WISCONSIN KNOWLEDGE AND CONCEPTS EXAMINATIONS

FALL 2006 WKCE TECHNICAL MANUAL

JUNE 2007

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

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).

Acknowledgments

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

CTB/McGraw-Hill:

Kim, Dong-In
Research Project Manager

Friedman, Greg
Statistical Analyst

Fatica, Kevin
Research Associate

Tully, Margie
Publishing Director

Hailes, Artamease
Development Manager, Publishing

Hughes, Robin
Scoring Project Manager

Dodge, Leslie
Program Manager

Wisconsin Department of Public Instruction:

Jason Engle
Consultant - Standards Based Assessment

Bradley Carl
School Administration Consultant for Statistics & Assessment

Jason Bierbrauer
School Administrative Consultant

TABLE OF CONTENTS

Foreword i

Acknowledgments ii

Part 1: Executive Summary 1

Part 2: Involvement of Wisconsin Educators 6

2.1 Establishing Test Content 6
2.2 Writing and Developing Assessment Materials 7
2.3 Descriptor Writing 11

Part 3: Test Design 12

3.1 Test Specifications 12
3.1.1 Content Framework and Assessment Limits 12
3.1.2 Test Blueprint 15
3.1.3 Reading Passage Selection 19
3.1.4 Development and Alignment of Items 20
3.2 Test Blueprints 21
3.3 Description of the WKCE 2006 Tests 21
3.3.1 Reading 22
3.3.2 Mathematics 22
3.3.3 Language Arts 22
3.3.4 Social Studies 23
3.3.5 Science 23

Part 4: Test Development 24

4.1 Overall Test Book Creation and Editing Process 24
4.1.1 Solution Management 24
4.1.2 Documents and Materials Development 24
4.1.3 Item Development and Editing 25
4.1.4 Quality Reviews 27
4.2 Item Pool 28
4.2.1 Item Writing 28
4.2.2 Content/Bias Review 29
4.2.3 Item Alignment 32
4.3 Item Selection of 2005 WKCE 34
4.3.1 WKCE Item Selection 34
4.3.2 WKCE Field Test Item Selection 36
4.4 Style and Format Decisions 36
4.5 Customer Approvals 37
4.5.1 Item Content and Bias Review 37
4.5.2 Item Selection Approval 37
4.5.3 Manuscript Approvals 37
4.5.4 Second Pages Approvals 38
4.5.5 Final Pages Sign-off 38

Part 5: Test Administration 39

5.1 Accommodations 39
5.1.1 Standard Accommodations 40
5.2 Reporting Results of Assessments Taken with Accommodations 42
5.3 Test Security 42

TABLE OF CONTENTS CONT'D

5.4 Test Administration....42

Part 6: Scoring....44

6.1 Rubric Description....44
6.2 Processing and Scoring of Test Materials....44
6.3 Scoring Process Overview....45
6.4 CTB Scoring System....45
6.4.1 Prework....46
6.4.2 Receiving....46
6.4.3 Login....47
6.4.4 Scanning....48
6.4.5 Editing/Updates....48
6.4.6 Document Retention....49
6.5 Handscoring Process....49
6.5.1 Anchor Papers and Training Papers....50
6.5.2 Training....50
6.5.3 Intra-Rater Reliability....51
6.5.4 Scoring Personnel - Constructed Response Scorers....51
6.6 Quality Assurance of Custom Software....52
6.7 Advanced Function Printing (AFP)....53
6.8 Distribution of CR Items....54

Part 7: Data and Analysis....56

7.1 14 Data: Calibration Sample and Census Data....56
7.1.1 Proportion of Students by the 5 NCLB Groupings....56
7.1.2 Comparison of Scale Score....58
7.2 Valid Records in Calibration Sample....58
7.3 Descriptive Statistics by Test....58
Reading....60
Mathematics....60
Language Arts....61
Social Studies....61
Science....61
7.4 Classical Item Analysis....62
Reading....64
Mathematics....64
Language Arts....65
Social Studies....65
Science....65
7.4.1 Speededness....66

Part 8: Calibration and Scaling....67

8.1 Calibration Methods....67
8.1.1 Calibration Models....67
8.1.2 Calibration Software....68
8.2 Scaling Procedures....69
8.2.1 Reading and Mathematics....69
8.2.2 Language Arts, Social Studies, and Science....69
8.3 Calibration and Scaling Results....70
8.3.1 IRT Item Parameters....70
8.3.2 IRT Item Fit....71

TABLE OF CONTENTS CONT'D

8.3.3 Scoring and Standard Error of Measurement 72

Part 9: Test Results 74

9.1 Summary Statistics for Scale Scores 74
Reading 75
Mathematics 75
Language Arts 75
Social Studies 76
Science 76
9.1.1 LOSS and HOSS 76
9.2 Performance Level Information 77
Reading 77
Mathematics 78
Language Arts 78
Social Studies 79
Science 79
9.3 Standard Performance Indicator (SPI) for Content Standard 80
Reading 81
Mathematics 82
Language Arts 82
Social Studies 83
Science 83

Part 10: Reliability 84

10.1 Measures of Internal Consistency and SEM 85
10.2 Classification Consistency and Accuracy 87
10.3 Inter-rater Reliability for CR Items and Writing Prompts 91
Reading 93
Mathematics 94
Writing 94

Part 11: Validity 95

11.1 Differential Item Functioning 97
11.2 Construct Validity 102
11.3 Erasure Analysis 103

Part 12: Cut Scores and Descriptor Writing 104

12.1 Cut Scores 104
12.2 Descriptor Writing 105

Tables and Figures 109

References 440

TABLE OF TABLES

PART 2

Table 2-1 Wisconsin WKCE-CRT Assessment Limits Meeting, August 25–28, 2003 110
Table 2-2 Wisconsin WKCE-CRT Reading Passage Review Meeting..... 112
Table 2-3 Checklist for the Content Reviewer 114
Table 2-4 Checklist for the Sensitivity Reviewer..... 115
Table 2-5 November 2005 Item Selection Review Participants..... 116
Table 2-6 January 2007 Item Selection Review Participants 117
Table 2-7 Item Content Review Results, December 2003 118
Table 2-8 Item Content Review Results, April 2004 119
Table 2-9 Item Content Review Results, March 2005 120
Table 2-10 Item Content Review Results, November 2005..... 121
Table 2-11 Item Content Review Results, January 2007 122

PART 3

Table 3-1 Reading Passage Review Results, September, 2003..... 123
Table 3-2 Reading Passage Review Results, December 2004 123
Table 3-3 Reading Passage Review Results, August 2005 123
Table 3-4 Reading Test Blueprint: Grades 3–8, 10..... 124
Table 3-5 Mathematics Test Blueprint: Grades 3–8, 10..... 126
Table 3-6 Language Arts Test Blueprint: Grades 4, 8, 10..... 128
Table 3-7 Science Test Blueprint: Grades 4, 8, 10..... 128
Table 3-8 Social Studies Test Blueprint: Grades 4, 8, 10 128
Table 3-9 Reading Test Structure..... 129
Table 3-10 Reading Test Structure Cont'd..... 130
Table 3-11 Mathematics Test Structure 130
Table 3-12 Mathematics Test Structure Cont'd 131
Table 3-13 Language Arts Test Structure 132
Table 3-14 Social Studies Test Structure 133
Table 3-15 Science Test Structure 134

PART 4

Table 4-1 Item Development Each Year and Total to Date 135
Table 4-2 Unique Items Field Tested Each Year and Total to Date..... 136
Table 4-3 Reading Grade 3 2005 Item Dev. and November 2005 Item Rev. Results..... 137
Table 4-4 Reading Grade 4 2005 Item Dev. and November 2005 Item Rev. Results..... 138
Table 4-5 Reading Grade 5 2005 Item Dev. and November 2005 Item Rev. Results..... 139
Table 4-6 Reading Grade 6 2005 Item Dev. and November 2005 Item Rev. Results..... 140
Table 4-7 Reading Grade 7 2005 Item Dev. and November 2005 Item Rev. Results..... 141
Table 4-8 Reading Grade 8 2005 Item Dev. and November 2005 Item Rev. Results..... 142
Table 4-9 Mathematics Grade 3 2005 Item Dev. and November 2005 Item Rev. Results 143
Table 4-10 Mathematics Grade 4 2005 Item Dev. and November 2005 Item Rev. Results 145
Table 4-11 Mathematics Grade 5 2005 Item Dev. and November 2005 Item Rev. Results 147
Table 4-12 Mathematics Grade 6 2005 Item Dev. and November 2005 Item Rev. Results 149
Table 4-13 Mathematics Grade 7 2005 Item Dev. and November 2005 Item Rev. Results 151
Table 4-14 Mathematics Grade 8 2005 Item Dev. and November 2005 Item Rev. Results 153
Table 4-15 Science Grade 4 2005 Item Dev. and November 2005 Item Rev. Results..... 155
Table 4-16 Science Grade 8 2005 Item Dev. and November 2005 Item Rev. Results..... 158
Table 4-17 Science Grade 10 2005 Item Dev. and November 2005 Item Rev. Results..... 161
Table 4-18 Reading: 2006 Item Dev. Plan by Grade, Reporting Category, and Item Type..... 162
Table 4-19 Reading: 2006 Item Dev. Plan by Rationale and Item Type..... 163
Table 4-20 Mathematics: 2006 Item Dev. Plan by Grade, Category, and Item Type 164

TABLE OF TABLES CONT'D

PART 4 CONT'D

Table 4-21 Mathematics: 2006 Item Dev. Plan by Rationale and Item Type 166
Table 4-22 Mathematics Grade 3 2006 Item Dev. and January 2007 Item Rev. Results 167
Table 4-23 Mathematics Grade 4 2006 Item Dev. and January 2007 Item Rev. Results 169
Table 4-24 Mathematics Grade 5 2006 Item Dev. and January 2007 Item Rev. Results 171
Table 4-25 Mathematics Grade 6 2006 Item Dev. and January 2007 Item Rev. Results 173
Table 4-26 Mathematics Grade 7 2006 Item Dev. and January 2007 Item Rev. Results 175
Table 4-27 Mathematics Grade 8 2006 Item Dev. and January 2007 Item Rev. Results 177
Table 4-28 Reading Grade 3 2006 Item Dev. and January 2007 Item Rev. Results 179
Table 4-29 Reading Grade 4 2006 Item Dev. and January 2007 Item Rev. Results 180
Table 4-30 Reading Grade 5 2006 Item Dev. and January 2007 Item Rev. Results 181
Table 4-31 Reading Grade 6 2006 Item Dev. and January 2007 Item Rev. Results 182
Table 4-32 Reading Grade 7 2006 Item Dev. and January 2007 Item Rev. Results 183
Table 4-33 Reading Grade 8 2006 Item Dev. and January 2007 Item Rev. Results 184

PART 6

Table 6-1 Reading Rubric, Grades 3–8 and 10 185
Table 6-2 Mathematics Rubric, Grades 3–8 and 10 186
Table 6-3 Writing Rubric, Conventions of Written English, Grade 4 187
Table 6-4 Writing Rubric, Composing, Grade 4 188
Table 6-5 Writing Rubric, Conventions of Written English, Grade 8 189
Table 6-6 Writing Rubric, Composing, Grade 8 190
Table 6-7 Writing Rubric, Conventions of Written English, Grade 10 192
Table 6-8 Writing Rubric, Composing, Grade 10 193
Table 6-9 Score Distribution for Reading* 195
Table 6-10 Score Distribution for Mathematics* 196
Table 6-11 Score Distribution by for Grade 4, 8, and 10 Writing Composing 199
Table 6-12 Percentage Distribution of Scores, Grade 4, 8, and 10 Writing Composing 199
Table 6-13 Score Distribution, Grade 4, 8, and 10 Writing Convention 200
Table 6-14 Percentage Distribution of Scores for Grade 4, 8, and 10 Writing Convention 200
Table 6-15 Score Distribution for Grade 4, 8, and 10 Writing Total Score 201
Table 6-16 Percent for Grade 4, 8, and 10 Writing Total Score 201

PART 7

Table 7-1 The Current 14 Calibration Districts 202
Table 7-2 Number and Percent of Census and 14 CD Students, by Gender* 203
Table 7-3 Number and Percent of Census and 14 CD Students, by Ethnicity* 204
Table 7-4 Number and Percent of Census and 14 CD Students, by SES 205
Table 7-5 Number and Percent of Census and 14 CD Students, by Disability 206
Table 7-6 Number and Percent of Census and 14 CD Students, by ELP 207
Table 7-7 Scale Score Mean and Standard Deviation for Census and 14 CD Results 208
Table 7-8 Total Number of Students for Census and 14 CD 209
Table 7-9 Test Configuration of Operational (OP) and Field Test (FT) Items 210
Table 7-10 Raw Score Descriptive Statistics based on Census Data 211
Table 7-11 Raw Score Descriptive Statistics based on 14 Districts 212
Table 7-12 Raw Score Descriptive Statistics by Gender 213
Table 7-13 Raw Score Descriptive Statistics for Reading by Ethnicity 214
Table 7-14 Raw Score Descriptive Statistics for Mathematics by Ethnicity 215
Table 7-15 Raw Score Descriptive Statistics for Language Arts by Ethnicity 216
Table 7-16 Raw Score Descriptive Statistics for Social Studies by Ethnicity 216
Table 7-17 Raw Score Descriptive Statistics for Science by Ethnicity 217
Table 7-18 Raw Score Descriptive Statistics by Socioeconomic Status 218

TABLE OF TABLES CONT'D

PART 7 CONT'D

Table 7-19 Raw Score Descriptive Statistics by Disability 219
Table 7-20 Raw Score Descriptive Statistics by English Language Proficiency 220
Table 7-21 Item Analysis Grade 3 Reading 221
Table 7-22 Item Analysis Grade 4 Reading 225
Table 7-23 Item Analysis Grade 5 Reading 229
Table 7-24 Item Analysis Grade 6 Reading 234
Table 7-25 Item Analysis Grade 7 Reading 238
Table 7-26 Item Analysis Grade 8 Reading 243
Table 7-27 Item Analysis Grade 10 Reading 247
Table 7-28 Item Analysis Grade 3 Mathematics 249
Table 7-29 Item Analysis Grade 4 Mathematics 253
Table 7-30 Item Analysis Grade 5 Mathematics 257
Table 7-31 Item Analysis Grade 6 Mathematics 261
Table 7-32 Item Analysis Grade 7 Mathematics 265
Table 7-33 Item Analysis Grade 8 Mathematics 269
Table 7-34 Item Analysis Grade 10 Mathematics 273
Table 7-35 Item Analysis Grade 4 Language Arts* 276
Table 7-36 Item Analysis Grade 8 Language Arts* 278
Table 7-37 Item Analysis Grade 10 Language Arts 280
Table 7-38 Item Analysis Grade 4 Social Studies 282
Table 7-39 Item Analysis Grade 8 Social Studies 284
Table 7-40 Item Analysis Grade 10 Social Studies 286
Table 7-41 Item Analysis Grade 4 Science 288
Table 7-42 Item Analysis Grade 8 Science 292
Table 7-43 Item Analysis Grade 10 Science 296
Table 7-44 The Number of Items Flagged* 299

PART 8

Table 8-1 Item Flagged Based on Yen's Q_1 300
Table 8-2 Scoring Table for Reading Grade 3 301
Table 8-3 Scoring Table for Reading Grade 4 302
Table 8-4 Scoring Table for Reading Grade 5 303
Table 8-5 Scoring Table for Reading Grade 6 304
Table 8-6 Scoring Table for Reading Grade 7 305
Table 8-7 Scoring Table for Reading Grade 8 306
Table 8-8 Scoring Table for Reading Grade 10 307
Table 8-9 Scoring Table for Mathematics Grade 3 308
Table 8-10 Scoring Table for Mathematics Grade 4 309
Table 8-11 Scoring Table for Mathematics Grade 5 310
Table 8-12 Scoring Table for Mathematics Grade 6* 311
Table 8-13 Scoring Table for Mathematics Grade 7 312
Table 8-14 Scoring Table for Mathematics Grade 8 313
Table 8-15 Scoring Table for Mathematics Grade 10 314
Table 8-16 Scoring Table for Language Arts Grade 4 315
Table 8-17 Scoring Table for Language Arts Grade 8 316
Table 8-18 Scoring Table for Language Arts Grade 10 317
Table 8-19 Scoring Table for Social Studies Grade 4 318
Table 8-20 Scoring Table for Social Studies Grade 8 319
Table 8-21 Scoring Table for Social Studies Grade 10 320
Table 8-22 Scoring Table for Science Grade 4 321
Table 8-23 Scoring Table for Science Grade 8 322
Table 8-24 Scoring Table for Science Grade 10 323

TABLE OF TABLES CONT'D

PART 9

Table 9-1 Scale Score Descriptive Statistics based on Census Data 324
Table 9-2 Scale Score Descriptive Statistics based on 14 Districts 325
Table 9-3 Scale Score Descriptive Statistics by Gender 326
Table 9-4 Scale Score Descriptive Statistics for Reading by Ethnicity 327
Table 9-5 Scale Score Descriptive Statistics for Mathematics by Ethnicity 328
Table 9-6 Scale Score Descriptive Statistics for Language Arts by Ethnicity 329
Table 9-7 Scale Score Descriptive Statistics for Social Studies by Ethnicity 330
Table 9-8 Scale Score Descriptive Statistics for Science by Ethnicity 330
Table 9-9 Scale Score Descriptive Statistics by Socioeconomic Status 331
Table 9-10 Scale Score Descriptive Statistics by Disability 332
Table 9-11 Scale Score Descriptive Statistics by English Language Proficiency 333
Table 9-12 The Number of Students and Percents at LOSS and HOSS 334
Table 9-13 Summary Statistics for Reading Content Standards Raw and SPI Scores 335
Table 9-14 Summary Statistics for Mathematics Content Standards Raw and SPI Scores 337
Table 9-15 Summary Statistics for Language Arts Content Standards Raw and SPI Scores 339
Table 9-16 Summary Statistics for Social Studies Content Standards Raw and SPI Scores 340
Table 9-17 Summary Statistics for Science Content Standards Raw and SPI Scores 341
Table 9-18 Performance Level Cut Scores for all Contents 342
Table 9-19 Pct. of Students in Each Performance Level by Sub-Group (Reading) 343
Table 9-20 Pct. of Students in Each Performance Level by Sub-Group (Mathematics) 346
Table 9-21 Pct. of Students in Each Performance Level by Sub-Group (Language Arts) 349
Table 9-22 Pct. of Students in Each Performance Level by Sub-Group (Social Studies) 350
Table 9-23 Pct. of Students in Each Performance Level by Sub-Group (Science) 351

PART 10

Table 10-1 Reliability for Total Group and Subgroups 352
Table 10-2 Standard Error of Measurement for Total Group and Subgroups 353
Table 10-3 Cronbach Alpha Reliability Coefficients for Content Standards 354
Table 10-4 Standard Error of Measurement per Content Standard 355
Table 10-5 Classif. Consistency and Classif. Accuracy for Reading Grade 3 356
Table 10-6 Classif. Consistency and Classif. Accuracy for Reading Grade 4 357
Table 10-7 Classif. Consistency and Classif. Accuracy for Reading Grade 5 358
Table 10-8 Classif. Consistency and Classif. Accuracy for Reading Grade 6 359
Table 10-9 Classif. Consistency and Classif. Accuracy for Reading Grade 7 360
Table 10-10 Classif. Consistency and Classif. Accuracy for Reading Grade 8 361
Table 10-11 Classif. Consistency and Classif. Accuracy for Reading Grade 10 362
Table 10-12 Classif. Consistency and Classif. Accuracy for Mathematics Grade 3 363
Table 10-13 Classif. Consistency and Classif. Accuracy for Mathematics Grade 4 364
Table 10-14 Classif. Consistency and Classif. Accuracy for Mathematics Grade 5 365
Table 10-15 Classif. Consistency and Classif. Accuracy for Mathematics Grade 6 366
Table 10-16 Classif. Consistency and Classif. Accuracy for Mathematics Grade 7 367
Table 10-17 Classif. Consistency and Classif. Accuracy for Mathematics Grade 8 368
Table 10-18 Classif. Consistency and Classif. Accuracy for Mathematics Grade 10 369
Table 10-19 Classif. Consistency and Classif. Accuracy for Language Arts Grade 4 370
Table 10-20 Classif. Consistency and Classif. Accuracy for Language Arts Grade 8 371
Table 10-21 Classif. Consistency and Classif. Accuracy for Language Arts Grade 10 372
Table 10-22 Classif. Consistency and Classif. Accuracy for Social Studies Grade 4 373
Table 10-23 Classif. Consistency and Classif. Accuracy for Social Studies Grade 8 374
Table 10-24 Classif. Consistency and Classif. Accuracy for Social Studies Grade 10 375
Table 10-25 Classif. Consistency and Classif. Accuracy for Science Grade 4 376
Table 10-26 Classif. Consistency and Classif. Accuracy for Science Grade 8 377
Table 10-27 Classif. Consistency and Classif. Accuracy for Science Grade 10 378

TABLE OF TABLES CONT'D

PART 10 CONT'D

Table 10-28 Inter-Rater Reliability Reading 379
Table 10-29 Inter-Rater Reliability Mathematics 382
Table 10-30 Inter-Rater Reliability Language Arts* 391

PART 11

Table 11-1 Items Flagged for DIF, By Gender 392
Table 11-2 Items Flagged for DIF, By Ethnicity, African American 393
Table 11-3 Items Flagged for DIF, By Ethnicity, Hispanic 394
Table 11-4 Items Flagged for DIF, By Ethnicity, Asian 395
Table 11-5 Items Flagged for DIF, By Ethnicity, American Indian 398
Table 11-6 Items Flagged for DIF, By English Language Proficiency 399
Table 11-7 Items Flagged for DIF, By Disability Status 401
Table 11-8 Correlations among Reading Objectives 402
Table 11-9 Correlations among Mathematics Objectives 403
Table 11-10 Correlations among Language Arts Objectives 404
Table 11-11 Correlations among Social Studies Objectives 404
Table 11-12 Correlations among Science Objectives 405
Table 11-13 Factor Analysis 406

PART 12

Table 12-1 WKCE-CRT Descriptor Writing Agenda, June 20–22, 2006 407
Table 12-2 Cut scores and Associated Impact Data for WKCE-CRT Reading 409
Table 12-3 Cut scores and Associated Impact Data for WKCE-CRT Mathematics 409
Table 12-4 Cut scores and Associated Impact Data for WKCE-CRT Language Arts 410
Table 12-5 Cut scores and Associated Impact Data for WKCE-CRT Social Studies 410
Table 12-6 Cut scores and Associated Impact Data for WKCE-CRT Science 411

TABLE OF FIGURES

PART 4

Figure 4-1 CTB's Item Development Process 412

PART 8

Figure 8-1 TCC Curve for Reading Grades 3-8, 10 415
Figure 8-2 TCC Curve for Mathematics Grades 3-8, 10 416
Figure 8-3 TCC Curve for Language Arts Grades 4, 8, 10 417
Figure 8-4 TCC Curve for Social Studies Grades 4, 8, 10 418
Figure 8-5 TCC Curve for Science Grades 4, 8, 10 419
Figure 8-6 SEM Curves, Reading Grades 3-6 420
Figure 8-7 SEM Curves, Mathematics Grades 3-6 422
Figure 8-8 SEM Curves, Language Arts Grades 4, 8, 10 424
Figure 8-9 SEM Curves, Social Studies Grades 4, 8, 10 425
Figure 8-10 SEM Curves, Science Grades 4, 8, 10 426

PART 11

Figure 11-1 Reading Indices for Classif. Consistency and Classif. Accuracy 427
Figure 11-2 Mathematics Indices for Classif. Consistency and Classif. Accuracy 427
Figure 11-3 Language Arts Indices for Classif. Consistency and Classif. Accuracy 428

TABLE OF FIGURES CONT'D

PART 11 CONT'D

Figure 11-4 Social Studies Indices for Classif. Consistency and Classif. Accuracy 428
Figure 11-5 Science Indices for Classif. Consistency and Classif. Accuracy 429

PART 12

Figure 12-1 Cut Scores for Reading 430
Figure 12-2 Cut Scores for Mathematics 431
Figure 12-3 Cut Scores for Language Arts 432
Figure 12-4 Cut Scores for Social Studies 433
Figure 12-5 Cut Scores for Science 434
Figure 12-6 Percent of Students for Reading 435
Figure 12-7 Percent of Students for Mathematics 436
Figure 12-8 Percent of Students for Language Arts 437
Figure 12-9 Percent of Students for Social Studies 438
Figure 12-10 Percent of Students for Science 439

Part 1: Executive Summary

The present report documents the processes and procedures applied in the Fall 2006 WKCE, and the results thereof. This report shows how the processes, procedures, and the results relate to the issues of validity and reliability, and to the *Standards for Educational and Psychological Testing* (American Education Research Association, American Psychological Association National Council on Measurement in Education, 1999). This report demonstrates that the processes and procedures applied in the Fall 2006 WKCE adhered to 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 the Fall 2006 WKCE.

The entire WCKE program and the Technical Report are constructed with reference to a specific policy environment. Perhaps the most fundamental feature of that environment is the *Improving America's Schools Act* of 1994, which requires that states establish challenging academic standards as well as aligned annual assessments to evaluate them. 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. The Technical Report is an important part of the state's reporting efforts within that context.

Wisconsin students in grades 4, 8, and 10 began taking the Wisconsin Knowledge and Concepts (WKCE) norm-referenced assessments in the 1997 school year. The assessments used at that time were *TerraNova* tests developed by CTB/McGraw-Hill. The selection of those tests was partly predicated on an awareness of 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 Model Academic Standards*, chaired by then-current Lieutenant Governor McCallum and the Wisconsin Department of Public Instruction (WDPI). The *Model Academic Standards* would measure student performance in the same subjects as the *TerraNova* tests.

Beginning in the 2005–2006 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.

Specifically, *No Child Left Behind* requires that, beginning with the 2005-06 school year, each state must measure every child's progress in Reading and Mathematics in each of grades 3 through 8 and at least once during grades 10 through 12. By school year 2007-2008, states must also have in place Science assessments to be administered at least once during grades 3-5, grades 6-9, and grades 10-12. Students in grades 4, 8 and 10 are, and will continue to be assessed in Language Arts, Science and Social Studies as required by s. 118.30 Wisconsin Statutes.

It is within this policy context that the WKCE was reconstructed, as a criterion-referenced test, for the Fall 2005 administration, replacing the previously existing norm-referenced WKCE

Reading and Mathematics tests. The new criterion-referenced WKCE was designed specifically for Wisconsin students, and specifically to measure their performance on the Wisconsin Model Academic Standards adopted by the state. The new criterion-referenced WKCE assessments were designed to evaluate students' knowledge and to measure achievement in the basic skills taught in schools at grades 3-8 and 10. The Fall 2006 WKCE is the second administration of this new criterion-referenced assessment.

The WKCE tests consist of criterion-referenced items written by Wisconsin teachers and items from CTB/McGraw-Hill's norm-referenced test, *TerraNova, The Second Edition*® (*TerraNova*; CTB/McGraw-Hill, 2001). The WKCE tests include criterion-referenced tests in Mathematics and Reading at grades 3-8 and 10. At grades 4, 8 and 10 students are also tested in Science, Social Studies and Language Arts (including Writing) which are *TerraNova* tests or, as in the case of Science, the items are sourced from various *TerraNova* Science Assessments. The Writing tests are single prompt essay tests scored using a holistic rubric for composition and an analytic rubric for convention.

The structure of the present Technical Report mirrors the testing cycle. The WKCE begins with and is rooted in the input of WI educators and the *Wisconsin Model Academic Standards*. On that basis, a design was derived, developed, and administered. A scale score was then developed using models derived from item-response theory, the test was scored, and the results were analyzed. The reliability and validity of all applied processes, procedures, and the results were evaluated. The entire testing cycle was continuously monitored for any issues that might be investigated further, and recommendations for future years were considered. The Technical Report documents all of these aspects of the testing cycle, and others, in the subsequent chapters. A brief content summary of the Technical Report is provided below.

Design

- Wisconsin educators were involved in design at every step to insure 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 to review reading passage and page specifications. The test specifications documents created and later approved by WDPI serve as a foundation for item and test development through 2007-2008 and will continue through 2008-2009.
- Wisconsin Model Academic Standards were translated into the grade-level content frameworks, which in turn, formed the basis for test blueprints and item specifications.

Item and Form Development

- Item development was based on the approved test blueprints, with a sufficient quantity of items written to develop multiple operational test forms.
- Items were edited, reviewed, and re-reviewed at sessions involving Wisconsin educators.
- The resulting items underwent field testing in May and December 2004 and embedded field testing in Fall 2005 and Fall 2006. A committee of educators met in October 2004 to review items field tested in May 2004 that were flagged for differential item functioning.
- At of the end of 2005, a total of 5,248 MC and 797 CR items (6,045 total) had been developed, and by Fall 2005 3,266 MC and 299 CR (3,565 total) items had been field tested. At of the end of 2006 3,726 MC and 339 CR (4,065 total) had been field tested.
- Selection of the Fall 2006 operational forms was done using the ITEMWIN software similar to Fall 2005 for all grades and content areas.

Administration

- The Test Administration Window was October 23 - November 24, 2006. Delivery of materials was handled through the district and school assessment coordinators.
- Two books were used for the Grades 4, 8 and 10 for the Fall 2006 administration year where there were five content areas tested. In the past administrations, only one book was used per grade. The change here related to manufacturing recommendations for the maximum number of pages per book, and nicely segregated the NCLB content areas. CTB recommended that using two books could avoid manufacturing defects that occurred in Fall 2005.

Scoring

- CTB's primary goal in scoring and processing test documents is to deliver quality results to Wisconsin educators, students and parents. The processes and procedures applied to control the flow of testing materials from WI students though the scoring facilities, the application of quality assurance measures, the machine scanning process, and the hand scoring process supported that goal.
- Handscoring processes applied in the WKCE program, including the development and review of the scoring rubrics, anchor papers, and writing prompts, as well as the training of scoring personnel, ongoing handscoring quality assurance, the application of an inter-rater reliability assessment, and a systematic review of the resulting score distributions supported the development of reliable scores for constructed response items.
- Since the WKCE includes constructed response (CR) items, which are hand-scored, scoring rubrics (which are tools or sets of specific criteria used to evaluate a student's

performance) were required in order to standardize the handscoring process for the CR items. The rubrics used in the WKCE are presented in detail for all relevant content areas, i.e. Reading, Math and Writing. The rubric incorporates a holistic scoring scale. The Writing rubric indicates how scoring covers both convention and composing.

Data and Item Analysis

- Scoring and all norms for the Fall 2006 WKCE test were based on the 14 calibration districts. These districts are identified in the Technical Report. These districts were successfully used for the same purposes in previous administrations. The characteristics of the 5 NCLB groupings (gender, ethnicity, socioeconomic status, disability, and English Language Proficiency) as derived from the 14 calibration districts were compared to those of the total WI student population, and were demonstrated to be a representative sample of the state's students. Mean scale scores and standard deviations were also derived from the calibration district data, compared to the results from the entire student population, and further demonstrated the validity of using these 14 districts. The methods determining when student records are valid, and should be included in the sample data, are explained.
- Summary statistics for raw scores are presented based on all students that took the Fall 2006 assessment.
- Evidence in support of reliability of the test is established through an item analysis which includes Cronbach's alpha, standard error of measurement (SEM), and the omit rate. The item analysis proceeds through the items in the order they appeared in the test booklets.

Calibration and Scaling

- Item-pattern scoring was applied to the Fall 2006 WKCE, and is generally recommended over raw scores because it produces more accurate scores for individual students; the methodology is described in this section.
- The Fall 2006 WKCE was calibrated and scaled using methods and models derived from item response theory which is the basis of most large scale standardized testing. The items for Fall 2006 were calibrated for each grade and content and the procedures applied are discussed.
- A new Wisconsin scale was constructed in Fall 2005. There is a vertical relationship of the scales across grades. Two different item response theory models were used, one for CR items and one for and MC items.
- Calibration and scaling results and scoring tables which include standard error of measurement are also presented.

Test Results

- Summary descriptive statistics for scale score were reported for all students and the 5 NCLB groupings.
- The percentage of students at each performance level was analyzed.
- Summary descriptive statistics for a Standardized Performance Indicator (SPI) were reported.

Reliability

- Internal consistency was assessed for all MC and CR items using Chronbach's alpha.
- Standard error of measurement (SEM) was assessed for raw score and scale score.
- Classification consistency and accuracy were estimated for performance classification.
- Inter-rater reliability was estimated for all of the CR items.

