRT Language Arts

	Score Range				Impact Data				
Grade	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
4	140-251	252-276	277-307	308-420	4.3%	14.8%	44.5%	36.4%	*80.9%*
8	250-357	358-384	385-417	418-520	11.5%	22.3%	39.9%	26.3%	*66.2%*
10	290-392	393-427	428-483	484-630	8.6%	19.0%	53.0%	19.4%	*72.4%*

Table 11-4
Cut scores and Associated Impact Data for WKCE-CRT Social Studies

	Score Range				Impact Data				
Grade	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
4	170-241	242-262	263-287	288-400	1.8%	5.2%	28.5%	64.6%	*93.1%*
8	230-333	334-363	364-402	403-530	3.9%	11.2%	40.0%	44.9%	*84.9%*
10	240-407	408-419	420-454	455-620	16.9%	6.8%	30.7%	45.6%	*76.3%*

Table 11-5
Cut scores and Associated Impact Data for WKCE-CRT Science

	Score Range				Impact Data				
Grade	Minimal	Basic	Proficient	Advanced	Minimal	Basic	Proficient	Advanced	Proficient +Advanced
4	170-248	249-278	279-319	320-440	4.8%	15.7%	57.6%	21.9%	*79.5%*
8	230-348	349-374	375-418	419-560	8.9%	15.7%	46.8%	28.6%	*75.4%*
10	240-410	411-428	429-465	466-610	16.6%	10.6%	35.5%	37.2%	*72.8%*

Table 11-6
A Crosswalk Table for Reading Grade 4 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
433-557	1	280-338	642	45	476
558-569	2	339-369	643	46	477
570-577	3	370-384	644	47	478
578-583	4	385-394	645	48	479
584-588	5	395-402	646	49	480
589-592	6	403-408	646	50	481
593-595	7	409-413	647	51	482
596-598	8	414-417	648	52	483
599-601	9	418-421	649	53	484
602-603	10	422-424	650	54	485
604-605	11	425-427	650	55	486
606-607	12	428-430	651	56	487
608-609	13	431-432	652	57	488
610-611	14	433-434	653	58	489
612	15	435-436	654	59	490
613-614	16	437-438	654	60	491
615	17	439-440	655	61	492
616-617	18	441-442	656	62	493
618	19	443-444	657	63	494
619	20	445	658	64	495
620	21	446-447	659	65	496
621	22	448	659	66	497
622	23	449-450	660	67	498
623-624	24	451	661	68	499
625	25	452-453	662	69	500
626	26	454	663	70	501
627	27	455	664	71	502
628	28	456-457	665	72	503
629	29	458	666	73	504
630	30	459	667	74	505-506
630	31	460	668	75	507
631	32	461	669	76	508
632	33	462-463	670	77	509
633	34	464	671	78	510
634	35	465	672	79	511
635	36	466	673	80	512-513
636	37	467	674	81	514
637	38	468	675	82	515
638	39	469-470	676-677	83	516
638	40	471	678	84	517-518
639	41	472	679	85	519-520
640	42	473	680-681	86	521
641	43	474	682	87	522-523
642	44	475	683-684	88	524

Table 11-6
A Crosswalk Table for Reading Grade 4 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
685-686	89	525-526
687-688	90	527-528
689-690	91	529-530
691-692	92	531-533
693-696	93	534-535
697-698	94	536-538
699-702	95	539-541
703-707	96	542-545
708-714	97	546-551
715-723	98	552-558
724-780	99	559-650

Table 11-7
A Crosswalk Table for Reading Grade 8 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
507-604	1	330-373	687	45	525
605-614	2	374-408	687	46	526
615-621	3	409-424	688	47	527
622-627	4	425-435	689	48	528
628-631	5	436-443	690	49	529
632-634	6	444-450	691	50	530
635-637	7	451-455	692	51	531
638-640	8	456-459	692	52	532
641-643	9	460-463	693	53	533
644-645	10	464-467	694	54	534
646-647	11	468-471	695	55	535
648-649	12	472-474	696	56	536
650-651	13	475-476	696	57	537
652-653	14	477-479	697	58	538
654	15	480-481	698	59	539
655-656	16	482-484	699	60	540
657-658	17	485-486	700	61	541-542
659	18	487-488	700	62	543
660-661	19	489-490	701	63	544
662	20	491-492	702	64	545
663	21	493	703	65	546
664	22	494-495	704	66	547
665-666	23	496-497	705	67	548
667	24	498	705	68	549
668	25	499-500	706	69	550
669	26	501	707	70	551-552
670	27	502-503	708	71	553
671	28	504	709	72	554
672	29	505-506	710	73	555
673	30	507	711	74	556
674	31	508	712	75	557
675	32	509-510	713	76	558-559
676	33	511	714	77	560
677	34	512	715	78	561-562
678	35	513	716	79	563
679	36	514	717	80	564
680	37	515-516	718	81	565-566
681	38	517	719	82	567
682	39	518	720	83	568-569
683	40	519	721-722	84	570-571
683	41	520	723	85	572
684	42	521	724	86	573-574
685	43	522	725-726	87	575-576
686	44	523-524	727	88	577-578

Table 11-7
A Crosswalk Table for Reading Grade 8 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
728-729	89	579-581
730-731	90	582-583
732-733	91	584-585
734-735	92	586-588
736-737	93	589-591
738-740	94	592-595
741-744	95	596-598
745-748	96	599-603
749-754	97	604-609
755-764	98	610-618
765-820	99	619-790

Table 11-8
A Crosswalk Table for Reading Grade 10 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
529-638	1	350	711	45	538-539
639-652	2	351-378	712	46	540
653-658	3	379-403	713	47	541-542
659-663	4	404-418	714	48	543
664-666	5	419-429	715	49	544
667-669	6	430-437	716	50	545-546
670-671	7	438-444	716	51	547
672-673	8	445-451	717	52	548
674-675	9	452-456	718	53	549-550
676-677	10	457-461	719	54	551
678	11	462-466	720	55	552
679	12	467-470	720	56	553-554
680-681	13	471-474	721	57	555
682	14	475-477	722	58	556
683	15	478-481	723	59	557-558
684-685	16	482-484	724	60	559
686	17	485-487	725	61	560-561
687	18	488-489	726	62	562
688	19	490-492	726	63	563
689	20	493-495	727	64	564-565
690	21	496-497	728	65	566
691	22	498-500	729	66	567
692	23	501-502	730	67	568-569
693	24	503-504	731	68	570
694	25	505-506	732	69	571-572
695	26	507-508	733	70	573
696	27	509-510	734	71	574-575
697	28	511-512	735	72	576
698	29	513-514	736	73	577-578
699	30	515-516	737	74	579
700	31	517-518	738	75	580-581
701	32	519	739	76	582-583
701	33	520-521	740	77	584
702	34	522-523	741	78	585-586
703	35	524	742-744	79	587-588
704	36	525-526	745	80	589-590
705	37	527	746-747	81	591-592
706	38	528-529	748	82	593-594
706	39	530	749	83	595-596
707	40	531-532	750	84	597-598
708	41	533	751	85	599-600
709	42	534	752-753	86	601-602
710	43	535-536	754	87	603-605
710	44	537	755	88	606-608

Table 11-8
A Crosswalk Table for Reading Grade 10 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
756	89	609-610
757	90	611-613
758-776	91	614-617
777	92	618-620
777	93	621-624
778-837	94	625-629
838	95	630-634
838	96	635-640
838	97	641-648
838	98	649-661
838	99	662-820

Table 11-9
A Crosswalk Table for Mathematics Grade 4 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
403-556	1	240-356	629	45	458
557-565	2	357-369	630	46	459-460
566-571	3	370-377	631	47	461
572-576	4	378-383	632	48	462
577-579	5	384-388	633	49	463
580-582	6	389-393	633	50	464
583-585	7	394-397	634	51	465
586-587	8	398-400	635	52	466
588-590	9	401-403	636	53	467
591	10	404-406	636	54	468
592-593	11	407-409	637	55	469
594-595	12	410-411	638	56	470
596-597	13	412-413	639	57	471-472
598	14	414-416	640	58	473
599	15	417-418	640	59	474
600-601	16	419-420	641	60	475
602	17	421-422	642	61	476
603	18	423-424	643	62	477
604	19	425	644	63	478
605-606	20	426-427	645	64	479
607	21	428-429	645	65	480
608	22	430	646	66	481
609	23	431-432	647	67	482
610	24	433	648	68	483-484
611	25	434-435	649	69	485
612	26	436	650	70	486
613	27	437-438	651	71	487
614	28	439	652	72	488
615	29	440	653	73	489-490
616	30	441	653	74	491
617	31	442-443	654	75	492
618	32	444	655	76	493
619	33	445	656-657	77	494-495
620	34	446	658	78	496
621	35	447	659	79	497-498
622	36	448-449	660	80	499
623	37	450	661	81	500-501
623	38	451	662	82	502
624	39	452	663-664	83	503-504
625	40	453	665	84	505-506
626	41	454	666-667	85	507
627	42	455	668	86	508-509
628	43	456	669-670	87	510-511
628	44	457	671	88	512-513

Table 11-9
A Crosswalk Table for Mathematics Grade 4 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
672-673	89	514-515
674-675	90	516-518
676-677	91	519-520
678-679	92	521-523
680-681	93	524-526
682-684	94	527-530
685-686	95	531-535
687-690	96	536-540
691-695	97	541-547
696-702	98	548-556
703-770	99	557-650

Table 11-10
A Crosswalk Table for Mathematics Grade 8 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
502-609	1	350-410	701	45	538
610-621	2	411-429	702	46	539-540
622-628	3	430-441	703	47	541
629-634	4	442-449	704	48	542
635-639	5	450-456	705	49	543
640-643	6	457-462	705	50	544
644-646	7	463-467	706	51	545
647-649	8	468-471	707	52	546
650-652	9	472-474	708	53	547
653-655	10	475-478	709	54	548
656-657	11	479-481	710	55	549
658-659	12	482-484	711	56	550-551
660-662	13	485-487	712	57	552
663-664	14	488-490	713	58	553
665-666	15	491-492	713	59	554
667	16	493-495	714	60	555
668-669	17	496-497	715	61	556
670-671	18	498-499	716	62	557
672	19	500-501	717	63	558
673-674	20	502-503	718	64	559
675	21	504-505	719	65	560-561
676-677	22	506-507	720	66	562
678	23	508-509	721	67	563
679	24	510-511	722	68	564
680	25	512	723	69	565
681-682	26	513-514	724	70	566-567
683	27	515-516	725	71	568
684	28	517	726	72	569
685	29	518-519	727	73	570
686	30	520	728	74	571-572
687	31	521	729	75	573
688	32	522-523	730	76	574
689	33	524	731	77	575-576
690-691	34	525	732-733	78	577
692	35	526	734	79	578
692	36	527-528	735	80	579-580
693	37	529	736	81	581
694	38	530	737-738	82	582-583
695	39	531	739	83	584-585
696	40	532-533	740-741	84	586
697	41	534	742	85	587-588
698	42	535	743-744	86	589-590
699	43	536	745	87	591-592
700	44	537	746-747	88	593-594

Table 11-10
A Crosswalk Table for Mathematics Grade 8 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
748-749	89	595-596
750-751	90	597-598
752-754	91	599-600
755-756	92	601-603
757-759	93	604-605
760-763	94	606-609
764-767	95	610-612
768-771	96	613-616
772-779	97	617-622
780-788	98	623-629
789-872	99	630-730

Table 11-11
A Crosswalk Table for Mathematics Grade 10 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
530-631	1	410	741	45	563
632-647	2	410	742	46	564
648-658	3	411-446	743	47	565
659-665	4	447-467	744	48	566
666-671	5	468-478	745	49	567
672-676	6	479-486	746	50	568
677-680	7	487-491	746	51	569
681-684	8	492-496	747	52	570
685-687	9	497-500	748	53	571
688-691	10	501-504	749	54	572
692-693	11	505-507	750	55	573
694-696	12	508-510	751	56	574
697-698	13	511-513	752	57	575
699-700	14	514-515	752	58	576-577
701-702	15	516-518	753	59	578
703-704	16	519-520	754	60	579
705-706	17	521-522	755	61	580
707-708	18	523-524	756	62	581
709-710	19	525-527	757	63	582
711-712	20	528-529	758	64	583
713	21	530-531	759	65	584
714-715	22	532	759	66	585
716-717	23	533-534	760	67	586
718	24	535-536	761	68	587
719	25	537	762	69	588-589
720-721	26	538-539	763	70	590
722	27	540	764	71	591
723	28	541-542	765-766	72	592
724	29	543	767	73	593
725-726	30	544-545	768	74	594-595
727	31	546	769	75	596
728	32	547	770	76	597
729	33	548-549	771	77	598-599
730	34	550	772-773	78	600
731	35	551	774	79	601
732	36	552-553	775	80	602-603
733	37	554	776-777	81	604
734	38	555	778	82	605-606
735	39	556	779-780	83	607-608
736	40	557	781	84	609
737	41	558	782-783	85	610-611
738	42	559	784-785	86	612-613
739	43	560-561	786-787	87	614
740	44	562	788-789	88	615-616

Table 11-11
A Crosswalk Table for Mathematics Grade 10 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
790-792	89	617-618
793-794	90	619-621
795-796	91	622-623
797-800	92	624-625
801-803	93	626-628
804-807	94	629-632
808-812	95	633-635
813-817	96	636-640
818-827	97	641-645
828-849	98	646-654
850-892	99	655-750

Table 11-12
A Crosswalk Table for Language Arts Grade 4 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
465-569	1	140-218	643	45	295
570-579	2	219-238	644	46	296
580-586	3	239-246	645	47	297
587-590	4	247-251	645	48	297
591-594	5	252-255	646	49	298
595-597	6	256-258	647	50	299
598-600	7	259-260	647	51	299
601-602	8	261-262	648	52	300
603-605	9	263-264	649	53	300
606-607	10	265-266	650	54	301
608	11	267	650	55	302
609-610	12	268	651	56	302
611-612	13	269-270	652	57	303
613	14	271	653	58	304
614	15	272	653	59	304
615-616	16	273	654	60	305
617	17	274	655	61	306
618	18	275	656	62	306
619-620	19	276	656	63	307
621	20	277	657	64	307
622	21	278	658	65	308
623	22	279	659	66	309
624	23	280	659	67	309
625	24	280	660	68	310
626	25	281	661	69	311
627	26	282	662	70	312
628	27	283	663	71	312
629	28	284	663	72	313
630	29	284	664	73	314
631	30	285	665	74	315
632	31	286	666	75	315
633	32	287	667	76	316
633	33	287	668	77	317
634	34	288	669	78	318
635	35	289	670	79	318
636	36	289	671	80	319
637	37	290	672	81	320
638	38	291	673	82	321
638	39	291	674	83	322
639	40	292	675	84	323
640	41	293	676	85	324
641	42	293	677-678	86	325
641	43	294	679	87	326
642	44	295	680-681	88	327-328

Table 11-12
A Crosswalk Table for Language Arts Grade 4 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
682	89	329
683-684	90	330-331
685-686	91	332
687-688	92	333-334
689-690	93	335-336
691-693	94	337-339
694-696	95	340-342
697-701	96	343-346
702-706	97	347-353
707-715	98	354-370
716-757	99	371-420

Table 11-13
A Crosswalk Table for Language Arts Grade 8 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
523-606	1	250-306	676	45	393
607-615	2	307-321	677	46	394
616-621	3	322-330	677	47	395
622-625	4	331-336	678	48	396
626-629	5	337-340	679	49	397
630-631	6	341-344	680	50	397
632-634	7	345-347	681	51	398
635-636	8	348-350	681	52	399
637-638	9	351-352	682	53	400
639-640	10	353-354	683	54	400
641-642	11	355-356	684	55	401
643	12	357-358	685	56	402
644-645	13	359-360	685	57	403
646	14	361-362	686	58	404
647-648	15	363	687	59	404
649	16	364-365	688	60	405
650	17	366	689	61	406
651	18	367-368	690	62	407
652	19	369	690	63	408
653-654	20	370	691	64	408
655	21	371	692	65	409
656	22	372	693	66	410
657	23	373-374	694	67	411
658	24	375	695	68	412
659	25	376	696	69	413
660	26	377	696	70	414
660	27	378	697	71	415
661	28	379	698	72	415
662	29	380	699	73	416
663	30	381	700	74	417
664	31	382	701	75	418
665	32	383	702	76	419
666	33	383	703	77	420
667	34	384	704	78	421
668	35	385	705-706	79	422
668	36	386	707	80	423
669	37	387	708	81	424
670	38	388	709	82	425-426
671	39	388	710	83	427
672	40	389	711-712	84	428
673	41	390	713	85	429-430
673	42	391	714-715	86	431
674	43	392	716	87	432
675	44	392	717-718	88	433-434