Validity

- Differential item functioning was performed for all 5 NCLB groupings.
- Factor analysis and correlations among content standards were presented as construct validity.
- Erasure analysis was performed to identify high erasure rates.
- The chapter on validity also reviews the main validity issues discussed in all prior chapters.

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?

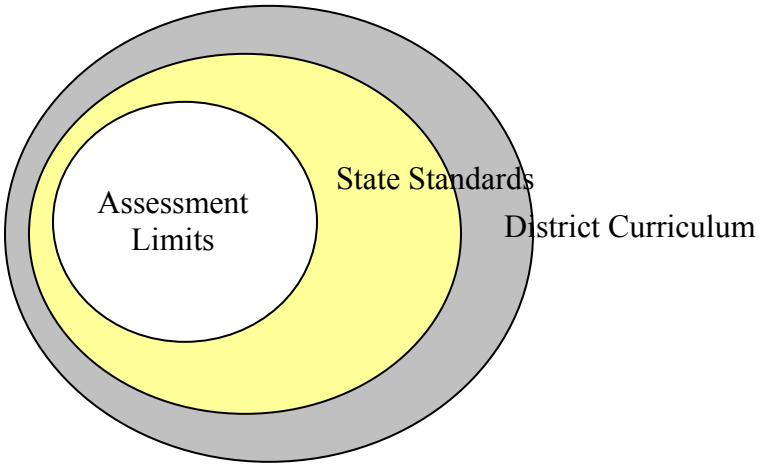
Establishing the content framework and eligible test content for Reading and Mathematics began in August 2003 with a workshop with Wisconsin Educators. This meeting is described in detail in section 2.1. At this workshop facilitated by CTB and WDPI staff, educators considered what summative test information 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 5: Test Administration

Part 5 of the Technical Report describes a set of standardized procedures and policies applied to administer the WKCE. This section includes a discussion of the accommodations provided to examinees through the alternative assessment, the Wisconsin Student Assessment System's (WSAS) Alternative Assessments (WAA). Eligibility criteria for the alternative assessment are discussed, and relevant aspects of the policy environment are also cited. Test security is also discussed briefly in Part 6. The issue of test security has important implications for the integrity of the results and thus the validity of the 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 here. The following AERA/ APA/ NCME standards are addressed here: 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.

5.1 Accommodations

The Wisconsin Department of Public Instruction (WDPI) 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. For the overwhelming majority of students, a major component of accountability is achieved through administration of the Wisconsin Student Assessment System (WSAS) which includes the WKCE. For a small group of English language learners and special education students, however, assessment of progress using WSAS may be inappropriate. An alternate system of assessment directly aligned with Wisconsin's Model Academic Standards is required to meet both the spirit and letter of the federal No Child Left Behind Act (NCLB) Act of 2001 and PI 13, Wisconsin Administrative Code.

Wisconsin Student Assessment System (WSAS) Alternate Assessments are standards-based alternatives to WKCE tests at grades 3–8 and 10 and consist of WDPI-approved protocols and rubrics for the local collection and local scoring of student work. Nearly all students at grades 3–8 and 10, including most students with disabilities and students with limited English proficiency are expected to take the WKCE, with allowable accommodations as needed. WSAS Alternate Assessments are given in lieu of the specific WKCE subject area tests to two student groups: (1) students with more severe disabilities if the local Individualized Education Program (IEP) team determines that the students are not able to demonstrate at least some of the knowledge and skills on the WKCE subject area tests and (2) students whose first language is not English and whose academic English skills are at a beginning level as determined by the individual school districts.

Students with disabilities who have an IEP, or who have a 504 plan, may be considered for standard accommodations. Also, students identified as Limited English Proficient (LEP) and students who have been identified as Fluent English Proficient (FEP) for no more than two years may be considered for standard accommodations.

For the purposes of assessment, a special education student is eligible to receive services under the Individuals with Disabilities Education Act–1997 and has an Individualized Education

Program (IEP). For the purposes of assessment, a 504 student is eligible under Section 504 of the Rehabilitation Act of 1973 and has a 504 Accommodation Plan.

An English Language Learner (ELL) is a student whose native language is other than English and is learning English as a second language. By definition, the English language proficiency of these students precludes them from meaningful participation in the WKCE. The WSAS Alternative Assessment for English Language Learners (WAA-ELL) allows these students access to the same concepts as their English proficient peers while minimizing the influence of language. Alternate assessment for English language learners entails the collection, analysis, and interpretation of original student work in Reading and Mathematics at grade levels 3–8 and 10. In addition, comprehensive assessment of Language Arts, Science and Social Studies occurs in grades 4, 8 and 10. WAA-ELL is based on the identical set of state content standards in Reading, Mathematics, English Language Arts, Science and Social Studies that exists for all students, and it includes alternate performance indicators (APIs) aligned to those standards for English language learners.

5.1.1 Standard Accommodations

Standard accommodations are provisions made in how a student accesses and demonstrates learning. These should not substantially change the instructional level, the content, or the performance criteria. The changes are made in order to provide a student equal access to learning and equal opportunity to demonstrate what is known. Standard accommodations are changes in the routine conditions under which students take assessments, and involve changes in:

- Timing or scheduling of the test (i.e., administration of the test in short intervals or at a time of day that takes into account a student's medical needs);
- Test setting (i.e., administration of the test individually or in a small group setting, under special lighting, or using special furniture);
- Test presentation (i.e., test questions presented in large print or Braille, repeated directions, or explanation of directions); or
- How the student responds to test questions (i.e., the student points to answers or records answers in the test booklet instead of the answer booklet).

In addition, a standard accommodation that is available to English Learners is limited oral translation in the student's native language. When this accommodation is provided on state assessments only the verbal directions stated by the Test Administrator and the written directions that the student is expected to read may be orally translated into the student's native language. The translation must be an exact translation which is as close to verbatim as possible, and translation is to be provided on an as-needed basis only. Translating any test item or translations that paraphrase, simplify, or clarify directions, or written translations are not permitted.

Table 5-1 describes in detail all standard accommodations made available for the Fall 2006 WKCE assessments.

Explanations of asterisks (from Assessment Accommodations for ELL students and students with disabilities/504)

*** Denotes an accommodation involving the use of highlighters. Highlighters may only be used by ELL students and students with disabilities.** Please note: carefully supervise the use of highlighters because they may cause smudging of pencil marks and bubbles and, therefore, could affect reliability of scoring. If highlighters are used, the following guidelines must be followed:

Guidelines for Highlighters (CTB/McGraw-Hill):

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

****Denotes an accommodation for which test security should be considered.**
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 the proctor or a test examiner. In addition, students must not be allowed to use any form of wireless communication during these breaks.

*****Parameters for marking test booklet with No. 2 Pencil**.
Do not mark in the bubble answer positions.
Do not mark in the student Pre-ID Barcode on 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 page of the test booklet).
Do not mark more than one answer bubble as the scanner cannot determine a response.

5.2 Reporting Results of Assessments Taken with Accommodations

Scores of assessments taken with standard accommodations will be included with the results of students who took these tests under standard conditions at the school, district, and state level.

5.3 Test Security

The primary goal of the WKCE test security was to protect the integrity of the examinations. To ensure that trends in achievement results can be calculated across years 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. WDPI ensured that all who had access to test materials understood the critical need for test security, presented during the 2005 and 2006 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 WDPI.

Wisconsin Student Assessment Security Warning: The following statement was directed by WDPI to appear on every test booklet beginning with the 2004-2005 school year:

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.

For additional information on test security, refer also to Part 10, section 10.2.3 for Erasure Analysis.

5.4 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 test coordinators (WDPI, 2006-2007). The Guide included the following topics:

- Test Security
- Schedule of Important Dates
- Responsibilities of District Assessment Coordinators (DACs)
- Responsibilities of School Assessment Coordinators (SACs)
- Scheduling Test Administration
- Preparations for Testing
- Students to be Tested
- Standard Accommodations
- WDPI Guidelines to Facilitate Participation of Students with Special Needs
- Data Collections and Reporting
- Word Processors, Scribes, Tape Recorders, and Large Print and Braille Tests
- Test Materials
- Receiving Test Materials
- Inventorying Test Materials
- Procedures During Test Administration
- Procedures Following Test Administration
- Returning Materials to CTB/McGraw-Hill

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

- Test Materials
- Test Security
- Testing Schedules
- Test Administration Guidelines
- Guidelines for Preparing Students to Take the Test
- Use of Appropriate Testing Procedures
- Student Identification Information
- Explanation of Symbols
- Detailed Scripts for Administration of Each Part of Each Test
- Procedures Following Test Administration

For specific information related to test administration, refer to the Test Coordinator's Manual and/or the Test Administration Manuals (http://WDPI.wi.us/oea/kce_publin.html).

Part 6: Scoring

Part 6 of the Technical Report serves to describe how multiple choice items and constructed response items were scored. As described below, the multiple-choice items were scored using optical scanning software and large scale scanning machines, and the constructed-response items were scored by human raters. Part 6 details how the completed testing materials from WI students flowed though the scoring facilities, how they were tracked and monitored, the application of several quality assurance measures along the way, the machine scanning process, and the hand scoring process as well. The scoring rubrics used for scoring the constructed response items are presented, the process of selecting evaluators is described (their background and training requirements are specified), and the hand scoring process itself is detailed. Finally, Part 6 also provides distributions of examinee responses across score levels, to check the reasonability of CR items.

The *Standards* emphasize that "[s]tandardized scoring procedures help to ensure accurate scoring and reporting…. When scoring is done by machine, the accuracy of the machine is at issue, including any scoring algorithm. When scoring is done by human judges, scorers require careful training" (p. 62). The fundamental purpose of Part 6 is to demonstrate adherence to AERA/APA/ NCME standards for scoring, and specifically standards 5.8 (on the documentation of scoring process) and 5.9 (on scoring rubrics).

6.1 Rubric Description

For the 2006 operational test administration, the Reading and Mathematics tested at grades 3–8 and had constructed response items. In addition, a Writing prompt was administered at grades 4, 8, and 10. The Writing prompt at grades 4, 8, and 10 was scored using two holistic rubrics: a 3-point rubric for Conventions of written English, and a 6-point rubric for Composing. Tables 6-1 to 6-8 present the rubrics.

6.2 Processing and Scoring of Test Materials

As described below, the processing and scoring of testing materials ensured that Wisconsin educators, students, and parents received accurate timely data. CTB's WKCE programs Scoring Team, which is based in Monterey, California, coordinated the Fall 2006 operational processing at the Indianapolis, Indiana Scoring facility and the handscoring services through the Mather, California, Delran, New Jersey and Orlando, Florida handscoring teams.

CTB applied handscoring processes, controls, and quality assurance practices to the Fall 2006 administration which are uniformly administered at all scoring facilities. A Scoring Team of trained technical specialists was responsible for coordinating all scoring and reporting activities related to the processing of the WKCE test documents. Document preparation, interdepartmental coordination and communication, processing specifications, and problem resolution were performed by the assigned Scoring Project Manager from the WKCE programs

team. The scoring team worked closely with all CTB departments to ensure the successful scoring and reporting of the WKCE program.

6.3 Scoring Process Overview

CTB conducted quality control checks at all phases of test processing, scoring, data analysis and reporting, and integrates the quality controls into a comprehensive end-to-end quality assurance program for all deliverables. Established procedures required all personnel and subject matter experts to complete their tasks according to strict quality assurance regulations within the data processing, scoring and reporting areas.

6.4 CTB Scoring System

In the Fall of 2005, CTB re-engineered the scoring operations process to organize previously separate functional work areas into organized workcells. The goal was to optimize efficiency and ensure accuracy through each step of the document handling and document scoring process. Each workcell, or self-contained, cross-functional team, was assembled to provide focused, individualized processing specifically designed to meet specific contract requirements. Each workcell was equipped with the skillsets (equipment and personnel) to efficiently and accurately organize, scan, edit and prepare for document retention a group (generally defined by District) of student documents in an efficient document flow/document management system.

The workcell structure or layout was made up of the necessary processing stations to complete the operational processing cycle for student test documents. Within the workcell, documents were organized or staged in preparation for scanning, image-scanned using customized scanning software and high-speed image scanners, and processed through a post-scanning editing cycle running CTB's proprietary Winscore system. Once complete, documents were prepared for secure storage and entered into CTB's document retention system. Documents moved directly from process to process, or sat only momentarily in mini-queues. The result was a significant increase in through-put rates, while increasing efficiency and quality of the process-focused communication. Each station worked together in a smoothly-functioning, self-regulating, continuously-improving cycle.

CTB used scanning and scoring technology to capture all multiple-choice and student demographic data, as well as data from all ancillary documents. Through use of high-performance equipment, combined with well-established processes and systems, CTB ensured the accurate and efficient capture of student data. Scanning production systems fully integrated with image-scoring systems, allowed imaged responses to be separated for handscoring and merged back into the scored multiple-choice data file for use in data analysis and report production.

This process was refined during the Fall of 2005 and the Spring of 2006. Within each step of the process cycle, CTB integrated procedures to ensure quality. Presented below, in order of occurrence, are quality assurance procedures of the Scoring and Reporting process.

6.4.1 Prework

Prior to document arrival at CTB, the scoring team utilized available customer data to prepared materials to expedite the document-handling process. Pre-work involved the delivery by WDPI of final enrollment data by mid-August and two bio-data update files on October 1st and November 1st. Using these files, Scoring created Test Coordinator Kids that were delivered to the districts by October 9th and two deliveries of student barcode labels that were onsite by October 18th and November 10th , respectively. The actual case count for Fall 2006 was 441,979 records. Team members verified the accuracy of the following materials:

- Expected number of students by grade and school
- Test date
- Student Barcode Labels
- Precoded headers generated from school/district enrollment files
- Return Shipping Labels
- Report services specifications
- Sample reports
- Report collation examples
- Report packing schematics
- Document type (i.e., multiple choice/constructed-response)
- Packing lists generated for report shipments
- Other requirements to meet WKCE specifications

Prior to receipt of answer documents, detailed scoring specifications for WKCE programs were distributed to all departments and workstations involved in the scoring and editing process.

6.4.2 Receiving

Shipments were tracked electronically, from the time of pickup at the sites, until delivery at CTB. After receipt, documents were organized by district and the following steps performed:

The box count was verified against the carrier's bill of lading and/or box count indicators as printed on the outside of the box. If a discrepancy was encountered, boxes were placed in a problem resolution area and discrepancy procedures were enforced. If missing boxes were not located within 24 hours, the Scoring Team was notified to contact the District for resolution. There was one case of missing documents in Manitoba School District, which was never found. The case count discrepancy was ultimately identified in our internal online systems, however the box was never received at CTB. The UPS records indicated that 10 boxes were picked up at Manitoba Elementary School on November 27, 2006, but that one box contained in that group was not labeled and ultimately was not delivered to our location. According to UPS, this box was

mistakenly included with a return services package picked up by UPS at another area location and may have been destroyed.

The shipment was checked for damaged materials. If the integrity of the documents was affected by any kind of damage, the Scoring Team was notified. Depending on the severity of the problem, the team member contacted the District for resolution. A record of all damaged materials was maintained. Per CTB's exception systems, four groups of books were identified as arriving in damaged boxes. All of the books were scannable or were "onlined" pursuant to CTB standard operating procedures.

- Before leaving the Receiving area, documents were logged into the computerized tracking system which provided real-time information regarding the status of the documents throughout the scoring and editing process. The electronic profile for each District was updated with at least the following information:
    - District name
    - Date of receipt
    - Box count
    - Shipping carrier
- CTB followed-up with each district if and when scorable test materials were not received in accordance with program timelines. CTB provided retrieval status information as requested to WDPI.

From there, documents moved to the WKCE workcells for the continuation of document process activities. CTB's Tracking and Management systems, including a Document Hold System and CTB's integrated scoring database, Monarch, provided Project Management and Scoring Management document and batch status information to ensure all documents were accounted for as they move through the scoring process.

6.4.3 Login

After the documents were logged into the computerized tracking system and released by Receiving, the Login stage in the process began where the following activities occurred:

- School Header Sheets were checked against School Group Lists to verify the number of students tested within each group, and to check for the required Principal or Designee signature and contact information. The Scoring Team was notified of any missing signatures, requiring a call to the district.
- Documents were grouped into manageable stacks and aligned to ensure proper scanning.
- Scannable headers were placed on top of each stack and a number assigned to identify each unique stack within a group.

6.4.4 Scanning

After all information was verified as received and prepared for scanning, the documents were placed in mini-queues for the next steps in the process, where they were cut into single sheets and electronically scanned. Scanners were calibrated periodically.

The scanners used by CTB have built-in checks for miscalibration. Hardware bias checking is used in real-time to verify that the scanner calibration is maintained during the scanning process. Additional checks were implemented by CTB to reinforce the built-in hardware checks and to ensure optimal scanner setup.

CTB's scanning software utilized the speed of the NCS 5000I optical scanners to capture document images and bubbled data without requiring specific document editing and resolution rules. Scanners were thus able to run at rated speed with no interruptions except for problems with the physical documents. All editing of the scanned documents was performed, in a subsequent step, in the raw scoring/editing system.

The scanning program evaluated every detectable mark on both sides of each page, and recorded the intensity and coordinates of solid marks for resolution in the subsequent raw scoring step. The form identification (i.e., "skunk marks") determined the type of document, and the headers determined customer identification and district, school, and class. Fall 2006 WKCE books were scanned using these scanning machines. Scanners functioned properly and normal processes went smoothly.

6.4.5 Editing/Updates

Raw scoring and editing of scanned data was performed in a client/server system (WinScore), where a sophisticated system of edits were invoked to review the integrity of each batch scanned and produce a list of error suspects. While the editors could view data from any document on-line, an error suspect list concentrated on the most likely problems based on pre-defined guidelines. This system reduced editing time and provided a high degree of quality control.

CTB continues to enhance the capability of editing software to simplify the detection and correction of errors. On-line editing screens focused an editor on potential problems and then provided related information. The actual scanned documents were always available to the editor, and the software supported the review and correction of any field in the scanned record. Entry and verification of the necessary corrections were enhanced to ensure each error is actually corrected.

As batches of the scanned documents were extracted for scoring, a final edit was performed to ensure all requirements for scoring were met. This automated final edit flagged a batch for further editing if any error, such as an unreadable or duplicate barcode/litho code, a student name coding issue, or a case count discrepancy, was still detected. A batch that contained

errors could not be extracted for reporting. This ensured a high level of accuracy of the scored data.

CTB maintained a professional staff of specialized data processing technicians to lead the verification process to ensure the integrity of the student response data at both group and individual levels. This process included the following error checks**:**

- Reliability. The reliability check ensured that the raw scores for each subtest were above chance levels. Scores not passing this edit were checked by a trained specialist to ensure that responses were being read correctly and that the correct form and level of the test was being used.

- Biographical data. Electronic edits were performed on elements such as student name to ensure obvious errors were corrected when possible.

- Student counts. Actual counts based on scanned records were electronically compared with expected counts, and discrepancies were flagged.

- School name/number**.** Pre-assigned school numbers and names were verified against an electronic file.

- Custom edits**.** Special edits can be performed using custom software that works in conjunction with our standard scoring process.

6.4.6 Document Retention

When the editing process was completed, documents were moved to a staging area to be prepared for retention. Bundles were caged, warehoused in a recoverable location, and retained for possible retrieval during the specified retention period. Once this period was over, documents will be destroyed according to procedures that ensure security is maintained. The typical retention period is one year after which time authorization is requested from WDPI to destroy the documents.

6.5 Handscoring Process

Handscore training for WKCE Fall 2006 occurred on 12/04/06. Readers were trained using customer approved training material (rubrics, anchors, horizontal, and two qualifying rounds). Once qualified, readers were required to maintain standards throughout the project. These standards were measured through checksets and targeted read behinds at the item level, for example, once the writing assessment was scored, a 100% second read was done as an additional check on inter-rater reliability. Data monitors generated reports daily that flagged any readers falling below standards and shared these reports with handscoring supervisors. Those readers falling below standards received additional coaching, more checksets and more read behinds.

Failure to improve led to dismissal of the reader from the project. This process was in place throughout the entire handscoring window, 12/06/06 - 01/05/07.

Electronic Handscoring System (EHS)

For WKCE programs, the Electronic Handscoring System (EHS) was used to score constructed response (CR) items. EHS presented images of scanned test books to trained readers, who assigned scores for constructed response items. Scanned output was viewed on high quality 19″ 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 routed to designated groups of readers trained to score these items. In addition to increased reader reliability, significant gains in reader productivity were evidenced following the implementation of EHS.

6.5.1 Anchor Papers and Training Papers

Prior to the actual scoring, the CTB Scoring Center created training materials. CR items for the WKCE were assessed using a WDPI holistic rubric with an X-point score scale. CTB randomly sampled student answer documents to ensure it was looking at a representative sample of the possible responses. Rangefinding meetings were held with WDPI staff and educators to select sample papers of each score point. Sample papers were used to construct scoring guides and training papers. CTB's Scoring Team collaborated with WDPI to make necessary revisions to the rubrics and selection of scoring guides and training papers.

This process included several presorting steps and subsequent iterative/consensus processes in order to achieve ever-increasing agreement and precision through a kind of "round robin" scoring, followed by discussion and selection.

Model responses/papers were selected for anchor training, horizontal training, and qualifying, and were consolidated into training formats. Once approved by WDPI, the Scoring Guides (consisting of rubrics, anchors, and annotations) served as a constant guide, setting the course for all subsequent training and scoring.

6.5.2 Training

Validation is a critical task in the assessment training process, and is the final determinant in reader readiness. All readers, including team leaders, must achieve 80 percent exact agreement on the qualifying round following training. Those readers not validating on the first attempt received further training prior to taking an additional qualifying round. Only those training who successfully validated were qualified as readers to score tests. Team leaders were required to complete two validation rounds with 80 percent exact agreement in each round.

6.5.3 Intra-Rater Reliability

Checksets

Throughout the course of the handscoring process, calibration sets of pre-scored papers (check-sets) were administered daily to the team leaders as well as to the readers, to monitor scoring accuracy and to maintain a consistent focus on the established rubric and guidelines. Imaging permitted this monitoring without reader knowledge of when a check-set was administered. Readers whose check-set scores fell below the qualifying level were removed from live scoring and were given additional training and another qualifying (validation) round. Readers unable to qualify were dismissed.

Read-Behinds

The "read-behind" is another valuable intra-rater reliability monitoring technique. On a daily basis, each team leader read a random selection of each reader's scored items, targeting at the item level or at the Rater Item Block level, where, potentially, more than one item was being scored. The scores were compared, and if they agreed, the team leader was able to offer feedback, which enhanced the reader's confidence and ability to score quickly and accurately. However, if a reader strayed from the standard established in the training and validation samples, the aberrant scoring was detected, and the team leader was able to offer the 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.

Double Reads

Where required, each constructed response was scored by two readers and inter-rater reliability was monitored throughout the scoring process. If the scores of the two assigned readers differed by one point, the student received the higher of the two scores. If the scores of the two readers differed by more than one point, a third rating was provided by an expert rater, who resolved the discrepancy and assigned a final score.

6.5.4 Scoring Personnel - Constructed Response Scorers

CTB recruited, trained, and managed a staff, over multiple handscoring sites, to complete all handscoring operations within the time lines of this contract. CTB's experience involved extensive consultations between CTB Scoring, Publishing, and the customer to review scoring rubrics, develop anchor papers and other reader training materials, and provide analyses of student responses to tryout forms.

Readers

Many CTB readers had years of classroom teaching experience. The CTB reader pool included editors, published authors, and individuals with advanced degrees. The minimum qualification for all Scoring Center readers was a Bachelor's degree. WKCE constructed response (CR) items were scored in Mather, California, Delran, New Jersey, and Orlando, Florida handscoring sites. Handscoring readers were recruited from the local city areas and were required to possess, and show evidence of, either, a BA or BS degree. Evaluator staff was comprised of individuals from many walks of life—from retired or current educators, to engineers, possessing bachelor's to doctorate degrees.

Team Leaders

Team leaders were selected on the basis of their ability to maintain a high degree of scoring accuracy and consistency, often across multiple subjects and grades. Team Leaders were also required to possess good interpersonal and leadership skills in order to be effective when training and counseling readers. While it is possible to conduct handscoring with more readers per team leader, it has been CTB's experience that inter-rater reliability and production goals are jeopardized unless a trained leader can frequently monitor all readers.

Scoring Supervisors

Scoring Supervisors were the core group at CTB scoring centers who directed and organized the assessment process and train team leaders and readers. Scoring Supervisors had extensive experience as Team Leaders prior to their qualification and selection. Scoring Supervisors were subject area experts in the content areas they supervised and train.

6.6 Quality Assurance of Custom Software

CTB integrated the quality controls into an end-to-end quality assurance program for all deliverables. Established procedures required all personnel and subject matter experts to complete their tasks according to strict quality assurance regulations. Within the data processing, scoring and reporting arenas, CTB has established an independent Quality Assurance department as an adjunct to both our larger Technology and Scoring departments. The Quality Assurance group, charged with the mission of ensuring accurate scoring and reporting software and systems, ensured accuracy of all scoring and reporting software. This group worked in tandem with the Technology and Scoring departments, shared their findings, and maintained an independent function and focus, thereby creating an internal quality assurance audit within the project. The Quality Assurance group developed a data verification plan for each administration of the WKCE. Stringent adherence to this plan and verification of data at each process ensured the quality of reported data.

CTB created a multiple step process spanning the period of time preceding test administration and document receipt through to the stage of final reporting. A brief overview of the process is as follows:

Step 1 – Test Data I – Months prior to CTB receiving documents from the WKCE administration, our Technology Department developed requirements and programming specifications that detailed the business rules and required scanning and scoring specifications for each administration. Scanning programs were then developed and verified to ensure accuracy and adherence to program requirements. After the development of scanning programs, the first step in the verification of accuracy of scanning programs and systems occurred within the Quality Assurance department. The Quality Assurance Department, comprised of Technology Quality Assurance and User Acceptance groups, conducted independent "end-to-end" production tests using answer documents that were hand-coded and scanned. This process has been designed to ensure the accuracy of the production system before live data are processed and results in a qualified production scoring system fully compliant with all WKCE program specifications. Each department created test decks for all scoring and reporting systems. Test decks are hand bubbled for all test sections and answer types, including multiple-choice and constructed-response items. Unusual bubbling patterns were tested including “all rights,” “only one correct response per test section,” “no correct response,” “double marks,” and many others. These patterns were used to verify test scores, invalidation, suppression, and omissions. Demographic grids and other contract-specific information were bubbled with a pre-identified set of criteria to ensure that biographical data are captured properly. This test deck proceeds through each subsystem before any production data are processed.

Step 2 – Test Data II – Next, the Quality Assurance Department created an electronic file to test all conditions of scoring and reporting. No physical test booklets were needed at this point. CTB specified all possible conditions to be verified and reviewed these specifications with ODE with the intent of identifying additional conditions. Simulated student test data were processed through the scoring software. Electronic report data and printed reports were compared to expected results using software utilities and manual verification procedures; and report production was tested before the release of scoring/reporting software to the scoring department. It was only after the successful completion of these “end-to-end” tests that CTB moved any system to a live production environment.

6.7 Advanced Function Printing (AFP)

The IBM Advanced Function Printing (AFP) system was a key factor in CTB’s ability to print large volumes of reports with varied content and sequences. CTB provided the functionality to print reports in the actual shipping sequence, with no manual sorting or collation required. In addition, each page contained complex graphics and the visual aids necessary to clearly convey the information to the wide variety of people who read the reports. CTB converted all mainframe systems to AFP and developed all new reports in this environment.

AFP operated on high-speed laser printers using large roll feeders for several hours of uninterrupted printing at a rate of over 200 pages per minute. The printers' output processors then separated packages, or sets, of reports.

AFP supported report collation. Reports were printed in any desired sequence, since the contents of each set of reports can be predefined. The sequence in which these packages were printed was also predefined. A "break page" of control and routing information preceded each package of reports. For example, for a district-wide school package, the break page may contain test, type of report, report level/grade, school name, principal's name and school address information. Packages were produced in the final order for quality checks and packaging for shipment.

With AFP graphic capabilities, CTB could design more meaningful reports. Form and content could be varied at any time while printing, graphics can be added, and complex graphics could be inserted to represent variable data.

CTB adopted procedures to provide unprecedented flexibility in the reporting software. In many cases, an application program need not be changed to modify or enhance a report; the much simpler AFP page definition can be changed, leaving the application program intact. Thus, programming, testing, and quality assurance were all simplified.

6.8 Distribution of CR Items

Tables 6-9 to 6-16 show the number of examinees at each score level (1 point, 2 points, etc) for each CR item. These distributions of examinee responses across score levels were provided to check the reasonability of CR items. The score distributions and the condition code counts yield several indications of reasonability. For example, a score distribution with a very high proportion of students scoring at either the highest or the lowest score level could indicate a need to investigate the reasonability of the item further. To take another example, the omit rate can also provide an indication of reasonability. A very high omit rate may indicate a very high degree of item difficulty, or it may suggest that a very high proportion of students lacked time to answer the item. While both high difficulty and a long test could be desired, they could also indicate a need to investigate the reasonability of the item further.

The tables also use four condition codes. Condition code "A" denotes no response or no attempt, "B" represents illegible, "C" means another language, and "D" denotes off-topic. Responses receiving condition codes were converted to zero score points. All responses for operational items were scored, but only a portion of the responses for field test items was scored. Operational and field test items can be identified based on the number of students (N, in the sixth column).

Tables 6-9 and 6-10 show the score distributions for Reading and Mathematics CR items. These tables show the score distribution of the first read. All Reading items have one part, and a maximum score of 3 points. As can be seen in Table 6-9 for Reading, while some items showed higher omit rates, most items split scores between either 1 score point or two, and relatively few

students scored at the 3-point level. All Mathematics items in grades 3-8 have two parts. Part A has one score point, and Part B has two score points. In Grade 10, the Mathematics items have one part. Looking across the Part A CR items in Table 6-10, one may observe that from approximately 15% to 91% of students obtained the one available score point. This suggests a wide range of difficulty levels across CR items for Part A. For the Part B items, there were two possible score points. As in Part A, the distribution of scores across the score levels suggests a wide range of difficulty for the CR items in Part B.

The most common condition code in Reading and Mathematics was "A," indicating that many students did not attempt the CR items. This can kind of result can be expected. Omit rates are often higher for CR items than MC items.

Tables 6-11 and 6-12 present the score distributions for the Writing Prompts. The tables for Writing are split between the 6-point Composing rubric and the 3-point Convention rubric. There is one writing prompt in each grade 4, 8, and 10. While only the score from the first read is used for Reading and Mathematics, the average of two raw scores is used for Writing. As such, the similarity of first rater's score and the second rater's score is important. To check, scores from the first read and second read are reported, and the difference between them is also presented. As can be observed in the tables, the distributions of the two reads are similar across all score levels for grades 4, 8, and 10. These tables are split between counts and percentages.

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

Tables 6-15 and 6-16 show the total score, combining both Composing and Convention. 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 0 score points, and a small proportion obtained the highest possible total score.

Part 7: Data and Analysis

Part 7 of the Technical Report serves to describe the data used for calibration and scaling, it describes how that data was determined to be sufficiently representative of WI students for calibration and scaling purposes, and it provides detailed comparisons of the data used for calibration and scaling to the census data, as evidence in support of the reliability and validity of the WKCE program. The calibration and scaling processes subsequently described in Chapter 8 depend on a representative sample of examinees in order to produce valid, accurate derived scores, for all examinees. The analysis presented here exists to demonstrate that the data used were sufficiently representative of WI students.

Part 7 also presents raw score descriptive statistics and classical item analysis. The raw score results and the item analysis are split by content and grade level. A raw score simply represents the number of correct items, such as 50 out of 54. Item analysis is a term broadly used to define the computation and examination of any statistical property of examinees' responses to an individual item (Crocker & Algina, 1986). The term "classical item analysis" indicates that the item analysis is based on the classical test theory, which assumes that the observed score is the sum of true score and error score. For further information on classical item analysis, consult Crocker & Algina (1986).

These three presentations (the data analysis, raw score results, and item analysis) serve as a foundation for subsequent analyses in Chapter 8, where calibration and scaling methods are discussed in detail, and Chapter 9, where the Technical Report moves on to present the resulting scale scores, performance level ratings, and scores on the sub-content level. Those discussions expand upon the basic raw score results and the item-level statistics presented here. Part 7 demonstrates adherence to AERA/APA/NCME standards 1.5, 1.13, 2.4, 2.8, 3.18, 6.5, and 7.1 in the WKCE program.