Table 11-13
A Crosswalk Table for Language Arts Grade 8 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
719-720	89	435
721	90	436-438
722-724	91	439-440
725-726	92	441-442
727-728	93	443-445
729-731	94	446-448
732-735	95	449-452
736-739	96	453-464
740-745	97	465
746-755	98	466-478
756-819	99	479-520

Table 11-14
A Crosswalk Table for Language Arts Grade 10 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
535-643	1	290-354	709	45	446
644-650	2	355-367	710	46	447
651-656	3	368-374	711	47	448
657-659	4	375-379	711	48	449
660-663	5	380-382	712	49	450
664-665	6	383-386	713	50	451
666-668	7	387-389	714	51	452
669-670	8	390-392	714	52	453
671-672	9	393-395	715	53	454
673	10	396-397	716	54	455
674-675	11	398-400	717	55	456
676	12	401-402	717	56	457
677-678	13	403-404	718	57	458
679	14	405-406	719	58	459
680-681	15	407-408	720	59	460
682	16	409-410	720	60	461
683	17	411-412	721	61	462
684	18	413	722	62	463
685-686	19	414-415	723	63	464
687	20	416-417	724	64	465
688	21	418	724	65	466
689	22	419-420	725	66	467
690	23	421	726	67	468
691	24	422-423	727	68	469
692	25	424	728	69	470
693	26	425	728	70	471
694	27	426-427	729	71	472
695	28	428	730	72	473
696	29	429	731	73	474
697	30	430	732	74	475
697	31	431-432	733	75	476
698	32	433	734	76	477-478
699	33	434	735	77	479
700	34	435	736	78	480
701	35	436	737	79	481
702	36	437	738	80	482-483
703	37	438	739	81	484
703	38	439	740	82	485
704	39	440	741	83	486-487
705	40	441	742	84	488
706	41	442	743-744	85	489-490
707	42	443-444	745-746	86	491
707	43	445	747	87	492-493
708	44	445	748	88	494-495

Table 11-14
A Crosswalk Table for Language Arts Grade 10 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
749-750	89	496-497
751	90	498-499
752-754	91	500-501
755-756	92	502-504
757-758	93	505-506
759-761	94	507-509
762	95	510-513
763-767	96	514-517
768-775	97	518-522
776-785	98	523-530
786-835	99	531-630

Table 11-15
A Crosswalk Table for Social Studies Grade 4 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
460-588	1	170-237	644	45	292
589-595	2	238-246	644	46	293
596-599	3	247-251	645	47	293
600-603	4	252-255	646	48	294
604-605	5	256-258	646	49	294
606-608	6	259-261	647	50	295
609-610	7	262	647	51	295
611-612	8	263-264	648	52	296
613-614	9	265-266	648	53	296
615	10	267	649	54	297
616-617	11	268	650	55	297
618	12	269	650	56	298
619	13	270-271	651	57	298
620-621	14	272	651	58	299
622	15	273	652	59	299
623	16	274	652	60	300
624	17	275	653	61	301
625	18	275	654	62	301
626	19	276	654	63	302
627	20	277	655	64	302
628	21	278	655	65	303
628	22	279	656	66	303
629	23	279	657	67	304
630	24	280	657	68	305
631	25	281	658	69	305
632	26	281	658	70	306
632	27	282	659	71	307
633	28	283	660	72	307
634	29	284	660	73	308
634	30	284	661	74	309
635	31	285	662	75	309
636	32	285	662	76	310
636	33	286	663	77	311
637	34	286	664	78	312-313
638	35	287	665	79	313
638	36	287	665	80	314
639	37	288	666	81	315
640	38	289	667	82	316
640	39	289	668	83	317-318
641	40	290	669	84	318
641	41	290	670	85	319
642	42	291	671	86	320
643	43	291	672	87	321-322
643	44	292	673	88	323-324

Table 11-15
A Crosswalk Table for Social Studies Grade 4 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
674-675	89	325-327
676	90	328
677-678	91	329-331
679	92	332
680-681	93	333
682-684	94	334-341
685-687	95	342-358
688-690	96	359-371
691-695	97	372-399
696-702	98	400
703-763	99	400

Table 11-16
A Crosswalk Table for Social Studies Grade 8 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
523-622	1	230-301	686	45	394
623-630	2	302-321	687	46	395
631-635	3	322-330	688	47	396
636-639	4	331-336	688	48	396
640-642	5	337-340	689	49	397
643-645	6	341-344	689	50	398
646-647	7	345-347	690	51	399
648-650	8	348-350	690	52	400
651-652	9	351-352	691	53	400
653	10	353-354	691	54	401
654-655	11	355-356	692	55	402
656	12	357-358	693	56	403
657-658	13	359-360	693	57	404
659	14	361-362	694	58	404
660-661	15	363	694	59	405
662	16	364-365	695	60	406
663	17	366	695	61	407
664	18	367	696	62	408
665-666	19	368-369	697	63	409
667	20	370	697	64	410
668	21	371	698	65	410
669	22	372	698	66	411
670	23	373	699	67	412
671	24	374	699	68	413
672	25	375	700	69	414
673	26	376-377	701	70	415
674	27	378	701	71	416
674	28	379	702	72	417
675	29	380	702	73	418
676	30	381	703	74	419
677	31	382	704	75	420
678	32	383	704	76	421
678	33	383	705	77	422
679	34	384	706	78	423
680	35	385	706	79	424
680	36	386	707	80	425
681	37	387	708	81	426-427
682	38	388	709	82	428
683	39	389	709	83	429
683	40	390	710	84	430-431
684	41	391	711	85	432
684	42	391	712	86	433-434
685	43	392	713	87	435
686	44	393	714	88	436-437

Table 11-16
A Crosswalk Table for Social Studies Grade 8 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
715	89	438-439
716	90	440
717-718	91	441-443
719	92	444-446
720-721	93	447-448
722-723	94	449-452
724-725	95	453-456
726-728	96	457-463
729-731	97	464-472
732-737	98	473-493
738-803	99	494-530

Table 11-17
A Crosswalk Table for Social Studies Grade 10 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
548-652	1	240	707	45	445
653-659	2	241-344	708	46	446
660-664	3	345-357	708	47	447
665-668	4	358-366	709	48	448
669-670	5	367-372	710	49	449
671-673	6	373-377	710	50	450
674-675	7	378-381	711	51	451
676-677	8	382-385	711	52	452
678-679	9	386-388	712	53	453
680	10	389-391	712	54	454
681	11	392-394	713	55	455
682-683	12	395-396	714	56	456
684	13	397-399	714	57	456
685	14	400-401	715	58	457
686	15	402-404	715	59	458
687	16	405-406	716	60	459
688	17	407-408	716	61	460
689	18	409-410	717	62	461
690	19	411	718	63	462
691	20	412-413	718	64	463
692	21	414-415	719	65	464
693	22	416-417	719	66	465
693	23	418	720	67	466
694	24	419-420	721	68	467
695	25	421	721	69	468
696	26	422-423	722	70	469
696	27	424	723	71	470
697	28	425-426	723	72	471
698	29	427	724	73	472
698	30	428	725	74	473-474
699	31	429-430	725	75	475
700	32	431	726	76	476
700	33	432	727	77	477
701	34	433	727	78	478
702	35	434	728	79	479
702	36	435-436	729	80	480-481
703	37	437	730	81	482
703	38	438	731	82	483
704	39	439	731	83	484-485
704	40	440	732	84	486
705	41	441	733	85	487-488
706	42	442	734	86	489
706	43	443	735	87	490-491
707	44	444	736	88	492-493

Table 11-17
A Crosswalk Table for Social Studies Grade 10 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
737-738	89	494-495
739	90	496-497
740-741	91	498-499
742	92	500-502
743-744	93	503-504
745-747	94	505-508
748-749	95	509-511
750-752	96	512-516
753-757	97	517-522
758-764	98	523-530
765-821	99	531-620

Table 11-18
A Crosswalk Table for Science Grade 4 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
421-553	1	170-225	636	45	297
554-564	2	226-234	637	46	298
565-570	3	235-241	638	47	299
571-576	4	242-246	639	48	299
577-580	5	247-251	639	49	300
581-584	6	252-254	640	50	301
585-587	7	255-257	641	51	301
588-591	8	258-259	642	52	302
592-593	9	260-261	642	53	302
594-595	10	262-263	643	54	303
596-597	11	264-265	644	55	304
598-599	12	266-267	645	56	304
600-601	13	268	645	57	305
602-603	14	269-270	646	58	305
604-605	15	271	647	59	306
606	16	272-273	647	60	307
607-608	17	274	648	61	307
609	18	275	649	62	308
610-611	19	276	650	63	309
612	20	277	650	64	309
613	21	278-279	651	65	310
614-615	22	280	652	66	310
616	23	281	653	67	311
617	24	282	653	68	312
618	25	282	654	69	312
619	26	283	655	70	313
620	27	284	656	71	314
621	28	285	656	72	314
622	29	286	657	73	315
623	30	287	658	74	316
624	31	288	659	75	317
625	32	288	660	76	317
626	33	289	660	77	318
627	34	290	661	78	319
628	35	291	662	79	320
629	36	291	663	80	321
630	37	292	664	81	322
631	38	293	665	82	322
632	39	293	666	83	323
632	40	294	667	84	324-325
633	41	295	668	85	326
634	42	295	669	86	327
635	43	296	670	87	328
635	44	297	671	88	329

Table 11-18
A Crosswalk Table for Science Grade 4 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
672-673	89	330-331
674	90	332
675	91	333-334
676-677	92	335-336
678-679	93	337-339
680-681	94	340-341
682-683	95	342-345
684-685	96	346-349
686-689	97	350-354
690-694	98	355-366
695-799	99	367-440

Table 11-19
A Crosswalk Table for Science Grade 8 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
487-619	1	230-308	697	45	394
620-631	2	309-319	698	46	395
632-638	3	320-327	698	47	396
639-644	4	328-332	699	48	397
645-648	5	333-337	700	49	398
649-651	6	338-340	700	50	399
652-654	7	341-343	701	51	400
655-657	8	344-346	702	52	401
658-659	9	347-349	702	53	402
660-661	10	350-351	703	54	402
662-663	11	352-353	704	55	403
664-665	12	354-355	704	56	404
666-667	13	356-357	705	57	405
668	14	358-359	706	58	406
669-670	15	360-361	706	59	407
671	16	362	707	60	408
672	17	363-364	708	61	409
673-674	18	365	708	62	409
675	19	366-367	709	63	410
676	20	368	710	64	411
677	21	369-370	711	65	412
678	22	371	711	66	413
679	23	372	712	67	414
680	24	373	713	68	415
681	25	374-375	714	69	416
682	26	376	714	70	417
683	27	377	715	71	418
684	28	378	716	72	419
685	29	379	717	73	420
686	30	380	717	74	421
687	31	381	718	75	422
687	32	382	719	76	423
688	33	383	720	77	424
689	34	384	721	78	425
690	35	385	722	79	426
690	36	386	723	80	427
691	37	387	724	81	428-429
692	38	388	725	82	430
693	39	389	726	83	431
693	40	390	727	84	432-433
694	41	391	728	85	434
695	42	392	729	86	435-436
695	43	393	730	87	437
696	44	394	731-732	88	438-439

Table 11-19
A Crosswalk Table for Science Grade 8 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
733	89	440-441
734-735	90	442-443
736	91	444-445
737-738	92	446-448
739-740	93	449-451
741-743	94	452-454
744-745	95	455-458
746-748	96	459-464
749-752	97	465-471
753-758	98	472-484
759-857	99	485-560

Table 11-20
A Crosswalk Table for Science Grade 10 Based on State Percentile

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT	Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
501-628	1	240	725	45	448
629-646	2	241-330	726	46	449
647-657	3	331-351	726	47	450
658-664	4	352-362	727	48	451
665-669	5	363-371	728	49	452
670-673	6	372-377	729	50	453
674-676	7	378-382	730	51	454
677-680	8	383-387	730	52	455
681-682	9	388-390	731	53	456
683-684	10	391-394	732	54	457
685-687	11	395-397	733	55	458
688-689	12	398-400	733	56	459
690	13	401-402	734	57	460
691-692	14	403-405	735	58	461
693-694	15	406-407	736	59	462
695	16	408-409	737	60	462
696-697	17	410-411	737	61	463
698	18	412-413	738	62	464
699	19	414-415	739	63	465
700-701	20	416-417	739	64	466
702	21	418-419	740	65	467
703	22	420	741	66	468
704	23	421-422	742	67	469
705	24	423	743	68	470
706	25	424-425	743	69	471
707-708	26	426	744	70	472
709	27	427-428	745	71	473
710	28	429	746	72	474
711	29	430-431	746	73	475
712	30	432	747	74	476
713	31	433	748	75	477
714	32	434	749	76	478
715	33	435-436	750	77	479
715	34	437	751	78	480-481
716	35	438	751	79	482
717	36	439	752	80	483
718	37	440	753	81	484
719	38	441	754	82	485-486
720	39	442	755	83	487
721	40	443	756-757	84	488
722	41	444	758	85	489-490
722	42	445	759	86	491
723	43	446	760	87	492-493
724	44	447	761	88	494-495

Table 11-20
A Crosswalk Table for Science Grade 10 Based on State Percentile (Cont.)

Fall 2004 WKCE	State Percentile	Fall 2005 WKCE-CRT
762-763	89	496
764	90	497-498
765-766	91	499-500
767-768	92	501-503
769-770	93	504-505
771-773	94	506-508
774-775	95	509-511
776-779	96	512-516
780-783	97	517-521
784-790	98	522-529
791-893	99	530-610

Figure 8-1
Non-converged Item: Reading Grade 7 Form A/B/C FT #82

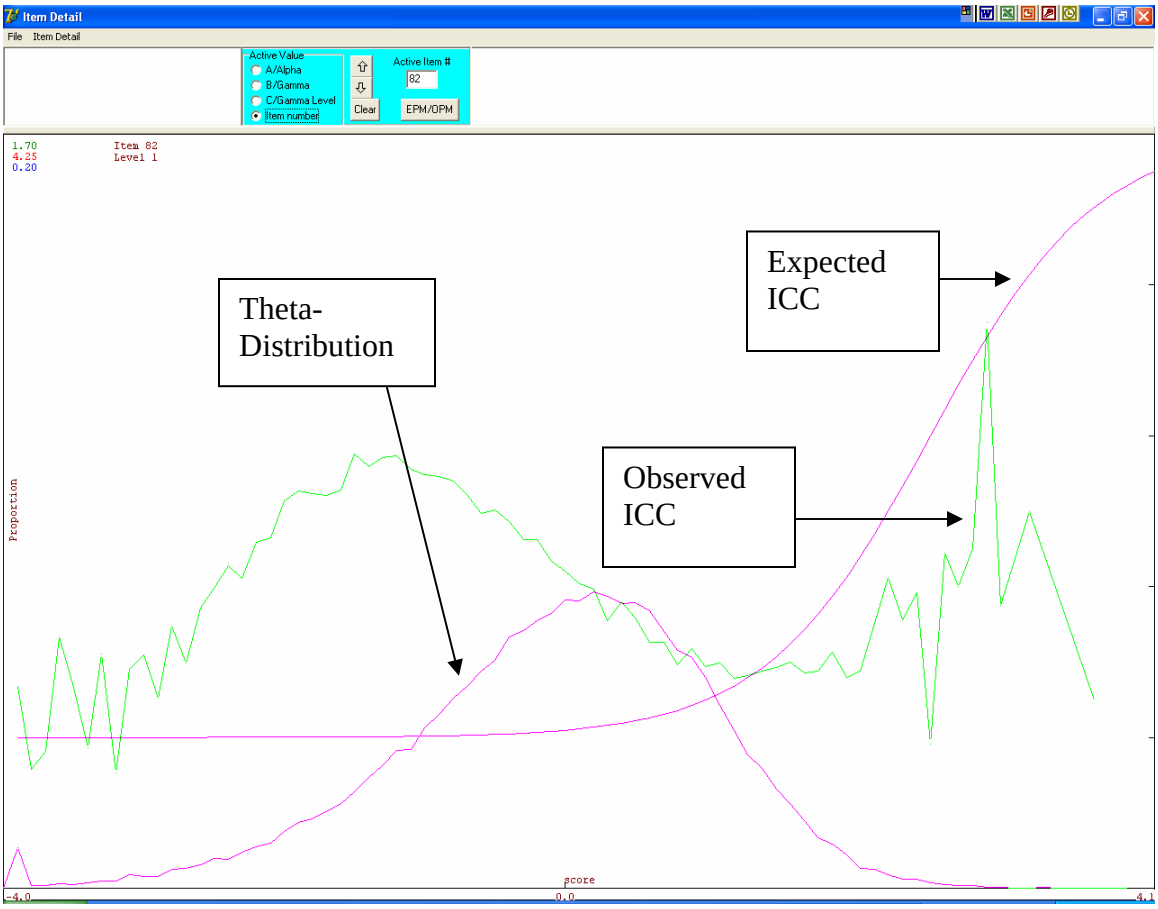


Figure 8-2
Non-converged Item: Mathematics Grade 7 Form B FT #73

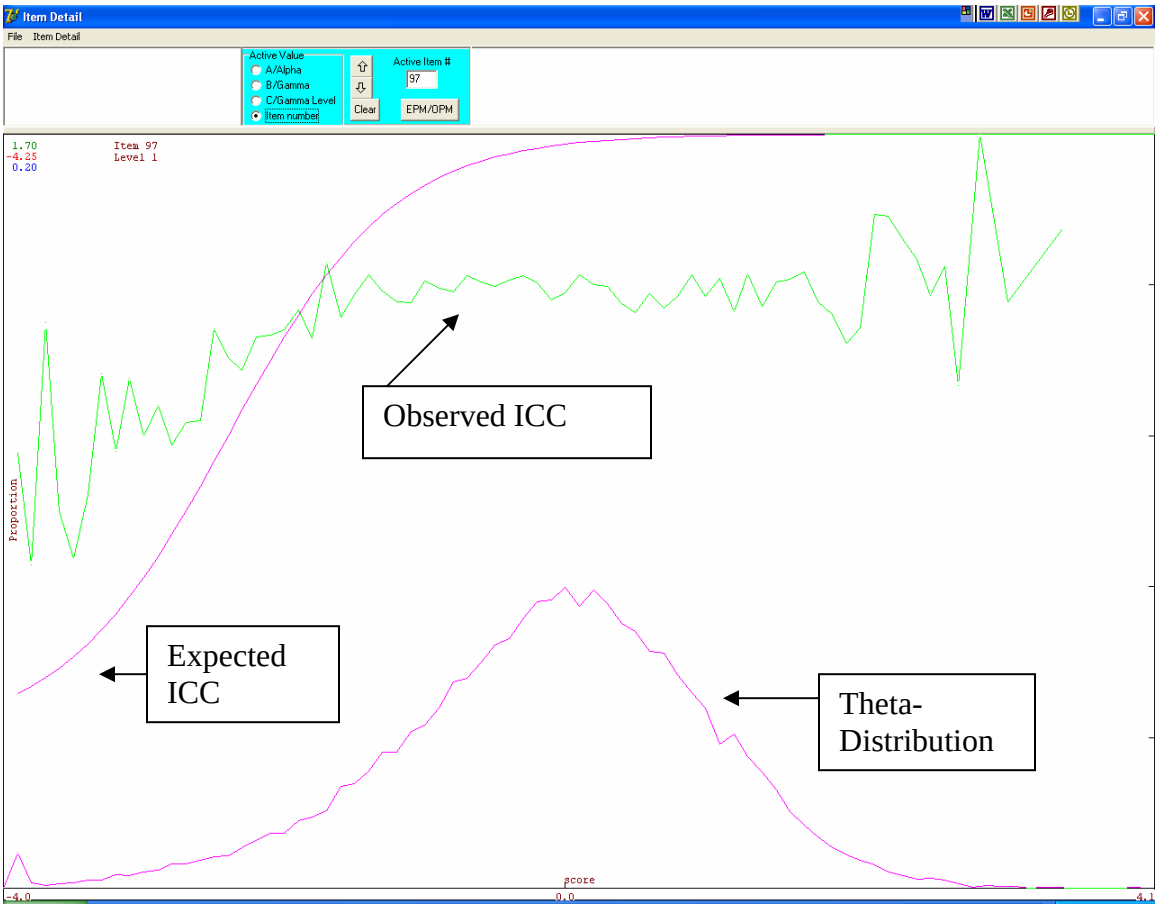


Figure 8-3
Non-converged Item: Language Arts Grade 8 item #32

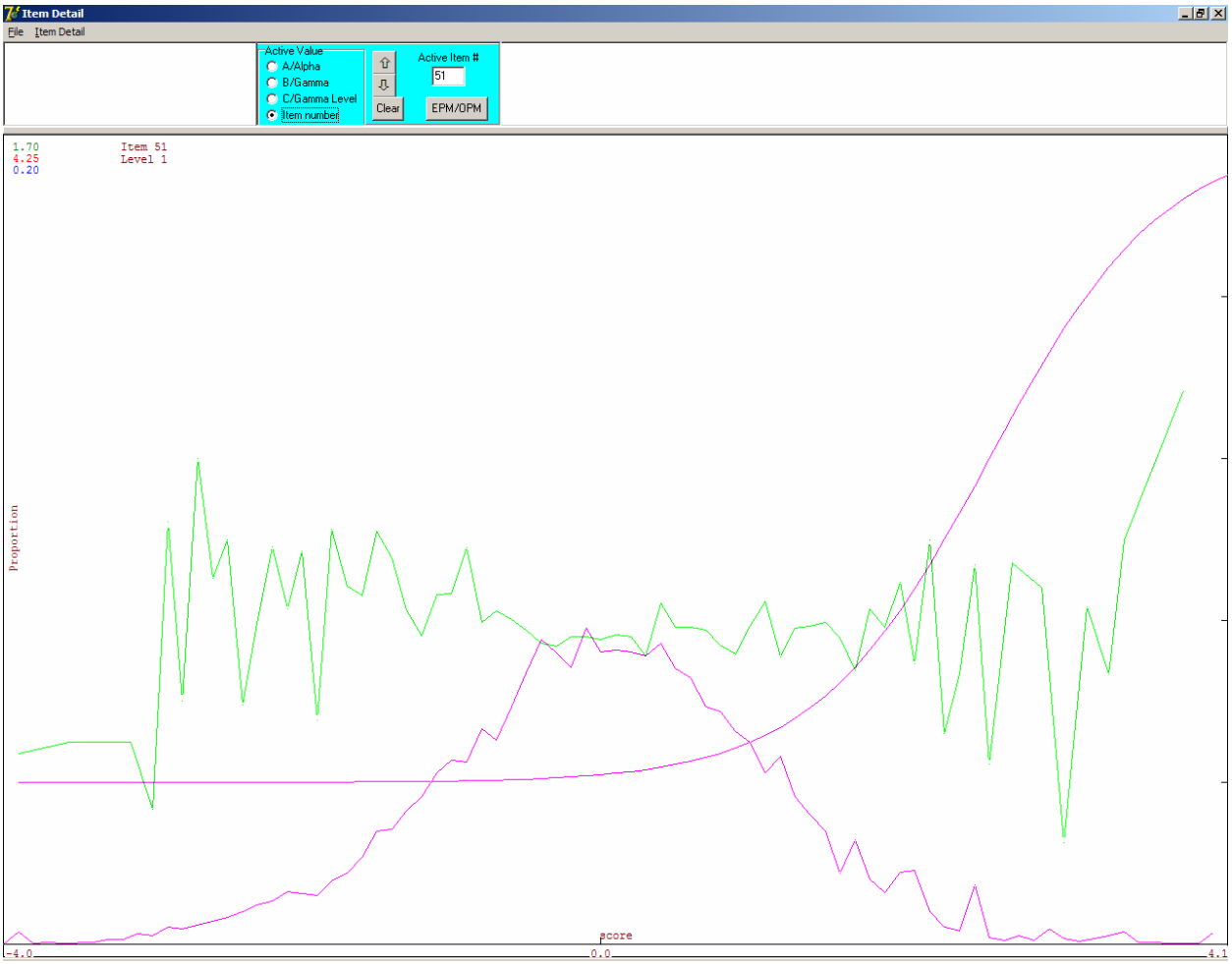


Figure 8-4
TCC Curve for Reading Grades 3-8, 10

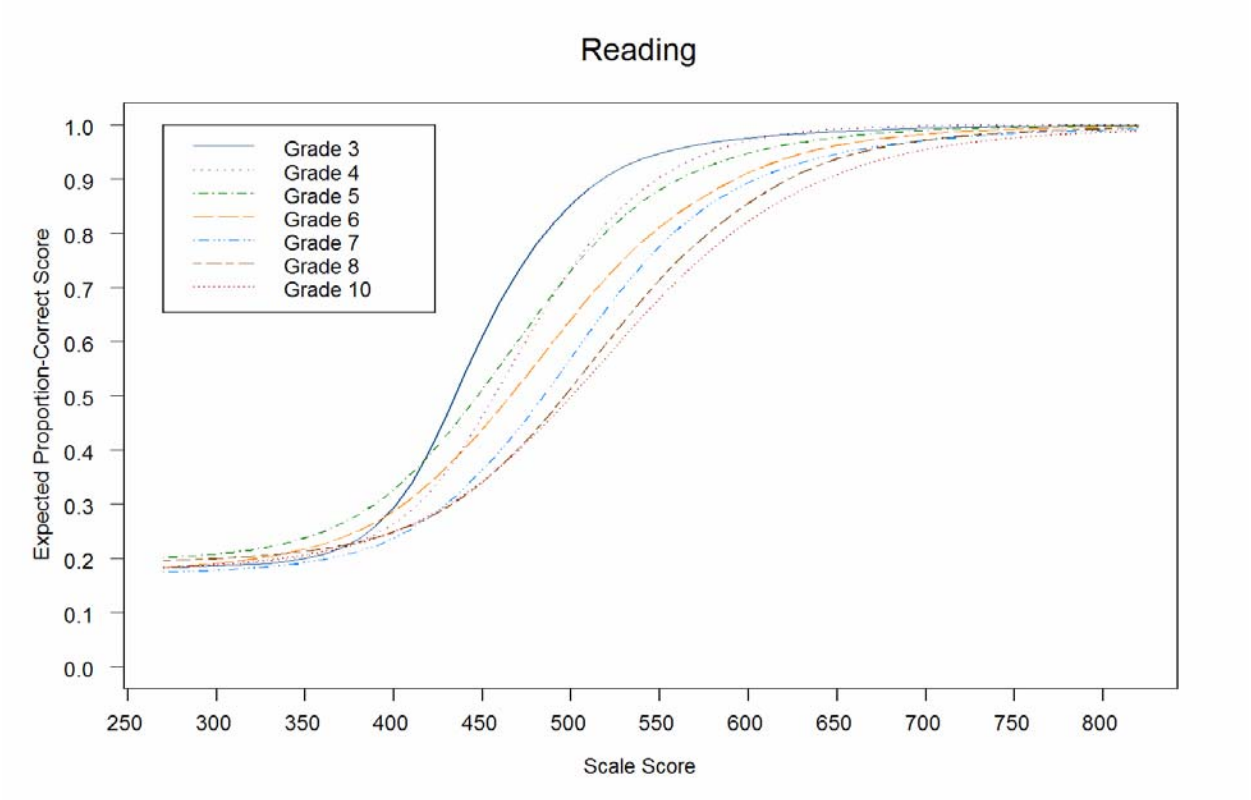


Figure 8-5
TCC Curve for Mathematics Grades 3-8, 10