7.1 14 Data: Calibration Sample and Census Data

Scoring and all norms for the Fall 2006 WKCE were based on the 14 calibration districts (CD), which have been used in previous administrations. Table 7-1 lists the 14 districts. The districts were determined to be a representative sample of all Wisconsin students, based on two analyses. First, the characteristics of the 5 NCLB groupings (gender, ethnicity, socioeconomic status, disability, and English Language Proficiency), as derived from the 14 CD sample data and the census population, were systematically compared. Second, the scale scores of the 14 CD and WI census test populations were compared. As detailed below, the results of these two analyses indicate that the 14 calibration districts formed a representative sample of all Wisconsin students.

7.1.1 Proportion of Students by the 5 NCLB Groupings

Side-by-side comparisons of the student populations as derived from the 14 CD data and the WI census data are presented in Tables 7-2 to Table 7-6. Note that the census data includes the 14 CD students. In other words, the 14 CD data is a subsample of the census data.

Table 7-2 shows a side-by-side comparison of gender characteristics from the two data sources. There were more male students in both the census data and the 14 CD data. In most grades, 51% of the students were male in the census data and 52% were male in the 14 CD data.

Table 7-3 presents ethnicity characteristics. Overall, the two data sources show the same picture: most students were White, and next were African American students, followed by Hispanic students, Asian students, and American Indian students. However, there were some differences in the census and the 14 CD data. The main difference was that the 14 CD data contained a smaller proportion of White students, and a larger proportion of students of every other ethnicity than was found in the census. In the census data, White students were, on average, 77% of the population, that is, on average across grades. In the 14 CD data, White students numbered fewer, at 66% of the population, on average across grades. African American, Hispanic, and Asian student populations appeared somewhat larger in the 14 CD data than in the census data, though the American Indian student population showed less difference between the census and the 14 CD data.

Table 7-4 shows socio-economic status (SES) as derived from the two data sources. Students were identified as "Economically Disadvantaged" or "Not Economically Disadvantaged." In the table, "Yes" means "Economically Disadvantaged." The census and the 14 CD data again show a similar picture, though there were some differences. Economically Disadvantaged students were a smaller proportion of the total student population in the census data than in the 14 CD data. The census data shows that, looking across grades, 26% to 36% of students were Economically Disadvantaged. The census data also shows that the proportion Economically Disadvantaged steadily declines from Grade 3 to Grade 10. As stated, the 14 CD data shows a similar picture. The main difference here was that the Economically Disadvantaged student population appears slightly larger in the 14 CD data than in the census. The 14 CD data shows that 31% to 41% of students were Economically Disadvantaged. Notably, the 14 CD data shows the same decline in the proportion Economically Disadvantaged from Grade 3 to Grade 10.

Table 7-5 presents the data on Disability status. In this table, "Yes" denotes "Disabled." The census data and the 14 CD data show very similar pictures here. In both cases, approximately 1 in 7 students were Disabled. The 14 CD data shows only a slightly larger proportion of the population as Disabled than the census data.

Table 7-6 shows a side-by-side comparison of English Language Proficiency (ELP) characteristics as derived from the census and the 14 CD data. In the table, "No" means "Not Proficient." The census data and the 14 CD data show similar images of the student population. Most students, by far, are Proficient. However, the Not Proficient population is larger in the 14 CD data than in the census data. The census data shows that from 3% to 7% of the students were Not Proficient. The 14 CD data shows that 10% to 14% of the students were Not Proficient. Both the census and the 14 CD data show more students in the lower grades and fewer students in the higher grades as Not Proficient.

7.1.2 Comparison of Scale Score

Scale score means and standard deviations as derived from the 14 CD data and the census data are compared for all grades and contents in Table 7-7. The table shows the mean scale score from the 14 CD data, the mean score from the census, and the difference between the two. The table also shows the standard deviation (SD) of the scale scores from the 14 CD data, the SD of the scale scores from the census, and the difference between these two SDs. In all grades and content areas, the difference in scale scores was small. The range between LOSS and HOSS for any grade and content area is hundreds of scale score points. The differences observed were mostly on the scale of one or two scale score points. The largest difference was in Reading Grade 10, where the difference was 3.50 scale score points. There were some differences in the distributions of scores observable through the standard deviation. In all cases, the standard deviations were smaller in the census than in the 14 CD data. The difference in SDs ranged from 0.47 to 6.91.

Generally speaking, a sample is most likely to fully represent all of the characteristics of the population from which it was drawn when the sample unit is small. In the present context, using sample units such as schools or classrooms would likely yield a sample that fully represented all of the characteristics of the Wisconsin student population. In the present context however, it would be very difficult to use schools or classrooms as sampling units because it would take a great deal of time to gather test answer sheets from every each school and classroom for the purposes of sampling. So, for the WKCE, the sampling unit employed was the district. That means all students in a sampled district were included in the 14 CD.

7.2 Valid Records in Calibration Sample

Only valid student records were used in the calibration sample. Two rules were used to exclude records. First, if a student did not attempt the first 5 MC items, the record was removed from the calibration sample. Second, if the student's responses for the first 5 MC items showed any multiple-marks, the record was excluded from the sample. If either of these response patterns occurred, the record was excluded. The calibration sample includes all valid cases, that is, it includes all cases where neither of these response patterns occurred.

Table 7-8 shows the total number of students in the census data and in the 14 calibration districts. As indicated in Table 7-2, students without gender information were excluded from the analysis by gender, and as indicated in Table 7-3, students without ethnicity information were excluded from the analysis by ethnicity.

7.3 Descriptive Statistics by Test

Table 7-9 shows the configuration of the test. The table is split into operational items (used for scoring), and field test items (not used for scoring). The table is also split by grade and content area. The number of MC and CR items is indicated by grade and content area, for both operational and field test items. For operational items, the number of score points associated with

each CR item and the total number of score points are also indicated. More information for test configuration can be found in Part 4: Test Design.

The raw score table (Table 7-10) includes the number of students, the mean number of items correct, test difficulty, the standard deviation (SD) of raw scores, skewness of the raw score distribution, kurtosis, the minimum observed score, the maximum observed score, maximum possible score, reliability (Cronbach's alpha), and the standard error of measurement for raw scores. Summary statistics for raw scores are based on all students that took the Fall 2006 assessment.

In terms of the measurements applied in the raw score table, note first that the mean number of items correct, or the mean raw score, should be understood by grade and content area, and specifically in the context of the maximum possible score points, which varies by grade and content area. 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 mean raw score / maximum possible score points, and ranges from 0 to 1.0. A larger test difficulty statistic indicates a mean raw score which is close to the maximum possible score, and it indicates, therefore, an easier test. A smaller test difficulty statistic indicates a mean raw score which is further from the maximum possible score, and it indicates therefore a more difficult test. Consider an example: the test difficulty statistic would be 0.90 if we obtained a mean score of 45.00 on a test with a maximum possible score of 50. This could be considered an easier test. On the other hand, test difficulty would be 0.50 if we obtained a mean raw score of 25.00 on the same test. This could be considered a more difficult test. In Reading Grade 3, the test difficulty (0.69) was obtained by taking the mean raw score of 41.10 and dividing it by 60.

Like the standard deviation (SD), skewness and kurtosis also describe the shape of a distribution. When a distribution is perfectly normal, skewness is 0. A negative skew indicates the presence of some extreme low scores and (because the mean is sensitive to extreme scores) a corresponding increase in the number of student scores above the mean. A positive skew indicates a distribution with some extreme high scores and a corresponding increase in scores below the mean. Kurtosis describes a distribution in terms of its degree of peakedness. When a distribution is perfectly normal, Kurtosis is 0. A negative Kurtosis statistic indicates a distribution which is flatter than a perfectly normal curve, and a positive Kurtosis statistic indicates a distribution which is more peaked than a perfectly normal curve.

Where any student failed all items on the test, the minimum observed score is 0. Where any student obtained the full scores for all items, the maximum observed score is equal to the maximum number of points possible on the test. For example, in Reading Grade 3, there is at least one student who failed all items, and at least one student who got the maximum score.

A reliable test is one with high reliability as represented by statistics such as Cronbach's alpha and a low standard error of measurement (SEM). 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 10: Reliability and Validity. However, 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.

Reading

Looking across all grades, test difficulty in Reading ranged from 0.63 to 0.69, indicating that each test was of approximately equal difficulty for each grade. Standard deviations, in all cases, indicated a moderate degree of dispersion in scores. The distributions of raw scores were approximately normal in all grades. For every grade, alpha was high (0.90 to 0.93). SEM ranged from 3.0 to 3.28. The raw score information for Reading is split by gender, ethnicity, socioeconomic status, disability, and English language proficiency. In terms of gender, female students, as a group, had a slightly higher score than male students, as a group, in every grade. There was also slightly less variation, or dispersion, in female scores, as indicated by the standard deviation. Grouped and compared by ethnicity, mean scores varied more widely. Aside from Grade 10, in every grade, White students had the highest score, followed by Asian students, American Indian students, Hispanic students, and African American students. The difference in Grade 10 was that American Indian students, as a group, had a slightly higher mean score than Asian students, as a group. Scores also varied by socio-economic status. Taken as a group, in every grade, those students who were Not Economically Disadvantaged scored higher than students who were Economically Disadvantaged. The mean Reading score was higher among Not Disabled students than among Disabled students in every grade. Students who were Proficient in English had a higher mean score, as a group, than those students who were Not Proficient in English.

Mathematics

In Mathematics, test difficulty ranged from 0.59 in grades 8 and 10, to 0.70 in grades 3 and 4. (See Test Construction in Part 4 Test Development for an explanation). Alpha was high for every grade (0.90 to 0.92). SEM ranged from 2.92 to 3.38. The raw scores were split by gender, ethnicity, socioeconomic status, disability status, and English language proficiency. Split and compared by gender, male students had a higher mean raw score in every grade, but that difference in mean score was very small, and the two mean scores were nearly equivalent in each grade. Scores varied more widely by ethnicity, where, in every grade, White students, as a group, had the highest mean score, followed by Asian students, American Indian students, Hispanic students, and African American students. Grouped and compared by SES, the mean score was higher among those Not Economically Disadvantaged than among those who were Economically Disadvantaged. The mean score among those students who were Not Disabled was higher than the mean score among students who were Disabled. Those students who were Proficient in English had a higher mean score, as a group, than students who were Not Proficient.

Language Arts

There were some differences in test difficulty in Language Arts, where test difficulty ranged from 0.63 to 0.72. Raw scores were approximately normally distributed in each grade. Note that there are 30 possible points in grades 4 and 8, but 39 in Grade 10, because in Grade 10, a 9-point writing prompt is part of the Language Arts score. Looking across all grades, alpha ranged from 0.82 to 0.84. SEM ranged from 2.14 to 2.52. This raw score data is also split by gender, ethnicity, socioeconomic status, disability status, and English language proficiency. As a group, female students scored higher than male students, in each grade. White students had the highest mean score, followed by Asian students, American Indian students, Hispanic students, and African American students. In every grade, the mean score was higher among those students who were Not Economically Disadvantaged than among students who were Economically Disadvantaged. Students who were Not Disabled had a higher mean raw score than students who were Disabled, in each grade. In every grade, the mean score was higher among students who were Proficient in English than among those students who were Not Proficient.

Social Studies

Looking across grades, there were some differences in test difficulty. Test difficulty ranged from 0.72 to 0.85. The distributions of raw scores were approximately normal in all grades. Note that there is a maximum of 38 possible points in Grade 4, 40 in Grade 8, and 50 in Grade 10, so the distributions should be assessed within context of those differences in range. Across all grades, alpha ranged from 0.87 to 0.91 and SEM ranged from 2.20 to 2.90. The raw score data for Social Studies is also split by gender, ethnicity, socioeconomic status, disability, and English language proficiency. Looking across gender, there were minute differences in mean raw scores. While in every grade the mean score was fractionally higher among female students, the scores were nearly equivalent. There were larger differences in mean scores by ethnicity. As a group, White students had the highest mean score, followed by Asian students, American Indian students, Hispanic students, and African American students. The mean score was higher among students who were Not Economically Disadvantaged than among students who were Economically Disadvantaged, in every grade. The mean raw score was higher among students who were Not Disabled than among students who were Disabled. Taken as a whole, students who were Proficient in English had a higher mean score in every grade than students who were Not Proficient in English.

Science

Test difficulty in Science ranged from 0.63 to 0.72. Raw scores in Science were approximately normally distributed in each grade. Note that the maximum possible number of score points is 40 in grades 4 and 8, but 50 in Grade 10. The score distributions should be evaluated within context of those ranges. Across all grades, alpha ranged from 0.86 to 0.89 and SEM ranged from 2.49 to 2.94. There were small differences in mean scores by gender. In each grade, male students, as a group, had a higher mean score than female students. However, the differences in grades 4 and 8 were very small, less than one raw score point. There were larger

differences in scores by ethnicity. As was generally the case in the other content areas, White students had the highest mean score, followed by Asian students, American Indian students, Hispanic students, and African American students. Science scores also differed by SES. In every grade, the mean Science score was higher among students who were Not Economically Disadvantaged than among students who were Economically Disadvantaged. Students who were Not Disabled scored higher, as a group, than students who were Disabled, in every grade. In every grade, the mean score in Science was higher among students who were Proficient in English than among students who were Not Proficient in English.

7.4 Classical Item Analysis

Tables 7-21 through 7-43 present item-level item analysis for all grades and contents. First, the tables make a distinction between operational and field test items. Next, the test book form and test book item number are indicated. The test book form and test book item number can be used to understand the location of test items as students actually encountered them in test booklets. Readers can locate items in test booklets using this information.

The item analysis tables also indicate item type (MC or CR). Multiple choice and constructed response items are fundamentally different, so this basic item type information is included in the tables and should be considered alongside the item statistics during item analysis.

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 .30. So, the lower the *p*-value, the more difficult the item is. Item *p*-value is a good indication of difficulty; it takes student performance into account, and it makes comparing items in terms of a common statistic very simple.

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 .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 .75, this would indicate a much easier item, where, on average, students scored 75% of the maximum possible points for the item. As such, for CR items as well, *p*-value indicates difficulty, and the lower the *p*-value, the more difficult the item is.

A point-biserial correlation between item score and the total score on the test was also computed for MC items. The point-biserial correlation indicates the correlation between the item score and the total score on the test. If an item were to show a correlation of .80, this would indicate a strong relationship between the item score and a total test score. If the correlation for a given item were only .10, this would indicate that the performance on the item is weakly related to the total test score.

The point-biserial correlation is only appropriate for dichotomous level data (yes/no, right/wrong), so for the CR items, a Pearson correlation between the item score and the total

score on the test was computed. The Pearson correlation can be interpreted the same way: it is a correlation between the score for a given CR item and the total test score.[*]

A formula similar to the point-biserial correlation was applied to compute the correlation between each distractor and the total score. In general, negative correlations are expected for all distractors when an item is good. However, a small positive correlation for a distractor can often mean that the distractor is very attractive for low performing students. The omit rate indicates those cases where students did not attempt the item. Negative correlations and high omit rates can be an indication of mis-keyed items. Any suspect item is further investigated.

Items were flagged for further investigation when certain thresholds were reached. The *p*-value was flagged when the statistic fell below 0.30 for MC items. This would indicate an especially difficult item, where fewer than 30% of students obtained the correct answer. The point-biserial correlation was flagged where the coefficient was below 0.15. This would indicate a weak correlation between the likelihood of a correct answer choice and the total test score. The omit-rate was flagged when it was above 5%. A high omit rate can indicate an especially difficult item, or if located near the end of the test, it can indicate a speeded test, where students lack time. Distractors were flagged when they had a positive correlation with the total test score. That is to say, they were flagged when a wrong answer choice was, for some reason, an attractive choice.

Note that the item analysis for operational items is based on the 14 calibration districts, and the item analysis for the field test items is based on the census data.

Now, paging through the summary analysis and tables, note first, that there were a wide range of *p*-values. This means that the assessment contains items at all difficulty levels. This is both desirable and intended. A test made up of too many easy items or too many difficult items would not give us an accurate picture of the ability of students. Second, note that the correlation column shows a wide range of values. This is also desirable and intended, as a full range is necessary to discriminate between students at all ability levels.

All flag types did occur. That means first, that for some items, less than 30% of students answered correctly, thereby generating the *p*-value flag. In other words, some items were difficult. Second, the item-to-total score was flagged, indicating that in some cases the item score was only weakly related to the total test score. For some FT items, the correlation was actually negative, indicating a correct answer choice is a negatively correlated with the total test score. These items were sent to Development staff for further examination. Third, some items were flagged for the omit rate. An omit rate flag implies that at least 5% of the students did not attempt the item. Note that in many cases, the omit rate was very low, less than 1%, indicating that many items were attempted by nearly every student. Also note that it is often found that the omit rates for CR items are higher than those for MC items.

---

[*] For item analysis, the studied item was 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.

When flags occur, we can use the item analysis information with the test book in hand to examine the specific content of the test item, and from that vantage point consider more fully why one item may have been omitted more often than another, or why one item was more difficult than another, and so on.

The presence of a flag is an indication that an item should be investigated further, but that does not mean it cannot be included in the test as an operational item for valid reasons. Where flags did occur, all items were investigated further in order to assess whether the item should be excluded from the item pool. We generally expect FT items to be flagged more often than operational items. Note that all operational items were field tested. For further discussion of the 2006 WKCE item selection process, see section 4.3.1 on Test Development.

Table 7-44 shows the number of items flagged for all contents and grades, distinguishing between operational items and field test items. Note that because item analysis was performed by form, and FT items can occur in more than one form, the same FT item can appear as flagged more than once. Operational items appear only once.

Reading

Speaking now with reference to operational items and all flagging criteria, in Reading Grade 3, two items were flagged. One was flagged for a distractor, the other for *p*-value. The same results occurred in Grade 4 and Grade 5: two items were flagged, one for a distractor, and one for *p*-value. In Grade 6, two items were flagged, both for a distractor. In Grade 7, four items were flagged. Two were flagged for distractors, and two were flagged for *p*-value. In Grade 8, two items were flagged, both for distractors. Three items were flagged in Grade 10, all for distractors.

Among field test items, in Reading Grade 3, looking across all criteria, 5 items were flagged. Two items were flagged for the point-biserial correlation, a distractor, and p-value. Two items were flagged for a distractor. One item was flagged for the point-biserial correlation and a distractor.

Mathematics

Among operational items, only one item in Mathematics Grade 3 was flagged for any reason. The one item flagged was flagged for p-value. In Grade 4, two items were flagged, both for p-value. In Grade 5, three items were flagged, all for p-value. In Grade 6, two items were flagged. One item was flagged for a distractor, one item was flagged for p-value. Four items were flagged in Grade 7. One item was flagged for a distractor and p-value, two items were flagged for only a distractor, and one item was flagged for only p-value. In Grade 8, two items were flagged. One was flagged for a distractor, the other for a distractor and p-value. Across all criteria, four items in Mathematics Grade 10 were flagged. Three were flagged for only a distractor, one was flagged for only p-value.

Speaking now with reference only to field test items, no items were flagged for any criteria in Grade 3 Mathematics. In Grade 4, two items were flagged, both for a distractor and p-value. In Grade 5, only one item was flagged, and only for a distractor. In Grade 6, nine items were flagged. Five were flagged only for distractors. Two were flagged only for p-value. One item was flagged for a distractor and for p-value. One items was flagged for a distractor, p-value, and the omit rate. In Grade 7, four items were flagged. Two items were flagged for distractors only, one item was flagged for p-value only and one item was flagged for both a distractor and p-value. In Grade 8, three items were flagged for distractors only, and three items were flagged for p-value only. One item in Grade 8 was flagged for the point-biserial correlation, a distractor, and for p-value. One item was also flagged for only the point-biserial correlation.

Language Arts

Looking across all flagging criteria in Grade 4 Language Arts, there is one item flagged. The flag is for a distractor. In Grade 8, three items were flagged for any of the criteria. Two items were flagged for the omit rate, and one item was flagged for the point-biserial correlation and for a distractor. In Grade 10, only one item was flagged for any criteria. The item was flagged for a distractor and for p-value. There were no FT items in Language Arts.

Social Studies

In Social Studies Grade 4, two items were flagged. One was flagged for a distractor, one was flagged for omit rate. In Grade 8, only one item was flagged for any criteria. The one item flagged was flagged for p-value. Three items were flagged in Grade 10 Social Studies. One item was flagged for the omit rate, one item was flagged for a distractor, and one item was flagged for the point-biserial correlation, a distractor, and for p-value. There were no FT items in SS.

Science

In Science, for Grade 4, only one item was flagged for any reason. The one item flagged was flagged for a distractor. In Grade 8, two items were flagged. Omit rate was flagged for one item, and a distractor was flagged for the other item. In Grade 10, two items were flagged. One item was flagged for a distractor, the other was flagged for p-value.

Now, speaking with reference to the FT items, in Grade 4, one item was flagged. The flagged item was flagged for the point-biserial correlation and for a distractor. In Grade 8, five items were flagged. Three were flagged for distractors only. One item was flagged for the point-biserial correlation, a distractor, and for p-value. One item was also flagged for a distractor and for p-value. No FT items were flagged in Grade 10.

7.4.1 Speededness

The degree to which a test is speeded can be evaluated by examining the percentage of students who fail to respond to the last items on a test, or the last items in a timed section. Preliminary research has not indicated that allotted time is insufficient for students to adequately complete questions. However, some anecdotal evidence from WI districts has suggested that some students, especially in earlier grades, may have more difficulty budgeting time, especially in the initial section of exams. As of this printing, CTB research is investigating speededness issues, as a potential threat to validity, in a special study to be completed by the end of summer 2007.

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 9: Test Results

Part 9 of the Technical Report contains information on the results of the Fall 2006 administration of the WKCE. The scale score results, as derived from the calibration and scaling procedures described in the previous chapter, are presented here. These results are split by content area and grade level, and they are presented for each of the 5 NCLB groupings (gender, ethnicity, socioeconomic status, disability, and English Language Proficiency). Performance level information is also provided. Presenting the results by performance level translates the quantitative scale provided through scale scores into a qualitative, meaningful, transparent description of student performance: Minimal Performance, Basic, Proficient, and Advanced. While the scale score provides an essential quantitative reference to student performance, the performance level information speaks directly to requirements of the NCLB Act, as well as to parents, students, and educators. The performance level information is also presented by content area, grade, and NCLB grouping (gender, ethnicity, socioeconomic status, disability, and English Language Proficiency). Part 9 also presents Standard Performance Indicator scores, or SPI scores. SPI scores probe further into student performance to provide a set of scores within each content area, assessing specific skills and abilities, within the broader content areas of, for example, Reading, or Mathematics. SPI scores speak to the need among teachers and educational decision makers for diagnostic information to inform instructional strategies and identify student strengths and weaknesses. Combined, scale scores, performance levels, and SPI scores provide a comprehensive set of tools to assess WI student performance by content, grade level, and NCLB grouping. The AERA/APA/NCME standards addressed in Part 9 include: 1.5, 4.3, 4.5, 4.6, 4.7, 7.1, 7.10, 13.15, and 13.19.

Results presented below are based on population data in the General Research File. The results presented here may differ slightly from final testing results presented on the Wisconsin Department of Public Instruction website due to slight differences in the application of exclusion rules. Official final results typically use more detailed school-level information for generating state summary reports than is used to conduct research analyses. The results in the following tables are presented as evidence of reliability and validity of the WKCE assessments, and should not be used for state accountability purposes.

9.1 Summary Statistics for Scale Scores

As indicated previously, in addition to reporting raw score statistics, the WKCE program also reports scale scores, based on IRT pattern scoring. Scale scores have been discussed earlier and are discussed further in a later section; however, for present purposes, the reader should note that a) the item-pattern scoring procedure produces scale scores (or maximum-likelihood ability estimates) based on patterns of item responses, as described by Lord (1974; 1980, 179-181) b) the item-pattern score is the maximum likelihood ability estimate for a student with a given response pattern, produced in a context where item parameter estimates are known, and c) scale scores are based on student performance on all the items in the test.

Table 9-1 shows the mean scale score, standard deviation, skewness, kurtosis, minimum observed scale score (min), and maximum observed scale score (max) for the total population,

plus LOSS and HOSS for all contents and grades. As reflected in Table 9-1, mean scale scores rise as grade level increases. This is the (intended) result of vertical scaling (See Part 8.2 and the subsections thereof for more on vertical scaling).

Reading

Scores in Reading are moderately dispersed in every grade, as indicated by the standard deviation. The standard deviations here should be understood within the context of range between the LOSS and HOSS, which generally increases as grade increases. The Minimum and Maximum columns indicate the minimum and maximum scores obtained. In every grade, at least one student scored at the LOSS and at least one student scored at the HOSS. The Reading scores showed consistent scoring patterns within subgroups. For example, in every grade, female students had a slightly higher mean score than male students. Every grade also showed a consistent pattern of scores by ethnicity. Setting aside Grade 10, in every grade level, White students, as a group, had the highest mean score, followed by Asian students, American Indian students, Hispanic students and African American students. The difference in Grade 10 was that American Indian students scored slightly higher than, rather than slightly lower than, Asian students, as a group. Taken as a group, those students who were Not Economically Disadvantaged scored higher than students who were Economically Disadvantaged. At every grade level, students who were Not Disabled had a higher mean Reading score than students who were Disabled. In every grade, students who were Not Proficient in English had a lower mean Reading score, as a group, than students who were Proficient.

Mathematics

As indicated by the standard deviations in Table 9-1, Mathematics scores were moderately dispersed in every grade. The mean score was higher among male students than among female students in every grade, but these differences were very small. There were consistent differences in mean scores by ethnicity in Mathematics. White students had the highest mean score followed by Asian students, American Indian students, Hispanic students and African American students. Socio-economic differences recurred as well. In every grade, the mean score among Economically Disadvantaged students was lower than the mean for students who were Not Economically Disadvantaged. As was observed in Reading, as a group, in every grade, those students who were Not Disabled, scored higher than students who were Disabled. Also as we saw in Reading, in every grade, students who were Not Proficient in English had a lower mean score than students who were Proficient.

Language Arts

Scale scores for Language Arts show a moderate degree of dispersion in each grade. In every grade, the mean score among female students was higher than that among male students. By ethnicity, White students had the highest mean score, followed by Asian students, American Indian Students, Hispanic students and African American students. As was observed in other

content areas, in every grade, those students who were Not Economically Disadvantaged, scored higher, as a group, than students who were Economically Disadvantaged. In Language Arts, as in other content areas, students who were Not Disabled had a higher mean score than students who were Disabled. Also as was observed in other content areas, students who were Proficient in English had a higher mean score, as a group, than those students who were Not Proficient in English, in every grade.

Social Studies

When we look at the distributions of scale scores for Social Studies, across grades we see standard deviations that indicate moderate dispersion. The range between LOSS and HOSS increases with each grade, so the rising standard deviations can be understood within that light. While some gender differences were observed in other content areas, in Social Studies, the differences in mean scores by gender are all within a single scale score point. Though, in every grade that small difference was such that female students had a very slightly higher mean score than did male students. Differences by ethnicity which were observed in other content areas recurred here as well: White students had the highest mean score, followed by Asian students, American Indian Students, Hispanic students, and African American students. In Social Studies, as elsewhere, in every grade, students who were Not Economically Disadvantaged scored higher, as a group, than students who were Economically Disadvantaged. Those students who were Not Disabled had a higher mean score than students who were Disabled, in every grade. In every grade, students who were Not Proficient in English had a lower mean score than students who were Proficient.

Science

Scale scores for Science show a moderate degree of dispersion in every grade. The range between the LOSS and HOSS increases with each grade, so the rising standard deviations can be understood within that context. In Science, male students had a higher mean score than female students, in each grade. As we saw in other content areas, grouped and compared by ethnicity, White students had the highest mean score, followed by Asian students, American Indian Students, Hispanic students, and African American students. As was also observed in other content areas, in every grade, those students who were Not Economically Disadvantaged scored higher, as a group, than students who were Economically Disadvantaged. In Science, as in other content areas, those students who were Not Disabled had a higher mean score than students who were Disabled, in every grade. As was also observed in other contents, students who were Not Proficient in English scored lower, as a group, than students who were Proficient.

9.1.1 LOSS and HOSS

Table 9-12 shows the number and percent of students at the lowest obtainable scale score (LOSS) and the highest obtainable scale score (HOSS). For Reading, the percent at the LOSS ranges from 0.65 to 1.38 across grades. Grade 8 and 10 show the largest two percentages at 1.38

and 1.36. The percent at HOSS for Reading ranges from 0.01 to 0.05. For Mathematics, the percent at the LOSS ranges from 0.10 to 1.22 across grades. Grade 10 shows the largest percentage at 1.22%. The percent as the HOSS ranges from 0.07 to 0.22. If we consider two percent as the flagging criterion for LOSS and HOSS, no items in Reading or Mathematics were flagged, and none are very close to the criterion. The percent at the LOSS in Language Arts ranges from 0.42 to 0.74 across grades, and the percent at the HOSS ranges from 0.01 to 1.40. For Social Studies, the percent at the LOSS ranges from 0.19 to 0.93, and the percent at the HOSS ranges from 0.31 to 2.12. This somewhat larger percentage of students at the HOSS for Grade 4 indicates that the 2007 WKCE needs some more difficult items for Grade 4 in order to reduce the proportion of students scoring at the HOSS. In Science, the percent at the LOSS ranges from 0.22 to 0.71, and the percent at the HOSS ranges from 0.07 to 0.67.

9.2 Performance Level Information

In addition to raw scores and scale scores, the present report also provides student performance level information. Students are classified in terms of a performance level based on their scale scores. The cut scores that define performance levels were established in a linking study between the 2004 WKCE and the Fall 2005 WKCE (See Part 11.1 Linking Study in the Fall 2005 WKCE Technical Manual,). There are four performance levels: Minimal Performance, Basic, Proficient, and Advanced. Table 9-18 shows the cut scores for each content and grade level. Tables 9-19 to 9-23 show the percentage of all students in each performance category, as well as subgroup comparisons by each NCLB grouping.

Reading

Most students were either Proficient or Advanced in Reading. The proportion Proficient and the proportion Advanced were generally approximately equal, but in grades 3 and 10 more students were Advanced. In all grades, students were much more likely to score above the Proficient level (at Advanced) than below the Proficient level (at either the Minimal Performance or Basic level). As a group, female students were more likely than males to be at the Advanced level and less likely than males to be at either Minimal Performance or Basic proficiency. Grouped and compared by ethnicity, looking either at the low end of performance or the high end, the prevailing tendency in performance was such that White students performed the best, followed by Asian Students, American Indian students, Hispanic students, and African American students. There were exceptions however in grades 4, 5, and 10, where American Indian students were less likely than Asian students to be at the Minimal Performance level of proficiency. The change here in relative performance highlights the fact that many of the differences in performance by ethnicity were small. Across all grades, those students Proficient in English were less likely to score at the Minimal Performance level than students who were Not Proficient in English. At the same time, and in every grade, those students who were Proficient in English were more likely to score at the Advanced level than students Not Proficient in English. As a group, Disabled students were more frequently at the Minimal Performance level than Not Disabled students. The Advanced proficiency level for Reading was more common among Not Disabled students than among Disabled Students. Students Not Economically Disadvantaged

were less likely to score at the Minimal Performance level and were more likely to score at the Advanced level in comparison to Economically Disadvantaged students.

Mathematics

The Mathematics performance level data show that most students were at the Proficient or Advanced performance level. Students were more likely to score above the Proficient level (at Advanced) than below the Proficient level (at Basic or Minimal Performance) except in Grade 10, and in Grade 8 that was only marginally so. Looking across subgroups, there were some small but interesting differences in performance by gender. Male students were slightly less likely than female students to score in the lowest performance category in grades 3–5, but were slightly more likely to do so in grades 6–8 and 10. Males were also more likely than females to score at the Advanced level in every grade. Grouped and compared by ethnicity, looking either at the low end of performance or the high end of performance, the tendency was such that White students performed the best, followed by Asian Students, American Indian students, Hispanic students, and African American students. As in Reading, those students Proficient in English were less likely score at the Minimal Performance level than students who were Not Proficient in English. At the same time, students Proficient in English were more likely to score at the Advanced level than students Not Proficient in English. Also as we saw in Reading, as a group, Disabled students were more frequently at the Minimal Performance level than Not Disabled students. The Advanced proficiency level in Mathematics was more common among Not Disabled students than among Disabled Students. In Mathematics, as elsewhere, those students Not Economically Disadvantaged were more likely to score at the Advanced level, and less likely to score at Minimal Performance level in comparison to Economically Disadvantaged students.