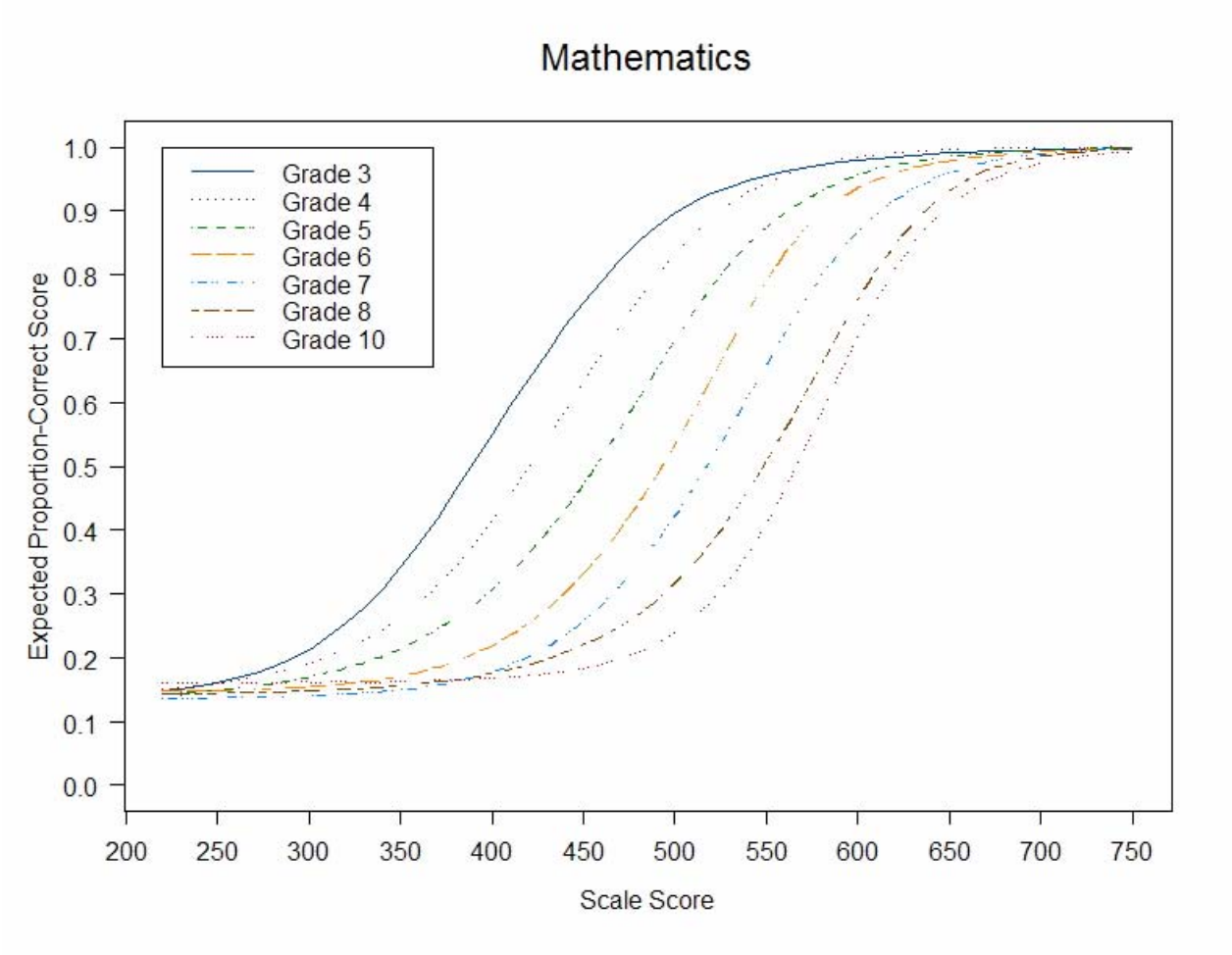


Figure 8-6
TCC Curve for Language Arts Grades 4, 8, 10

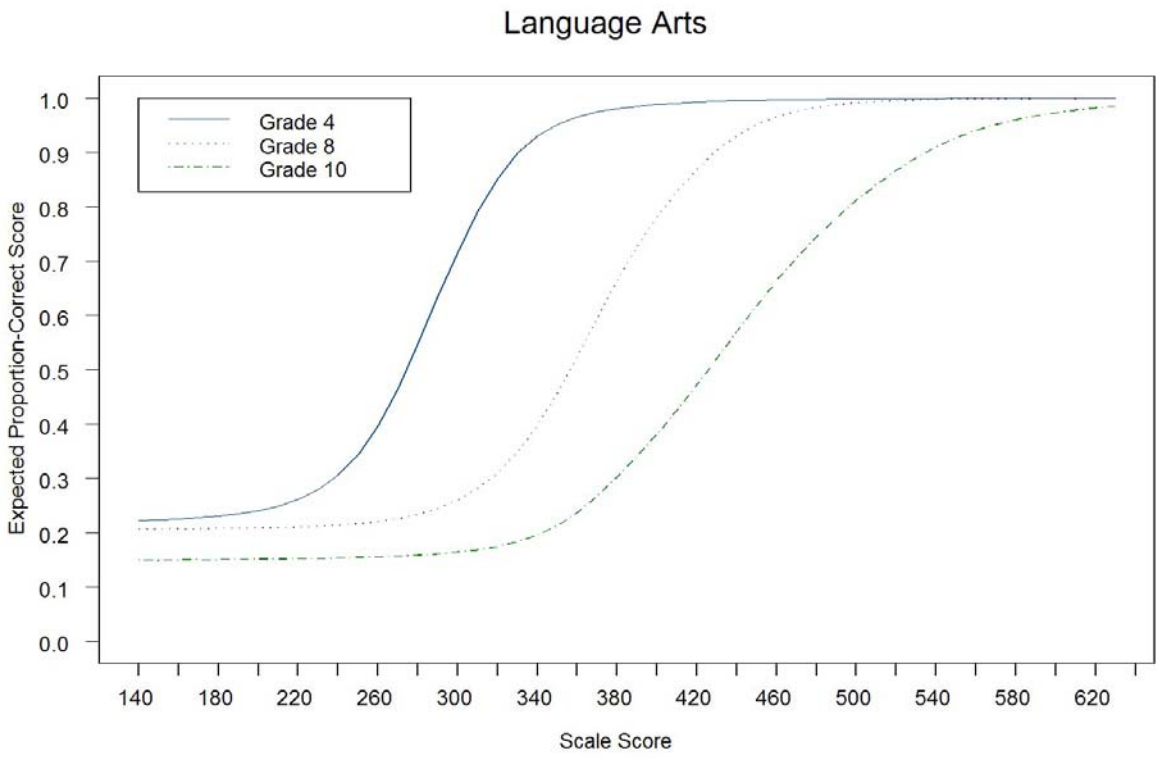


Figure 8-7
TCC Curve for Social Studies Grades 4, 8, 10

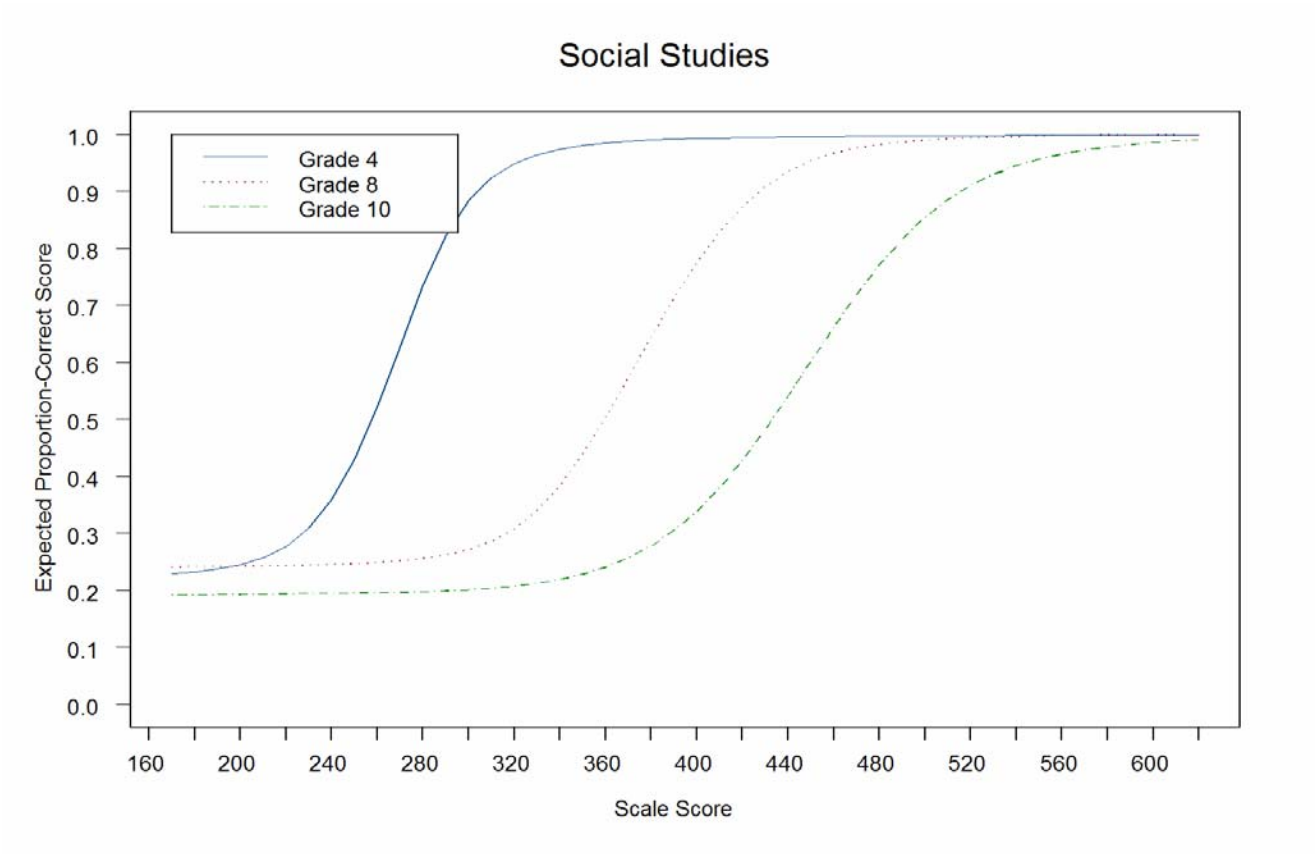


Figure 8-8
TCC Curve for Science Grades 4, 8, 10

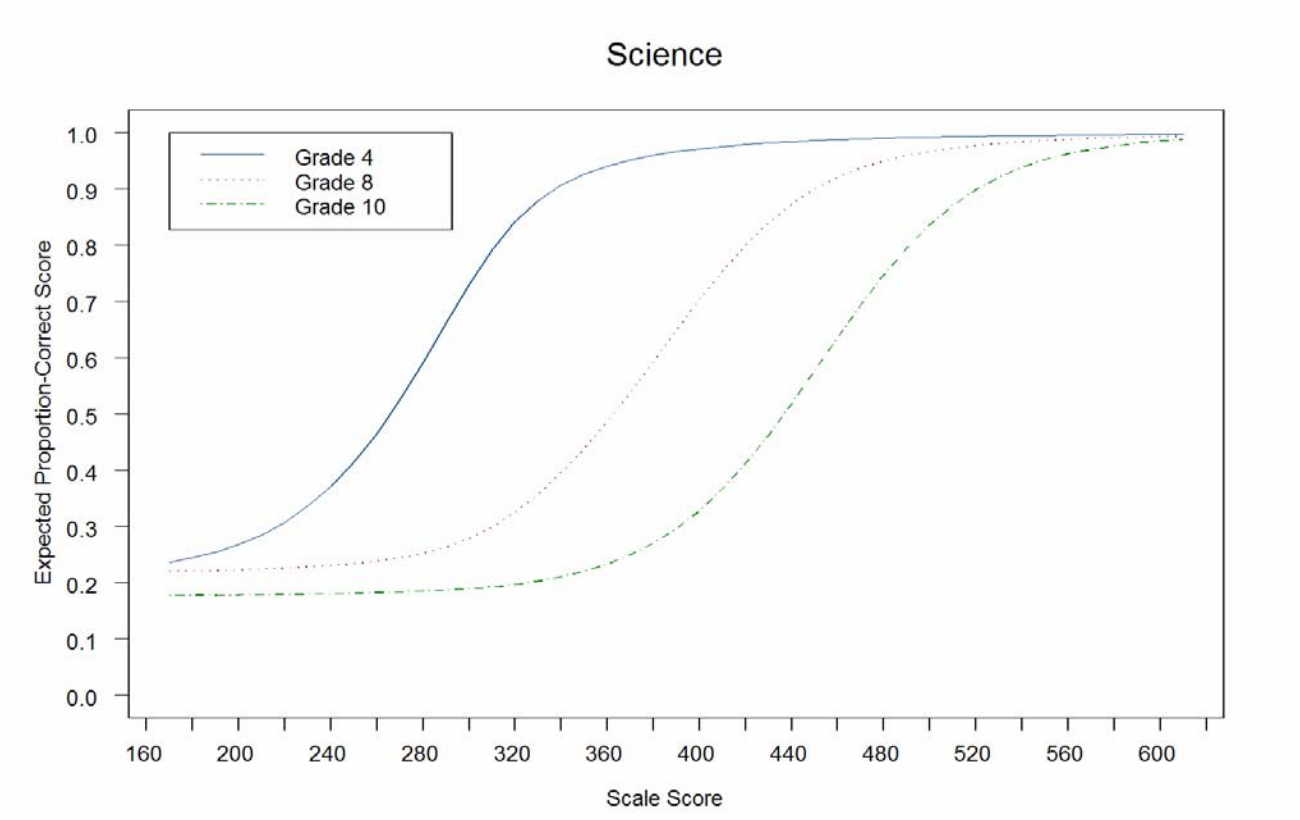


Figure 8-9
SEM Curves, Reading Grades 3-6

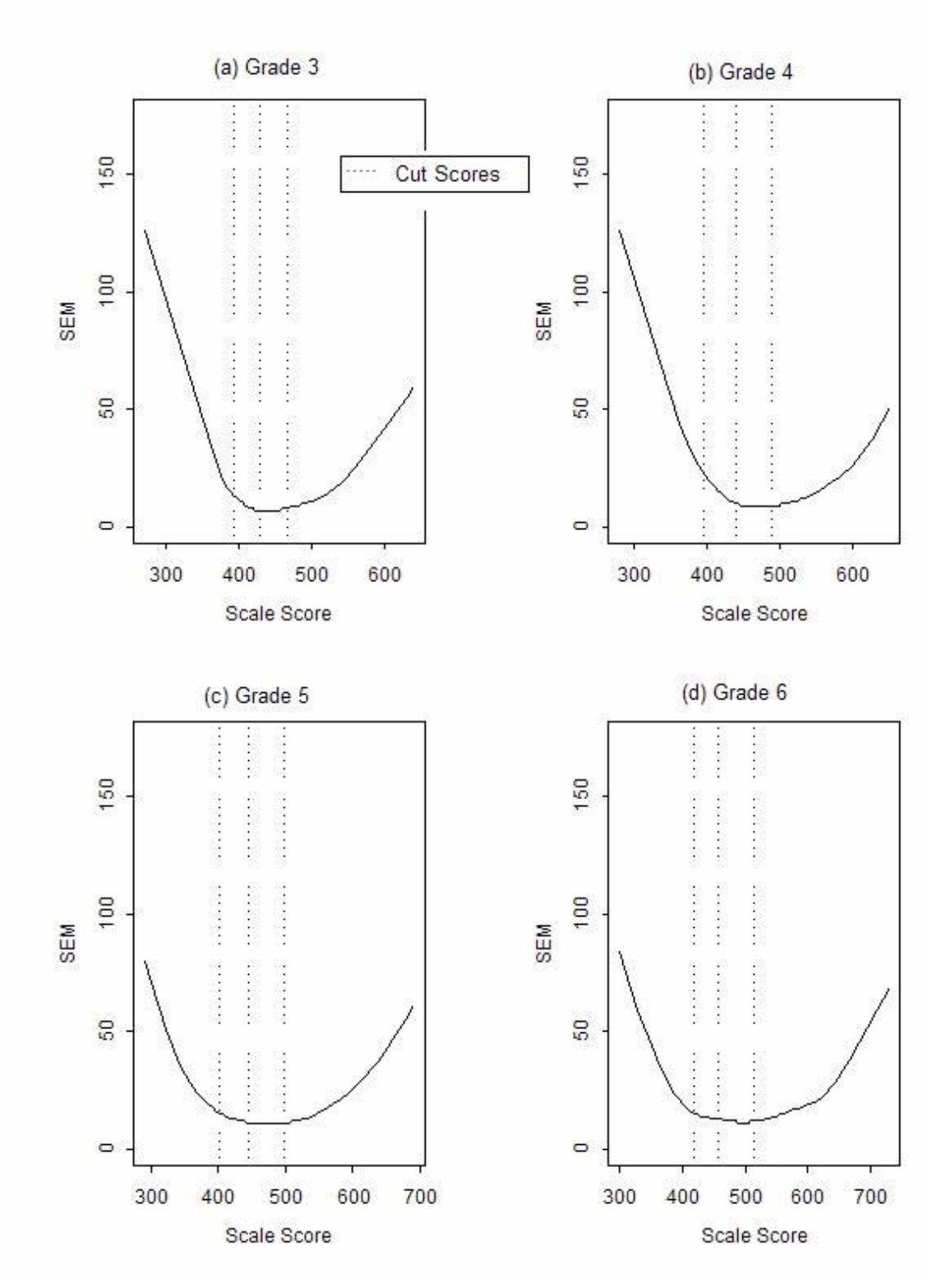


Figure 8-9
SEM Curves, Reading Grades 7, 8, 10

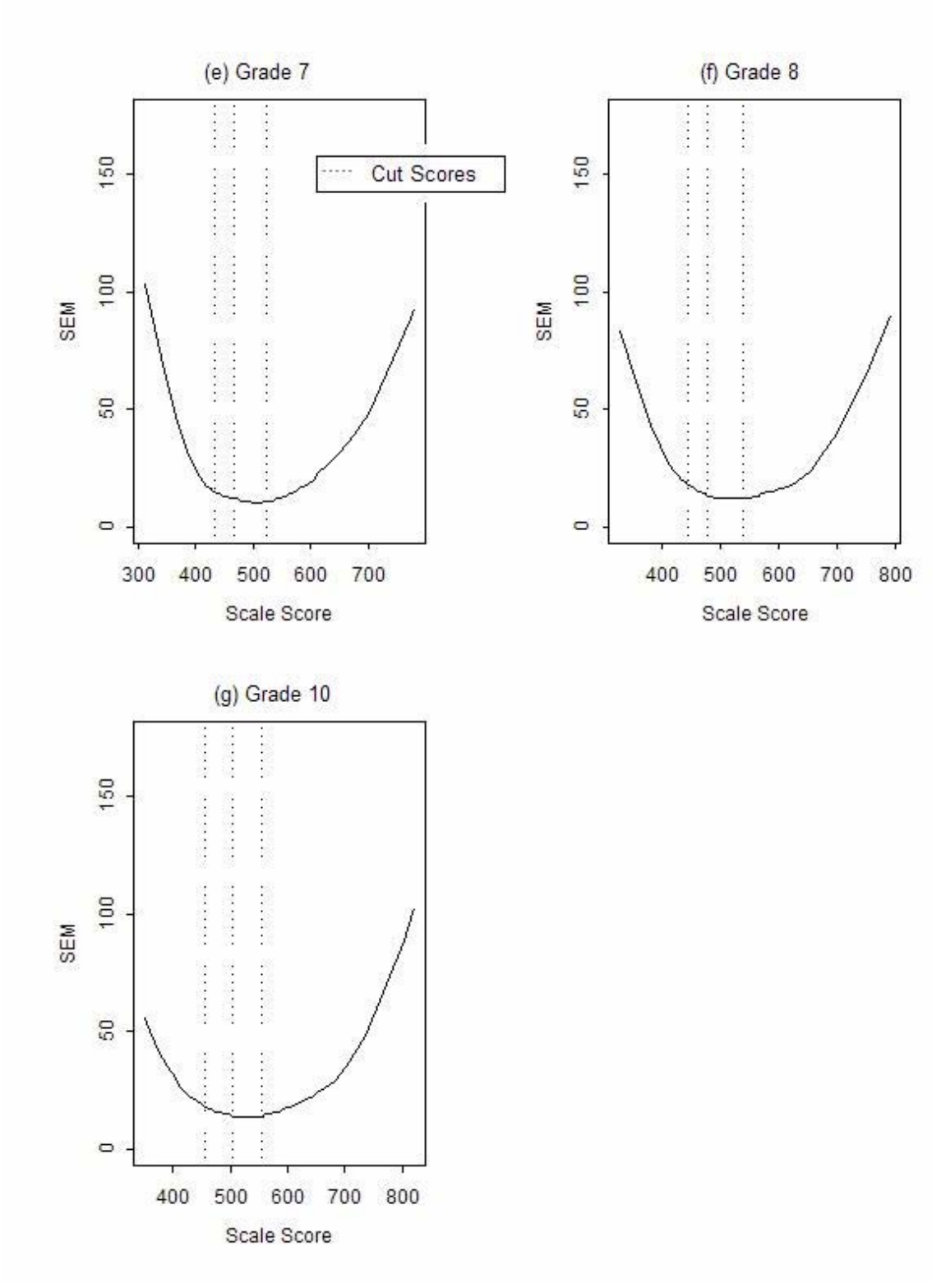


Figure 8-10
SEM Curves, Mathematics Grades 3-6

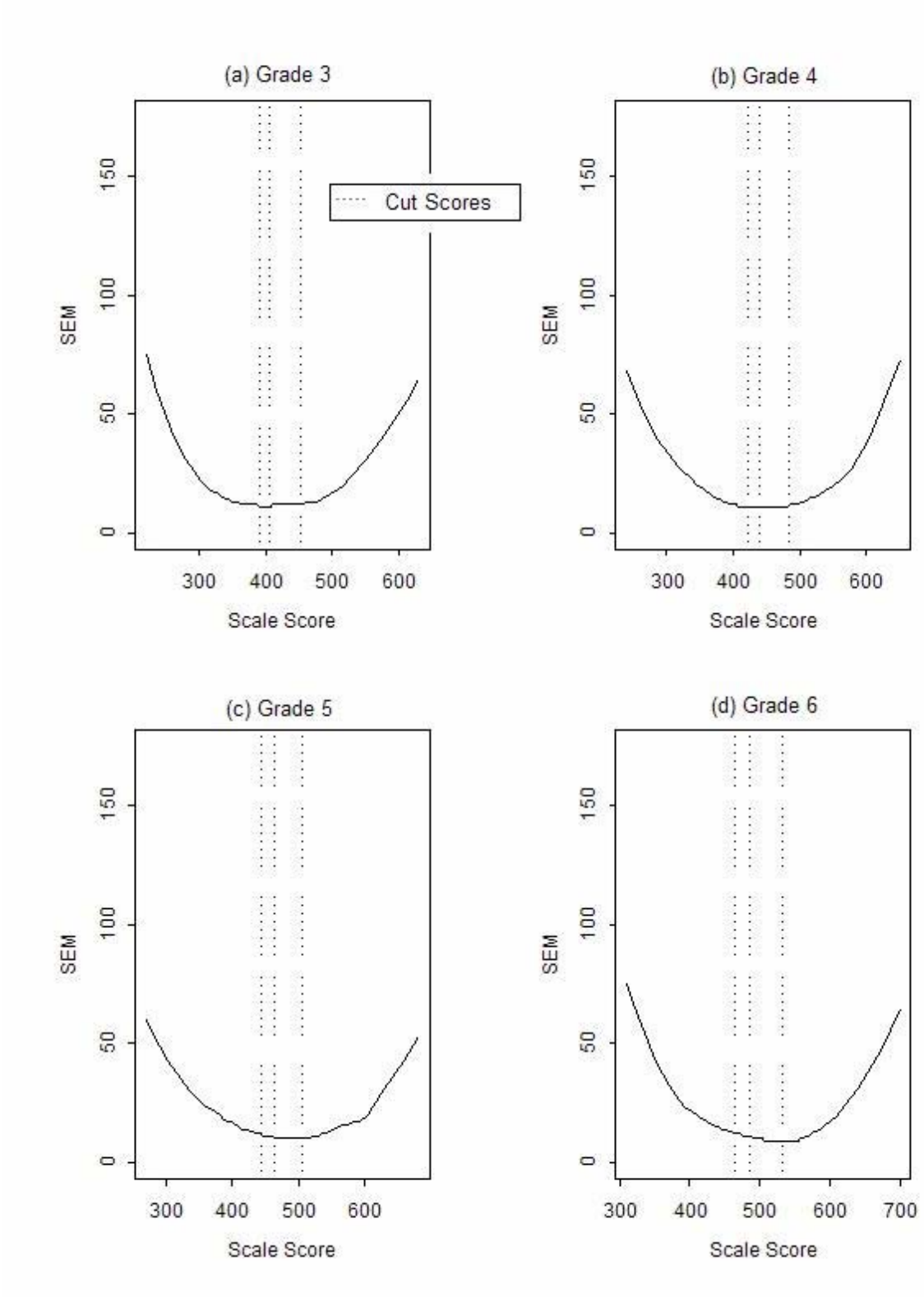


Figure 8-10 Cont'd
SEM Curves, Mathematics Grades 7, 8, 10

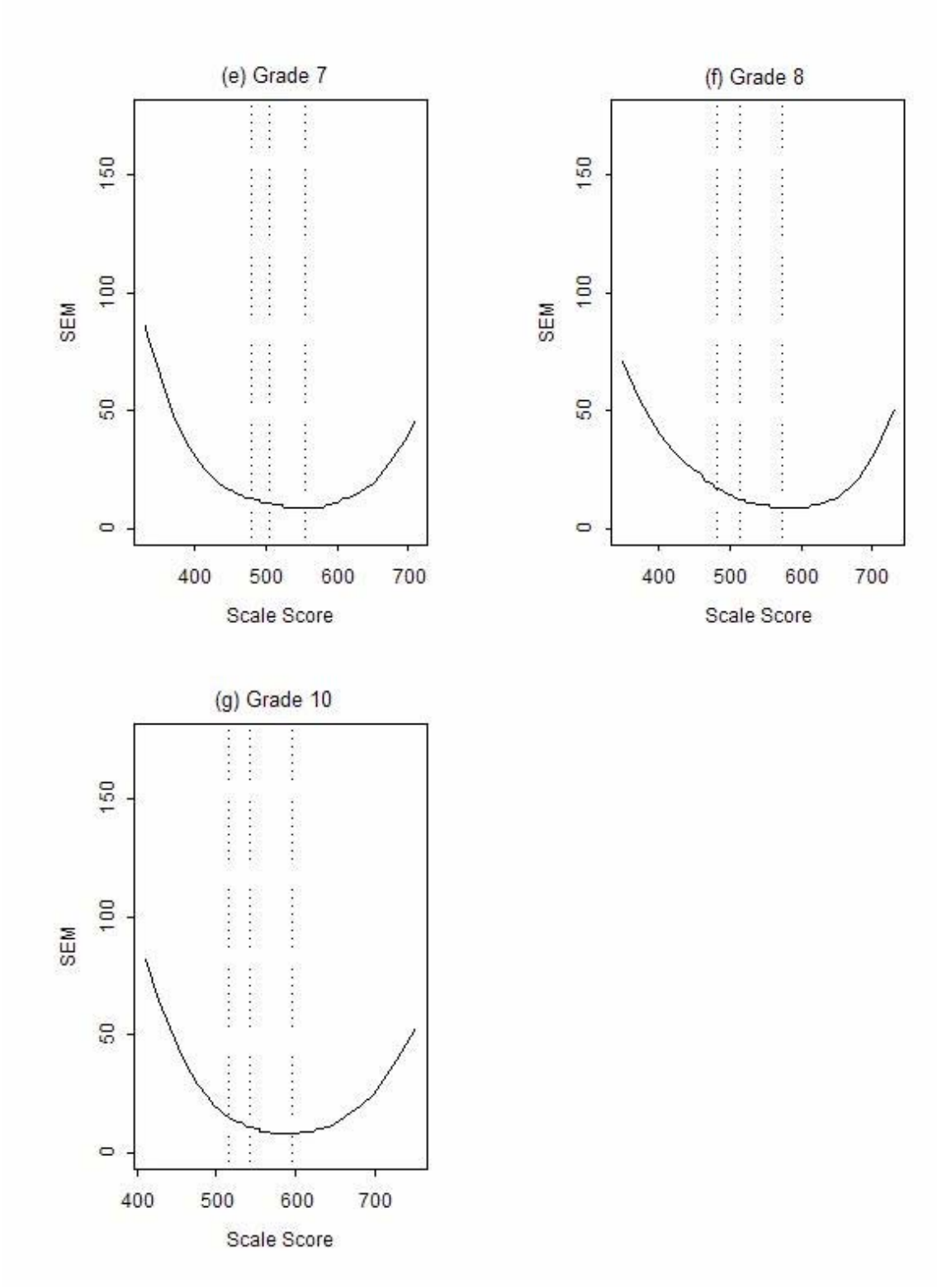


Figure 8-11
SEM Curves, Language Arts Grades 4, 8, 10

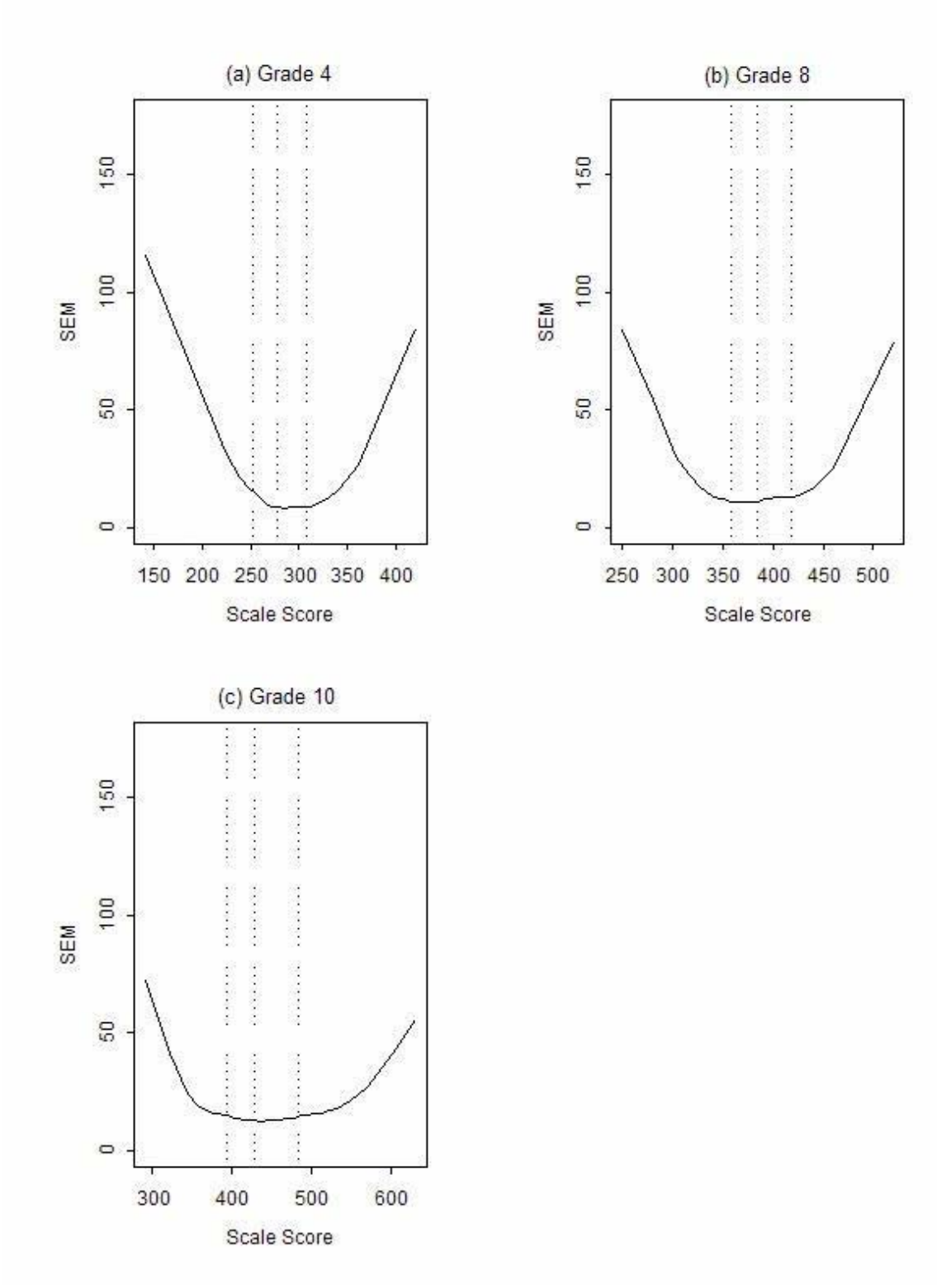


Figure 8-12
SEM Curves, Social Studies Grades 4, 8, 10

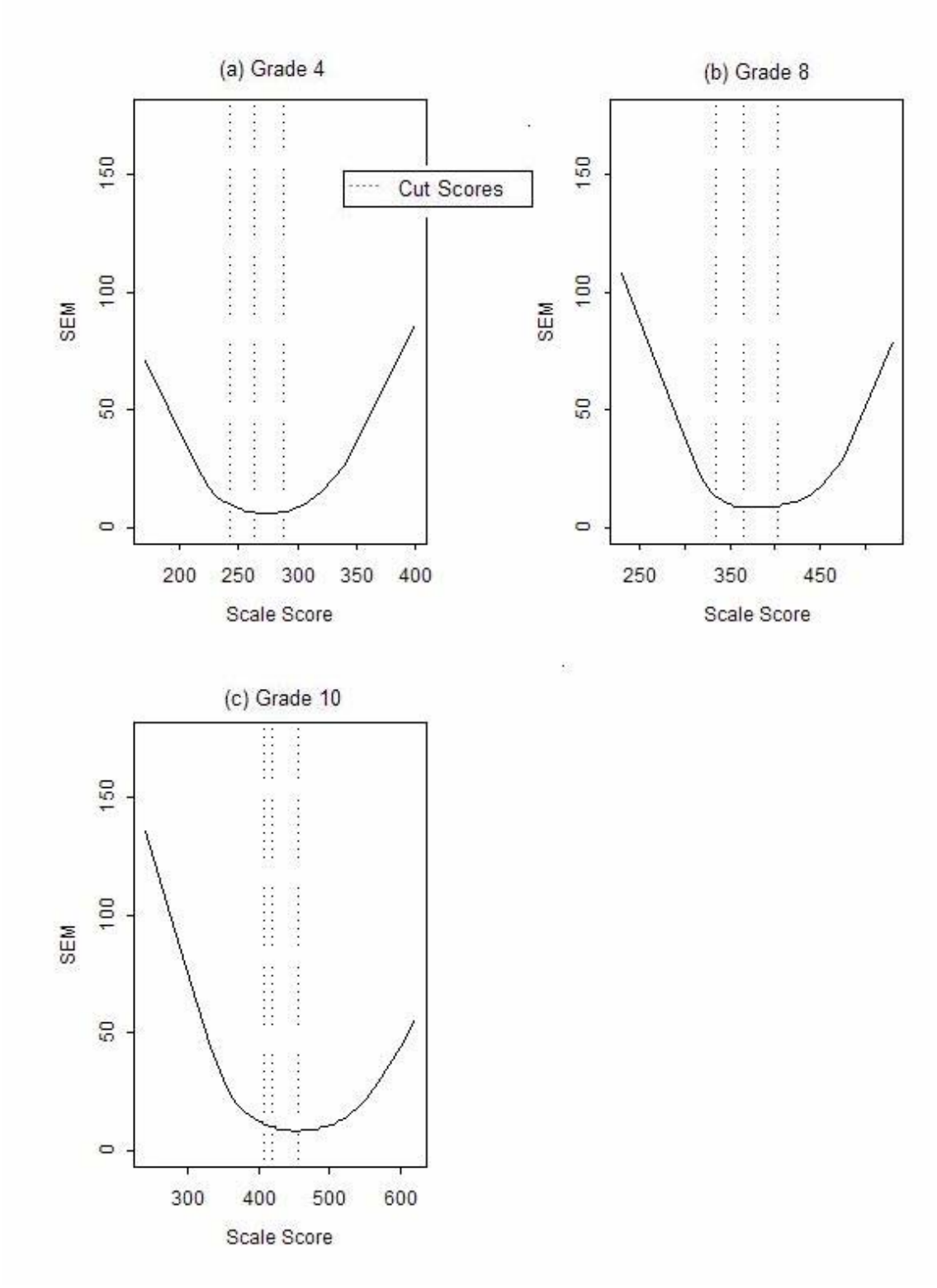