Language Arts

The Language Arts performance level data show that, in each grade, most students are Proficient. Looking beyond the issue of Proficiency, it may be interesting to observe that in Grade 4, students were much more likely to be above the Proficient level (at Advanced) than below it (at either the Basic or Minimal Performance level), but this was not the case in either Grade 8 or in Grade 10. In both Grade 8 and Grade 10, the proportion below the Proficient level (at either the Basic or Minimal Performance level) is greater than the proportion above the Proficient level (at Advanced). In terms of gender, male students were more likely than female students to score at the lowest level of performance and less likely to score at the highest level. As was observed in Reading and Mathematics, compared by ethnicity, the prevailing tendency in performance was such that White students performed the best, followed by Asian students, American Indian students, Hispanic students, and African American students. However, some of the differences between Hispanic and American Indian students were quite small, and in Grade 8 there was no difference at the Minimal Performance level. As in other content areas, those students Proficient in English were more likely to score at the Advanced level and less likely to score at the Minimal Performance level than students who were Not Proficient in English. Also as in other content areas, as a group, Disabled students were more frequently at the Minimal

Performance level than Not Disabled students. At the same time, the Advanced proficiency level was more common among Not Disabled students than among Disabled Students. In Language Arts, as elsewhere, those students Not Economically Disadvantaged were more likely to score at the Advanced level and less likely to score at the Minimal Performance level in comparison to their Economically Disadvantaged counterparts.

Social Studies

The performance level data for Social Studies shows that most students were either Advanced or Proficient in Social Studies. In fact, students were Advanced more often than they were Proficient in each grade. In all grades, students were much more likely to score above the Proficient level (at Advanced) than below it (at either Minimal Performance or Basic). In terms of gender, male students were more likely to score at the lowest level than female students, though these differences were very small, and, notably, male students were slightly more likely to score at the Advanced level in grades 8 and 10, though not in grade 4. As was observed in other content areas, looking at either the low end of performance or the high end, the prevailing tendency was such that White students performed the best, followed by Asian Students, American Indian students, Hispanic students, and African American students. However, as observed elsewhere, some differences between American Indian and Hispanic students were quite small, and in Social Studies, the proportion of Asian students scoring at the lowest performance level was actually between that of American Indian and Hispanic students. As in other areas, those students Proficient in English were more likely to score at the Advanced level and less likely to score at the lowest level than students who were Not Proficient in English. Also, as in other content areas, as a group, Disabled students were more frequently at the lowest performance level and less frequently at the highest performance level than their Not Disabled counterparts. In Social Studies, as elsewhere, those students Not Economically Disadvantaged were more likely to score at the Advanced level and less likely to score at the lowest level than Economically Disadvantaged students.

Science

For Science, the performance level data shows that in Grades 4 and 8 most students are Proficient, and in Grade 10, by a small margin, most students are Advanced. In all grades, students were, as a whole, more likely to score above Proficient (at Advanced) than below Proficient (at either Basic or Minimal Performance). That difference was small but noteworthy in grades 4 and 8, and more readily apparent in Grade 10. Looking across gender, male students and female students were approximately equally likely to score in the lowest performance level in every grade. Males were slightly more likely to score at the highest level than females, and that difference increased in each reported grade. Performance levels in Science varied by ethnicity. Looking at the low end of performance or the high end, White students performed the best, followed by Asian Students, American Indian students, Hispanic students, and African American students. As in other content areas, some of those differences were small. Here, for example, one can see these small differences between Asian and American Indian students. As in other content areas, students Proficient in English scored at the highest level more often and at

the lowest level less often than their Not Proficient counterparts. Also as in other content areas, as a group, Disabled students were more frequently at the Minimal Performance level than Not Disabled students, and the Advanced proficiency level was more common among Not Disabled students than among Disabled Students. In Science, as elsewhere, those students Not Economically Disadvantaged were more likely to score at the highest level and less likely to score at the lowest level than Economically Disadvantaged Students.

9.3 Standard Performance Indicator (SPI) for Content Standard

In addition to raw scores and scale scores, teachers and educational decision-makers frequently need diagnostic information to inform instructional strategies and help identify students' strengths and weaknesses. This kind of information can be derived from scores on subsets of test items which estimate how much a student knows in a clearly defined skill domain. These skill domains are called content standards (or simply standards, or objectives).

Scores for content standards are reported as Standard Performance Indicator (SPI) scores. The SPI is 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).

Because most standards are measured by a relatively small number of items, a Bayesian procedure that takes into account the overall test performance is used to improve the reliability of the within-standard scores. Given a student's scale score on the entire test, IRT is used, via the 3PL model for MC items and 2PPC model for CR items, to estimate a score for each content standard. This estimate provides the initial (Bayesian prior) estimate of the student's mastery score. If this initial estimate is consistent with the points the student actually obtained for the standard, as indicated by a chi-square test, the two scores are combined as a weighted average to produce the SPI score (the estimated true score). The appropriate weight for the Bayesian prior estimate is computed as a function of the standard error of the scale score on which it is based; the smaller the standard error, the larger the weight. If the prior estimate and the observed score differ significantly, the observed proportion of the maximum score is used, without the prior estimate, to compute the student's SPI score.

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 even the SPI can not be stable for any content standard with a small number of items.* Also, while the percentages belonging to each performance level based on SPI cut scores can be compared across years, the SPI value can not be compared across years. This approach, identifying student proficiency of each content standard, relates to the Wisconsin Model Academic Standards.

Note that the average difficulty of items belonging to each content standard within each grade was not considered when the test form was constructed. Also, note that the average

difficulty of items for each content standard across grades was not controlled. There has not been an effort to make the average difficulty of items for all content standards similar within each grade, or more difficult as grade level increases. The difficulty of items is determined, in part, by the difficulty of the content being measured. The difficulty of concepts varies across content standards within grades as well as across grades. The current test blueprints do not specify the average difficulty of items for each content standard within grades or across grades. The mean *p*-value and mean SPI scores cannot be compared across grades. The mean *p*-value or mean SPI scores simply show the relative difficulty among content standards within one grade and content.

Tables 9-13 to 9-17 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, mean *p*-value, standard deviation of the raw scores, as well as the mean SPI score, and the standard deviation of SPI scores, for all contents.

Reading

Looking at Reading SPI scores as a whole, across content standards and grades, one can observe that mean p-values and mean SPI scores varied across content standards. For example, in Grade 3, the mean p-value ranged from 0.61 to 0.76 and mean SPI ranged from 61.71 to 75.83. In standard 2, the SPI mean score of 75.83 and its standard deviation of 20.78 imply that if Grade 3 students were to take a test of 100 similar Reading items, their mean score would be 75.83, and the standard deviation of their scores would be 20.78. Now, one can compare that score to Standard 4, where, if Grade 3 students were to take a test of 100 similar items, their mean score would be 61.71, and the standard deviation of their scores would be 16.53. In other words, the score on standard 2 would be higher than the score on standard 4. The SPI for standard 2 was high relative to the other scores in Grade 3. In Grade 4, there were three tiers of scores. Standard 4 was on the low end, standard 3 was at a higher level, and standards 1 and 2 were at one level higher still. There were marked differences here among the four standards. Grade 5 showed a similar picture, with three levels of SPI scores. Standards 3 and 4 were at the low end, standard 1 was one step above, and standard 2 was one step further. Here as well, there were marked differences among the standards. Grade 6 showed two basic levels of SPI scores. Standards 3 and 4 were on the lower end, and standards 1 and 2 were on the higher end. The difference between the two sets of SPI scores would span the equivalent of approximately 10 score points on tests of 100 similar items per standard, with a maximum raw score of 100. In Grade 7, the SPI scores spanned a total range of the equivalent of approximately 10 score points. In Grade 8, standards 1 and 2 were higher than standards 3 and 4, and the total range spanned from 58.67 in standard 3 to 72.15 in standard 1, or the equivalent of approximately 14 points. In Grade 10, standard 1 (73.16) was higher than the other standards, which were more similar to one another (at 65.68, 62.33, and 65.00 respectively). This means that a test of 100 similar items would yield a higher mean score based on standard 1 than it would based on any other standard, that is, among Grade 10 students.

Mathematics

Mathematics p-values and SPI scores varied across content standards. For example, within Grade 3, mean SPI ranged from 38.32 to 77.86. Standard A was markedly lower than all of the other standards, which were similar to one another. Looking at standard A then, if Grade 3 students were to take a test of 100 similar items, their mean score would be 38.32 and the standard deviation of their scores would be 24.54. At the same time, if Grade 3 students were to take a test of 100 similar items in any other standard, their mean score would be above 70.00, and the standard deviation of those scores would be below 19.00. In other words, the mean score on standard A would be much lower than scores on any other standard. The SPI (and with it, the p-value) for standard A was, as stated, much lower than the SPI for all other standards. The same basic pattern appeared in grade 3, 4, and 5. One difference however, was that in grades 4 and 5, standard B was slightly higher than standards C, D, E, and F. In Grade 6, standard A again showed the lowest SPI (and mean p-value) but there were three basic levels of SPI scores here. Standard A was, as stated, on the low end, B, D, and E occupied a middle level, and standards C and F were slightly higher. Grade 7 showed a similar pattern. Here there were three levels of SPI scores (and mean p-values). Standard A showed the lowest SPI, but D was not far from it. Standards B, C, and E were one step up, and standard F was one step further. The SPI scores here spanned a wide range (46.93 to 81.30). In Grade 8, there are two levels of SPI scores. The SPI score for Standard F (75.91) was much higher than the respective statistics for the other standards, which were all clustered together in a smaller range (49.34 – 56.64). In Grade 10, the variation in SPI scores across standards would span the equivalent of approximately 20 score points on tests of 100 similar items per standard with a maximum raw score of 100. There were three loose levels of SPI scores here. Standards C and D were on the low end at 53.58, and 56.15. A second level of SPI scores begins a small step away at 58.65 for Standard F, and includes E, and A. Standard B showed an SPI higher than the other standards, at 69.16. A test of 100 similar items would yield a higher mean score based on standard B than it would based on any other standard, that is, among Grade 10 students.

Language Arts

The SPI data in Table 9-15 for Language Arts shows variation in mean p-values and mean SPI scores across content standards, for each grade. For example, within Grade 4, mean p-values ran from 0.57 to 0.71, and mean SPI ranged from 57.36 to 70.84. If Grade 4 students were to take a test of 100 similar items in standard C, their mean score would be 56.36. That score may suggest a difficult test. But if the test were in standard B, the mean score would be 70.84, which suggests less difficulty. Standard C stood out as uniquely low in Grade 4; standards B and D were close, at 70.84, and 67.52. For Grade 8, there were two tiers of SPI scores. Standard B was the highest, at 78.07. Standards D and F were lower at 67.30 and 62.57. In Grade 10 no standard was distinctly different, either higher or lower than the general tendency. SPI scores ranged from 61.51 to 66.36.

Social Studies

Social Studies SPI scores varied across content standards in each grade. For example, in Grade 4, standard D stood apart as uniquely high at 84.81. A test of 100 similar items would yield a higher mean score based on standard D than it would based on any other standard, that is, among Grade 4 students. The other standards were more similar, and were clustered together in a narrow range at 74.06 to 78.12. In Grade 8, the SPI score for standard E stood out as uniquely low at 62.74. The other scores are clustered from 69.44 to 77.478. In Grade 10, there were two basic levels of SPI scores. B, C, and E were on the lower end, while A and D were on the higher end. The total range in SPI scores in Grade 10 would be equivalent to approximately 12 score points in the mean scores from a set of tests consisting of 100 similar items per standard, and a maximum raw score of 100.

Science

As indicated in Table 9-17, SPI scores in Science varied, in each grade. Within Grade 4, for example, the mean SPI rangds from 67.01 to 85.12 or the equivalent of approximately 20 score points across standards. There were three groups of scores in Grade 4. Standard A was uniquely high at 85.12, and standard B (78.96) was also a slightly higher than the other standards, which spanned a range from 67.01 to 73.28. There was a wide range of SPI scores (and p-values) in Grade 8. SPI scores here spanned a range from 54.44 to 82.29. There were four distinct levels of SPI scores in Grade 8. Standards D and E were on the low end at 54.44 and 59.48 respectively. Standard B was one step higher than this, at 65.68. Standards C, F, G, and H were clustered together in a relatively narrow range from 72.29 to 77.39. Grade 10 also showed a wide range of SPI scores. Here, two standards, D and F, occupied a lower level of SPI scores. The p-values here were 0.50 and 0.52 respectively. Standards A and H showed slightly higher p-values and SPI scores at 60.65 and 59.58 respectively. SPI scores in standards E and G were one level above these at 66.44 and 63.09. The highest SPI scores (and p-values) were found in standards B and C. Here, the SPI scores were 70.16 and 76.27 respectively. A test of 100 similar items would yield a higher mean score based on standard B or C than it would based on any other standard, that is, among Grade 10 students.

Part 10: Reliability

The previous chapter provided results, in several ways. The present chapter now builds upon the previous one by providing estimates of the reliability of those results. Several measures of reliability will be discussed here. The present chapter thus serves to demonstrate adherence to AERA/APA/NCME standards: 2.1, 2.2, 2.10, 2.11, 2.14, and 2.15.

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 a 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 2006 WKCE was estimated in four ways:

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

Standard 2.1 advises providing reliability estimates and the SEM for scores reported. Accordingly, Part 10 presents both a reliability analysis (using Chronbach's alpha) and SEM. Reliability and SEM are presented by both content area and sub-content area, as well as for both the total group of examinees and for each of the 5 NCLB groupings (gender, ethnicity, socioeconomic status, disability, and English Language Proficiency), thereby speaking to standard 2.11, which advises that reliability and SEM should be assessed for all population groups.

Standard 2.2 is also addressed. This standard advises reporting SEM for both raw score and scale score. The SEM is presented in Parts 7 (for raw score), 8 (for scale score), and in the present chapter. In the present chapter however, the SEM is presented by content area, grade level, the 5 NCLB groupings, and on the sub-content level as well.

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. The WKCE program adhered to this standard through a detailed analysis of classification consistency and classification accuracy, two related measures with bearing upon the reliability of the performance level classifications. This analysis also addresses Standard 2.14, by providing a conditional SEM for the cut scores, which 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 6, CR items were scored by (human) raters; the process thus involves subjective judgment. As Part 10 will show, a detailed assessment of inter-rater consistency was applied to the WKCE. The assessment conducted is termed inter-rater reliability. Inter-rater reliability measures the reliability of human raters as they score CR items.

Combined, Chronbach's alpha, SEM, classification consistency, classification accuracy, and inter-rater reliability provide several ways of looking at the reliability of the WKCE. Chronbach's alpha and SEM operate at the content level, they provide estimates of reliability for examinee 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. *

10.1 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 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 follows:

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

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

Tables 10-1 and 10-2 show Cronbach's alpha and the standard error of measurement (SEM) for all students and each NCLB grouping. As indicated in Table 10-1, reliability was highest in Reading and Mathematics. Looking at all examinees together, in the "Total" column, one can observe that reliability ranges from 0.90 to 0.93 across grades for Reading, from 0.90 to 0.92 for Mathematics, from 0.82 to 0.84 for Language Arts, from 0.87 to 0.91 for Social Studies, and from 0.86 to 0.89 for Science. If we consider 0.90 as a conservative criterion for an acceptable level of reliability, as measured by Chronbach's alpha, Language Arts, Social Studies grades 4 and 8, and Science would not meet the criterion. These content areas have a relatively small maximum number of items (and score points), and, notably, fewer items (and score points) than either Reading or Mathematics (this can be observed in Table 7-9). The number of items (and score points) has a close relationship with reliability, so these results, both for Reading and Mathematics, as well as Language Arts, Social Studies, and Science, can be understood within that context.

* Note that the field test items were not used for reliability and validity evidence.

Now, looking across the subgroup reliability statistics in Table 10-1, one may observe that there were some differences by gender: reliability for female students was often slightly lower than for male students. Reliability was often higher for Asian students than for other ethnicities in Reading, Mathematics, and Science. In all grades and content areas, reliability was higher for the Proficient group than it was for the Not Proficient group. There were also some differences in reliability between Disabled and Not Disabled students, but those differences were not consistent: in some cases reliability was higher for the Disabled group, in other cases reliability was higher for the Not Disabled group. Differences were also mixed in the comparison by SES, but these differences were consistently very small.

Table 10-2 shows the SEM for all examines and each NCLB grouping. Reading and Mathematics both produced a larger SEM than the other content areas. As we have observed, Reading and Mathematics have more items than the other content areas. SEM tends to increase as the number of items (or score points) increases because, in general, the standard deviation is large for tests with large numbers of items or score points, and the SD is a part of the calculation for SEM. There were some observable differences in SEM looking across the five NCLB groupings. For example, in some cases there was no difference by gender, but where there were differences, approximately half of the cases showed a higher SEM for females and the other half showed a SEM that was higher for males. Looking at the SEM by ethnicity, one may note that SEM was smaller for White students than for all other ethnicities in all grades and contents, except Grade 3 Mathematics. The SEM was smaller among students Proficient in English than among students Not Proficient in English, in all grades and contents. The SEM was larger for students who were Disabled than those Not Disabled. In all grades and content areas, the SEM was larger for Economically Disadvantaged students than for those students who were Not Economically Disadvantaged.

Table 10-3 shows reliability by content standard. The last column presents the reliability for the total test per 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 small number of items (or score points) is often associated with lower reliability. The lower level of reliability statistics obtained per content standard can be understood within context of the smaller number of items per standard than for the total test, per content area. The lower reliability statistics 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, reliability ranged from 0.51 to 0.85 in Reading. The lowest reliability coefficients in Reading were for standard D in grades 3 and 4, and for standard B in Grade 10. In Mathematics, reliability by content standard ranged from 0.51 to 0.76. Here, the lowest reliability statistics were found in standard A for Grade 4, standard C for grades 4 and 5, and in standard D for Grade 5. Reliability by content standard ranged from 0.43 to 0.75 in Language Arts. Here the lowest reliability statistics were in standard D for Grade 4, and in standard F for Grade 8 and 10. In Social Studies, reliability by content standard ranged from 0.45 to 0.73. Lower reliability statistics were found in standards C and D in Grade 4, and in standard D in grade 8. Science showed the lowest reliability statistics of all content areas, ranging from

0.25 to 0.65. Uniquely low reliability coefficients were obtained for standard B in grades 4 and 8, in standard D for Grade 4, and in standard G for Grade 4.

Table 10-4 shows the SEM for each content standard for each content area and grade. As indicated by the discussion above, these SEMs are smaller compared to those for the total test because of the small number of items within each content standard. Some differences in SEM by content standard can be observed. For example, in Reading, standard C showed a higher SEM than the other standards. In Mathematics, the SEM was consistent across standards. In Language Arts, the SEM for standard B was higher than for other standards. The SEM for Social Studies was higher in standard A, B, and C than in standard D and E. The SEM in Science was higher in standards C, D, E, and F than in standards A, B, G, and H.

10.2 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 states. Because of a heavy emphasis on moving all students to or above the "Proficient" category by year 2014, the consistency and accuracy of the classification of students into these performance categories is of particular interest.

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 $(H+1) \times (H+1)$, where H is the number of cut scores. Three cut scores yield a 4×4 contingency table, as can be seen below (see Table 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 sum of diagonal values of the contingency table:

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

Table a
Contingency Table with 3 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):

$$\text{kappa} = \frac{P - P_c}{1 - P_c},$$

where P_c is the chance probability of a consistent classification under two completely random assignments. This 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:

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

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.

Kolen and Kim's Method for Pattern Scoring

When item response theory (IRT) is applied to score examinees' responses, two types of scoring are available: number-correct scoring and pattern scoring. WKCE is an example of a program that has applied 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 pattern scoring based on IRT. Kolen and Kim (2004) suggested 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 b using the two ability estimates obtained from Step 2.

Table b
Classification Table for One Cut Point (C_1)[1]

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

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

Step 5: Multiply distribution weight ($\hat{g}(\theta)$) by 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).

As can be seen in Table 10-5, for Reading Grade 3, there are two tables for each grade and content area. The first table is a contingency table with all three cut scores. This table 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 procedure by Kolen and Kim, the score distributions for the first administration and the second administration are estimated using simulation. So, the

[1] This table is constructed for each quadrature point and replication. One, and only one, cell will have a value of 1 and zeros elsewhere.

value in each cell represents the probability of belonging to a certain performance level in the first administration and the second administration. For example, 0.03 represents the probability of belonging to "Minimal Performance" in the both first and second administrations. The 0.06 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.

The second table shows indices for classification consistency and classification accuracy. Because there are three cuts for the 2006 WKCE, four performance levels exist (See Part 11 on cut scores). The values in "All cuts" were obtained by applying all three cuts together. In Table 10-5 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 is 0.71, and classification accuracy is 0.87. The values for cut 1 were obtained by applying only the first cut score. There are two levels whenever only one cut is applied. It is clear that the values for P, kappa, and classification accuracy with all three cuts are smaller than those with only one cut. 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 the AYP report, the reliability values for this second cut need to be considered carefully. In Table 10-5, for example, the P for the second cut, which establishes the Proficient performance level, was 0.94, k was 0.82, and classification accuracy was 0.96. The interpretation of the table values is the same for Tables 10-6 to 10-27.

When only the Proficient cut score was applied, P was equal to or larger than 0.90, and kappa was equal to or larger than 0.75 for Reading and Mathematics. For Language Arts, the lowest P was 0.87 and the lowest kappa was 0.70. For Social Studies, the lowest P was 0.91 and the lowest kappa was 0.73. For Science, the lowest P was 0.89 and the lowest kappa was 0.73. If the criterion value of 0.90 is applied for P, Language Arts was lower than the criterion, as was Science (grades 4 and 8), but in both cases the value for P was larger than 0.80. Like inter-rater reliability, 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." According to Landis and Koch's criteria for Kappa, all tests for Reading and Mathematics showed "excellent agreement." For the remaining three contents, Kappa represented "good agreement."

Figures 10-1 through 10-5 also show P, kappa, and classification accuracy when students were classified based on the total three cuts. These values are in tables 10-5 to 10-27, but the plots are also provided for ease of understanding. As can be seen in the plots, for Reading and Mathematics, the values of classification consistency (P) were over 0.70, and the values of k were over 0.60 across all grades. The values for Language Arts were the smallest among the five contents. Unlike other content areas, a Writing prompt of 9 points contributed to the Language Arts Grade 10 scores, for which the total number of score points is 39. The impact of this Writing prompt may need to be examined. The values of k for Science grades 4 and 8 were lower than 0.60. Based on Landis and Koch' criteria then, all test forms showed "good agreement."

The SEMs for the cut scores can be found in Part 8: Scoring and Standard Error of Measurement. The scoring tables (Tables 8-2 to 8-25) show the SEMs around all scores,

including cut scores. Also, the SEM for each grade and content area is plotted in Figures 8-9 to 8-13, and the location of the cut scores is provided in each plot.

10.3 Inter-rater Reliability for CR Items and Writing Prompts

The reliability of handscoring should 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 indexes for inter-rater reliability are presented here: intraclass correlation, and weighted kappa. The formula for intraclass correlation and weighted kappa are from Rich Patz's 1998 unpublished paper, "Calculating Handscoring Reliability Coefficients."

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

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

Response #	Score1	Score2	Mean Score
1	X_{11}	X_{12}	$\overline{X}_{1.}$
2	X_{21}	X_{12}	$\overline{X}_{2.}$
.	.	.	
.	.	.	
N	X_{N1}	X_{11}	$\overline{X}_{N.}$
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 Score1 over all responses (similarly for Score2), and

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

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

Data Structure 2: Cross-tabulation of Score1 and Score2. As an alternative, we may create a square table of counts for each Score1 by Score2 (i.e., $X_{n1} \times X_{n2}$) combination:

Data Structure 2: Cross-tabulation of Score1 and Score2

		Score2				Row Total
		0	1	…	K	
Score 1	0	n_{00}	n_{01}	…	n_{0K}	n_{0+}
	1	n_{10}	n_{11}	…	n_{1K}	n_{1+}
	.	.	.	…		
	.	.	.	…		
	K	n_{K0}	n_{K1}	…	n_{KK}	n_{K+}
Column Total		n_{+0}	n_{+1}	…	n_{+K}	n_{++}

where K is the maximum score (for a rubric including zero) obtainable for the item, n_{ij} is the number of responses for which Score1 = i and Score2 = j, n_{i+} is the number of responses for which Score1 = i, and n_{+j} is the number of responses for which Score2 = j.
Formulas for the two reliability coefficients of interest are now 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:

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

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. In general, Kappa will have values equal to or smaller than the intraclass correlation. If agreement is perfect, then Kappa is +1. If agreement is at chance levels, Kappa is 0. Landis and Koch (1977) suggest that values of Kappa greater than .75 indicate "excellent agreement", values between .40 and .74 represent "good agreement" beyond chance, and values below .40 denote "poor agreement". Criteria for intraclass correlation or weighed kappa are not certain.

Tables 10-9 through Table 10-11 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 indexes of inter-reliability, and the distributions of scores across score levels. "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 receive the condition codes, A, B, C, or D. For example, in Table 10-12 for Reading Grade 3, item #31 shows that the perfect agreement, adjacent agreement, discrepant agreement, and codes are 71%, 23%, 1%, and 5%, respectively. The percentage values for agreement (all forms of agreement) were computed after the students with condition codes were dropped. The 5% of the responses with condition codes means that about 78 students out of 1514 received a condition code from either one of the two readers, or both readers. The condition codes were transformed to the raw score of 0 when the three indexes for inter-reliability were estimated.

For Reading and Mathematics, all responses were read by a single rater, and approximately 3,000 cases of the operational items, and 500 cases of the field test items were submitted to a second rater for scoring. All Writing prompts were scored by two readers.

Reading

As indicated in Table 10-9, the maximum score for any CR item in Reading is 3 score points. By far the most common mode of agreement was perfect agreement. The percent of perfect agreement ranges from 57% (Grade 6 Form D item #19) to 86% (Grade 4 Form B item #33). Adjacent agreement occurred in approximately 24% to 29% of responses in each grade, except in Grade 4 where a higher level of perfect agreement reduced the level of adjacent

agreement. Discrepant agreement occurred far less often than any other mode of agreement. The percent of discrepant agreement ranges from less than 1% to 4%. If we consider a difference of more than 1 score point between raters as critical, the rate of discrepant agreement indicates there were few critical differences. The percentage of responses generating codes ranged from 2% to 11% across grades. Intraclass correlation ranges from 0.77 (in Grade 6 Form D item #19) to 0.96 (Grade 4 Form B, C, and D, item #33). Although the intraclass correlation was low for Grade 6 item #19, the percent of discrepant agreement there was only 3%. Fifty seven percent (57%) of the responses here generated perfect agreement, and 36% generated adjacent agreement. Weighted Kappa in Reading ranged from 0.54 to 0.92.

Mathematics

Mathematics CR items have a maximum score of either of 1 or 2 score points. In general, a lower level of maximum of score points produces better inter-rater agreement and reliability than a higher maximum score will. As may be expected then, compared to Reading, Mathematics produced, on average, a larger percent of perfect agreement and smaller percent of discrepant agreement. Perfect agreement ranged from 70% to 98% across grades, adjacent agreement ranged from less than 1% to 25% across grades, and discrepant agreement ranged from less than 1% to 6%. Looking across all grades and items, from 2% to 14% of the responses generated codes. Intraclass correlation ranges from 0.85 (in Grade 3 Form D, item 19B, Grade 4 Form C item 26B, and Grade 4 Form D item 26B) to 1.00 in many instances. Weighted Kappa ranged from 0.71 to 1.00. Both intraclass correlation and weighted kappa were, in general, higher for Mathematics than for Reading. As suggested above, better results for Mathematics do not mean that the raters for Mathematics did better than the raters for Reading. In many large scale assessments, it is often found that inter-reliability is higher for Mathematics than for Reading. Scoring Rubrics for Mathematics can be much clearer for raters than those for Reading.

Writing

As indicated above, Writing prompts were scored on two rubrics, one for Composing (6 points) and one for Convention (3 points). The inter-rater reliability tables show that the rate of perfect agreement was lower for the 6-point Composing element than for the 3-point Convention element. The difference here is due to the difference in score points. Perfect agreement is, as discussed above, 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 here ranged from 59% to 60% for Composing, but from 86% to 92% for Convention. Adjacent agreement ranged from 33% to 35% for Composing and from 5% to 9% for Convention. Notably, the sum of perfect agreement and adjacent agreement was similar for both Composing and Convention, and discrepant agreement was rare in both for both elements of the scoring rubric (less than 1% to 3%). Approximately 4% of the responses generated condition codes. Intraclass correlation here ranges from 0.89 to 0.96, and weighted kappa ranges from 0.78 to 0.84.

Part 11: Validity

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 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 uses. 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 and their relationship to the *Standards.* Part 11 now addresses three final issues in validity; specifically, the issues of bias, construct validity, and test integrity. The analyses presented here add to the perspectives provided in Chapters 2 through 10. Below is a brief review:

Part 2 of the Technical Report described the involvement of Wisconsin educators 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 content formed a legitimate basis upon which to valid derive conclusions about student performance. This part of the Technical Report thus addressed standard 3.5 of the *Standards* (AERA, APA, NCME, 1999).

Part 3 of the Technical Report addressed the issue of Test Design. Part 3 showed that the test design process and the participation of Wisconsin educators in that process provide a solid rationale for having confidence in the content and design of the WKCE as a tool from which to derive valid inferences about Wisconsin student performance. Part 3 addressed 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.

Part 4 of the Technical Report brought the discussion begun in Part 2 on the Participation of Wisconsin Educators and Part 3 on Test Design to a conclusion by detailing the final phases of test development. Part 4 showed how test specification documents derived from earlier developmental activities guided the final phases of test development, and ultimately yielded the test booklets that students encountered. This section of the Technical Report thus addressed 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.

Part 5 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 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.

Part 6 of the Technical Report demonstrated adherence to AERA/APA/NCME standards 5.8 and 5.9. It described how multiple choice items and constructed response items were scored, it described the hand scoring process, the training and selection of evaluators, the scoring rubrics used for scoring the constructed response items, and the resulting score distributions.

Part 7 of the Technical Report described the data used for calibration and scaling. It showed how the data upon which the validity of calibration and scaling depend was determined to be sufficiently representative of WI students. Several comparisons of sample and population characteristics were presented. Raw score results, and a classical item analysis were also provided, and these served as a foundation for subsequent analyses. Part 7 thereby demonstrated adherence to AERA/APA/NCME standards 1.5, 1.13, 2.4, 2.8, 3.18, 6.5, and 7.1.

Part 8 of the Technical Report described the calibration and scaling processes, procedures, and results. Some references to introductory and advanced discussions of IRT were provided. Several axes upon which to evaluate the calibration and scaling procedures, such as the data used, software applied, the vertical relationship across grades, the successful estimation of parameters, item fit, and the standard error of measurement were all discussed. Part 8 of this report thereby addressed 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.

Part 9 presented 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 spoke directly to the performance level requirements of the NCLB policy environment, as well as to 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 WI student performance by content, grade level, and NCLB grouping (gender, ethnicity, socioeconomic status, disability, and English Language Proficiency). Part 9 thus addressed AERA/APA/NCME standards 1.5, 4.3, 4.5, 4.6, 4.7, 7.1, 7.10, 13.15, and 13.19.

Part 10 demonstrated adherence to AERA/APA/NCME standards through several analyses of the reliability of the Fall 2006 WKCE. It presented a reliability analysis using Chronbach's alpha, SEM results, a detailed analysis of classification consistency and classification accuracy, and a full analysis of inter-rater reliability. The Fall 2006 WKCE thereby addressed AERA/APA/NCME standards: 2.1, 2.2, 2.10, 2.11, 2.14, and 2.15.

In the subsequent pages, Part 11 will, as stated, present additional axes upon 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 five NCLB groupings. This analysis specifically addresses standards 7.1, 7.2, and 7.3.

The present chapter will also provide estimations of construct validity. Two measures are provided: correlations among content area objectives, and a factor 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 where scaling and analyzing results through IRT models. As such, these analyses are essential elements in assessing the validity of the WKCE.

11.1 Differential Item Functioning

The Fall 2006 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. Thus, an empirical differential item functioning (DIF) approach was used to examine potential item bias. The approach applied included systematic item analyses to determine if examinees with the same underlying level of ability had the same probability of correctly responding to the item. Items identified with DIF were examined to determine if item performance differences between identifiable subgroups of the population were due to extraneous or construct irrelevant information making the items unfairly difficult for one of the subgroups.

As indicated above, DIF analyses were conducted for the 5 NCLB groups: gender (male and female), ethnicity (White, African American, Hispanic, Asian, American Indian), socioeconomic status (Economically Disadvantaged, Not Economically Disadvantaged), disability (Disabled, Not Disabled), and ELP (Proficient, and Not Proficient).