Figure 8-13
SEM Curves, Science Grades 4, 8, 10

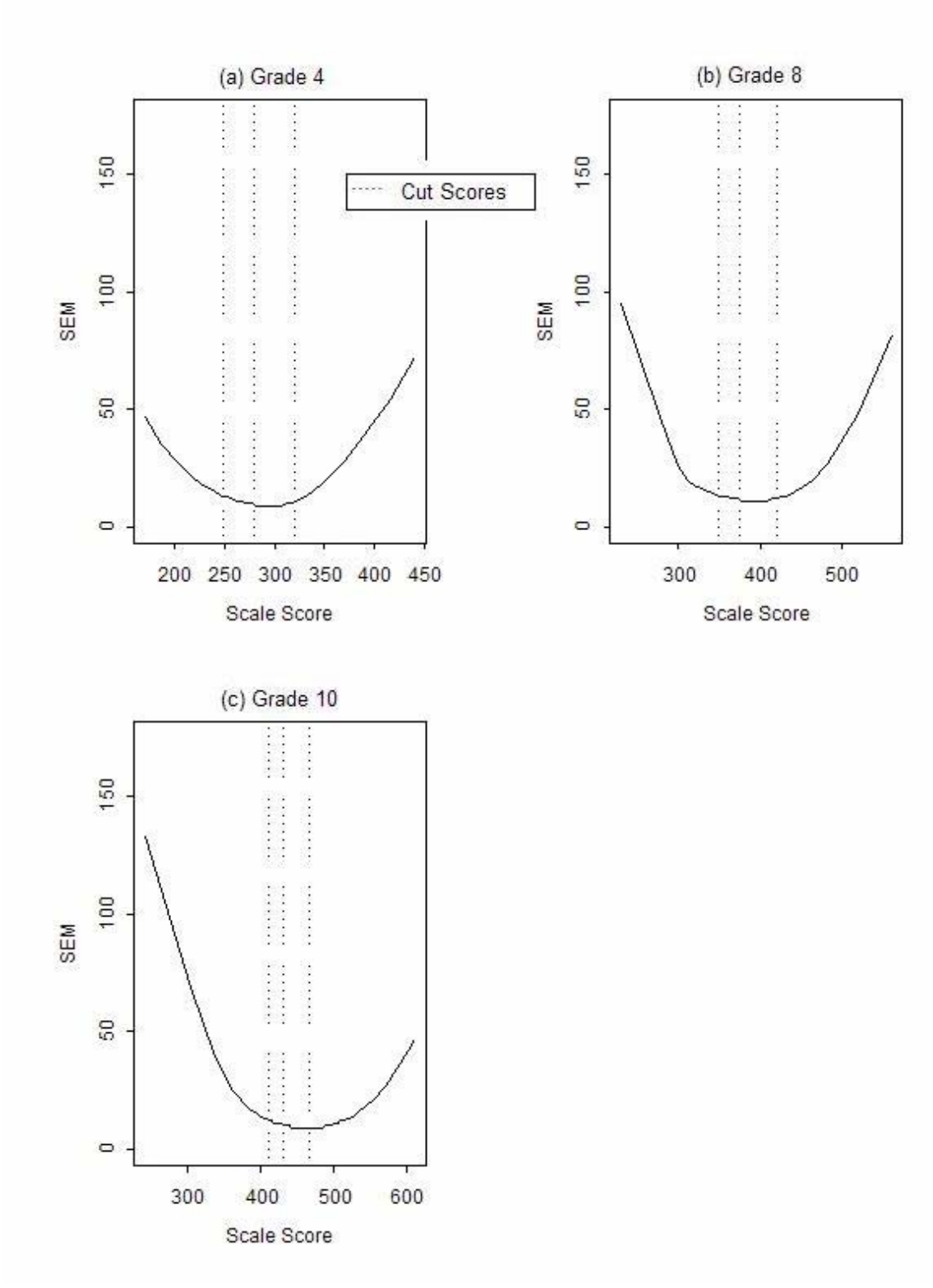


Figure 11-1
Cut Scores for Reading

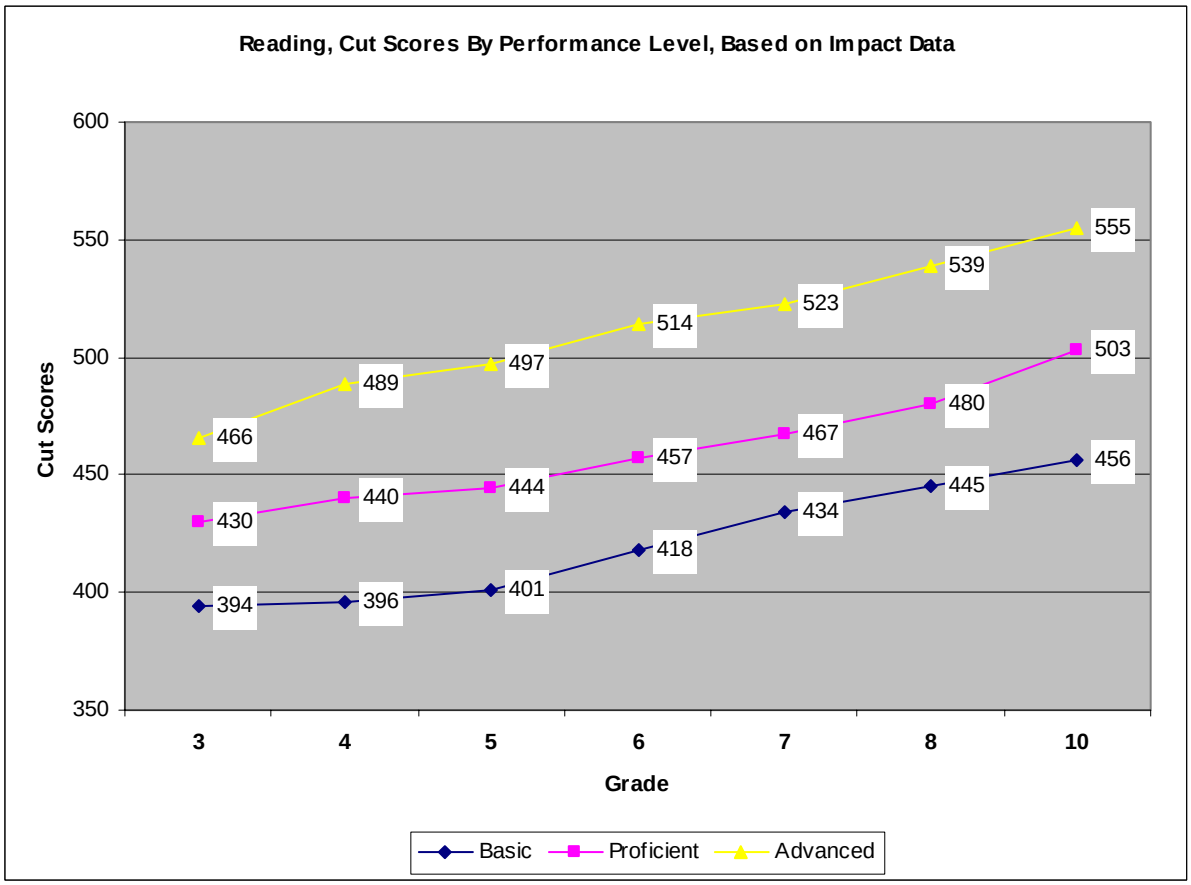


Figure 11-2
Cut Scores for Mathematics

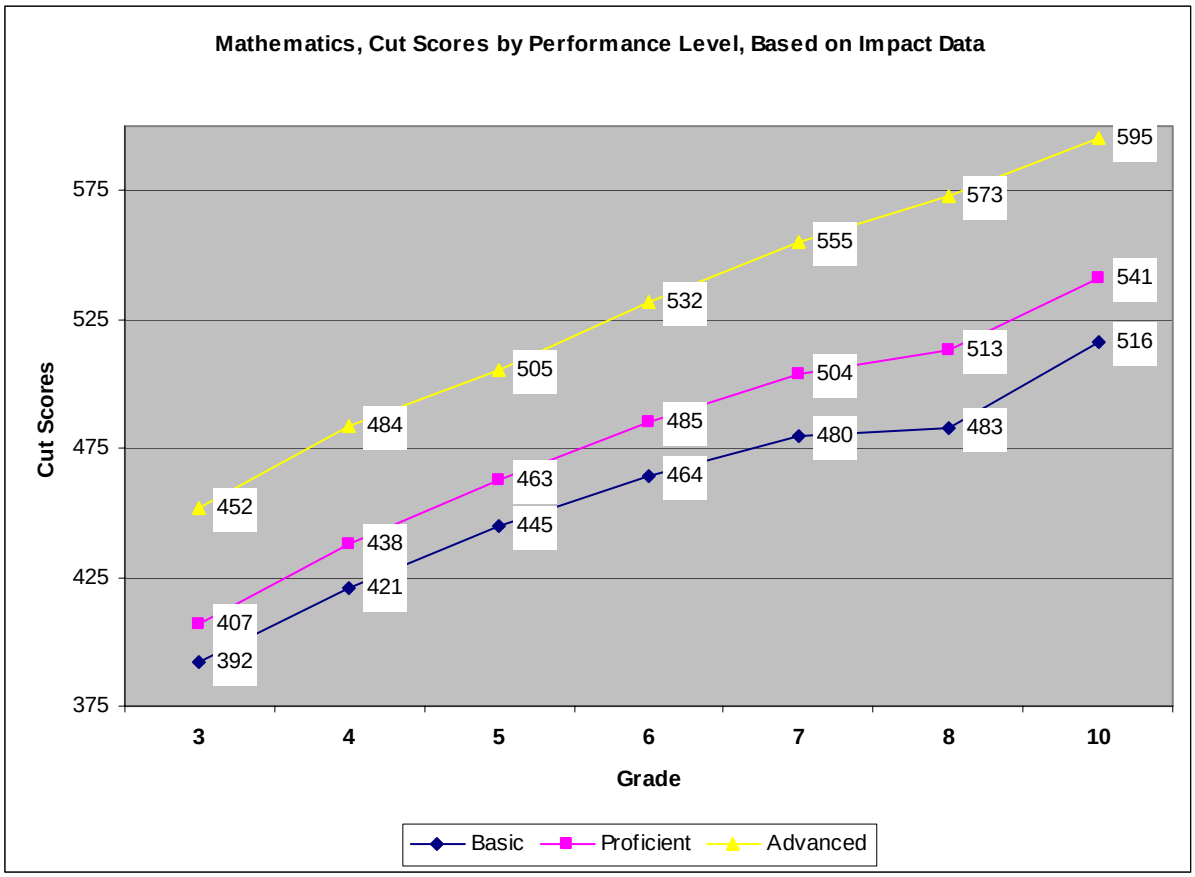


Figure 11-3
Cut Scores for Language Arts

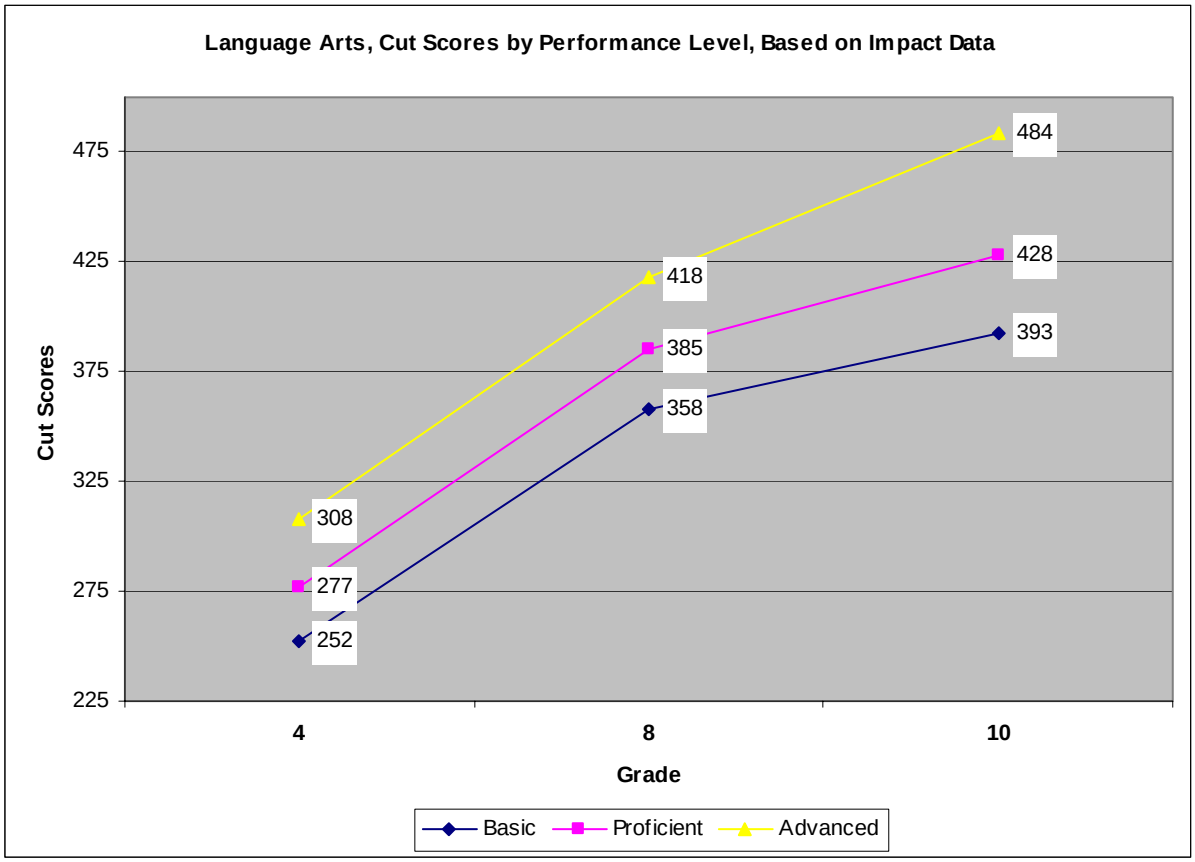


Figure 11-4
Cut Scores for Social Studies

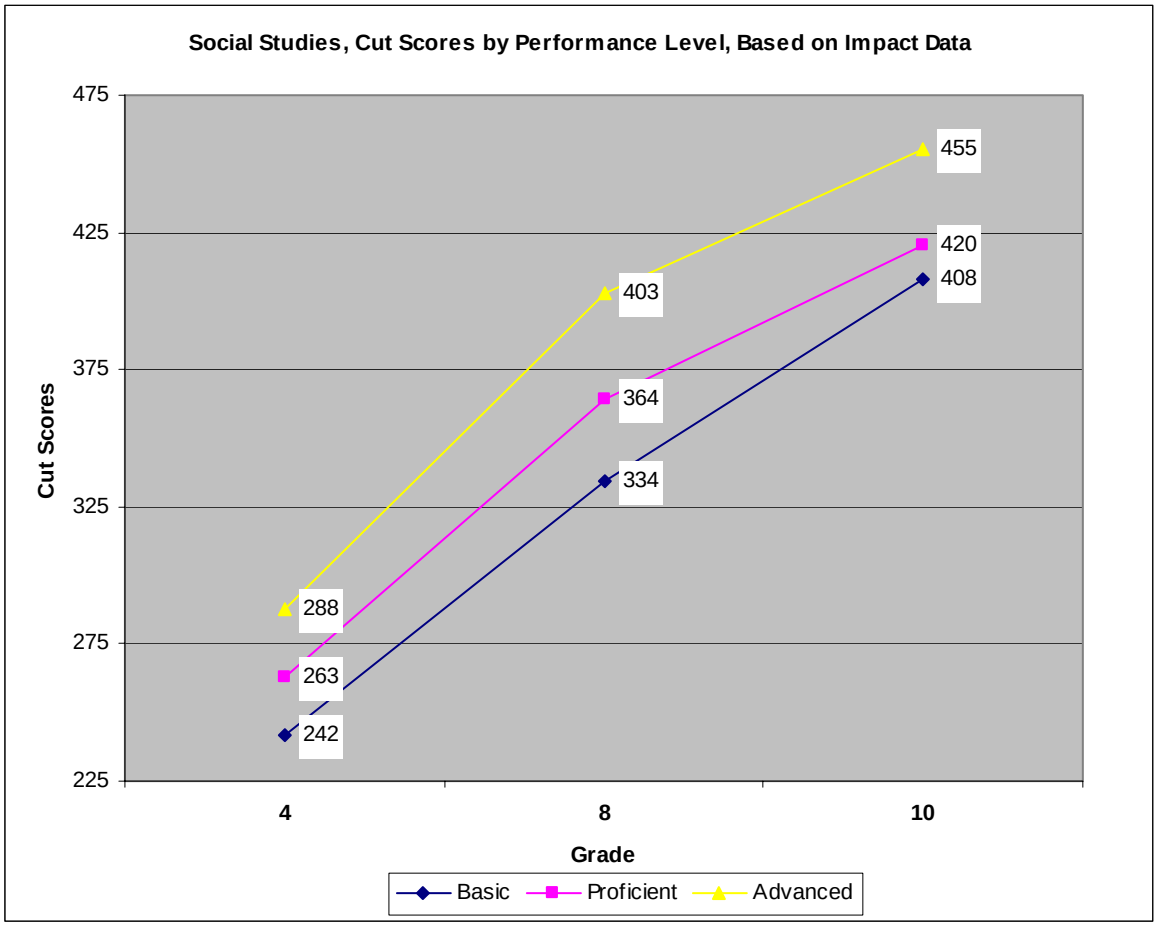


Figure 11-5
Cut Scores for Science

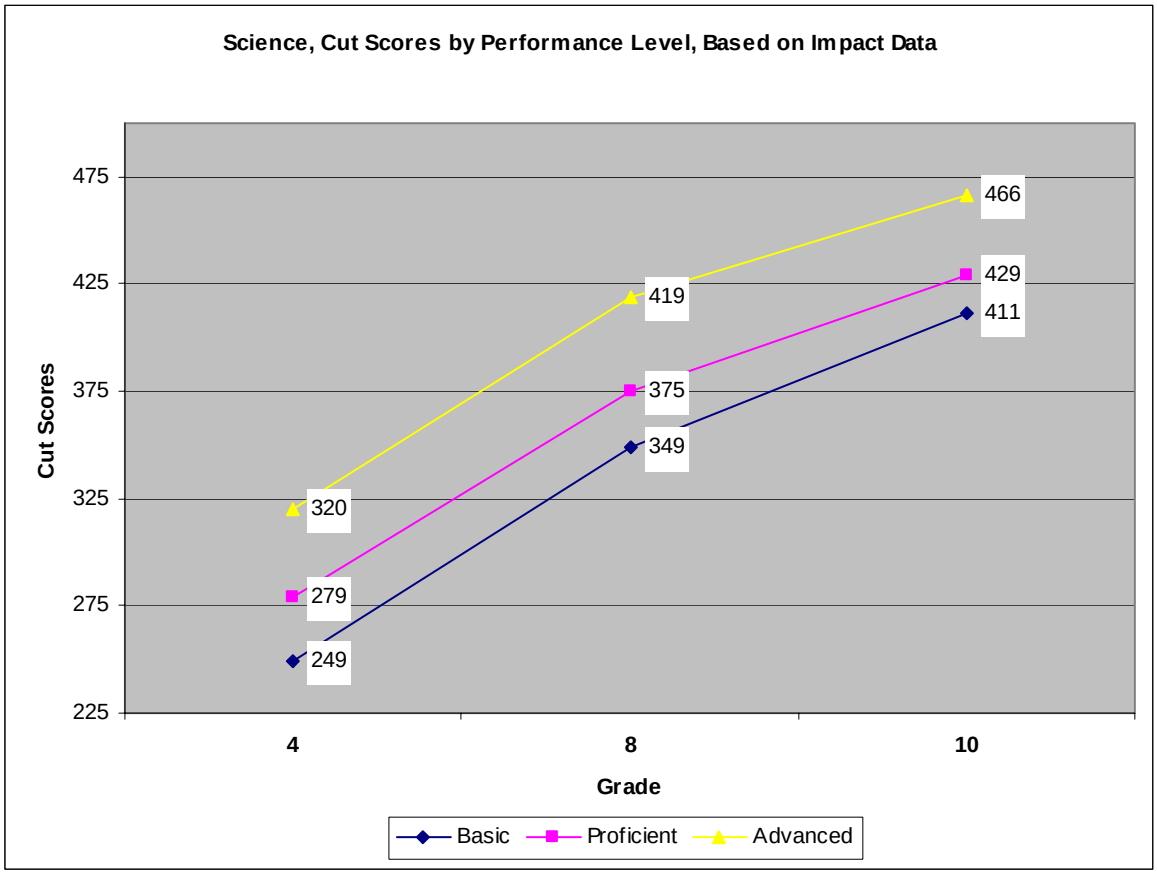


Figure 11-6
Percent of Students for Reading

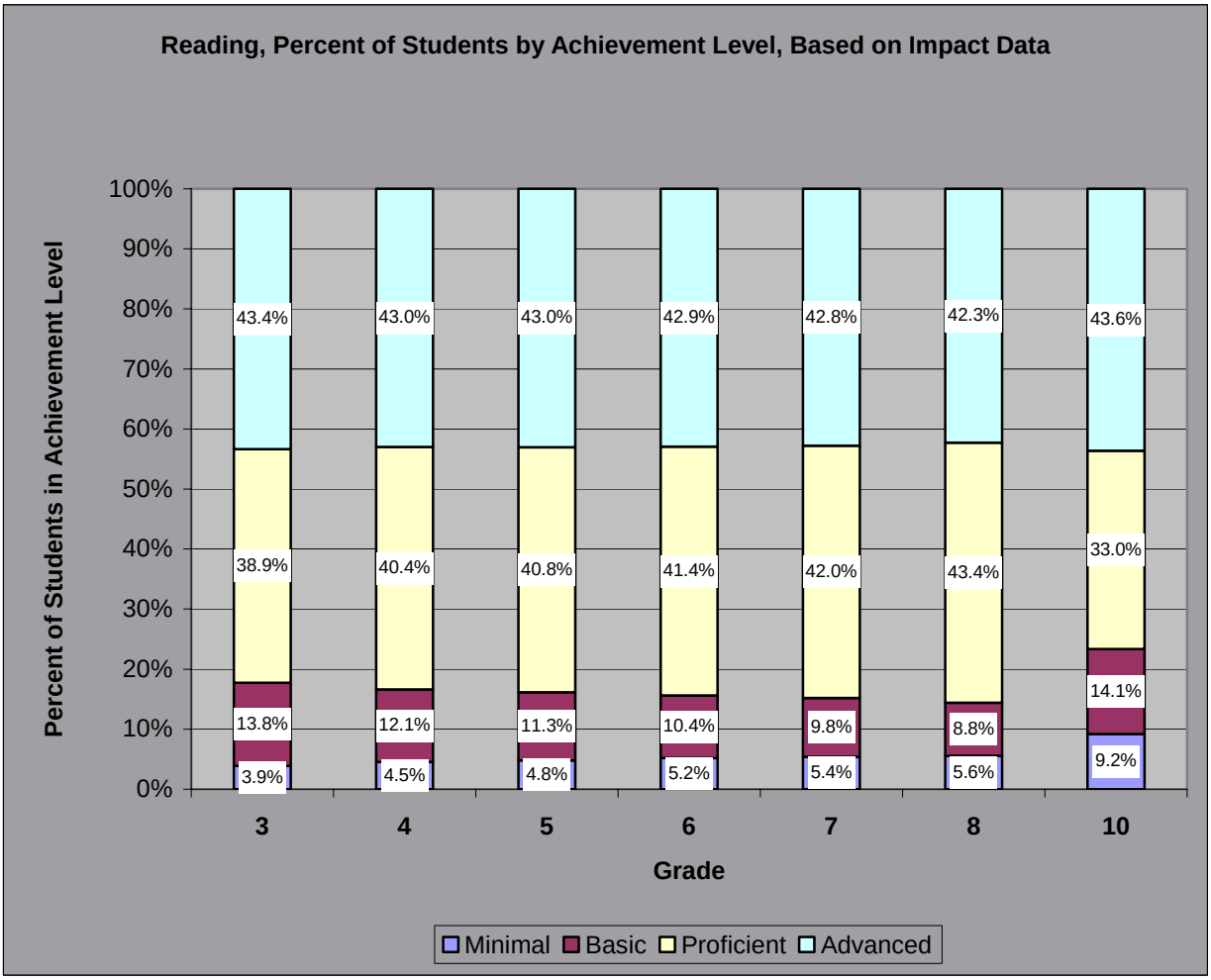


Figure 11-7
Percent of Students for Mathematics

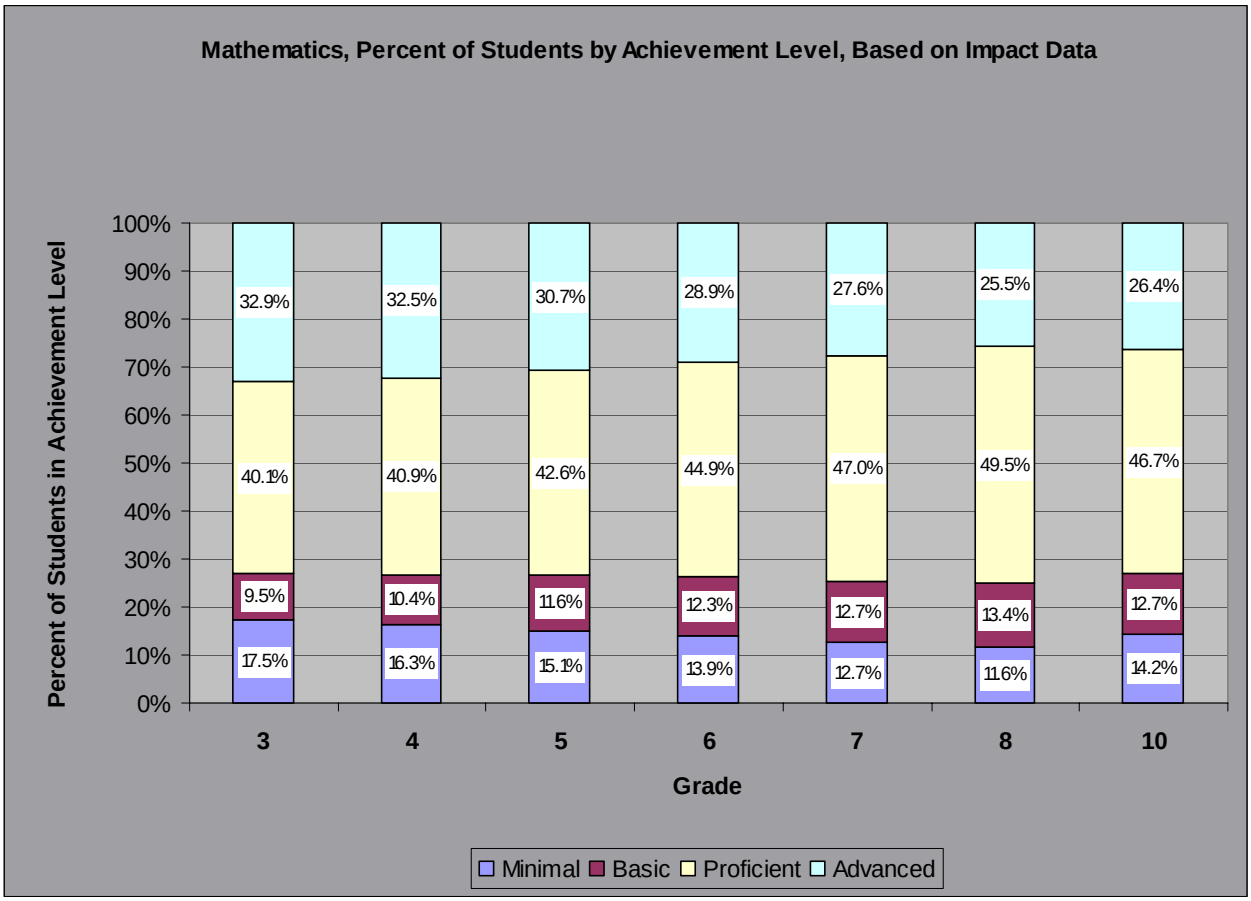


Figure 11-8
Percent of Students for Language Arts

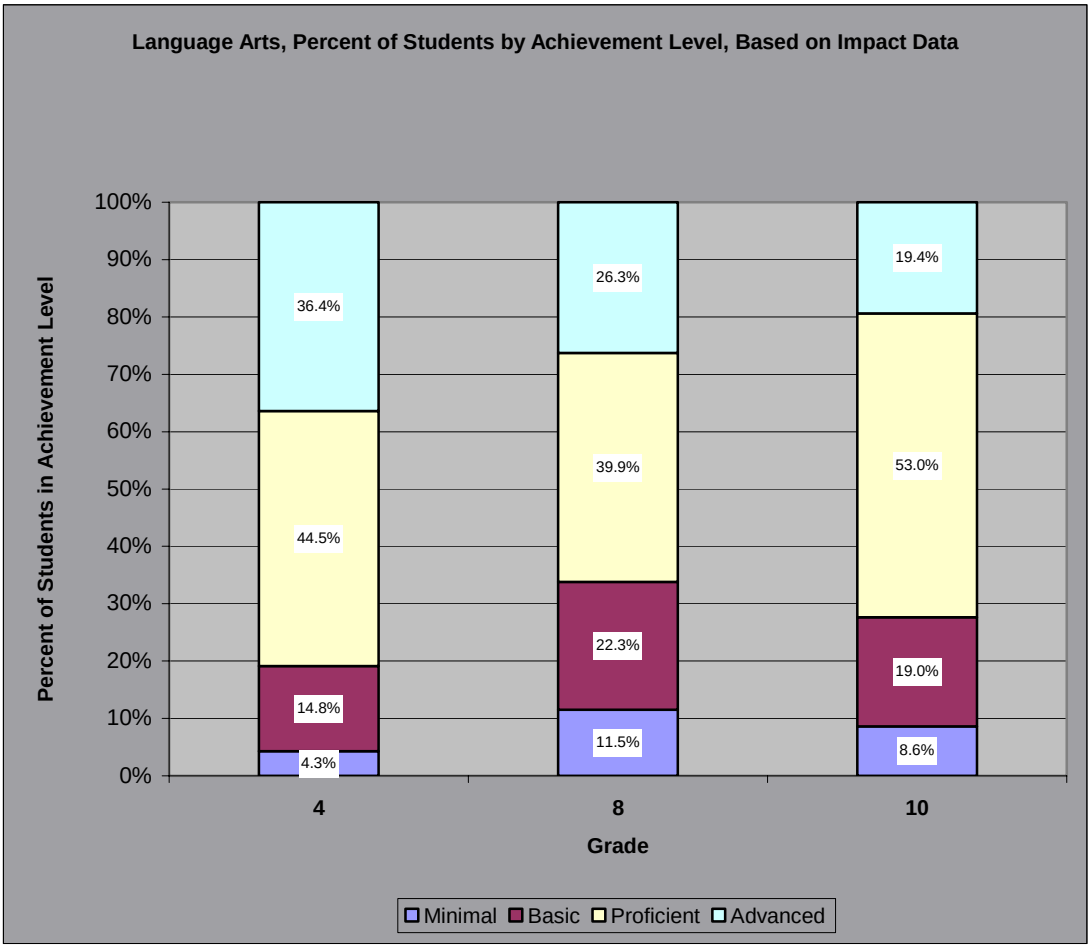


Figure 11-9
Percent of Students for Social Studies

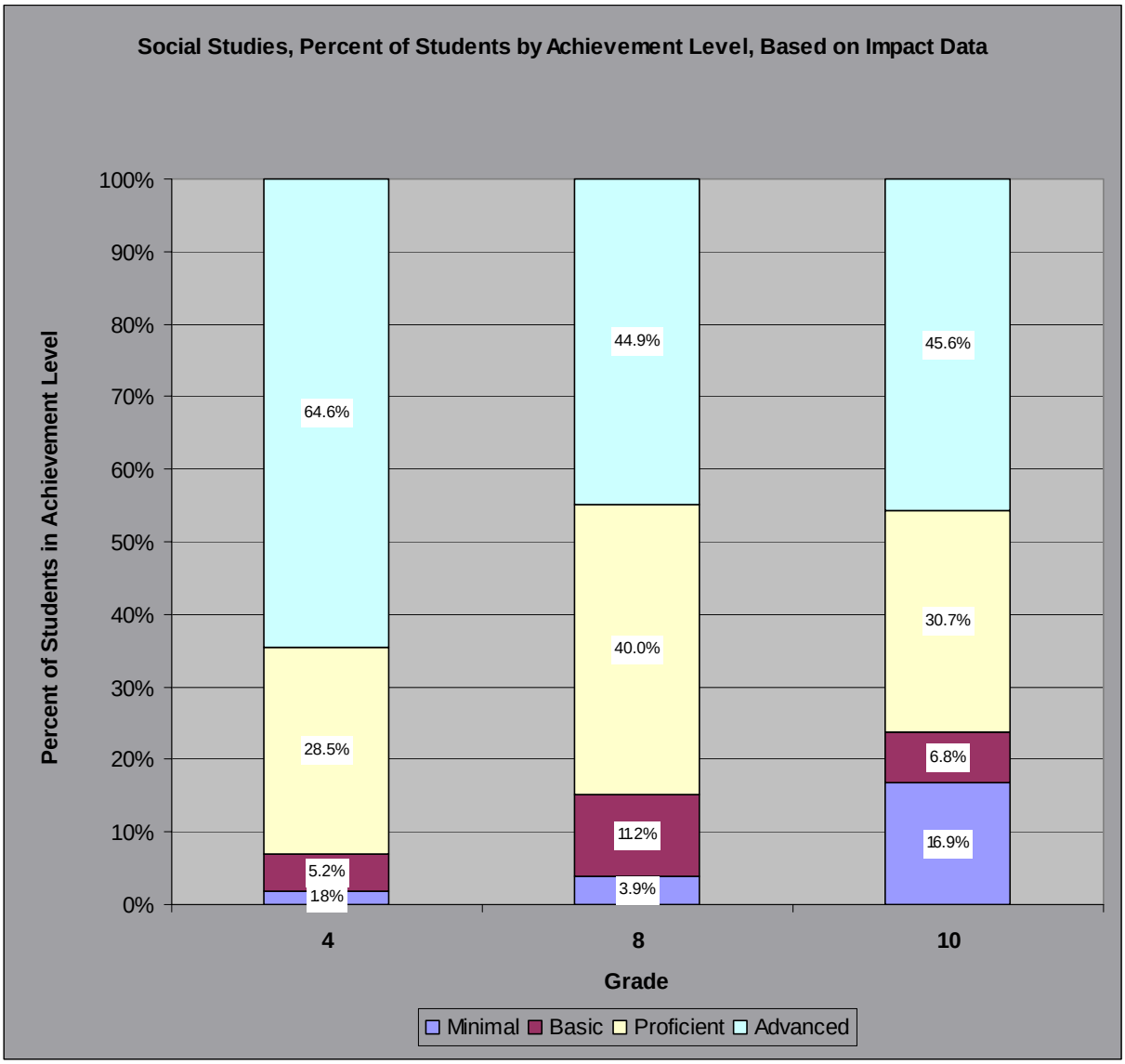
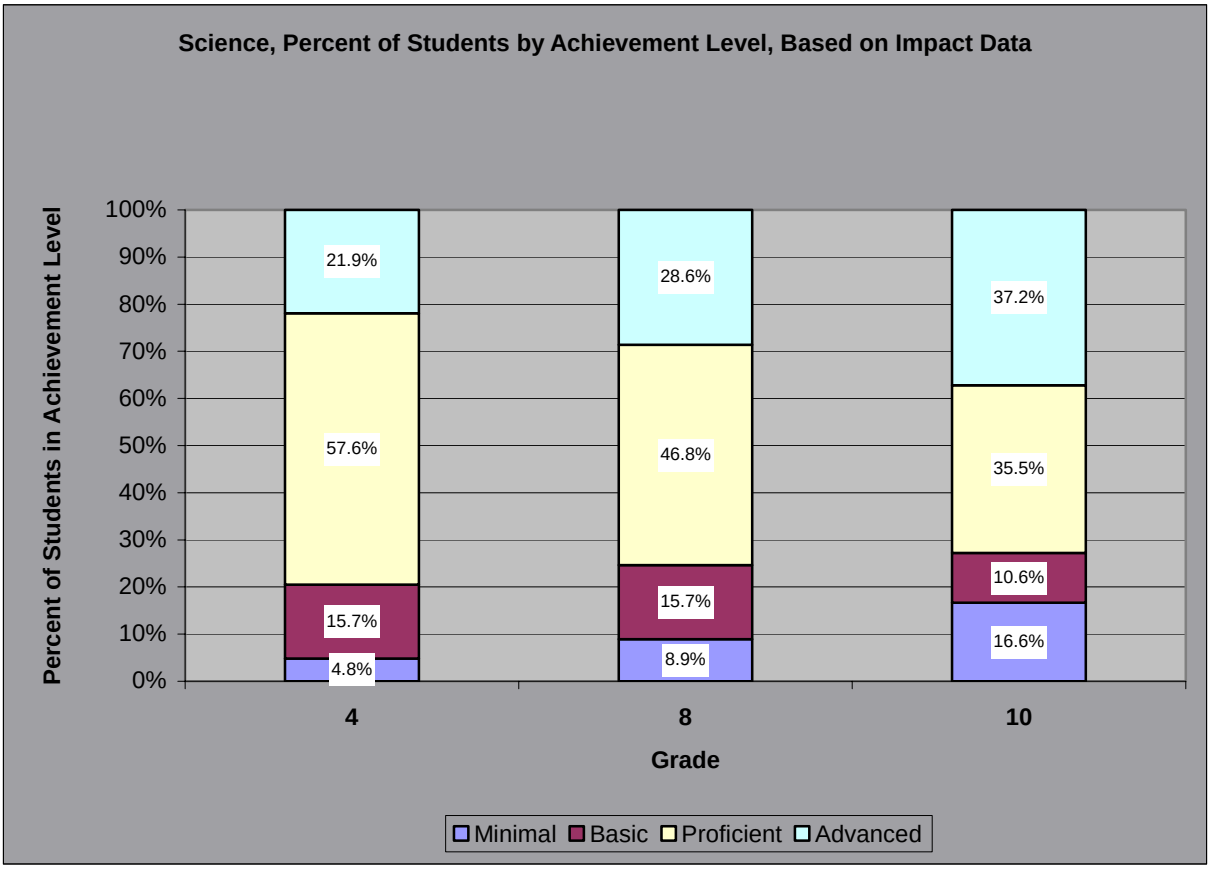