Three kinds of DIF statistics were used in this study, Mantel-Haenszel, standardized mean difference, and Linn-Harnisch. Mantel-Haenszel is applied to MC items and standardized mean difference is applied to CR items. Linn-Harnisch is used for both MC and CR items. An item was flagged for DIF when any of these indexes crossed their respective thresholds.

(1) Linn-Harnisch (L-H)

Because the WKCE was built using item response theory (IRT), an appropriate procedure for examining item bias would need to 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 area between the item characteristic curves (e.g., Linn, Levine, Hastings, & Wardrop, 1981). However, these procedures require a minimum of 800 to 1000 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).

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. Note that the item parameters were based on data from the total sample of examinees, which includes all subgroups. 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 on the scale score (θ) scale. 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 people 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 people expected to answer item i correctly (over all deciles) for a specific subgroup (e.g., African American), the following statistic was computed:

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

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 people 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:

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

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 observed and expected proportion correctly answering item i in decile g is:

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

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

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

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) scale. 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 (Z_i) for the subsample (see Linn & Harnisch, 1981, p. 112). 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 target subsample. If D_i is negative, the item is biased against the target subsample.

(2) Mantel-Haenszel (M-H)

The Mantel-Haenszel statistic is computed as (Zwick, Donoghue, & Grima, 1993):

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

where F_k is the sum of scores for the focal group at the kth level of the matching variable. Note that the Mantel-Haenszel statistic is sensitive to N such that larger sample sizes increase the value of chi square.

In addition to the Mantel-Haenszel chi-square statistic, the delta statistic (MH-D DIF) was computed for all items. Educational Testing Service (ETS) first developed the MH-D DIF statistic. To compute delta, alpha (the odds ratio) 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. MH-D DIF is then computed as:

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

Positive values of MH-D DIF indicate items that favor the focal group, whereas negative values of MH-D DIF indicate items that favor the reference group.

An item was flagged based on the delta statistic. An item is flagged when:

$$| MHD - DIF | \geq 1.5$$

Note that this procedure is applied to MC items only.

(3) Standardized Mean Difference (SMD)

The SMD is an effect size index of DIF which is relatively easy to interpret (Zwick et al., 1993). The SMD compares the mean of the reference and focal group, adjusting for the distribution of reference and focal group members on the conditioning variable (Zwick et al., 1993). SMD is computed as (Zwick et al., 1993):

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

where p_{fk}= proportion of the focal group members at the kth level of the matching variable, m_{Fk}=1/N_{F1k} and m_{Rk}=1/N_{R1k}. A negative SMD value indicates an item on which the focal group has a lower mean than the reference group. A positive SMD value indicates an item on which the reference group has a lower mean than the focal group.

An item is flagged when:

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

Results

Tables 11-1 to 11-7 show items flagged based on the criteria described above. As indicated, an item flagged for differential item functioning (DIF) is more difficult for a particular group of students than would be expected based on their total test scores. DIF analyses were

conducted by gender, ethnicity, English Language Proficiency (ELP), disability status, and socio-economic status. The tables present Linn-Harnisch statistics (D+, D-, and Z) first, then the standardized mean difference is presented, and finally the Mantel-Haenszel statistic (Delta) is presented. 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 Mantel-Haenszel or the standardized mean difference produced a flag. MH is only computed for the focal group.

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. For ethnicity, the reference group is white. This means that the performance of other ethnic groups is considered with reference to the performance of white students. No items are flagged for White students. The DIF analysis investigating item-functioning with reference to socioeconomic status defines Economically Disadvantaged students as the focal group, and Not Economically Disadvantaged students as the reference group. The DIF analysis for Disability uses the Not Disabled student population as a reference group to assess DIF within the Disabled student population. The DIF analysis for ELP compares item-functioning among students Proficient and Not Proficient in English. Here, the Proficient group is the reference group, and the Not Proficient group is the focal group.

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) indicate that an item disadvantages the focal group, such as female, African American, or Economically Disadvantaged. A positive flag indicates that the item favors the focal group.

Note that a single item can be flagged for multiple NCLB groupings, such as for African American students, and for Hispanic students, and for Economically Disadvantaged students. Also note that the tables, and the analysis below, reflect DIF information for both operational and field test items.

The DIF analysis produced flags on both MC and CR items for all NCLB groupings. Both operational and field test items were flagged. Overall, socioeconomic status, gender, and disability were flagged less often than ethnicity and ELP. Eighteen (18) items were flagged for gender. Notably, within gender not all items flagged disadvantaged female students. Several items were flagged because they disadvantaged male students, and some items were flagged because they favored female students. In the DIF analysis for ethnicity, 24 items were flagged for DIF with reference to African American students. All but four of the flags here were negative, meaning that nearly all items flagged with reference to African American students were flagged for disadvantage. Fifteen (15) items were flagged for DIF with reference to Hispanic students. All but three of these items disadvantaged Hispanic students. More items were flagged for Asian students than for any other ethnicity. Seventy-three (73) items were flagged. Most were flagged as disadvantage, but about one-third were flagged for favor, and most of these items were in Reading. Eighteen (18) items were flagged for DIF with reference to American Indian students. All but three of the twelve flags indicated disadvantage for American Indian students. For the ELP analysis, many items were flagged (38). Notably, some items favored students that were Not Proficient. Only one item was also flagged for SES. This was item 31 in Reading,

Grade 10. The item showed a Linn-Harnisch –CC flag for both Economically Disadvantaged students and for Not Economically Disadvantaged students. This single item for SES is not included the tables below. In Disability, 17 items were flagged. Three (3) of the 17 items indicated favor rather than disadvantage.

11.2 Construct Validity

Construct validity is an indication of how well tests measure the skills or constructs they intend to measure, and it is the central concept underlying the 2006 WKCE assessment validation process. Tests designed to measure a set of similar skills or constructs should correlate more highly than tests designed to measure distinctly different skills or constructs. For example, scores derived from two tests for Reading, designed to measure similar skills, should correlate more highly than scores for Reading and Mathematics, for example. By the same logic, scores on the sub-content level should be correlated. We expect, for example, that the specific reading sub-skill behind standard 3 in Reading on the WKCE, which is "Analyzes Text," correlates with other sub-skills for Reading, such as standard 1, which is "Determines Meaning." Although a well-designed achievement test might encompass several sub-content areas, the test as a whole should coherently assess a single construct, e.g., reading achievement, and as such, we expect scores on the sub-content level to be correlated. Achievement tests are typically designed to measure student proficiency on a single continuum (or unidimensional construct). For a test to be scalable and adequately analyzed using a unidimensional Item Response Theory (IRT) model, like what was used for the WKCE, the test should be essentially unidimensional.

To be meaningful then, for a given content area, the test should have appropriate correlation coefficients within the content standards. If the correlation coefficient is very high between two content standards, it indicates that the two standards measure the same trait, while low correlation coefficients indicate two standards measure traits which are different to some degree. In general, the size of the correlation coefficient is influenced by the length of test, the number of items, or score points. Tables 11-8 to 11-12 show these correlations among content standards by content area.

Across all contents, correlations within standards were generally highest in Reading. The correlations in Reading ranged from a low of 0.58 to a high of 0.78 and more frequently between 0.65 and 0.78. The correlation for standard 4 was often lower than for other standards, but not always. This indicates that in some cases standard 4 may be slightly different than the other standards for Reading. The correlations for 4-1 in Grade 3 and 4 stand out as uniquely low, as does the correlation for 4-2 in Grade 4. The correlation for 2-1 in Grade 10 was also uniquely low. Correlation coefficients in Mathematics tended to be lower than Reading, indicating that standards A-E for Mathematics were more unique than the set of standards in Reading. Correlations among standards for Mathematics ranged from 0.45 to 0.73 though most values ranged from 0.53 to 0.68. Grade 8 and 10 standards showed a higher degree of correlation. Correlations for standard C in Grades 5, 6, and to a lesser extent, Grade 4, were lower than correlations for other standards. In some cases, such as Grade 3 and Grade 5, correlations for standard E tended to be higher than others. Language Arts had a small number of content standards and the correlations here ranged from 0.42 to 0.69; if the D-F correlation is excluded,

the range is from 0.52 to 0.69. The D-B correlation in Grade 8 and 10 was stronger than the other correlations. Correlations were often stronger in the higher grades in Social Studies. For example, the range in Grade 4 is 0.50 to 0.61 and in Grade 8 it is 0.51 to 0.64, but in Grade 10, the range jumps from 0.60 to 0.70. In Science, correlations with standard B tended to be lower than others in Grade 4 and Grade 8. In Grade 10, the same pattern was not evident. Overall, the correlations among standards in Science ranged from 0.25 to 0.52 in Grade 4, from 0.33 to 0.54 in grade 8, and from 0.43 to 0.57 in Grade 10.

As stated above, in the context of IRT, for a test score to be meaningful for a given content area, the test should be essentially unidimensional. Factor analysis is a statistical technique commonly used to evaluate dimensionality. Factor analysis can identify the latent constructs underlying test items and indicate the presence of a single construct underlying a test. Table 11-13 displays a summary of factor analysis results. Previous research shows that the examination of first two Eigenvalues can be useful in determining the existence of a dominant factor. The ratios of the first two Eigenvalues range from 4.15 to 14.16. That is, the variance of the first factor is approximately 4 to 14 times larger than the variance of the second largest factor. In Reading, ratios ranged from 11.42 to 14.16. Mathematics ratios were smaller; there the range was from 4.15 to 9.44. Language Arts ratios ranged from 7.08 to 11.83, Social Studies from 7.96 to 12.25, and Science from 9.40 to 10.91. Within the context of the strength of IRT as a unidimensional model, in general, these ratios can be understood as indicating that the content assessments in the WKCE were sufficiently unidimensional.

11.3 Erasure Analysis

Research used a special program to analyze erasures in multiple choice items. A high rate of erasures may identify situations in which test integrity needs to be examined further. Erasure analysis was performed separately by grade and content area, per school. The list of flagged schools will be released to WDPI.

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.

Tables and Figures

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 in 2004	CR items in 2004	MC items in 2005	CR items in 2005	MC items in 2006	CR items in 2006	MC items in 2007	CR items in 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	10
Science	0	0	0	0	123	34			123	34
Total	**645**	**91**	**77**	**22**	**123**	**34**	**0**	**0**	**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	10
Science	0	0	0	0	125	34			125	34
Total	**654**	**95**	**77**	**39**	**125**	**34**	**0**	**0**	**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	255	188	24	207	25	181	25	3,035	389
Mathematics	1,791	246	248	98	120	10	170	25	2,329	379
Science					266	76				
Grand Total	**4,250**	**557**	**436**	**142**	**476**	**101**	**247**	**38**	**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	**318**	**14**
Math	231	29	15	2	32	4	**246**	**31**
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**	**40**	**0**	**604**	**51**
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**	**40**	**0**	**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**	**323**	**25**	**3,662**	**332**

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 6-1
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 *failed to respond to the task.*

Table 6-2
Mathematics Rubric, Grades 3–8 and 10

Mathematics constructed response operational items each have two parts. Part A is scored as correct/incorrect. Part B is scored using a 2-point holistic rubric.

2 points

- The student demonstrates a thorough understanding of the mathematical concepts and/or procedures represented in the problem. The student responds correctly to the problem, uses mathematical procedures and/or concepts, and provides clear and complete explanations and interpretations containing words, diagrams, or calculations unless otherwise specified. The response may contain minor flaws that do not detract from the demonstration of a thorough understanding of the problem.

1 point

- The student provides a response that is only partially correct. The student provides a correct solution, but may demonstrate a misunderstanding of the underlying mathematical concepts and/or procedures. The student provides a correct solution, but in place of showing his/her work writes, “I used my calculator.” The student provides a thorough demonstration of understanding the problem, but states an incorrect solution or conclusion.

0 points

- The student provides a completely incorrect solution, a response that cannot be interpreted, or no response at all.

Table 6-3
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 usually on more 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

0 points **Off Topic, No Response, Illegible, Another Language**

Table 6-4
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

Table 6-5
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

0 points **Off topic, No response, Illegible, Another language**

Table 6-6
Writing Rubric, Composing, Grade 8

Wisconsin Writing Rubric Grade 8 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, cliched, awkward, imprecise, vague wordy

Table 6-6 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

Table 6-7
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

0 points **Off Topic, No Response, Another Language, Illegible**

Table 6-8
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, cliched, awkward, imprecise, vague wordy

Table 6-9 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

Table 6-9
Score Distribution for Reading*

Grade	Form	Test Book Item No.	N	Scores				Condition Codes**			
				0	1	2	3	A	B	C	D
3		31	59635	5790	24477	25217	1523	2517	50	23	38
		48	59635	6680	18512	29803	1980	2544	47	27	42
	A	63	1818	499	749	333	81	136	2	12	6
	B	63	1764	527	680	337	87	128	3	0	2
	C	67	1806	278	696	554	126	134	4	0	14
	D	67	1764	338	603	506	174	130	4	1	8
4		14	59450	10526	29155	14407	2718	2494	27	7	116
		33	59450	26409	20000	8377	1661	2854	21	6	122
	A	62	3177	1022	1260	610	33	248	0	2	2
	B	62	3179	520	1328	1049	110	163	3	2	4
	C	66	1782	633	688	339	41	77	2	0	2
	D	66	1730	515	701	380	41	92	0	0	1
5		13	60318	3609	21762	26969	5592	2184	23	11	168
		48	60318	4738	27801	22143	3547	1957	22	13	97
	A	71	1649	493	435	491	131	91	2	5	1
	B	71	1679	468	460	542	142	66	0	0	1
	C	71	1645	488	612	436	41	67	0	1	0
	D	71	1622	482	591	436	46	67	0	0	0
6		19	61337	6959	33569	17106	1654	1933	50	7	59
		27	61337	21586	23393	12334	1434	2434	48	8	100
	A	63	1657	263	670	519	107	92	0	0	6
	B	63	1670	257	678	561	116	55	0	1	2
	C	67	3165	922	1243	565	179	249	0	2	5
	D	67	3191	790	1199	710	238	246	1	5	2
7		37	63958	14605	9536	27840	9200	2611	14	3	149
		56	63958	26900	21456	10795	1726	3010	13	4	54
	A	72	2785	710	1152	628	76	219	0	0	0
	B	72	1588	337	630	465	98	56	0	0	2
	C	72	2776	649	1169	692	90	175	0	0	1
	D	72	1569	305	656	431	102	75	0	0	0
8		31	65642	19027	18364	20295	4577	3337	19	2	21
		56	65642	8042	15050	29749	9566	3148	23	5	59
	A	63	3317	503	1458	918	92	341	0	3	2
	B	63	3355	464	867	1500	268	248	0	0	8
	C	67	3378	600	1956	428	85	307	0	0	2
	D	67	3359	805	1043	916	282	311	0	0	2
10		31	71682	2450	19751	30508	14399	4499	0	6	69
		52	71682	6364	32561	18800	8984	4909	0	4	60

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

Table 6-10
Score Distribution for Mathematics*

Grade	Form	Test Book Item No.	Part	N	Scores			Condition Codes**			
					0	1	2	A	B	C	D
3		8	A	59635	28302	29428		1887	1	8	9
		8	B	59635	24634	21861	10042	3013	6	48	31
		19	A	59635	18475	39196		1923	10	22	9
		19	B	59635	23765	24964	8156	2643	12	69	26
		26	A	59635	15527	42319		1761	5	14	9
		26	B	59635	34846	7077	14853	2756	8	66	29
		42	A	59635	17178	40805		1608	4	29	11
		42	B	59635	22692	8095	26407	2301	9	93	38
	A	52	A	3206	1210	1815		176	3	2	0
	A	52	B	3206	1331	577	1005	269	4	20	0
	B	52	A	3223	1623	1416		181	3	0	0
	B	52	B	3223	1603	510	849	248	8	4	1
	C	58	A	3197	1899	1162		131	2	1	2
	C	58	B	3197	1034	1435	498	208	4	9	9
	D	58	A	3199	170	2905		122	0	1	1
	D	58	B	3199	723	542	1775	143	7	4	5
4		12	A	59450	19851	38031		1560	5	2	1
		12	B	59450	23474	16232	17480	2193	6	43	22
		18	A	59450	28624	29532		1277	4	11	2
		18	B	59450	39306	3620	14696	1747	10	63	8
		26	A	59450	30510	27574		1363	2	1	0
		26	B	59450	30161	22431	4149	2629	7	60	13
		47	A	59450	13109	45013		1328	0	0	0
		47	B	59450	34206	11842	11280	2033	8	69	12
	A	52	A	3178	1300	1772		106	0	0	0
	A	52	B	3178	1026	659	1325	151	2	15	0
	B	52	A	3181	1409	1685		87	0	0	0
	B	52	B	3181	1407	1023	605	135	3	6	2
	C	52***	A	3120	0	0	0	3120	0	0	0
	C	52***	B	3120	0	0	0	3120	0	0	0
	D	58	A	3120	1183	1833		104	0	0	0
	D	58	B	3120	1010	1238	739	129	1	3	0
5		10	A	60318	27784	31083		1398	35	12	6
		10	B	60318	14104	26261	18329	1565	21	29	9
		19	A	60318	8821	50086		1377	16	6	12
		19	B	60318	38238	11832	8026	2157	20	40	5
		29	A	60318	45121	13611		1545	22	9	10
		29	B	60318	7533	36951	13734	2038	22	28	12
		41	A	60318	12805	46160		1279	18	8	48
		41	B	60318	22675	22832	13066	1645	39	36	25

*This is the score distribution of the first read
**A: No response or no attempt, B: Illegible, C: Another Language, D: Off-topic.
***This FT item was suppressed during range finding.

Table 6-10 (cont.)
Score Distribution for Mathematics*

Grade	Form	Test Book Item No.	Part	N	Scores			Condition Codes**			
					0	1	2	A	B	C	D
5	A	57	A	3151	1870	1198		81	0	2	0
	A	57	B	3151	1107	1095	823	115	3	6	2
	B	57	A	3194	1058	2043		93	0	0	0
	B	57	B	3194	1925	401	724	136	2	4	2
	C	57	A	3165	1638	1401		121	2	3	0
	C	57	B	3165	1189	383	1427	157	0	8	1
	D	63	A	3139	1354	1664		121	0	0	0
	D	63	B	3139	1721	419	842	154	1	2	0
6		11	A	61337	29577	29602		2087	62	5	4
		11	B	61337	21910	8345	28391	2631	21	27	12
		22	A	61337	7427	52024		1870	10	5	1
		22	B	61337	29398	23220	6101	2541	27	42	8
		33	A	61337	18998	40535		1784	10	6	4
		33	B	61337	18795	23696	16584	2192	25	35	10
		46	A	61337	45356	14030		1924	20	4	3
		46	B	61337	26729	17141	14940	2467	18	32	10
	A	56	A	3176	2437	591		148	0	0	0
	A	56	B	3176	1404	1161	407	203	0	1	0
	B	56	A	3179	1626	1401		152	0	0	0
	B	56	B	3179	1210	1188	596	183	0	2	0
	C	63	A	3165	1812	1181		172	0	0	0
	C	63	B	3165	1509	431	1010	213	0	2	0
	D	63	A	3192	1172	1856		164	0	0	0
	D	63	B	3192	1547	322	1105	214	0	4	0
7		7	A	63958	4171	56886		2879	22	0	0
		7	B	63958	15330	7942	38398	2238	31	16	3
		19	A	63958	30080	31681		2159	22	13	3
		19	B	63958	33872	19936	6638	3447	37	23	5
		31	A	63958	28663	33339		1940	13	2	1
		31	B	63958	20755	4949	35143	3053	31	19	8
		47	A	63958	36147	25696		2108	4	1	2
		47	B	63958	21318	35356	4330	2876	39	26	13
	A	63	A	3281	1366	1625		290	0	0	0
	A	63	B	3281	1996	942	90	252	0	1	0
	B	56***	A	3328	0	0	0	3303	0	0	0
	B	56***	B	3328	0	0	0	3301	0	0	0
	C	56	A	3302	838	2262		202	0	0	0
	C	56	B	3302	778	485	1803	233	0	2	1
	D	56	A	3307	2517	510		280	0	0	0
	D	56	B	3307	2558	282	99	367	0	1	0
8		9	A	65642	7656	56400		1564	2	19	1

*This is the score distribution of the first read
**A: No response or no attempt, B: Illegible, C: Another Language, D: Off-topic.
***This FT item was suppressed during range finding.

Table 6-10 (cont.)
Score Distribution for Mathematics*

Grade	Form	Test Book Item No.	Part	N	Scores			Condition Codes**			
					0	1	2	A	B	C	D
8		9	B	65642	11066	11026	41371	2115	9	50	5
		20	A	65642	13880	50084		1670	4	4	0
		20	B	65642	16207	6386	40822	2162	5	54	6
		32	A	65642	41890	18546		5200	1	2	3
		32	B	65642	24342	27480	8553	5187	9	54	17
		44	A	65642	33116	29235		3278	4	2	7
		44	B	65642	21577	15684	24095	4226	6	44	10
	A	57	A	1843	437	1318		86	1	0	1
	A	57	B	1843	468	961	324	87	1	1	1
	B	57	A	1845	452	1340		52	0	0	1
	B	57	B	1845	473	934	384	53	0	0	1
	C	57	A	3382	2303	666		407	0	0	6
	C	57	B	3381	1441	904	569	462	0	1	4
	D	65	A	3360	2082	990		287	0	0	1
	D	65	B	3360	1573	457	971	356	0	2	1
10		16		71682	37534	25531	2050	6554	2	4	7
		30		71682	33132	8824	22212	7486	1	20	7
		36		71682	29067	16299	20190	6099	5	13	9
		47		71682	27381	22519	16301	5419	15	38	9

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

Table 6-11
Score Distribution by for Grade 4, 8, and 10 Writing Composing

Grade	Form	Total N	Scores						Condition Codes**			
			1	2	3	4	5	6	A	B	C	D
4	Rater1	59440	1206	10086	26611	15902	2116	294	1795	38	11	1381
	Rater2	59440	1258	10073	26501	16085	2032	297	1794	35	13	1352
	Diff*	0	-52	13	110	-183	84	-3	1	3	-2	29
8	Rater1	65631	769	8204	24900	23804	4697	500	1978	13	13	753
	Rater2	65631	765	8064	24709	24066	4763	474	1982	10	11	787
	Diff*	0	4	140	191	-262	-66	26	-4	3	2	-34
10	Rater1	71592	1207	9500	27257	24650	4925	496	3362	15	2	178
	Rater2	71592	1199	9611	27029	24660	5013	525	3356	15	4	180
	Diff*	0	8	-111	228	-10	-88	-29	6	0	-2	-2

Table 6-12
Percentage Distribution of Scores, Grade 4, 8, and 10 Writing Composing

Grade	Form	Total N	Scores						Condition Codes**			
			1	2	3	4	5	6	A	B	C	D
4	Rater1	59440	2.03	16.97	44.77	26.75	3.56	0.49	3.02	0.06	0.02	2.32
	Rater2	59440	2.12	16.95	44.58	27.06	3.42	0.50	3.02	0.06	0.02	2.27
8	Rater1	65631	1.17	12.50	37.94	36.27	7.16	0.76	3.01	0.02	0.02	1.15
	Rater2	65631	1.17	12.29	37.65	36.67	7.26	0.72	3.02	0.02	0.02	1.20
10	Rater1	71592	1.69	13.27	38.07	34.43	6.88	0.69	4.70	0.02	0.00	0.25
	Rater2	71592	1.67	13.42	37.75	34.45	7.00	0.73	4.69	0.02	0.01	0.25

Table 6-13
Score Distribution, Grade 4, 8, and 10 Writing Convention

Grade		Total N	Scores			Condition Codes**		
			1	2	3	A	B	C
4	Rater1	59440	3339	53790	467	1795	38	11
	Rater2	59440	3427	53705	466	1794	35	13
	Diff*	0	-88	85	1	1	3	-2
8	Rater1	65631	1174	61100	1353	1978	13	13
	Rater2	65631	1190	61085	1353	1982	10	11
	Diff	0	-16	15	0	-4	3	2
10	Rater1	71592	1328	63131	3754	3362	15	2
	Rater2	71592	1366	63050	3801	3356	15	4
	Diff	0	-38	81	-47	6	0	-2

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

Table 6-14
Percentage Distribution of Scores for Grade 4, 8, and 10 Writing Convention

Grade		Total N	Scores			Condition Codes**		
			1	2	3	A	B	C
4	Rater1	59440	5.62	90.49	0.79	3.02	0.06	0.02
	Rater2	59440	5.77	90.35	0.78	3.02	0.06	0.02
8	Rater1	65631	1.79	93.10	2.06	3.01	0.02	0.02
	Rater2	65631	1.81	93.07	2.06	3.02	0.02	0.02
10	Rater1	71592	1.85	88.18	5.24	4.70	0.02	0.00
	Rater2	71592	1.91	88.07	5.31	4.69	0.02	0.01

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

Table 6-15
Score Distribution for Grade 4, 8, and 10 Writing Total Score

Grade		Total N	Scores									
			0	1	2	3	4	5	6	7	8	9
4	Rater1	59440	1844	169	2290	1712	8953	26182	15821	1954	363	152
	Rater2	59440	1842	176	2301	1750	8906	26093	15979	1862	380	151
	Diff*	0	2	-7	-11	-38	47	89	-158	92	-17	1
8	Rater1	65631	2004	70	1159	810	7774	24775	23656	4180	801	402
	Rater2	65631	2003	76	1176	813	7660	24559	23882	4315	760	387
	Diff	0	1	-6	-17	-3	114	216	-226	-135	41	15
10	Rater1	71592	3379	54	862	906	9137	27009	23490	4418	1957	380
	Rater2	71592	3375	58	853	946	9205	26813	23453	4469	2036	384
	Diff	0	4	-4	9	-40	-68	196	37	-51	-79	-4

*Diff = N of Rater1 – N of Rater 2.

Table 6-16
Percent for Grade 4, 8, and 10 Writing Total Score

Grade		Total N	Scores									
			0	1	2	3	4	5	6	7	8	9
4	Rater1	59440	3.10	0.28	3.85	2.88	15.06	44.05	26.62	3.29	0.61	0.26
	Rater2	59440	3.10	0.30	3.87	2.94	14.98	43.90	26.88	3.13	0.64	0.25
8	Rater1	65631	3.05	0.11	1.77	1.23	11.85	37.75	36.04	6.37	1.22	0.61
	Rater2	65631	3.05	0.12	1.79	1.24	11.67	37.42	36.39	6.57	1.16	0.59
10	Rater1	71592	4.72	0.08	1.20	1.27	12.76	37.73	32.81	6.17	2.73	0.53
	Diff*	71592	4.71	0.08	1.19	1.32	12.86	37.45	32.76	6.24	2.84	0.54

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

Table 7-1
The Current 14 Calibration Districts

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

Table 7-2
Number and Percent of Census and 14 CD Students, by Gender*

Grade	Census		14 CD	
	M	F	M	F
3	30565	29047	3408	3197
4	30430	29017	3396	3187
5	31028	29278	3388	3286
6	31493	29821	3472	3228
7	32693	31238	3574	3317
8	33875	31773	3654	3372
10	36843	34866	4072	3761

* Students of unspecified gender or ethnicity are not counted

Grade	Census		14 CD	
	M	F	M	F
3	51	49	52	48
4	51	49	52	48
5	51	49	51	49
6	51	49	52	48
7	51	49	52	48
8	52	48	52	48
10	51	49	52	48

* Students of unspecified gender or ethnicity are not counted

Table 7-3
Number and Percent of Census and 14 CD Students, by Ethnicity*

Grade	Census					14 CD				
	W	B	H	A	AI	W	B	H	A	AI
3	45050	6594	5011	2103	853	4205	977	802	536	85
4	45172	6469	4814	2160	832	4312	875	719	602	75
5	46188	6447	4567	2244	860	4425	841	729	607	72
6	47079	6657	4472	2232	873	4433	861	720	595	91
7	49441	7045	4211	2260	974	4535	965	672	607	112
8	51059	6995	4217	2358	1018	4785	862	675	618	86
10	57688	6565	3782	2591	1081	5462	864	635	759	113

* Students of unspecified ethnicity are not counted

Grade	Census**					14 CD				
	W	B	H	A	AI	W	B	H	A	AI
3	76	11	8	4	1	64	15	12	8	1
4	76	11	8	4	1	66	13	11	9	1
5	77	11	8	4	1	66	13	11	9	1
6	77	11	7	4	1	66	13	11	9	1
7	77	11	7	4	2	66	14	10	9	2
8	78	11	6	4	2	68	12	10	9	1
10	80	9	5	4	2	70	11	8	10	1

* Students of unspecified ethnicity are not counted
**There could be a round error so that the total percent may not be 100%

Table 7-4
Number and Percent of Census and 14 CD Students, by SES

Grade	Census		14 CD	
	Yes	No	Yes	No
3	21475	38141	2723	3882
4	20734	38716	2550	4033
5	20518	39788	2520	4154
6	20655	40660	2512	4188
7	20785	43147	2653	4239
8	20859	44792	2481	4545
10	18393	53316	2413	5420

Grade	Census**		14 CD	
	Yes	No	Yes	No
3	36	64	41	59
4	35	65	39	61
5	34	66	38	62
6	34	66	37	63
7	33	67	38	62
8	32	68	35	65
10	26	74	31	69

**There could be a round error so that the total percent may not be 100%

Table 7-5
Number and Percent of Census and 14 CD Students, by Disability

Grade	Census		14 CD	
	Yes	No	Yes	No
3	8088	51528	938	5667
4	8327	51123	957	5626
5	8663	51643	977	5697
6	8559	52756	977	5723
7	9149	54783	1080	5812
8	9555	56096	1119	5907
10	9761	61948	1203	6630

Grade	Census**		14 CD	
	Yes	No	Yes	No
3	14	86	14	86
4	14	86	15	85
5	14	86	15	85
6	14	86	15	85
7	14	86	16	84
8	15	85	16	84
10	14	86	15	85

**There could be a round error so that the total percent may not be 100%

Table 7-6
Number and Percent of Census and 14 CD Students, by ELP

Grade	Census		14 CD	
	Yes	No	Yes	No
3	55528	4088	5695	910
4	55489	3959	5709	874
5	56551	3755	5776	898
6	57919	3396	5835	865
7	60773	3159	6060	832
8	62767	2882	6278	748
10	69202	2506	7012	820

Grade	Census**		14 CD	
	Yes	No	Yes	No
3	93	7	86	14
4	93	7	87	13
5	94	6	87	13
6	94	6	87	13
7	95	5	88	12
8	96	4	89	11
10	97	3	90	10

**There could be a round error so that the total percent may not be 100%

Table 7-7
Scale Score Mean and Standard Deviation for Census and 14 CD Results

Content	Grade	14 CD Mean	Census Mean	Diff = 14 CD-Census	14 CD Standard deviation	Census Standard deviation	Diff = 14 CD-Census
Reading	3	457.30	458.53	-1.23	41.50	38.93	2.57
	4	477.44	477.33	0.11	48.71	45.17	3.54
	5	485.79	484.76	1.03	51.84	46.95	4.89
	6	502.60	503.10	-0.50	54.51	49.52	4.99
	7	510.44	512.86	-2.42	51.78	47.77	4.01
	8	527.66	526.73	0.92	56.86	52.38	4.47
	10	535.86	539.36	-3.50	68.91	62.01	6.91
Mathematics	3	434.01	434.33	-0.32	47.28	46.81	0.47
	4	468.47	466.31	2.16	46.31	43.12	3.19
	5	490.88	489.39	1.49	46.53	44.09	2.44
	6	515.83	513.18	2.65	47.31	45.46	1.85
	7	535.07	535.01	0.06	45.83	43.51	2.32
	8	545.40	542.58	2.82	51.17	48.36	2.81
	10	562.24	563.40	-1.16	54.19	49.26	4.93
Language Arts	4	295.37	295.79	-0.42	32.20	29.04	3.16
	8	396.23	394.95	1.28	41.62	38.60	3.02
	10	447.82	448.68	-0.85	41.81	38.73	3.09
Social Studies	4	298.01	298.00	0.00	28.93	27.54	1.39
	8	397.56	397.14	0.42	40.29	36.80	3.48
	10	447.82	447.66	0.16	50.03	45.73	4.30
Science	4	298.71	299.59	-0.88	33.97	31.23	2.74
	8	396.94	398.09	-1.15	43.60	40.87	2.73
	10	448.57	451.06	-2.49	46.82	43.08	3.74

Table 7-8
Total Number of Students for Census and 14 CD

Grade	Census	14 CD
3	59616	6605
4	59450	6583
5	60306	6674
6	61315	6700
7	63932	6892
8	65651	7026
10	71709	7833

Table 7-9
Test Configuration of Operational (OP) and Field Test (FT) Items

Content	Grade	N of OP MC Items	N of OP CR Items					Total Score Points	N of FT MC Items	N of FT CR Items
			1 point	2 point	3 point	4 point	6 point			
Reading	3	54			2			60	10	1
	4	53			2			59	10	1
	5	54			2			60	14	1
	6	54			2			60	10	1
	7	54			2			60	15	1
	8	54			2			60	10	1
	10	50			2			56		
Mathematics	3	45	4	4				57	8	2
	4	45	4	4				57	8	2
	5	50	4	4				62	8	2
	6	50	4	4				62*	8	2
	7	50	4	4				62	8	2
	8	50	4	4				62	10	2
	10	50		4				58		
Language Arts	4	30						30		
	8	30						30		
	10	30			1		1	39		
Social Studies	4	38						38		
	8	40						40		
	10	50						50		
Science	4	40						40	10	
	8	40						40	10	
	10	50						50	10	

*A suppressed item in Grade 6 Mathematics reduces the effective total of score points from 62 to 61.