Figure 11-10
Percent of Students for Science



Appendix 4:
Fall 2007 WKCE Technical Report, Section 2.4

Copyright © 2009 by the Wisconsin Department of Public Instruction

(Prior sections of the Fall 2007 WKCE Technical Report removed here).

2.4 Item Development and Editing

This section describes the process and results of developing test items during 2007 for field testing in Fall 2008. The development of items included as embedded field test items on the Fall 2007 forms is described in the 2006 Technical Report, Section 4.2.

In August 2007, CTB editors developed plans for new item development for items to be field tested on the Fall 2008 forms. In addition to developing new items to meet DOK needs, other goals for item development included creating additional items for certain subskills to

Copyright © 2008 by the Wisconsin Department of Public Instruction

increase the item pool, providing flexibility in meeting the blueprint, and increasing overall flexibility in selecting items for forms.

An alignment study was conducted by Dr. Norman Webb of the University of Wisconsin-Madison using the Fall 2005 test forms. The results of the alignment study were reviewed by DPI's Technical Advisory Committee (TAC) in their December 2005 meeting. It was determined that the DOK level of the items should be reviewed for Reading, Math and Science. The TAC recommended, based on the alignment study, test forms should have 65% or more items at or above the DOK level of the objective to have a fully aligned test. As a result of the item alignment study, item development in 2006 and 2007 focused on creating items at higher levels of DOK for targeted objectives and subskills. The 2006 Technical Report, Section 4.2.3, describes the implications of the alignment study on item development in greater detail.

The Mathematics and Reading item development plans for grades 3–8 were determined by content editors and supervisors who analyzed the Mathematics and Reading item banks. Careful evaluation of the banks was conducted to determine if a sufficient number of items existed to meet test development criteria for each grade and content area. Criteria included meeting blueprints in terms of content diversity and the DOK requirement for each objective. Content editors also evaluated item banks relative to Reading and Mathematics blueprint changes requested by DPI in 2007 (see Section 2.2).

The Reading item development plan for grades 3–8 was based on creating additional items for existing passages in order to have more items measuring higher levels of DOK. Additional items for existing passages were warranted in order to expand the item set for the passage, and to allow greater flexibility when selecting operational forms.

The plan for Mathematics focused primarily on developing DOK level 3 items for statistics and probability, and on creating multiple-choice items that measure objective A, "Mathematical Processes." This was done in response to DPI's request to include multiple-choice items for this standard on the test blueprint. Additional items for remaining objectives were developed to broaden content diversity and flexibility of the item bank.

The development plans were presented to DPI in August 2007. Tables 2-11 and 2-12 present the Reading and Mathematics item development plans for the items to be field tested in Fall 2008; these plans represent the minimum number of items to be developed in 2007. Tables 2-13 and 2-14 show the number of items CTB developed prior to the Item Selection Review meeting (January 9–11, 2008), the number of items written by the committees during the review meeting, and the total number of items reviewed. Tables 2-15 and 2-16 show the number of items developed for Reading and Mathematics by grade level, reporting category, and item format. The number of items developed exceeded the number of items proposed in the plans. Increased development was a result of continued evaluation of the item banks by the Content Development team. Table 2-17 shows how many multiple-choice, constructed-response and total items have been written to date.

A staff of professional item writers, many of them experienced teachers, wrote the WKCE test items developed in 2007. Item writers adhered to the item specifications as they

Copyright © 2008 by the Wisconsin Department of Public Instruction

drafted and revised items. CTB assessment editors also used the item specifications during editorial reviews and revisions of the items. The item specifications provide detailed information regarding the following:

- item type
- content strand, standard, objective, subskills to be measured
- clarification statement of the task students will perform when answering each item type
- assessment limits
- stimulus attributes (stems, graphics, narratives)
- response attributes (general, correct response, acceptable distractors, unacceptable
- distractors)
- scoring rubric attributes (general or item/task specific)
- sample items

Throughout the item development and review process, the alignment between the item and the content standard/subskill/assessment limit was checked during each editing phase. All test items were carefully reviewed for content and style by test development specialists, Wisconsin educators, and the content specialists from DPI. All test items developed in 2007 were reviewed internally by CTB supervisors familiar with the Wisconsin content frameworks and item specifications. During all item reviews, careful attention was paid to verifying that each item measured the intended objective, subskill, and assessment limit. If any misalignment was found, the item was either rejected, edited to achieve greater alignment, or a different subskill or assessment limit was assigned.

Copyright © 2008 by the Wisconsin Department of Public Instruction

Table 2-11
Reading: 2007 Item Development Plan (for Fall 2008)

Grade	Reporting Category								Total	
	1		2		3		4			
	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR
3	10		10		26	2	6	2	52	4
4	7		7		26	2	14	2	54	4
5	6		5		26	2	7	2	44	4
6	7		6		23	3	10	2	46	5
7	12		8		22	2	8	2	50	4
8	5		11		20	2	8	2	44	4
Total	47		47		143	12	53	12	290	25

Table 2-12
Mathematics: 2007 Item Development Plan (for Fall 2008)

Grade	Reporting Category												Total	
	A		B		C		D		E		F			
	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR
3	6	0	0	2	3	2	3	0	8	3	0	1	20	8
4	8	0	0	0	3	3	2	2	10	3	0	0	23	8
5	6	0	2	0	4	3	3	1	9	2	2	3	26	9
6	4	0	2	0	2	3	5	2	9	3	3	1	25	9
7	5	0	2	0	6	3	5	0	7	4	5	2	30	9
8	3	0	0	0	8	3	8	3	5	1	6	3	30	10
Total	32	0	6	2	26	17	26	8	48	16	16	10	154	53

Copyright © 2008 by the Wisconsin Department of Public Instruction

Table 2-13
Reading 2007 Item Development (for Fall 2008)

Grade	Items Brought to Review Meeting	Items Written at Review Meeting	Total Items Reviewed
3	56	0	56
4	58	0	58
5	48	0	48
6	51	0	51
7	54	0	54
8	48	0	48
Reading Total	315	0	315

Table 2-14
Mathematics 2007 Item Development (for Fall 2008)

Grade	Items Brought to Review Meeting	Items Written at Review Meeting	Total Items Reviewed
3	37	2	39
4	34	7	41
5	40	1	41
6	40	2	42
7	41	0	41
8	49	0	49
Mathematics Total	241	12	253

Table 2-15
Reading: 2007 Item Development by Reporting Category and Item Format (for Fall 2008)

Grade	Reporting Category								Total	
	1		2		3		4			
	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR
3	10	0	10	0	26	1	6	3	52	4
4	7	0	7	0	26	1	14	3	54	4
5	6	0	5	0	26	1	7	3	44	4
6	7	0	6	1	23	2	10	2	46	5
7	12	0	8	0	22	2	8	2	50	4
8	5	0	11	0	20	0	8	4	44	4
Total	47	0	47	1	143	7	53	17	290	25

Copyright © 2008 by the Wisconsin Department of Public Instruction

Table 2-16
Mathematics: 2007 Item Development by Reporting Category and Item Format (for Fall 2008)

Grade	Reporting Category												Total	
	A		B		C		D		E		F			
	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR	MC	CR
3	6	4	6	2	4	2	2	0	9	2	1	1	28	11
4	6	5	3	0	3	3	2	2	12	2	2	1	28	13
5	5	4	2	0	4	3	3	1	10	2	4	3	28	13
6	3	3	6	0	2	3	4	2	10	3	5	1	30	12
7	4	3	3	0	6	3	6	0	7	2	5	2	31	10
8	5	6	1	0	8	6	7	3	4	1	7	1	32	17
Total	29	25	21	2	27	20	24	8	52	12	24	9	177	76

Copyright © 2008 by the Wisconsin Department of Public Instruction

Table 2-17
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	Total MC to date	Total CR to date
Grade 3												
Reading	411	52	23	2	30	4	40	3	52	4	556	65
Math	317	36	33	14	18	2	30	4	28	11	426	67
Total	728	88	56	16	48	6	70	7	80	15	982	132
Grade 4												
Reading	380	56	32	3	34	3	25	4	54	4	525	70
Math	265	35	45	9	29	1	26	4	28	13	393	62
Language Arts	0	0	0	10	0	0	0	0	0	0	0	10
Science	0	0	0	0	123	34	0	0	0	0	123	34
Total	645	91	77	22	186	38	51	8	82	17	1041	176
Grade 5												
Reading	433	59	36	6	29	5	29	7	44	4	571	81
Math	305	49	38	11	26	3	30	5	28	13	427	81
Total	738	108	74	17	55	8	59	12	72	17	998	162
Grade 6												
Reading	511	56	32	5	42	5	37	6	46	5	668	77
Math	310	41	53	16	7	2	28	4	30	12	428	75
Total	821	97	85	21	49	7	65	10	76	17	1096	152
Grade 7												
Reading	359	44	35	4	38	4	25	5	50	4	507	61
Math	305	34	32	23	20	0	28	4	31	10	416	71
Total	664	78	67	27	58	4	53	9	81	14	923	132
Grade 8												
Reading	365	44	30	4	34	4	25	4	44	4	498	60
Math	289	51	47	25	20	2	28	4	32	17	416	99
Language Arts	0	0	0	10	0	0	0	0	0	0	0	10
Science	0	0	0	0	125	34	0	0	0	0	125	34
Total	654	95	77	39	179	40	53	8	76	21	1039	203
Grade 10												
Reading	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
Language Arts	0	0	0	0	0	0	0	0	0	0	0	0
Science	0	0	0	0	18	8	0	0	0	0	18	8
Total	0	0	0	0	18	8	0	0	0	0	18	8
TOTALS												
Reading	2,459	311	188	24	207	25	181	29	290	25	3325	414
Mathematics	1,791	246	248	98	120	10	170	25	177	76	2506	455
Language Arts	0	0	0	20	0	0	0	0	0	0	0	20
Science	0	0	0	0	266	76	0	0	0	0	266	76
Grand Total	4,250	557	436	142	593	111	351	54	467	101	6097	965

Copyright © 2008 by the Wisconsin Department of Public Instruction

THE ASSESSMENT ACCOMMODATIONS MATRIX FOR STUDENTS WITH DISABILITIES – UPDATED 2011

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 a 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 aloud all directions.*
D 3	Explain or clarify directions.[1]	✓	✓
D 4	Student rereads and/or restates directions.[1]	✓	✓
Content Presentation			
D 5	Turn pages for student.	✓	✓
D 6	Braille; student responses must be transcribed into scorable test book.[6, 14]	✓	✓
D 7	NEW 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 8	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 9	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 10	Sign language for test passages and questions **(Not allowed on Reading tests)**.[11]	✓	✓
D 11	Text talker for test passages and questions **(Not allowed on Reading tests)**.[4]	✓	*N/A: Test administrator reads WAA-SwD aloud.*
D 12	Student reads aloud to self.	✓	✓
D 13	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 14	Student records him/herself reading aloud and plays back recording.[4]	✓	✓
D 15	Audio recording of test passages and questions in English **(Not allowed on WKCE Reading test or WAA-SwD "Read-by-Student" items)**.[4, 9]	✓	*N/A: Test administrator reads WAA-SwD aloud.*

Copyright © 2011 by the Wisconsin Department of Public Instruction

THE ASSESSMENT ACCOMMODATIONS MATRIX FOR STUDENTS WITH DISABILITIES - UPDATED 2011

Accommodation Description for *Students with Disabilities* (D)		WKCE	WAA-SwD
Content Presentation (cont.)			
D 16	Read the Reading test ONLY in the following scenarios as described in *Form I-7-B*:[8, 9] a) For a student with visual impairments who is not yet proficient in contracted Braille, the WKCE Reading test passages and questions may be read aloud. b) For a student with visual impairments who is not yet proficient in un-contracted Braille, the WAA-SwD "Read-by-Student" Reading test items may be read aloud.	✓	✓
Response			
D 17	Manipulatives, base-ten blocks, 3-D shapes, 100's chart (not multiplication table), whole integer number lines, 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 18	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 19	Braille writer; transcribe student responses into scorable test book.[4, 6, 14]	✓	✓
D 20	Student indicates responses orally to scribe.[5]	✓	*N/A: Test administrator records all student responses.*
D 21	Student signs responses to scribe. For the Writing test, no translation from American Sign Language (ASL) is allowed; student must sign in exact English to scribe.[5]	✓	✓
D 22	Student records responses using an audio or video device:[4] 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.[6, 14]	✓	*N/A: Student is allowed to communicate responses in whichever mode is best for the student. Test administrator records student responses.*
D 23	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 24	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			
D 25	Student moves, stands, or paces during individual administration.	✓	✓
Timing/Scheduling			
D 26	Extra time; test session must be completed within the same day the student started the session.[7]	✓	✓

Copyright © 2011 by the Wisconsin Department of Public Instruction

THE ASSESSMENT ACCOMMODATIONS MATRIX FOR STUDENTS WITH DISABILITIES - UPDATED 2011

Other Accommodations for Students with Disabilities	
D 27	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. ○ 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/dacforms.html. ○ 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 © 2011 by the Wisconsin Department of Public Instruction

THE ASSESSMENT ACCOMMODATIONS MATRIX FOR STUDENTS WITH DISABILITIES - UPDATED 2011

Explanation of Footnotes

[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.
- 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 the proctor or a test examiner. In addition, a student must not be allowed to use any form of wireless communication during these breaks.

Copyright © 2011 by the Wisconsin Department of Public Instruction

[8]**Student who is 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 un-contracted 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 D16):**

- 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/doc/translator_guidelines.doc) should have a Bachelor's Degree in Modern Languages. When this is not possible, be sure that a translator or interpreter has the following qualifications:
  1. Mastery of the target language
  2. Familiarity with both cultures
  3. Extensive general 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 Signing Exact English [SEE]. 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 with visual impairments 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 © 2011 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.

HOSS: Highest obtainable scale score. The highest possible scale score on a test.

Copyright © 2011 by the Wisconsin Department of Public Instruction

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.

SMD: Standardized mean difference. SMD is commonly used DIF statistic for constructed-choice items.

Copyright © 2011 by the Wisconsin Department of Public Instruction

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 © 2011 by the Wisconsin Department of Public Instruction