Table 7-10
Raw Score Descriptive Statistics based on Census Data

Content	Grade	N Count	Mean # of Items Correct	Test Difficulty	SD	Skewness	Kurtosis	Min Obtained	Max Obtained	Max Possible	Alpha	SEM
Reading	3	58187	41.10	0.69	11.40	-0.79	-0.16	0	60	60	0.93	3.00
	4	57788	37.67	0.64	10.70	-0.61	-0.36	0	59	59	0.91	3.14
	5	59019	39.00	0.65	10.62	-0.58	-0.34	0	60	60	0.91	3.15
	6	60066	39.43	0.66	9.79	-0.73	0.04	0	60	60	0.90	3.16
	7	62610	37.78	0.63	10.57	-0.49	-0.44	0	60	60	0.90	3.26
	8	64080	39.36	0.66	10.86	-0.61	-0.32	0	60	60	0.91	3.28
	10	69693	36.21	0.65	10.33	-0.55	-0.40	0	56	56	0.91	3.14
Mathematics	3	58719	40.13	0.70	9.79	-0.64	-0.12	0	57	57	0.91	3.00
	4	58446	40.02	0.70	9.14	-0.63	-0.08	0	57	57	0.90	2.92
	5	59415	41.12	0.66	10.14	-0.36	-0.43	0	62	62	0.90	3.18
	6	60156	39.13	0.64	11.69	-0.37	-0.60	0	61	61	0.92	3.28
	7	62855	38.76	0.63	11.05	-0.24	-0.62	0	62	62	0.91	3.29
	8	64353	36.56	0.59	11.51	0.00	-0.80	0	62	62	0.91	3.38
	10	69956	34.21	0.59	11.85	-0.09	-0.94	0	58	58	0.92	3.25
Language Arts	4	57763	20.24	0.67	5.20	-0.49	-0.33	0	30	30	0.82	2.21
	8	64046	21.59	0.72	5.44	-0.73	-0.04	0	30	30	0.84	2.14
	10	69225	24.46	0.63	6.31	-0.54	-0.22	0	39	39	0.84	2.52
Social Studies	4	58032	29.72	0.78	6.01	-1.09	0.92	0	38	38	0.87	2.20
	8	63999	28.94	0.72	7.09	-0.69	-0.18	0	40	40	0.88	2.43
	10	69258	32.99	0.66	9.59	-0.45	-0.62	0	50	50	0.91	2.90
Science	4	58091	28.90	0.72	6.62	-0.70	0.01	0	40	40	0.86	2.49
	8	64085	27.60	0.69	7.04	-0.54	-0.35	0	40	40	0.87	2.56
	10	69351	31.33	0.63	8.82	-0.33	-0.57	0	50	50	0.89	2.94

Table 7-11
Raw Score Descriptive Statistics based on 14 Districts

Content	Grade	N Count	Mean # of Items Correct	Test Difficulty	SD	Skewness	Kurtosis	Min Obtained	Max Obtained	Max Possible	Alpha	SEM
Reading	3	6451	40.59	0.68	11.87	-0.72	-0.35	0	60	60	0.94	3.01
	4	6424	37.62	0.64	11.29	-0.61	-0.45	0	58	59	0.92	3.14
	5	6547	39.30	0.65	11.29	-0.54	-0.51	4	59	60	0.92	3.17
	6	6575	39.34	0.66	10.45	-0.69	-0.14	2	60	60	0.91	3.19
	7	6769	37.16	0.62	11.18	-0.39	-0.64	1	60	60	0.91	3.27
	8	6883	39.52	0.66	11.51	-0.59	-0.45	2	60	60	0.92	3.29
	10	7586	35.61	0.64	11.20	-0.50	-0.59	1	56	56	0.92	3.18
Mathematics	3	6500	39.94	0.70	9.89	-0.58	-0.27	2	57	57	0.91	2.99
	4	6491	40.32	0.71	9.56	-0.64	-0.15	6	57	57	0.91	2.92
	5	6580	41.41	0.67	10.46	-0.33	-0.52	1	62	62	0.91	3.18
	6	6609	39.62	0.65	11.94	-0.38	-0.65	4	61	61	0.92	3.27
	7	6795	38.80	0.63	11.42	-0.23	-0.67	1	62	62	0.92	3.29
	8	6903	37.18	0.60	12.03	0.01	-0.88	1	62	62	0.92	3.38
	10	7670	34.01	0.59	12.52	-0.06	-1.03	0	58	58	0.93	3.25
Language Arts	4	6400	20.16	0.67	5.45	-0.51	-0.35	1	30	30	0.84	2.20
	8	6853	21.66	0.72	5.67	-0.70	-0.23	1	30	30	0.86	2.13
	10	7451	24.25	0.62	6.65	-0.48	-0.34	2	39	39	0.85	2.56
Social Studies	4	6417	29.59	0.78	6.26	-1.07	0.77	1	38	38	0.88	2.20
	8	6864	28.91	0.72	7.43	-0.64	-0.35	2	40	40	0.89	2.42
	10	7449	32.90	0.66	10.12	-0.42	-0.72	1	50	50	0.92	2.88
Science	4	6424	28.51	0.71	7.02	-0.64	-0.13	0	40	40	0.87	2.50
	8	6862	27.27	0.68	7.39	-0.45	-0.58	1	40	40	0.88	2.56
	10	7453	30.78	0.62	9.29	-0.25	-0.68	1	50	50	0.90	2.95

Table 7-12
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	29665	39.93	0.67	11.85	0.93	28521	42.32	0.71	10.78	0.93
	4	29426	37.03	0.63	10.87	0.92	28362	38.34	0.65	10.48	0.91
	5	30228	38.01	0.63	10.77	0.91	28791	40.03	0.67	10.36	0.91
	6	30695	38.41	0.64	10.03	0.90	29370	40.51	0.68	9.41	0.89
	7	31896	37.22	0.62	10.82	0.91	30713	38.37	0.64	10.27	0.90
	8	32944	38.31	0.64	11.16	0.91	31135	40.47	0.67	10.41	0.90
	10	35673	35.22	0.63	10.57	0.91	34020	37.25	0.67	9.98	0.90
Mathematics	3	29996	40.30	0.71	9.83	0.91	28722	39.95	0.70	9.75	0.90
	4	29832	40.32	0.71	9.13	0.90	28614	39.71	0.70	9.14	0.90
	5	30483	41.18	0.66	10.23	0.90	28932	41.05	0.66	10.04	0.90
	6	30776	39.17	0.64	11.95	0.93	29379	39.07	0.64	11.41	0.92
	7	32015	38.97	0.63	11.28	0.92	30840	38.55	0.62	10.82	0.91
	8	33103	36.97	0.60	11.68	0.92	31249	36.12	0.58	11.32	0.91
	10	35806	34.62	0.60	12.08	0.93	34150	33.78	0.58	11.58	0.92
Language Arts	4	29412	19.58	0.65	5.30	0.82	28351	20.92	0.70	5.01	0.81
	8	32905	20.81	0.69	5.71	0.85	31139	22.42	0.75	5.00	0.83
	10	35359	23.40	0.60	6.53	0.85	33800	25.58	0.66	5.85	0.82
Social Studies	4	29597	29.63	0.78	6.14	0.87	28435	29.81	0.78	5.88	0.86
	8	32894	28.94	0.72	7.31	0.89	31103	28.95	0.72	6.85	0.87
	10	35420	32.94	0.66	10.00	0.92	33838	33.05	0.66	9.13	0.90
Science	4	29638	29.14	0.73	6.66	0.86	28453	28.65	0.72	6.57	0.86
	8	32950	27.78	0.69	7.28	0.88	31134	27.41	0.69	6.78	0.86
	10	35476	32.05	0.64	9.17	0.90	33875	30.58	0.61	8.36	0.87

Table 7-13
Raw Score Descriptive Statistics for Reading by Ethnicity

Content	Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Reading	W	3	44380	42.98	0.72	10.54	0.92
		4	44396	39.42	0.67	9.99	0.90
		5	45513	40.81	0.68	9.83	0.90
		6	46385	41.24	0.69	8.85	0.88
		7	48781	39.80	0.66	9.64	0.89
		8	50211	41.10	0.69	10.11	0.90
		10	56715	37.90	0.68	9.48	0.89
	AA	3	6387	33.35	0.56	12.01	0.92
		4	6244	30.17	0.51	10.84	0.90
		5	6247	31.00	0.52	10.59	0.90
		6	6470	31.37	0.52	10.27	0.89
		7	6794	28.56	0.48	10.23	0.89
		8	6711	30.94	0.52	10.88	0.90
		10	6020	26.67	0.48	10.23	0.89
	H	3	4621	35.15	0.59	11.53	0.92
		4	4325	32.25	0.55	10.54	0.90
		5	4280	33.34	0.56	10.69	0.90
		6	4229	33.93	0.57	9.95	0.89
		7	3935	31.68	0.53	10.44	0.89
		8	3920	33.68	0.56	10.98	0.90
		10	3461	29.66	0.53	10.70	0.90
	A	3	1969	39.41	0.66	11.76	0.93
		4	2013	35.52	0.60	11.22	0.92
		5	2138	36.68	0.61	11.24	0.92
		6	2126	37.03	0.62	10.06	0.89
		7	2143	34.17	0.57	11.03	0.91
		8	2246	37.41	0.62	10.72	0.90
		10	2462	31.62	0.56	10.81	0.91
	AI	3	828	37.48	0.62	11.32	0.92
		4	810	34.04	0.58	10.17	0.90
		5	841	34.90	0.58	9.81	0.89
		6	854	35.86	0.60	9.61	0.88
		7	956	33.72	0.56	10.19	0.89
		8	990	34.84	0.58	10.84	0.90
		10	1033	32.04	0.57	9.84	0.89

Table 7-14
Raw Score Descriptive Statistics for Mathematics by Ethnicity

Content	Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Mathematics	W	3	44493	41.92	0.74	8.88	0.89
		4	44530	41.69	0.73	8.30	0.88
		5	45619	42.87	0.69	9.43	0.89
		6	46268	41.28	0.68	10.77	0.91
		7	48783	40.84	0.66	10.25	0.90
		8	50255	38.56	0.62	10.95	0.91
		10	56736	36.21	0.62	11.16	0.92
	AA	3	6420	31.71	0.56	10.55	0.91
		4	6294	31.89	0.56	9.63	0.90
		5	6272	32.40	0.52	9.75	0.89
		6	6469	28.39	0.47	10.99	0.90
		7	6787	28.24	0.46	9.81	0.88
		8	6707	26.35	0.43	9.38	0.86
		10	6016	22.02	0.38	9.21	0.87
	H	3	4915	35.51	0.62	9.39	0.89
		4	4691	35.74	0.63	8.79	0.88
		5	4481	36.01	0.58	9.55	0.88
		6	4373	33.03	0.54	10.87	0.90
		7	4115	32.91	0.53	9.88	0.88
		8	4102	30.15	0.49	9.88	0.88
		10	3634	26.64	0.46	10.30	0.90
	A	3	2054	40.12	0.70	9.71	0.91
		4	2117	39.61	0.69	9.25	0.90
		5	2200	41.24	0.67	10.29	0.90
		6	2192	39.44	0.65	11.63	0.92
		7	2219	38.03	0.61	11.18	0.91
		8	2294	36.13	0.58	11.65	0.92
		10	2536	31.66	0.55	11.86	0.92
	AI	3	835	36.87	0.65	9.33	0.89
		4	814	37.23	0.65	8.27	0.87
		5	843	37.74	0.61	9.17	0.87
		6	852	33.92	0.56	11.26	0.91
		7	951	34.33	0.55	10.27	0.89
		8	993	31.41	0.51	10.27	0.89
		10	1032	28.09	0.48	10.36	0.90

Table 7-15
Raw Score Descriptive Statistics for Language Arts by Ethnicity

Content	Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Language Arts	W	4	44443	20.96	0.70	4.94	0.81
		8	50285	22.40	0.75	5.09	0.83
		10	56529	25.38	0.65	5.91	0.83
	AA	4	6233	16.81	0.56	5.28	0.80
		8	6631	17.67	0.59	5.65	0.82
		10	5822	18.98	0.49	6.27	0.81
	H	4	4306	18.16	0.61	5.06	0.79
		8	3891	19.04	0.63	5.44	0.82
		10	3351	20.79	0.53	6.19	0.81
	A	4	1975	20.02	0.67	5.23	0.82
		8	2246	20.48	0.68	5.36	0.83
		10	2375	22.85	0.59	6.16	0.82
	AI	4	806	18.40	0.61	5.06	0.79
		8	990	19.15	0.64	5.41	0.82
		10	1014	21.38	0.55	6.14	0.81

Table 7-16
Raw Score Descriptive Statistics for Social Studies by Ethnicity

Content	Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Social Studies	W	4	44610	30.75	0.81	5.35	0.84
		8	50231	30.21	0.76	6.42	0.87
		10	56615	34.48	0.69	8.94	0.90
	AA	4	6265	24.58	0.65	6.97	0.87
		8	6628	22.39	0.56	7.38	0.86
		10	5858	23.68	0.47	9.05	0.88
	H	4	4354	27.29	0.72	6.22	0.85
		8	3901	25.18	0.63	7.23	0.86
		10	3374	27.57	0.55	9.44	0.89
	A	4	1990	28.86	0.76	6.15	0.86
		8	2246	27.92	0.70	6.89	0.87
		10	2386	30.25	0.60	9.00	0.89
	AI	4	813	27.73	0.73	5.99	0.84
		8	990	25.96	0.65	6.86	0.85
		10	1024	28.28	0.57	9.36	0.89

Table 7-17
Raw Score Descriptive Statistics for Science by Ethnicity

Content	Ethnicity	Grade	N Count	Mean	Test Difficulty	SD	Alpha
Science	W	4	44629	30.19	0.75	5.92	0.83
		8	50297	28.98	0.72	6.36	0.85
		10	56653	32.90	0.66	8.09	0.87
	AA	4	6285	22.62	0.57	6.96	0.84
		8	6639	20.76	0.52	6.78	0.83
		10	5880	21.49	0.43	7.76	0.84
	H	4	4369	25.57	0.64	6.46	0.83
		8	3904	23.39	0.58	6.84	0.84
		10	3395	25.64	0.51	8.22	0.86
	A	4	1993	28.10	0.70	6.58	0.85
		8	2254	25.42	0.64	7.11	0.86
		10	2395	28.34	0.57	8.64	0.87
	AI	4	815	26.32	0.66	6.52	0.84
		8	989	24.73	0.62	6.77	0.84
		10	1026	27.24	0.54	8.06	0.85

Table 7-18
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	20546	35.95	0.60	11.88	0.93	37641	43.92	0.73	10.08	0.92
	4	19652	32.57	0.55	10.83	0.91	38136	40.30	0.68	9.63	0.90
	5	19736	33.55	0.56	10.65	0.90	39283	41.74	0.70	9.49	0.90
	6	19947	34.34	0.57	10.14	0.89	40119	41.97	0.70	8.54	0.87
	7	19967	31.99	0.53	10.60	0.90	42643	40.50	0.67	9.40	0.89
	8	20020	33.82	0.56	11.08	0.90	44060	41.88	0.70	9.77	0.89
	10	17307	30.12	0.54	10.54	0.90	52386	38.22	0.68	9.44	0.89
Mathematics	3	20970	35.59	0.62	10.05	0.90	37749	42.65	0.75	8.68	0.89
	4	20192	35.64	0.63	9.40	0.90	38254	42.33	0.74	8.09	0.88
	5	20050	36.03	0.58	9.93	0.89	39365	43.71	0.70	9.22	0.89
	6	20143	32.94	0.54	11.45	0.91	40013	42.24	0.69	10.51	0.91
	7	20190	32.86	0.53	10.54	0.90	42665	41.55	0.67	10.16	0.90
	8	20238	30.42	0.49	10.44	0.89	44115	39.37	0.64	10.87	0.91
	10	17508	27.20	0.47	10.84	0.91	52448	36.55	0.63	11.23	0.92
Language Arts	4	19618	18.05	0.60	5.21	0.80	38145	21.36	0.71	4.83	0.80
	8	19927	18.95	0.63	5.59	0.83	44119	22.78	0.76	4.93	0.83
	10	16653	20.99	0.54	6.33	0.82	52572	25.56	0.66	5.89	0.83
Social Studies	4	19792	26.95	0.71	6.56	0.86	38240	31.15	0.82	5.16	0.84
	8	19941	25.11	0.63	7.35	0.87	44058	30.68	0.77	6.23	0.86
	10	17010	27.43	0.55	9.55	0.89	52248	34.81	0.70	8.88	0.90
Science	4	19831	25.62	0.64	6.90	0.85	38260	30.59	0.76	5.79	0.83
	8	19963	23.71	0.59	7.16	0.85	44122	29.36	0.73	6.23	0.84
	10	17061	26.08	0.52	8.69	0.87	52290	33.05	0.66	8.16	0.87

Table 7-19
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	7094	31.03	0.52	12.85	0.93	51093	42.50	0.71	10.45	0.92
	4	7262	27.82	0.47	11.62	0.92	50526	39.09	0.66	9.77	0.90
	5	7698	28.46	0.47	11.04	0.90	51321	40.58	0.68	9.60	0.90
	6	7654	28.74	0.48	10.50	0.89	52412	40.99	0.68	8.63	0.87
	7	8204	26.45	0.44	10.18	0.89	54406	39.49	0.66	9.52	0.89
	8	8541	27.17	0.45	10.70	0.89	55539	41.23	0.69	9.59	0.89
	10	8629	24.67	0.44	9.54	0.88	61064	37.84	0.68	9.36	0.89
Mathematics	3	7302	34.39	0.60	10.52	0.91	51417	40.95	0.72	9.40	0.90
	4	7490	33.56	0.59	10.28	0.91	50956	40.97	0.72	8.56	0.89
	5	7862	33.14	0.53	10.44	0.90	51553	42.33	0.68	9.52	0.89
	6	7692	28.20	0.46	11.65	0.91	52464	40.73	0.67	10.80	0.91
	7	8229	27.82	0.45	10.12	0.89	54626	40.41	0.65	10.22	0.90
	8	8587	25.62	0.41	9.32	0.86	55766	38.24	0.62	10.88	0.91
	10	8654	22.22	0.38	9.15	0.87	61302	35.91	0.62	11.19	0.92
Language Arts	4	7326	16.85	0.56	5.26	0.79	50437	20.73	0.69	5.01	0.81
	8	8519	15.99	0.53	5.63	0.81	55527	22.45	0.75	4.87	0.82
	10	8226	17.45	0.45	5.68	0.77	60999	25.41	0.65	5.77	0.82
Social Studies	4	7582	26.45	0.70	6.86	0.87	50450	30.21	0.80	5.72	0.86
	8	8580	22.14	0.55	7.43	0.86	55419	30.00	0.75	6.42	0.86
	10	8544	23.68	0.47	9.15	0.88	60714	34.31	0.69	8.90	0.90
Science	4	7601	25.21	0.63	7.27	0.86	50490	29.45	0.74	6.34	0.85
	8	8598	21.33	0.53	7.18	0.84	55487	28.57	0.71	6.50	0.85
	10	8577	23.27	0.47	8.23	0.85	60774	32.47	0.65	8.29	0.88

Table 7-20
Raw Score Descriptive Statistics by English Language Proficiency

Content	Grade	Proficient					Not Proficient				
		N Count	Mean	Test Difficulty	SD	Alpha	N Count	Mean	Test Difficulty	SD	Alpha
Reading	3	54562	41.63	0.69	11.23	0.93	3625	33.20	0.55	11.12	0.91
	4	54417	38.14	0.65	10.55	0.91	3371	30.08	0.51	10.17	0.89
	5	55630	39.51	0.66	10.43	0.91	3389	30.57	0.51	10.13	0.89
	6	56985	39.88	0.66	9.61	0.89	3081	31.12	0.52	9.33	0.86
	7	59768	38.22	0.64	10.41	0.90	2842	28.65	0.48	9.85	0.88
	8	61528	39.73	0.66	10.73	0.91	2552	30.41	0.51	10.12	0.88
	10	67519	36.58	0.65	10.15	0.90	2174	24.59	0.44	9.30	0.87
Mathematics	3	54725	40.47	0.71	9.75	0.91	3994	35.44	0.62	9.18	0.89
	4	54603	40.34	0.71	9.07	0.90	3843	35.39	0.62	8.87	0.88
	5	55757	41.49	0.67	10.07	0.90	3658	35.41	0.57	9.47	0.88
	6	56848	39.52	0.65	11.62	0.92	3308	32.33	0.53	10.74	0.90
	7	59759	39.11	0.63	11.00	0.91	3096	32.08	0.52	9.89	0.88
	8	61561	36.90	0.60	11.48	0.91	2792	29.07	0.47	9.56	0.87
	10	67536	34.55	0.60	11.77	0.92	2420	24.64	0.42	9.72	0.88
Language Arts	4	54434	20.39	0.68	5.17	0.82	3329	17.70	0.59	5.00	0.78
	8	61509	21.76	0.73	5.39	0.84	2537	17.47	0.58	5.02	0.77
	10	66048	24.70	0.63	6.21	0.84	1914	19.05	0.49	5.48	0.75
Social Studies	4	54653	29.91	0.79	5.95	0.87	3379	26.59	0.70	6.13	0.84
	8	61462	29.16	0.73	7.02	0.88	2537	23.71	0.59	6.73	0.83
	10	67196	33.25	0.67	9.51	0.91	2062	24.64	0.49	8.07	0.84
Science	4	54695	29.14	0.73	6.56	0.86	3396	24.99	0.62	6.36	0.82
	8	61533	27.85	0.70	6.95	0.87	2552	21.47	0.54	6.34	0.80
	10	67269	31.58	0.63	8.74	0.89	2082	23.33	0.47	7.40	0.82

Table 7-21
Item Analysis Grade 3 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.71	0.52	0.03%				
OP		2	MC	0.88	0.56	0.12%				
OP		3	MC	0.84	0.59	0.17%				
OP		4	MC	0.64	0.52	0.23%				
OP		5	MC	0.67	0.50	0.48%				
OP		6	MC	0.72	0.55	0.51%				
OP		7	MC	0.88	0.56	0.20%				
OP		8	MC	0.82	0.55	0.25%				
OP		9	MC	0.73	0.33	1.01%				
OP		10	MC	0.66	0.50	0.85%				
OP		11	MC	0.56	0.40	1.58%				
OP		12	MC	0.79	0.40	2.61%				
OP		13	MC	0.69	0.58	0.62%				
OP		14	MC	0.84	0.55	1.12%				
OP		15	MC	0.42	0.40	1.52%				
OP		16	MC	0.65	0.54	1.78%				
OP		17	MC	0.63	0.38	1.61%				
OP		18	MC	0.62	0.45	1.75%				
OP		19	MC	0.73	0.53	2.11%				
OP		20	MC	0.58	0.45	0.31%				
OP		21	MC	0.83	0.45	0.36%				
OP		22	MC	0.76	0.55	0.79%				
OP		23	MC	0.57	0.39	0.62%				
OP		24	MC	0.69	0.44	0.70%				
OP		25	MC	0.62	0.54	1.15%				
OP		26	MC	0.75	0.56	0.17%				

Table 7-21 Cont'd
Item Analysis Grade 3 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.43	0.27	0.34%		+		
OP		28	MC	0.79	0.51	1.02%				
OP		29	MC	0.71	0.57	0.64%				
OP		30	MC	0.68	0.41	1.16%				
OP		31	CR	0.45	0.51	0.00%				
OP		32	MC	0.95	0.30	1.86%				
OP		33	MC	0.77	0.49	2.27%				
OP		34	MC	0.82	0.53	2.75%				
OP		35	MC	0.84	0.47	2.79%				
OP		36	MC	0.68	0.41	2.51%				
OP		37	MC	0.69	0.40	2.62%				
OP		38	MC	0.45	0.43	2.75%				
OP		39	MC	0.80	0.59	0.25%				
OP		40	MC	0.80	0.43	0.34%				
OP		41	MC	0.75	0.61	0.43%				
OP		42	MC	0.89	0.50	0.68%				
OP		43	MC	0.81	0.64	0.36%				
OP		44	MC	0.37	0.20	1.07%				
OP		45	MC	0.88	0.47	0.73%				
OP		46	MC	0.90	0.43	2.13%				
OP		47	MC	0.77	0.45	0.54%				
OP		48	CR	0.48	0.56	0.00%				
OP		49	MC	0.69	0.53	1.13%				
OP		50	MC	0.74	0.27	1.23%				
OP		51	MC	0.74	0.54	1.58%				
OP		52	MC	0.51	0.45	2.13%				

Table 7-21 Cont'd
Item Analysis Grade 3 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		53	MC	0.66	0.45	2.70%				
OP		54	MC	0.29	0.30	1.57%				+
OP		55	MC	0.49	0.41	1.86%				
OP		56	MC	0.65	0.52	2.11%				
FT	A	57	MC	0.59	0.20	1.78%				
FT	A	58	MC	0.78	0.51	0.35%				
FT	A	59	MC	0.62	0.53	0.67%				
FT	A	60	MC	0.74	0.58	2.05%				
FT	A	61	MC	0.54	0.45	0.94%				
FT	A	62	MC	0.66	0.41	1.53%				
FT	A	63	CR	0.33	0.45	0.61%				
FT	A	64	MC	0.73	0.44	1.06%				
FT	A	65	MC	0.44	0.37	2.04%				
FT	A	66	MC	0.66	0.39	1.66%				
FT	A	67	MC	0.89	0.47	1.68%				
FT	B	57	MC	0.76	0.47	0.54%				
FT	B	58	MC	0.60	0.51	0.64%				
FT	B	59	MC	0.75	0.53	0.35%				
FT	B	60	MC	0.78	0.53	0.76%				
FT	B	61	MC	0.68	0.43	0.61%				
FT	B	62*	MC							
FT	B	63	CR	0.33	0.48	0.49%				
FT	B	64	MC	0.76	0.44	1.18%				
FT	B	65	MC	0.72	0.53	1.44%				
FT	B	66	MC	0.94	0.42	1.46%				

*This item did not converge, see Part 8 for more information

Table 7-21 Cont'd
Item Analysis Grade 3 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	B	67	MC	0.22	0.09	1.66%	+	+		+
FT	C	57	MC	0.59	0.19	0.38%				
FT	C	58	MC	0.52	0.33	0.48%				
FT	C	59	MC	0.54	0.24	0.33%		+		
FT	C	60	MC	0.42	0.36	0.69%		+		
FT	C	61	MC	0.32	0.29	0.89%				
FT	C	62	MC	0.59	0.39	2.07%				
FT	C	63	MC	0.75	0.43	0.90%				
FT	C	64	MC	0.47	0.39	1.90%				
FT	C	65	MC	0.68	0.37	1.65%				
FT	C	66	MC	0.92	0.45	1.03%				
FT	C	67	CR	0.44	0.51	0.56%				
FT	D	57	MC	0.75	0.47	1.51%				
FT	D	58	MC	0.80	0.52	0.28%				
FT	D	59	MC	0.32	0.31	0.50%				
FT	D	60	MC	0.67	0.40	1.18%				
FT	D	61	MC	0.71	0.56	1.02%				
FT	D	62	MC	0.61	0.51	1.67%				
FT	D	63	MC	0.75	0.43	0.91%				
FT	D	64	MC	0.73	0.52	1.07%				
FT	D	65	MC	0.94	0.41	1.19%				
FT	D	66	MC	0.21	0.09	1.52%	+	+		+
FT	D	67	CR	0.44	0.50	0.53%				

Table 7-22
Item Analysis Grade 4 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.64	0.34	0.09%				
OP		2	MC	0.60	0.57	0.11%				
OP		3	MC	0.76	0.43	0.20%				
OP		4	MC	0.72	0.49	0.50%				
OP		5	MC	0.76	0.55	0.19%				
OP		6	MC	0.58	0.39	0.62%				
OP		7	MC	0.79	0.44	2.27%				
OP		8	MC	0.79	0.51	0.20%				
OP		9	MC	0.63	0.39	0.51%				
OP		10	MC	0.70	0.42	1.12%				
OP		11	MC	0.34	0.28	0.19%				
OP		12	MC	0.37	0.36	0.30%		+		
OP		13	MC	0.67	0.36	0.73%				
OP		14	CR	0.39	0.57	0.00%				
OP		15	MC	0.72	0.41	1.17%				
OP		16	MC	0.74	0.35	1.54%				
OP		17	MC	0.81	0.62	1.46%				
OP		18	MC	0.82	0.28	1.77%				
OP		19	MC	0.74	0.54	1.77%				
OP		20	MC	0.41	0.42	0.22%				
OP		21	MC	0.80	0.53	0.33%				
OP		22	MC	0.57	0.32	0.25%				
OP		23	MC	0.66	0.26	0.76%				
OP		24	MC	0.83	0.43	1.63%				
OP		25	MC	0.49	0.28	0.19%				
OP		26	MC	0.57	0.45	0.36%				

Table 7-22 Cont'd
Item Analysis Grade 4 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.87	0.53	0.68%				
OP		28	MC	0.75	0.45	0.25%				
OP		29	MC	0.41	0.29	0.78%				
OP		30	MC	0.46	0.39	0.62%				
OP		31	MC	0.45	0.32	0.61%				
OP		32	MC	0.35	0.32	0.87%				
OP		33	CR	0.26	0.56	0.00%				+
OP		34	MC	0.56	0.49	1.71%				
OP		35	MC	0.69	0.30	1.91%				
OP		36	MC	0.79	0.56	1.99%				
OP		37	MC	0.51	0.45	2.37%				
OP		38	MC	0.73	0.55	2.69%				
OP		39	MC	0.83	0.42	0.23%				
OP		40	MC	0.79	0.58	0.30%				
OP		41	MC	0.79	0.63	0.44%				
OP		42	MC	0.74	0.56	0.56%				
OP		43	MC	0.88	0.52	1.01%				
OP		44	MC	0.72	0.59	2.16%				
OP		45	MC	0.64	0.45	0.26%				
OP		46	MC	0.67	0.52	0.39%				
OP		47	MC	0.80	0.36	1.56%				
OP		48	MC	0.74	0.45	0.37%				
OP		49	MC	0.84	0.56	0.44%				
OP		50	MC	0.73	0.40	1.46%				
OP		51	MC	0.81	0.51	1.31%				
OP		52	MC	0.56	0.33	1.43%				

Table 7-22 Cont'd
Item Analysis Grade 4 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		53	MC	0.76	0.51	0.34%				
OP		54	MC	0.74	0.50	0.86%				
OP		55	MC	0.58	0.45	0.62%				
FT	A	56	MC	0.72	0.59	0.28%				
FT	A	57	MC	0.54	0.38	0.34%				
FT	A	58	MC	0.75	0.47	0.94%				
FT	A	59	MC	0.56	0.55	1.47%				
FT	A	60	MC	0.81	0.50	0.40%				
FT	A	61	MC	0.60	0.42	0.67%				
FT	A	62	CR	0.29	0.49	0.75%				+
FT	A	63	MC	0.78	0.47	0.81%				
FT	A	64	MC	0.68	0.27	1.21%				
FT	A	65	MC	0.69	0.39	1.04%				
FT	A	66	MC	0.48	0.27	1.45%				
FT	B	56	MC	0.87	0.44	0.21%				
FT	B	57	MC	0.81	0.59	0.26%				
FT	B	58	MC	0.67	0.48	0.60%				
FT	B	59	MC	0.63	0.50	0.40%				
FT	B	60	MC	0.62	0.52	0.85%				
FT	B	61	MC	0.58	0.35	1.51%				
FT	B	62	CR	0.42	0.52	0.48%				
FT	B	63	MC	0.86	0.22	0.60%				
FT	B	64	MC	0.94	0.35	0.75%				
FT	B	65	MC	0.63	0.36	0.74%				
FT	B	66	MC	0.88	0.44	0.92%				
FT	C	56	MC	0.81	0.45	0.28%				

Table 7-22 Cont'd
Item Analysis Grade 4 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	C	57	MC	0.77	0.54	0.40%				
FT	C	58	MC	0.65	0.48	0.54%				
FT	C	59	MC	0.84	0.44	1.39%				
FT	C	60	MC	0.58	0.36	0.30%				
FT	C	61	MC	0.58	0.55	0.53%				
FT	C	62	MC	0.78	0.43	0.48%				
FT	C	63	MC	0.65	0.38	1.77%				
FT	C	64	MC	0.72	0.38	0.76%				
FT	C	65	MC	0.50	0.29	2.11%				
FT	C	66	CR	0.29	0.41	0.33%				+
FT	D	56	MC	0.70	0.49	0.24%				
FT	D	57	MC	0.73	0.45	0.31%				
FT	D	58	MC	0.74	0.60	0.32%				
FT	D	59	MC	0.72	0.56	0.94%				
FT	D	60	MC	0.88	0.53	0.25%				
FT	D	61	MC	0.84	0.53	0.48%				
FT	D	62	MC	0.80	0.06	0.34%	+	+		
FT	D	63	MC	0.94	0.37	0.43%				
FT	D	64	MC	0.62	0.36	0.50%				
FT	D	65	MC	0.88	0.45	1.18%				
FT	D	66	CR	0.32	0.46	0.22%				

Table 7-23
Item Analysis Grade 5 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.86	0.46	0.12%				
OP		2	MC	0.87	0.44	0.11%				
OP		3	MC	0.80	0.54	0.18%				
OP		4	MC	0.66	0.47	0.52%				
OP		5	MC	0.29	0.40	0.57%				+
OP		6	MC	0.68	0.41	0.43%				
OP		7	MC	0.78	0.53	0.61%				
OP		8	MC	0.80	0.48	1.10%				
OP		9	MC	0.35	0.25	0.23%		+		
OP		10	MC	0.65	0.54	0.32%				
OP		11	MC	0.56	0.42	0.57%				
OP		12	MC	0.85	0.54	0.20%				
OP		13	CR	0.54	0.57	0.00%				
OP		14	MC	0.85	0.44	2.02%				
OP		15	MC	0.78	0.44	2.11%				
OP		16	MC	0.53	0.41	2.43%				
OP		17	MC	0.68	0.45	2.46%				
OP		18	MC	0.81	0.49	2.71%				
OP		19	MC	0.52	0.23	2.97%				
OP		20	MC	0.77	0.45	0.23%				
OP		21	MC	0.75	0.45	0.14%				
OP		22	MC	0.76	0.51	0.18%				
OP		23	MC	0.48	0.30	0.21%				
OP		24	MC	0.69	0.46	0.37%				
OP		25	MC	0.76	0.56	0.20%				
OP		26	MC	0.56	0.44	0.41%				

Table 7-23 Cont'd
Item Analysis Grade 5 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.60	0.30	0.99%				
OP		28	MC	0.47	0.44	1.62%				
OP		29	MC	0.72	0.58	0.20%				
OP		30	MC	0.43	0.39	0.41%				
OP		31	MC	0.47	0.36	0.73%				
OP		32	MC	0.76	0.38	0.70%				
OP		33	MC	0.63	0.35	0.87%				
OP		34	MC	0.61	0.54	1.25%				
OP		35	MC	0.77	0.39	0.70%				
OP		36	MC	0.56	0.36	0.87%				
OP		37	MC	0.62	0.44	1.42%				
OP		38	MC	0.45	0.37	0.23%				
OP		39	MC	0.97	0.28	0.21%				
OP		40	MC	0.88	0.41	0.23%				
OP		41	MC	0.86	0.43	0.38%				
OP		42	MC	0.86	0.51	0.49%				
OP		43	MC	0.83	0.33	0.67%				
OP		44	MC	0.57	0.34	0.32%				
OP		45	MC	0.75	0.46	0.72%				
OP		46	MC	0.74	0.49	0.49%				
OP		47	MC	0.63	0.57	2.40%				
OP		48	CR	0.50	0.55	0.00%				
OP		49	MC	0.73	0.51	0.60%				
OP		50	MC	0.53	0.44	0.61%				
OP		51	MC	0.57	0.41	1.02%				
OP		52	MC	0.83	0.41	0.63%				

Table 7-23 Cont'd
Item Analysis Grade 5 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		53	MC	0.58	0.34	0.95%				
OP		54	MC	0.49	0.39	0.67%				
OP		55	MC	0.52	0.47	0.75%				
OP		56	MC	0.68	0.51	0.89%				
FT	A	57	MC	0.77	0.34	0.20%				
FT	A	58	MC	0.70	0.47	0.25%				
FT	A	59	MC	0.67	0.57	0.27%				
FT	A	60	MC	0.81	0.52	0.27%				
FT	A	61	MC	0.61	0.32	0.33%				
FT	A	62	MC	0.57	0.49	0.59%				
FT	A	63	MC	0.75	0.47	0.38%				
FT	A	64	MC	0.68	0.54	0.52%				
FT	A	65	MC	0.70	0.59	0.61%				
FT	A	66	MC	0.49	0.39	0.60%				
FT	A	67	MC	0.69	0.55	0.57%				
FT	A	68	MC	0.51	0.39	0.76%				
FT	A	69	MC	0.62	0.41	0.21%				
FT	A	70	MC	0.75	0.43	0.35%				
FT	A	71	CR	0.39	0.63	0.39%				
FT	B	57	MC	0.79	0.35	0.15%				
FT	B	58	MC	0.72	0.47	0.20%				
FT	B	59	MC	0.70	0.58	0.20%				
FT	B	60	MC	0.84	0.49	0.20%				
FT	B	61	MC	0.64	0.30	0.28%				
FT	B	62	MC	0.59	0.48	0.49%				
FT	B	63	MC	0.77	0.44	0.39%				

Table 7-23 Cont'd
Item Analysis Grade 5 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	B	64	MC	0.70	0.52	0.55%				
FT	B	65	MC	0.73	0.57	0.60%				
FT	B	66	MC	0.51	0.37	0.51%				
FT	B	67	MC	0.71	0.53	0.50%				
FT	B	68	MC	0.52	0.39	0.68%				
FT	B	69	MC	0.64	0.42	0.26%				
FT	B	70	MC	0.77	0.40	0.27%				
FT	B	71	CR	0.41	0.67	0.24%				
FT	C	57	MC	0.60	0.42	0.18%				
FT	C	58	MC	0.79	0.39	0.24%				
FT	C	59	MC	0.59	0.32	0.39%				
FT	C	60	MC	0.20	0.25	0.57%				+
FT	C	61	MC	0.86	0.34	0.65%				
FT	C	62	MC	0.32	0.33	0.60%		+		
FT	C	63	MC	0.76	0.42	0.46%				
FT	C	64	MC	0.53	0.42	0.66%				
FT	C	65	MC	0.81	0.48	0.41%				
FT	C	66	MC	0.72	0.53	0.58%				
FT	C	67	MC	0.81	0.51	0.54%				
FT	C	68	MC	0.74	0.43	0.48%				
FT	C	69	MC	0.38	0.29	0.88%				
FT	C	70	MC	0.68	0.50	1.63%				
FT	C	71	CR	0.34	0.50	0.24%				
FT	D	57	MC	0.60	0.42	0.17%				
FT	D	58	MC	0.80	0.38	0.20%				
FT	D	59	MC	0.60	0.31	0.29%				

Table 7-23 Cont'd
Item Analysis Grade 5 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	D	60	MC	0.21	0.25	0.48%		+		+
FT	D	61	MC	0.86	0.34	0.59%				
FT	D	62	MC	0.33	0.34	0.58%		+		
FT	D	63	MC	0.76	0.42	0.46%				
FT	D	64	MC	0.53	0.41	0.65%				
FT	D	65	MC	0.82	0.47	0.39%				
FT	D	66	MC	0.72	0.53	0.59%				
FT	D	67	MC	0.82	0.49	0.56%				
FT	D	68	MC	0.75	0.44	0.52%				
FT	D	69	MC	0.38	0.29	0.94%				
FT	D	70	MC	0.69	0.50	1.64%				
FT	D	71	CR	0.34	0.49	0.24%				

Table 7-24
Item Analysis Grade 6 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.55	0.32	0.09%				
OP		2	MC	0.91	0.45	0.09%				
OP		3	MC	0.63	0.47	0.17%				
OP		4	MC	0.79	0.56	0.17%				
OP		5	MC	0.68	0.27	0.56%				
OP		6	MC	0.68	0.50	0.18%				
OP		7	MC	0.46	0.31	0.17%				
OP		8	MC	0.83	0.48	2.33%				
OP		9	MC	0.56	0.22	3.33%				
OP		10	MC	0.37	0.29	0.29%				
OP		11	MC	0.81	0.40	0.32%				
OP		12	MC	0.67	0.40	0.47%				
OP		13	MC	0.45	0.39	0.49%				
OP		14	MC	0.81	0.41	0.17%				
OP		15	MC	0.80	0.50	0.29%				
OP		16	MC	0.81	0.39	0.65%				
OP		17	MC	0.71	0.47	0.18%				
OP		18	MC	0.76	0.45	0.33%				
OP		19	CR	0.43	0.48	0.00%				
OP		20	MC	0.82	0.42	0.17%				
OP		21	MC	0.76	0.53	0.29%				
OP		22	MC	0.76	0.31	0.56%				
OP		23	MC	0.96	0.41	0.17%				
OP		24	MC	0.55	0.47	0.56%				
OP		25	MC	0.82	0.52	0.33%				
OP		26	MC	0.77	0.47	0.47%				

Table 7-24 Cont'd
Item Analysis Grade 6 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	CR	0.31	0.49	0.00%				
OP		28	MC	0.53	0.38	1.16%		+		
OP		29	MC	0.88	0.35	1.36%				
OP		30	MC	0.80	0.30	4.28%				
OP		31	MC	0.70	0.29	1.25%				
OP		32	MC	0.76	0.48	1.26%				
OP		33	MC	0.77	0.47	1.83%				
OP		34	MC	0.72	0.37	2.37%				
OP		35	MC	0.83	0.39	1.31%				
OP		36	MC	0.79	0.36	1.58%				
OP		37	MC	0.48	0.28	1.45%				
OP		38	MC	0.73	0.53	1.96%				
OP		39	MC	0.60	0.32	0.44%				
OP		40	MC	0.85	0.40	0.38%				
OP		41	MC	0.67	0.48	0.50%				
OP		42	MC	0.59	0.40	0.49%				
OP		43	MC	0.43	0.33	1.04%		+		
OP		44	MC	0.75	0.36	0.41%				
OP		45	MC	0.68	0.49	0.68%				
OP		46	MC	0.78	0.60	1.71%				
OP		47	MC	0.82	0.52	2.07%				
OP		48	MC	0.67	0.27	0.49%				
OP		49	MC	0.47	0.38	0.55%				
OP		50	MC	0.31	0.34	0.67%				
OP		51	MC	0.80	0.54	1.19%				
OP		52	MC	0.70	0.32	0.73%				

Table 7-24 Cont'd
Item Analysis Grade 6 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		53	MC	0.53	0.44	0.68%				
OP		54	MC	0.68	0.54	4.83%				
OP		55	MC	0.51	0.30	0.82%				
OP		56	MC	0.60	0.35	1.04%				
FT	A	57	MC	0.68	0.37	0.29%				
FT	A	58	MC	0.80	0.57	0.39%				
FT	A	59	MC	0.91	0.50	0.35%				
FT	A	60	MC	0.58	0.41	0.38%				
FT	A	61	MC	0.62	0.31	0.71%				
FT	A	62	MC	0.85	0.46	0.72%				
FT	A	63	CR	0.43	0.62	0.30%				
FT	A	64	MC	0.74	0.43	0.78%				
FT	A	65	MC	0.75	0.40	0.89%				
FT	A	66	MC	0.74	0.56	0.94%				
FT	A	67	MC	0.77	0.33	0.95%				
FT	B	57	MC	0.63	0.35	0.38%				
FT	B	58	MC	0.84	0.45	0.32%				
FT	B	59	MC	0.67	0.41	0.43%				
FT	B	60	MC	0.87	0.44	0.35%				
FT	B	61	MC	0.27	0.24	0.71%		+		+
FT	B	62	MC	0.85	0.49	0.73%				
FT	B	63	CR	0.44	0.58	0.26%				
FT	B	64	MC	0.04	-0.20	0.80%	+	+		+
FT	B	65	MC	0.03	-0.20	0.85%	+	+		+
FT	B	66	MC	0.76	0.49	1.00%				

Table 7-24 Cont'd
Item Analysis Grade 6 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	B	67*	MC							
FT	C	57	MC	0.78	0.42	0.39%				
FT	C	58	MC	0.70	0.47	0.46%				
FT	C	59	MC	0.87	0.56	0.46%				
FT	C	60	MC	0.71	0.50	0.41%				
FT	C	61	MC	0.69	0.30	0.67%				
FT	C	62	MC	0.51	0.43	0.51%				
FT	C	63	MC	0.76	0.40	0.45%				
FT	C	64	MC	0.82	0.50	0.74%				
FT	C	65	MC	0.75	0.51	0.67%				
FT	C	66	MC	0.56	0.33	0.68%				
FT	C	67	CR	0.33	0.55	0.95%				
FT	D	57	MC	0.87	0.44	0.38%				
FT	D	58	MC	0.78	0.48	0.44%				
FT	D	59	MC	0.58	0.36	0.50%				
FT	D	60	MC	0.88	0.53	0.40%				
FT	D	61	MC	0.78	0.36	0.51%				
FT	D	62	MC	0.40	0.35	1.18%		+		
FT	D	63	MC	0.68	0.31	0.56%				
FT	D	64	MC	0.81	0.46	0.78%				
FT	D	65	MC	0.77	0.56	0.94%				
FT	D	66	MC	0.85	0.34	0.70%				
FT	D	67	CR	0.38	0.64	1.02%				

* This item did not converge, see Part 8 for more information.

Table 7-25
Item Analysis Grade 7 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.70	0.58	0.24%				
OP		2	MC	0.70	0.37	0.16%				
OP		3	MC	0.70	0.41	0.13%				
OP		4	MC	0.56	0.45	0.13%				
OP		5	MC	0.66	0.33	0.19%				
OP		6	MC	0.88	0.45	0.15%				
OP		7	MC	0.33	0.37	0.24%				
OP		8	MC	0.85	0.51	0.34%				
OP		9	MC	0.61	0.32	0.86%				
OP		10	MC	0.70	0.40	0.98%				
OP		11	MC	0.70	0.32	1.51%				
OP		12	MC	0.76	0.40	0.38%				
OP		13	MC	0.32	0.23	0.71%				
OP		14	MC	0.70	0.44	0.89%				
OP		15	MC	0.59	0.40	1.01%				
OP		16	MC	0.30	0.37	1.18%				+
OP		17	MC	0.57	0.34	1.05%				
OP		18	MC	0.48	0.39	1.35%				
OP		19	MC	0.62	0.51	0.27%				
OP		20	MC	0.34	0.35	0.34%		+		
OP		21	MC	0.43	0.40	0.58%				
OP		22	MC	0.78	0.34	0.87%				
OP		23	MC	0.84	0.53	0.31%				
OP		24	MC	0.70	0.55	0.41%				
OP		25	MC	0.52	0.39	2.09%				
OP		26	MC	0.93	0.36	2.35%				

Table 7-25 Cont'd
Item Analysis Grade 7 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.69	0.41	0.50%				
OP		28	MC	0.82	0.50	0.53%				
OP		29	MC	0.76	0.40	0.43%				
OP		30	MC	0.48	0.35	1.39%				
OP		31	MC	0.83	0.50	0.36%				
OP		32	MC	0.79	0.44	0.43%				
OP		33	MC	0.71	0.38	0.78%				
OP		34	MC	0.76	0.46	0.40%				
OP		35	MC	0.49	0.39	0.80%				
OP		36	MC	0.63	0.40	0.70%				
OP		37	CR	0.46	0.64	0.00%				
OP		38	MC	0.79	0.56	0.38%				
OP		39	MC	0.75	0.22	0.46%				
OP		40	MC	0.71	0.32	0.43%				
OP		41	MC	0.78	0.55	0.77%				
OP		42	MC	0.93	0.38	0.50%				
OP		43	MC	0.83	0.39	0.73%				
OP		44	MC	0.67	0.48	0.99%				
OP		45	MC	0.65	0.57	2.13%				
OP		46	MC	0.49	0.35	0.89%				
OP		47	MC	0.32	0.30	1.07%		+		
OP		48	MC	0.63	0.36	1.60%				
OP		49	MC	0.50	0.45	1.63%				
OP		50	MC	0.66	0.53	1.85%				
OP		51	MC	0.70	0.40	1.95%				
OP		52	MC	0.42	0.31	0.98%				

Table 7-25 Cont'd
Item Analysis Grade 7 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		53	MC	0.52	0.40	1.17%				
OP		54	MC	0.78	0.51	1.14%				
OP		55	MC	0.64	0.44	1.78%				
OP		56	CR	0.26	0.59	0.00%				+
FT	A	57	MC	0.66	0.47	0.66%				
FT	A	58	MC	0.61	0.34	0.78%				
FT	A	59	MC	0.78	0.54	0.91%				
FT	A	60	MC	0.75	0.42	1.73%				
FT	A	61	MC	0.91	0.44	1.70%				
FT	A	62	MC	0.89	0.47	0.64%				
FT	A	63	MC	0.74	0.52	0.85%				
FT	A	64	MC	0.56	0.46	1.23%				
FT	A	65	MC	0.63	0.39	0.81%				
FT	A	66	MC	0.67	0.39	1.55%				
FT	A	67	MC	0.48	0.30	0.62%				
FT	A	68*	MC							
FT	A	69	MC	0.41	0.12	0.82%	+			
FT	A	70	MC	0.80	0.50	1.04%				
FT	A	71	MC	0.53	0.40	1.30%				
FT	A	72	CR	0.34	0.57	0.76%				
FT	B	57	MC	0.67	0.46	0.42%				
FT	B	58	MC	0.63	0.33	0.49%				
FT	B	59	MC	0.81	0.52	0.68%				
FT	B	60	MC	0.77	0.41	1.59%				
FT	B	61	MC	0.92	0.44	1.56%				

* This item did not converge, see Part 8 for more information.

Table 7-25 Cont'd
Item Analysis Grade 7 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	B	62	MC	0.91	0.46	0.40%				
FT	B	63	MC	0.77	0.50	0.63%				
FT	B	64	MC	0.58	0.45	0.85%				
FT	B	65	MC	0.65	0.38	0.56%				
FT	B	66	MC	0.68	0.34	1.39%				
FT	B	67	MC	0.73	0.35	0.47%				
FT	B	68*	MC							
FT	B	69	MC	0.41	0.28	0.55%				
FT	B	70	MC	0.82	0.50	1.01%				
FT	B	71	MC	0.55	0.40	1.06%				
FT	B	72	CR	0.40	0.47	0.24%				
FT	C	57	MC	0.68	0.44	0.40%				
FT	C	58	MC	0.63	0.34	0.46%				
FT	C	59	MC	0.81	0.53	0.62%				
FT	C	60	MC	0.77	0.40	1.64%				
FT	C	61	MC	0.92	0.43	1.67%				
FT	C	62	MC	0.90	0.46	0.40%				
FT	C	63	MC	0.76	0.50	0.62%				
FT	C	64	MC	0.59	0.43	0.84%				
FT	C	65	MC	0.65	0.39	0.51%				
FT	C	66	MC	0.68	0.37	1.27%				
FT	C	67	MC	0.51	0.29	0.42%				
FT	C	68*	MC							
FT	C	69	MC	0.41	0.11	0.70%	+			
FT	C	70	MC	0.81	0.49	1.03%				

* This item did not converge, see Part 8 for more information.

Table 7-25 Cont'd
Item Analysis Grade 7 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	C	71	MC	0.55	0.39	1.16%				
FT	C	72	CR	0.36	0.57	0.65%				
FT	D	57	MC	0.68	0.46	0.48%				
FT	D	58	MC	0.64	0.32	0.58%				
FT	D	59	MC	0.81	0.52	0.60%				
FT	D	60	MC	0.77	0.41	1.77%				
FT	D	61	MC	0.92	0.42	1.65%				
FT	D	62	MC	0.91	0.45	0.45%				
FT	D	63	MC	0.77	0.50	0.72%				
FT	D	64	MC	0.59	0.44	0.98%				
FT	D	65	MC	0.66	0.39	0.66%				
FT	D	66	MC	0.69	0.35	1.33%				
FT	D	67	MC	0.74	0.37	0.52%				
FT	D	68*	MC							
FT	D	69	MC	0.41	0.27	0.60%				
FT	D	70	MC	0.82	0.49	0.98%				
FT	D	71	MC	0.55	0.40	1.16%				
FT	D	72	CR	0.41	0.51	0.21%				

* This item did not converge, see Part 8 for more information.

Table 7-26
Item Analysis Grade 8 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.78	0.37	0.04%				
OP		2	MC	0.47	0.28	0.00%				
OP		3	MC	0.72	0.41	0.28%				
OP		4	MC	0.75	0.39	0.17%				
OP		5	MC	0.52	0.44	0.25%				
OP		6	MC	0.70	0.45	0.12%				
OP		7	MC	0.63	0.47	0.22%				
OP		8	MC	0.80	0.50	0.54%				
OP		9	MC	0.44	0.35	1.11%				
OP		10	MC	0.72	0.45	0.13%				
OP		11	MC	0.80	0.54	0.31%				
OP		12	MC	0.59	0.50	0.44%				
OP		13	MC	0.61	0.41	1.28%				
OP		14	MC	0.54	0.33	1.29%				
OP		15	MC	0.68	0.24	1.32%				
OP		16	MC	0.55	0.38	1.72%				
OP		17	MC	0.57	0.38	1.16%				
OP		18	MC	0.61	0.49	1.61%				
OP		19	MC	0.60	0.39	0.26%				
OP		20	MC	0.88	0.52	0.16%				
OP		21	MC	0.57	0.49	0.20%				
OP		22	MC	0.86	0.49	0.52%				
OP		23	MC	0.77	0.47	0.28%				
OP		24	MC	0.84	0.46	0.31%				
OP		25	MC	0.85	0.51	0.35%				
OP		26	MC	0.62	0.45	0.45%				

Table 7-26
Item Analysis Grade 8 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.73	0.51	0.25%				
OP		28	MC	0.55	0.49	0.92%				
OP		29	MC	0.72	0.48	0.39%				
OP		30	MC	0.55	0.36	0.90%				
OP		31	CR	0.42	0.55	0.00%				
OP		32	MC	0.79	0.55	0.33%				
OP		33	MC	0.75	0.51	0.51%				
OP		34	MC	0.76	0.34	0.39%				
OP		35	MC	0.77	0.33	0.48%				
OP		36	MC	0.69	0.49	0.47%				
OP		37	MC	0.76	0.47	0.61%				
OP		38	MC	0.71	0.42	0.35%				
OP		39	MC	0.66	0.51	0.26%				
OP		40	MC	0.55	0.44	0.60%				
OP		41	MC	0.79	0.48	0.73%				
OP		42	MC	0.76	0.56	0.32%				
OP		43	MC	0.45	0.19	0.48%		+		
OP		44	MC	0.75	0.35	0.65%				
OP		45	MC	0.64	0.16	0.47%		+		
OP		46	MC	0.62	0.35	0.36%				
OP		47	MC	0.83	0.57	0.42%				
OP		48	MC	0.47	0.37	0.55%				
OP		49	MC	0.89	0.50	0.97%				
OP		50	MC	0.81	0.56	0.44%				
OP		51	MC	0.83	0.47	0.63%				
OP		52	MC	0.76	0.34	0.81%				

Table 7-26 Cont'd
Item Analysis Grade 8 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		53	MC	0.55	0.49	1.18%				
OP		54	MC	0.45	0.41	0.67%				
OP		55	MC	0.58	0.32	0.73%				
OP		56	CR	0.55	0.58	0.00%				
FT	A	57	MC	0.88	0.55	0.25%				
FT	A	58	MC	0.79	0.48	0.30%				
FT	A	59	MC	0.84	0.44	0.30%				
FT	A	60	MC	0.58	0.48	0.46%				
FT	A	61	MC	0.69	0.45	0.43%				
FT	A	62	MC	0.65	0.38	0.48%				
FT	A	63	CR	0.40	0.59	1.00%				
FT	A	64	MC	0.76	0.55	0.70%				
FT	A	65	MC	0.78	0.39	0.76%				
FT	A	66	MC	0.90	0.46	0.79%				
FT	A	67	MC	0.78	0.49	0.92%				
FT	B	57	MC	0.92	0.45	0.27%				
FT	B	58	MC	0.51	0.11	0.34%	+			
FT	B	59	MC	0.90	0.51	0.43%				
FT	B	60	MC	0.44	0.41	0.43%				
FT	B	61	MC	0.60	0.33	0.59%				
FT	B	62	MC	0.73	0.46	0.72%				
FT	B	63	CR	0.50	0.63	0.95%				
FT	B	64	MC	0.76	0.53	0.92%				
FT	B	65	MC	0.56	0.30	1.02%				
FT	B	66	MC	0.27	0.18	1.03%				+
FT	B	67	MC	0.38	0.37	1.17%				

Table 7-26 Cont'd
Item Analysis Grade 8 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	C	57	MC	0.81	0.53	0.24%				
FT	C	58	MC	0.89	0.48	0.27%				
FT	C	59	MC	0.67	0.37	0.26%				
FT	C	60	MC	0.45	0.33	0.38%				
FT	C	61	MC	0.84	0.56	0.46%				
FT	C	62	MC	0.79	0.43	0.62%				
FT	C	63	MC	0.76	0.52	1.15%				
FT	C	64	MC	0.91	0.42	1.37%				
FT	C	65	MC	0.76	0.49	1.17%				
FT	C	66	MC	0.46	0.38	1.20%				
FT	C	67	CR	0.33	0.51	1.19%				
FT	D	57	MC	0.79	0.52	0.25%				
FT	D	58	MC	0.86	0.39	0.27%				
FT	D	59	MC	0.71	0.49	0.28%				
FT	D	60	MC	0.50	0.42	0.49%				
FT	D	61	MC	0.88	0.51	0.39%				
FT	D	62	MC	0.52	0.40	0.37%				
FT	D	63	MC	0.62	0.34	1.40%				
FT	D	64	MC	0.74	0.48	1.51%				
FT	D	65	MC	0.65	0.40	1.44%				
FT	D	66	MC	0.45	0.42	1.84%				
FT	D	67	CR	0.41	0.62	1.24%				

Table 7-27
Item Analysis Grade 10 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.84	0.45	0.04%				
OP		2	MC	0.56	0.56	0.03%				
OP		3	MC	0.82	0.48	0.07%				
OP		4	MC	0.64	0.52	0.28%				
OP		5	MC	0.74	0.52	0.83%				
OP		6	MC	0.56	0.46	0.13%				
OP		7	MC	0.77	0.52	0.15%				
OP		8	MC	0.60	0.40	0.24%				
OP		9	MC	0.55	0.51	0.40%				
OP		10	MC	0.53	0.16	0.26%		+		
OP		11	MC	0.86	0.51	0.28%				
OP		12	MC	0.45	0.32	0.26%				
OP		13	MC	0.88	0.46	0.29%				
OP		14	MC	0.62	0.46	2.09%				
OP		15	MC	0.53	0.45	2.22%				
OP		16	MC	0.49	0.36	2.34%				
OP		17	MC	0.33	0.18	2.22%		+		
OP		18	MC	0.49	0.37	2.22%				
OP		19	MC	0.84	0.49	0.26%				
OP		20	MC	0.54	0.39	0.41%				
OP		21	MC	0.68	0.44	0.28%				
OP		22	MC	0.66	0.47	0.36%				
OP		23	MC	0.65	0.41	0.90%				
OP		24	MC	0.68	0.43	0.29%				
OP		25	MC	0.58	0.22	0.28%				
OP		26	MC	0.67	0.56	0.24%				

Table 7-27 Cont'd
Item Analysis Grade 10 Reading

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.58	0.49	0.24%				
OP		28	MC	0.74	0.56	0.33%				
OP		29	MC	0.82	0.51	0.63%				
OP		30	MC	0.75	0.42	0.48%				
OP		31	CR	0.62	0.64	0.00%				
OP		32	MC	0.63	0.27	2.46%				
OP		33	MC	0.64	0.31	2.46%				
OP		34	MC	0.72	0.53	2.54%				
OP		35	MC	0.64	0.43	0.77%				
OP		36	MC	0.62	0.48	0.57%				
OP		37	MC	0.55	0.45	0.81%				
OP		38	MC	0.45	0.37	0.70%				
OP		39	MC	0.71	0.52	0.67%				
OP		40	MC	0.95	0.35	0.49%				
OP		41	MC	0.59	0.34	0.71%				
OP		42	MC	0.82	0.52	0.51%				
OP		43	MC	0.70	0.44	0.77%				
OP		44	MC	0.90	0.51	0.57%				
OP		45	MC	0.58	0.45	0.67%				
OP		46	MC	0.50	0.44	0.66%				
OP		47	MC	0.70	0.61	0.67%				
OP		48	MC	0.57	0.41	1.12%				
OP		49	MC	0.52	0.33	1.04%		+		
OP		50	MC	0.63	0.56	0.94%				
OP		51	MC	0.51	0.33	0.90%				
OP		52	CR	0.48	0.64	0.00%				

Table 7-28
Item Analysis Grade 3 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.68	0.41	0.58%				
OP		2	MC	0.79	0.48	0.22%				
OP		3	MC	0.83	0.49	0.26%				
OP		4	MC	0.94	0.27	0.17%				
OP		5	MC	0.93	0.40	0.29%				
OP		6	MC	0.53	0.47	0.37%				
OP		7	MC	0.80	0.46	0.40%				
OP		8A	CR	0.49	0.59	0.00%				
OP		8B	CR	0.31	0.56	0.00%				
OP		9	MC	0.89	0.30	1.17%				
OP		10	MC	0.78	0.44	1.54%				
OP		11	MC	0.91	0.35	1.48%				
OP		12	MC	0.62	0.42	2.48%				
OP		13	MC	0.61	0.38	2.62%				
OP		14	MC	0.80	0.49	2.68%				
OP		15	MC	0.90	0.27	0.15%				
OP		16	MC	0.90	0.35	0.20%				
OP		17	MC	0.88	0.41	0.08%				
OP		18	MC	0.70	0.39	0.52%				
OP		19A	CR	0.71	0.57	0.00%				
OP		19B	CR	0.38	0.57	0.00%				
OP		20	MC	0.79	0.37	0.23%				
OP		21	MC	0.50	0.47	0.51%				
OP		22	MC	0.94	0.33	1.49%				
OP		23	MC	0.80	0.29	1.56%				
OP		24	MC	0.75	0.45	1.56%				

Table 7-28 Cont'd
Item Analysis Grade 3 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		25	MC	0.69	0.48	0.35%				
OP		26A	CR	0.70	0.55	0.00%				
OP		26B	CR	0.28	0.52	0.00%				+
OP		27	MC	0.86	0.44	1.35%				
OP		28	MC	0.94	0.35	1.19%				
OP		29	MC	0.56	0.42	1.66%				
OP		30	MC	0.77	0.44	1.57%				
OP		31	MC	0.97	0.15	0.62%				
OP		32	MC	0.88	0.37	0.77%				
OP		33	MC	0.51	0.40	0.95%				
OP		34	MC	0.62	0.36	0.94%				
OP		35	MC	0.61	0.37	0.92%				
OP		36	MC	0.46	0.52	1.40%				
OP		37	MC	0.76	0.33	1.71%				
OP		38	MC	0.78	0.32	0.58%				
OP		39	MC	0.89	0.38	0.83%				
OP		40	MC	0.58	0.42	0.46%				
OP		41	MC	0.53	0.30	0.86%				
OP		42A	CR	0.67	0.53	0.00%				
OP		42B	CR	0.47	0.58	0.00%				
OP		43	MC	0.93	0.33	0.77%				
OP		44	MC	0.91	0.29	1.17%				
OP		45	MC	0.75	0.40	0.80%				
OP		46	MC	0.89	0.41	1.25%				
OP		47	MC	0.74	0.46	1.19%				
OP		48	MC	0.71	0.53	1.59%				

Table 7-28 Cont'd
Item Analysis Grade 3 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		49	MC	0.93	0.22	1.31%				
FT	A	50	MC	0.45	0.41	0.37%				
FT	A	51	MC	0.89	0.39	0.53%				
FT	A	52A	CR	0.60	0.62	0.70%				
FT	A	52B	CR	0.44	0.59	1.32%				
FT	A	53	MC	0.96	0.28	0.47%				
FT	A	54	MC	0.89	0.43	0.93%				
FT	A	55	MC	0.55	0.36	1.15%				
FT	A	56	MC	0.78	0.46	0.51%				
FT	A	57	MC	0.89	0.40	0.81%				
FT	A	58	MC	0.79	0.44	2.18%				
FT	B	50	MC	0.79	0.54	0.36%				
FT	B	51	MC	0.88	0.37	0.46%				
FT	B	52A	CR	0.47	0.56	0.75%				
FT	B	52B	CR	0.37	0.51	1.21%				
FT	B	53	MC	0.93	0.39	0.44%				
FT	B	54	MC	0.85	0.37	0.50%				
FT	B	55	MC	0.87	0.45	1.26%				
FT	B	56	MC	0.69	0.34	1.10%				
FT	B	57	MC	0.94	0.23	0.56%				
FT	B	58	MC	0.75	0.48	0.63%				
FT	C	50	MC	0.81	0.45	0.40%				
FT	C	51	MC	0.76	0.47	0.60%				
FT	C	52	MC	0.93	0.43	0.62%				
FT	C	53	MC	0.41	0.33	1.35%				
FT	C	54	MC	0.61	0.54	0.51%				

Table 7-28 Cont'd
Item Analysis Grade 3 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	C	55	MC	0.60	0.48	0.81%				
FT	C	56	MC	0.62	0.39	0.35%				
FT	C	57	MC	0.93	0.35	0.40%				
FT	C	58A	CR	0.38	0.56	0.48%				
FT	C	58B	CR	0.41	0.56	1.00%				
FT	D	50	MC	0.91	0.37	0.30%				
FT	D	51	MC	0.85	0.47	0.68%				
FT	D	52	MC	0.41	0.42	0.96%				
FT	D	53	MC	0.49	0.41	1.65%				
FT	D	54	MC	0.63	0.41	0.66%				
FT	D	55	MC	0.48	0.47	1.06%				
FT	D	56	MC	0.74	0.55	0.40%				
FT	D	57	MC	0.79	0.47	1.76%				
FT	D	58A	CR	0.95	0.19	0.42%				
FT	D	58B	CR	0.67	0.38	0.56%				

Table 7-29
Item Analysis Grade 4 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.66	0.49	0.79%				
OP		2	MC	0.94	0.28	0.11%				
OP		3	MC	0.70	0.54	0.29%				
OP		4	MC	0.87	0.33	0.57%				
OP		5	MC	0.90	0.40	0.69%				
OP		6	MC	0.83	0.44	0.29%				
OP		7	MC	0.75	0.46	0.40%				
OP		8	MC	0.92	0.38	0.18%				
OP		9	MC	0.90	0.32	0.42%				
OP		10	MC	0.75	0.55	0.39%				
OP		11	MC	0.94	0.38	1.48%				
OP		13	MC	0.63	0.56	1.29%				
OP		14	MC	0.75	0.56	1.39%				
OP		15	MC	0.72	0.49	0.52%				
OP		16	MC	0.93	0.39	0.63%				
OP		17	MC	0.94	0.18	0.91%				
OP		19	MC	0.82	0.37	0.22%				
OP		20	MC	0.73	0.33	0.63%				
OP		21	MC	0.89	0.42	0.31%				
OP		22	MC	0.45	0.37	0.55%				
OP		23	MC	0.95	0.18	0.85%				
OP		24	MC	0.84	0.37	0.60%				
OP		25	MC	0.88	0.30	0.42%				
OP		26A	CR	0.50	0.59	0.00%				
OP		26B	CR	0.29	0.52	0.00%				+
OP		27	MC	0.64	0.41	0.63%				

Table 7-29 Cont'd
Item Analysis Grade 4 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		28	MC	0.71	0.42	0.75%				
OP		29	MC	0.59	0.50	0.63%				
OP		30	MC	0.43	0.47	0.86%				
OP		31	MC	0.93	0.30	0.26%				
OP		32	MC	0.95	0.17	0.29%				
OP		33	MC	0.93	0.30	0.39%				
OP		34	MC	0.95	0.24	0.71%				
OP		35	MC	0.87	0.32	0.57%				
OP		36	MC	0.93	0.40	1.63%				
OP		37	MC	0.52	0.40	0.66%				
OP		38	MC	0.73	0.54	4.65%				
OP		39	MC	0.66	0.49	0.45%				
OP		40	MC	0.96	0.19	0.37%				
OP		41	MC	0.58	0.48	0.39%				
OP		42	MC	0.90	0.21	0.48%				
OP		43	MC	0.74	0.54	0.42%				
OP		44	MC	0.40	0.50	0.31%				
OP		45	MC	0.51	0.49	0.22%				
OP		46	MC	0.75	0.58	0.59%				
OP		47A	CR	0.75	0.40	0.00%				
OP		47B	CR	0.32	0.54	0.00%				
OP		48	MC	0.90	0.44	0.79%				
OP		49	MC	0.86	0.37	0.68%				
FT	A	50	MC	0.67	0.46	0.31%				
FT	A	51	MC	0.57	0.48	0.43%				
FT	A	52A	CR	0.58	0.61	0.28%				

Table 7-29 Cont'd
Item Analysis Grade 4 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	A	52B	CR	0.55	0.65	0.58%				
FT	A	53	MC	0.92	0.29	0.35%				
FT	A	54	MC	0.80	0.38	0.76%				
FT	A	55	MC	0.27	0.30	0.37%		+		+
FT	A	56	MC	0.79	0.40	0.69%				
FT	A	57	MC	0.37	0.39	3.61%				
FT	A	58	MC	0.70	0.36	1.71%				
FT	B	50	MC	0.43	0.45	0.22%				
FT	B	51	MC	0.37	0.41	0.31%				
FT	B	52A	CR	0.54	0.58	0.27%				
FT	B	52B	CR	0.37	0.60	0.58%				
FT	B	53	MC	0.68	0.46	0.50%				
FT	B	54	MC	0.91	0.31	0.41%				
FT	B	55	MC	0.90	0.40	1.49%				
FT	B	56	MC	0.27	0.34	0.67%		+		+
FT	B	57	MC	0.81	0.40	0.78%				
FT	B	58	MC	0.69	0.43	1.03%				
FT	C	50	MC	0.74	0.42	0.33%				
FT	C	51	MC	0.48	0.47	0.34%				
FT	C	52A	CR	.	.					
FT	C	52B	CR	.	.					
FT	C	53	MC	0.97	0.22	0.28%				
FT	C	54	MC	0.80	0.31	0.51%				
FT	C	55	MC	0.74	0.27	0.67%				
FT	C	56	MC	0.68	0.43	0.65%				
FT	C	57	MC	0.77	0.36	0.45%				

Table 7-29 Cont'd
Item Analysis Grade 4 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	C	58	MC	0.78	0.41	1.49%				
FT	D	50	MC	0.83	0.43	0.25%				
FT	D	51	MC	0.83	0.42	0.49%				
FT	D	52	MC	0.88	0.32	1.43%				
FT	D	53	MC	0.73	0.50	1.93%				
FT	D	54	MC	0.85	0.43	0.36%				
FT	D	55	MC	0.58	0.39	0.62%				
FT	D	56	MC	0.77	0.29	0.33%				
FT	D	57	MC	0.61	0.40	0.58%				
FT	D	58A	CR	0.61	0.53	0.26%				
FT	D	58B	CR	0.45	0.57	0.43%				

Table 7-30
Item Analysis Grade 5 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.97	0.22	0.14%				
OP		2	MC	0.61	0.39	0.27%				
OP		3	MC	0.68	0.41	0.30%				
OP		4	MC	0.92	0.25	0.14%				
OP		5	MC	0.56	0.49	3.95%				
OP		6	MC	0.92	0.33	0.26%				
OP		7	MC	0.49	0.52	0.29%				
OP		8	MC	0.92	0.30	0.41%				
OP		9	MC	0.82	0.49	0.41%				
OP		10A	CR	0.54	0.54	0.00%				
OP		10B	CR	0.51	0.55	0.00%				
OP		11	MC	0.82	0.37	1.20%				
OP		12	MC	0.62	0.43	1.13%				
OP		13	MC	0.54	0.42	0.52%				
OP		14	MC	0.69	0.45	0.67%				
OP		15	MC	0.93	0.24	0.23%				
OP		16	MC	0.65	0.50	0.20%				
OP		17	MC	0.60	0.33	0.50%				
OP		18	MC	0.97	0.27	0.46%				
OP		19A	CR	0.85	0.26	0.00%				
OP		19B	CR	0.29	0.53	0.00%				+
OP		20	MC	0.64	0.28	0.67%				
OP		21	MC	0.66	0.49	0.90%				
OP		22	MC	0.26	0.30	0.68%				+
OP		23	MC	0.51	0.41	1.88%				
OP		24	MC	0.72	0.47	1.93%				

Table 7-30 Cont'd
Item Analysis Grade 5 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		25	MC	0.64	0.44	1.69%				
OP		26	MC	0.47	0.45	2.22%				
OP		27	MC	0.71	0.36	0.90%				
OP		28	MC	0.86	0.31	0.29%				
OP		29A	CR	0.25	0.54	0.00%				+
OP		29B	CR	0.53	0.56	0.00%				
OP		30	MC	0.45	0.48	1.20%				
OP		31	MC	0.65	0.35	0.65%				
OP		32	MC	0.80	0.19	1.78%				
OP		33	MC	0.85	0.28	2.49%				
OP		34	MC	0.85	0.44	1.03%				
OP		35	MC	0.67	0.38	0.74%				
OP		36	MC	0.90	0.36	0.74%				
OP		37	MC	0.61	0.49	0.81%				
OP		38	MC	0.73	0.42	0.23%				
OP		39	MC	0.67	0.54	0.32%				
OP		40	MC	0.50	0.34	0.64%				
OP		41A	CR	0.76	0.58	0.00%				
OP		41B	CR	0.42	0.56	0.00%				
OP		42	MC	0.91	0.18	0.43%				
OP		43	MC	0.48	0.43	0.67%				
OP		44	MC	0.75	0.19	0.55%				
OP		45	MC	0.62	0.40	0.85%				
OP		46	MC	0.84	0.43	0.87%				
OP		47	MC	0.93	0.30	0.50%				
OP		48	MC	0.73	0.36	0.87%				

Table 7-30 Cont'd
Item Analysis Grade 5 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		49	MC	0.96	0.26	0.55%				
OP		50	MC	0.59	0.32	0.61%				
OP		51	MC	0.46	0.50	0.87%				
OP		52	MC	0.81	0.31	0.74%				
OP		53	MC	0.59	0.52	0.43%				
OP		54	MC	0.97	0.24	0.49%				
FT	A	55	MC	0.51	0.38	0.24%				
FT	A	56	MC	0.73	0.48	0.40%				
FT	A	57A	CR	0.39	0.46	0.24%				
FT	A	57B	CR	0.45	0.61	0.47%				
FT	A	58	MC	0.65	0.30	0.25%				
FT	A	59	MC	0.82	0.39	0.26%				
FT	A	60	MC	0.36	0.40	1.12%				
FT	A	61	MC	0.73	0.42	1.64%				
FT	A	62	MC	0.78	0.46	0.24%				
FT	A	63	MC	0.51	0.47	0.29%				
FT	B	55	MC	0.82	0.43	0.22%				
FT	B	56	MC	0.45	0.52	0.27%				
FT	B	57A	CR	0.66	0.48	0.29%				
FT	B	57B	CR	0.30	0.57	0.58%				
FT	B	58	MC	0.81	0.39	0.56%				
FT	B	59	MC	0.70	0.46	0.33%				
FT	B	60	MC	0.67	0.51	0.37%				
FT	B	61	MC	0.87	0.36	0.37%				
FT	B	62	MC	0.96	0.24	0.37%				
FT	B	63	MC	0.71	0.36	0.43%				

Table 7-30 Cont'd
Item Analysis Grade 5 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	C	55	MC	0.81	0.49	0.20%				
FT	C	56	MC	0.63	0.48	0.15%				
FT	C	57A	CR	0.46	0.63	0.38%				
FT	C	57B	CR	0.54	0.69	0.63%				
FT	C	58	MC	0.64	0.52	0.40%				
FT	C	59	MC	0.38	0.39	0.30%				
FT	C	60	MC	0.37	0.40	0.41%		+		
FT	C	61	MC	0.82	0.45	0.26%				
FT	C	62	MC	0.83	0.41	0.67%				
FT	C	63	MC	0.83	0.45	0.47%				
FT	D	55	MC	0.67	0.51	0.41%				
FT	D	56	MC	0.75	0.45	0.30%				
FT	D	57	MC	0.86	0.33	0.53%				
FT	D	58	MC	0.95	0.39	0.30%				
FT	D	59	MC	0.51	0.35	0.86%				
FT	D	60	MC	0.44	0.32	1.13%				
FT	D	61	MC	0.37	0.44	0.32%				
FT	D	62	MC	0.43	0.43	0.26%				
FT	D	63A	CR	0.55	0.60	0.31%				
FT	D	63B	CR	0.35	0.52	0.53%				

Table 7-31
Item Analysis Grade 6 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.79	0.44	0.55%				
OP		2	MC	0.65	0.42	0.44%				
OP		3	MC	0.59	0.61	0.41%				
OP		4	MC	0.89	0.41	0.36%				
OP		5	MC	0.89	0.37	0.21%				
OP		6	MC	0.97	0.18	0.15%				
OP		7	MC	0.82	0.41	0.58%				
OP		8	MC	0.71	0.47	0.71%				
OP		9	MC	0.98	0.24	0.44%				
OP		10	MC	0.87	0.48	0.88%				
OP		11A	CR	0.51	0.63	0.00%				
OP		11B	CR	0.54	0.67	0.00%				
OP		12	MC	0.50	0.53	1.38%				
OP		13	MC	0.49	0.42	4.46%				
OP		14	MC	0.86	0.52	0.88%				
OP		15	MC	0.46	0.61	1.05%				
OP		16	MC	0.58	0.46	0.92%				
OP		17	MC	0.77	0.39	0.26%				
OP		18	MC	0.71	0.44	0.27%				
OP		19	MC	0.92	0.31	0.29%				
OP		20	MC	0.39	0.38	1.29%				
OP		21	MC	0.49	0.45	1.46%				
OP		22A	CR	0.86	0.36	0.00%				
OP		22B	CR	0.32	0.55	0.00%				
OP		23	MC	0.81	0.44	0.24%				
OP		24	MC	0.61	0.39	0.36%				

Table 7-31 Cont'd
Item Analysis Grade 6 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		25	MC	0.69	0.33	0.48%		+		
OP		26	MC	0.82	0.41	0.44%				
OP		27	MC	0.62	0.43	0.44%				
OP		28	MC	0.62	0.53	0.47%				
OP		29	MC	0.84	0.51	0.20%				
OP		30	MC	0.71	0.53	0.23%				
OP		31	MC	0.67	0.36	0.35%				
OP		32	MC	0.77	0.48	0.35%				
OP		33A	CR	0.67	0.50	0.00%				
OP		33B	CR	0.48	0.48	0.00%				
OP		34	MC	0.79	0.48	0.94%				
OP		35	MC	0.68	0.55	1.09%				
OP		36	MC	0.48	0.55	0.26%				
OP		37	MC	0.75	0.18	1.20%				
OP		38	MC	0.64	0.34	0.36%				
OP		39	MC	0.50	0.40	0.89%				
OP		40	MC	0.65	0.34	0.47%				
OP		41	MC	0.66	0.32	0.53%				
OP		42	MC	0.71	0.35	0.42%				
OP		43	MC	0.77	0.28	0.41%				
OP		44	MC	0.48	0.46	1.33%				
OP		45	MC	0.78	0.46	0.70%				
OP		46A	CR	0.28	0.52	0.00%				+
OP		46B	CR	0.45	0.63	0.00%				
OP		47	MC	0.54	0.50	0.70%				
OP		48	MC	0.72	0.30	1.02%				

Table 7-31 Cont'd
Item Analysis Grade 6 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		49	MC	0.57	0.44	0.92%				
OP		50*	MC							
OP		51	MC	0.73	0.50	0.68%				
OP		52	MC	0.46	0.39	1.24%				
OP		53	MC	0.81	0.50	1.29%				
OP		54	MC	0.58	0.36	1.70%				
FT	A	55	MC	0.64	0.44	0.39%				
FT	A	56A	CR	0.19	0.29	0.37%				+
FT	A	56B	CR	0.33	0.50	0.74%				
FT	A	57	MC	0.64	0.22	0.76%		+		
FT	A	58	MC	0.43	0.39	0.53%		+		
FT	A	59	MC	0.45	0.29	0.72%		+		
FT	A	60	MC	0.29	0.25	1.16%		+		+
FT	A	61	MC	0.27	0.21	11.20%		+	+	+
FT	A	62	MC	0.36	0.25	1.00%		+		
FT	A	63	MC	0.90	0.36	1.30%				
FT	B	55	MC	0.37	0.60	0.37%				
FT	B	56A	CR	0.46	0.45	0.63%				
FT	B	56B	CR	0.40	0.56	0.84%				
FT	B	57	MC	0.84	0.43	0.46%				
FT	B	58	MC	0.78	0.28	0.49%				
FT	B	59	MC	0.25	0.25	1.76%				+
FT	B	60	MC	0.60	0.28	2.25%				
FT	B	61	MC	0.55	0.53	0.86%				
FT	B	62	MC	0.95	0.22	0.41%				

*Mathematics Grade 6 item 50 was suppressed.

Table 7-31 Cont'd
Item Analysis Grade 6 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	B	63	MC	0.70	0.39	0.56%				
FT	C	55	MC	0.91	0.40	0.41%				
FT	C	56	MC	0.69	0.41	0.48%				
FT	C	57	MC	0.86	0.47	0.43%				
FT	C	58	MC	0.45	0.39	0.65%				
FT	C	59	MC	0.57	0.39	1.39%				
FT	C	60	MC	0.21	0.29	1.01%				+
FT	C	61	MC	0.90	0.38	1.65%				
FT	C	62	MC	0.45	0.19	0.58%				
FT	C	63A	CR	0.40	0.67	0.63%				
FT	C	63B	CR	0.42	0.72	0.90%				
FT	D	55	MC	0.31	0.56	0.40%				
FT	D	56	MC	0.61	0.50	0.47%				
FT	D	57	MC	0.62	0.62	1.15%				
FT	D	58	MC	0.48	0.31	1.32%				
FT	D	59	MC	0.63	0.53	0.45%				
FT	D	60	MC	0.90	0.40	0.47%				
FT	D	61	MC	0.88	0.29	0.67%				
FT	D	62	MC	0.40	0.27	0.84%		+		
FT	D	63A	CR	0.61	0.50	0.61%				
FT	D	63B	CR	0.43	0.60	0.94%				

Table 7-32
Item Analysis Grade 7 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.29	0.44	0.22%		+		+
OP		2	MC	0.94	0.35	0.01%				
OP		3	MC	0.93	0.34	0.06%				
OP		4	MC	0.62	0.47	0.37%				
OP		5	MC	0.80	0.23	0.07%				
OP		6	MC	0.67	0.34	0.26%				
OP		7A	CR	0.91	0.30	0.00%				
OP		7B	CR	0.70	0.57	0.00%				
OP		8	MC	0.89	0.39	0.22%				
OP		9	MC	0.94	0.17	0.15%				
OP		10	MC	0.55	0.49	0.44%				
OP		11	MC	0.77	0.41	0.34%				
OP		12	MC	0.85	0.40	0.44%				
OP		13	MC	0.56	0.58	0.44%				
OP		14	MC	0.89	0.33	0.49%				
OP		15	MC	0.78	0.43	0.51%				
OP		16	MC	0.81	0.43	0.22%				
OP		17	MC	0.46	0.42	0.16%				
OP		18	MC	0.79	0.36	0.35%				
OP		19A	CR	0.51	0.43	0.00%				
OP		19B	CR	0.27	0.56	0.00%				+
OP		20	MC	0.72	0.34	0.35%				
OP		21	MC	0.55	0.29	0.46%				
OP		22	MC	0.71	0.34	0.49%				
OP		23	MC	0.52	0.50	0.82%				
OP		24	MC	0.61	0.51	0.77%				

Table 7-32 Cont'd
Item Analysis Grade 7 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		25	MC	0.76	0.57	0.46%				
OP		26	MC	0.47	0.50	0.37%				
OP		27	MC	0.83	0.18	0.24%				
OP		28	MC	0.80	0.52	0.57%				
OP		29	MC	0.81	0.37	0.46%				
OP		30	MC	0.48	0.45	0.74%				
OP		31A	CR	0.46	0.61	0.00%				
OP		31B	CR	0.52	0.62	0.00%				
OP		32	MC	0.76	0.45	0.34%				
OP		33	MC	0.58	0.34	0.40%				
OP		34	MC	0.41	0.39	0.41%				
OP		35	MC	0.45	0.42	0.63%				
OP		36	MC	0.64	0.39	0.56%				
OP		37	MC	0.73	0.45	0.74%				
OP		38	MC	0.31	0.30	0.93%				
OP		39	MC	0.37	0.36	0.59%				
OP		40	MC	0.88	0.36	0.47%				
OP		41	MC	0.72	0.45	0.56%				
OP		42	MC	0.72	0.45	0.28%				
OP		43	MC	0.97	0.27	0.25%				
OP		44	MC	0.47	0.36	0.29%		+		
OP		45	MC	0.46	0.31	0.69%				
OP		46	MC	0.44	0.31	0.37%				
OP		47A	CR	0.43	0.54	0.00%				
OP		47B	CR	0.34	0.47	0.00%				
OP		48	MC	0.87	0.34	0.38%				

Table 7-32 Cont'd
Item Analysis Grade 7 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		49	MC	0.50	0.55	0.72%				
OP		50	MC	0.42	0.52	0.69%				
OP		51	MC	0.40	0.34	1.24%				
OP		52	MC	0.68	0.47	0.87%				
OP		53	MC	0.41	0.51	0.78%		+		
OP		54	MC	0.85	0.36	0.78%				
FT	A	55	MC	0.60	0.26	0.46%				
FT	A	56	MC	0.63	0.31	0.47%				
FT	A	57	MC	0.44	0.34	0.96%				
FT	A	58	MC	0.30	0.16	1.38%		+		+
FT	A	59	MC	0.74	0.29	0.42%				
FT	A	60	MC	0.31	0.49	0.68%				
FT	A	61	MC	0.40	0.52	0.95%				
FT	A	62	MC	0.63	0.32	0.48%				
FT	A	63A	CR	0.54	0.09	1.12%	+			
FT	A	63B	CR	0.19	0.41	0.88%				+
FT	B	55	MC	0.42	0.44	0.45%				
FT	B	56A	CR							
FT	B	56B	CR							
FT	B	57	MC	0.64	0.40	0.37%				
FT	B	58	MC	0.41	0.27	0.51%		+		
FT	B	59	MC	0.87	0.36	0.46%				
FT	B	60	MC	0.24	0.41	0.60%				+
FT	B	61	MC	0.84	0.40	0.53%				
FT	B	62	MC	0.44	0.39	1.81%		+		
FT	B	63	MC	0.57	0.34	1.06%				

Table 7-32 Cont'd
Item Analysis Grade 7 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	C	55	MC	0.66	0.48	0.32%				
FT	C	56A	CR	0.73	0.51	0.83%				
FT	C	56B	CR	0.67	0.53	1.02%				
FT	C	57	MC	0.91	0.23	0.33%				
FT	C	58	MC	0.58	0.51	0.51%				
FT	C	59	MC	0.54	0.48	0.51%				
FT	C	60	MC	0.79	0.31	0.50%				
FT	C	61	MC	0.38	0.35	0.47%				
FT	C	62	MC	0.95	0.31	0.47%				
FT	C	63	MC	0.46	0.42	0.54%				
FT	D	55	MC	0.48	0.28	0.29%				
FT	D	56A	CR	0.17	0.43	1.12%				+
FT	D	56B	CR	0.08	0.41	1.68%				+
FT	D	57	MC	0.69	0.30	0.50%				
FT	D	58	MC	0.54	0.56	0.60%				
FT	D	59	MC	0.69	0.43	1.83%				
FT	D	60	MC	0.40	0.43	0.62%				
FT	D	61	MC	0.72	0.46	0.51%				
FT	D	62	MC	0.76	0.38	0.88%				
FT	D	63	MC	0.82	0.44	1.00%				

Table 7-33
Item Analysis Grade 8 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.80	0.36	0.13%				
OP		2	MC	0.55	0.53	0.15%				
OP		3	MC	0.90	0.29	0.15%				
OP		4	MC	0.76	0.44	0.71%				
OP		5	MC	0.90	0.33	0.36%				
OP		6	MC	0.55	0.40	0.54%				
OP		7	MC	0.59	0.57	1.52%				
OP		8	MC	0.50	0.38	0.73%				
OP		9A	CR	0.87	0.37	0.00%				
OP		9B	CR	0.70	0.51	0.00%				
OP		10	MC	0.85	0.42	0.20%				
OP		11	MC	0.42	0.42	0.44%				
OP		12	MC	0.87	0.25	0.52%				
OP		13	MC	0.91	0.30	0.30%				
OP		14	MC	0.91	0.38	0.32%				
OP		15	MC	0.50	0.50	0.45%				
OP		16	MC	0.50	0.49	0.28%				
OP		17	MC	0.45	0.44	0.20%				
OP		18	MC	0.71	0.31	0.19%				
OP		19	MC	0.80	0.22	0.20%				
OP		20A	CR	0.77	0.31	0.00%				
OP		20B	CR	0.66	0.54	0.00%				
OP		21	MC	0.64	0.48	0.25%				
OP		22	MC	0.53	0.63	0.22%				
OP		23	MC	0.41	0.45	0.45%				
OP		24	MC	0.74	0.45	0.41%				

Table 7-33 Cont'd
Item Analysis Grade 8 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		25	MC	0.94	0.26	0.30%				
OP		26	MC	0.54	0.48	0.41%				
OP		27	MC	0.64	0.42	0.49%				
OP		28	MC	0.60	0.47	0.54%				
OP		29	MC	0.75	0.40	0.49%				
OP		30	MC	0.44	0.42	0.78%				
OP		31	MC	0.55	0.40	0.46%				
OP		32A	CR	0.32	0.52	0.00%				
OP		32B	CR	0.39	0.66	0.00%				
OP		33	MC	0.63	0.44	0.61%				
OP		34	MC	0.53	0.40	0.65%				
OP		35	MC	0.86	0.39	0.86%				
OP		36	MC	0.57	0.49	1.41%				
OP		37	MC	0.43	0.51	1.16%				
OP		38	MC	0.30	0.34	1.38%				
OP		39	MC	0.49	0.48	1.60%				
OP		40	MC	0.57	0.44	0.84%				
OP		41	MC	0.86	0.40	0.84%				
OP		42	MC	0.36	0.23	0.48%				
OP		43	MC	0.38	0.32	0.59%				
OP		44A	CR	0.49	0.68	0.00%				
OP		44B	CR	0.53	0.70	0.00%				
OP		45	MC	0.80	0.31	0.62%				
OP		46	MC	0.46	0.50	0.71%				
OP		47	MC	0.41	0.34	0.46%				
OP		48	MC	0.30	0.29	0.96%		+		

Table 7-33 Cont'd
Item Analysis Grade 8 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		49	MC	0.36	0.25	0.94%				
OP		50	MC	0.71	0.39	0.91%				
OP		51	MC	0.83	0.40	1.02%				
OP		52	MC	0.30	0.53	1.19%		+		+
OP		53	MC	0.36	0.23	1.31%				
OP		54	MC	0.43	0.50	0.90%				
FT	A	55	MC	0.79	0.50	0.40%				
FT	A	56	MC	0.32	0.44	0.66%				
FT	A	57A	CR	0.75	0.12	0.26%	+			
FT	A	57B	CR	0.46	0.41	0.26%				
FT	A	58	MC	0.90	0.31	0.58%				
FT	A	59	MC	0.40	0.34	0.81%				
FT	A	60	MC	0.31	0.38	1.16%				
FT	A	61	MC	0.88	0.44	1.13%				
FT	A	62	MC	0.74	0.53	0.51%				
FT	A	63	MC	0.89	0.34	0.54%				
FT	A	64	MC	0.77	0.52	0.79%				
FT	A	65	MC	0.78	0.36	1.04%				
FT	B	55	MC	0.56	0.42	0.39%				
FT	B	56	MC	0.53	0.30	0.50%				
FT	B	57A	CR	0.75	0.13	0.15%	+			
FT	B	57B	CR	0.48	0.44	0.15%				
FT	B	58	MC	0.92	0.20	0.65%				
FT	B	59	MC	0.34	0.25	0.51%		+		
FT	B	60	MC	0.53	0.31	1.60%				

* This item did not converge, see Part 8 for more information.

Table 7-33 Cont'd
Item Analysis Grade 8 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	B	61*	MC							
FT	B	62	MC	0.60	0.33	0.87%				
FT	B	63	MC	0.19	0.34	0.75%				+
FT	B	64	MC	0.65	0.42	2.34%				
FT	B	65	MC	0.33	0.22	2.53%		+		
FT	C	55	MC	0.64	0.36	0.35%				
FT	C	56	MC	0.82	0.39	0.59%				
FT	C	57A	CR	0.22	0.48	1.89%				+
FT	C	57B	CR	0.35	0.65	2.23%				
FT	C	58	MC	0.29	0.39	0.47%				+
FT	C	59	MC	0.58	0.48	0.49%				
FT	C	60	MC	0.74	0.44	1.68%				
FT	C	61	MC	0.69	0.42	0.52%				
FT	C	62	MC	0.29	0.28	1.17%				+
FT	C	63	MC	0.56	0.37	0.61%				
FT	C	64	MC	0.48	0.40	1.01%				
FT	C	65	MC	0.65	0.34	1.97%				
FT	D	55	MC	0.92	0.13	0.47%	+			
FT	D	56	MC	0.41	0.53	0.47%				
FT	D	57	MC	0.67	0.44	0.76%				
FT	D	58	MC	0.71	0.50	0.94%				
FT	D	59	MC	0.32	0.38	0.71%				
FT	D	60	MC	0.53	0.37	0.73%				
FT	D	61	MC	0.56	0.34	0.88%		+		
FT	D	62	MC	0.41	0.34	1.07%				

* This item did not converge. See Part 8 for more information.

Table 7-34
Item Analysis Grade 10 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	D	63	MC	0.73	0.48	0.57%				
FT	D	64	MC	0.65	0.38	0.47%				
FT	D	65A	CR	0.32	0.68	1.18%				
FT	D	65B	CR	0.40	0.72	1.60%				
OP		1	MC	0.89	0.28	0.21%				
OP		2	MC	0.66	0.45	0.23%				
OP		3	MC	0.64	0.34	0.46%				
OP		4	MC	0.44	0.49	0.52%				
OP		5	MC	0.48	0.44	0.16%				
OP		6	MC	0.59	0.50	0.74%				
OP		7	MC	0.68	0.47	0.47%				
OP		8	MC	0.48	0.23	0.55%				
OP		9	MC	0.40	0.36	0.73%		+		
OP		10	MC	0.46	0.37	0.90%				
OP		11	MC	0.68	0.43	1.03%				
OP		12	MC	0.80	0.40	0.30%				
OP		13	MC	0.56	0.54	0.29%				
OP		14	MC	0.46	0.43	0.50%		+		
OP		15	MC	0.76	0.40	0.30%				
OP		16	CR	0.22	0.45	0.00%				+
OP		17	MC	0.74	0.55	0.33%				
OP		18	MC	0.51	0.45	0.89%				
OP		19	MC	0.77	0.47	0.59%				
OP		20	MC	0.70	0.49	0.63%				
OP		21	MC	0.79	0.46	0.73%				
OP		22	MC	0.89	0.42	0.55%				

Table 7-34 Cont'd
Item Analysis Grade 10 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		23	MC	0.50	0.23	0.87%				
OP		24	MC	0.68	0.47	0.85%				
OP		25	MC	0.40	0.48	1.11%				
OP		26	MC	0.62	0.59	0.81%				
OP		27	MC	0.53	0.44	0.74%				
OP		28	MC	0.88	0.32	0.87%				
OP		29	MC	0.58	0.43	0.80%				
OP		30	CR	0.41	0.65	0.00%				
OP		31	MC	0.42	0.48	0.97%				
OP		32	MC	0.55	0.60	0.64%				
OP		33	MC	0.73	0.49	0.68%				
OP		34	MC	0.93	0.28	0.61%				
OP		35	MC	0.69	0.37	0.61%				
OP		36	CR	0.40	0.66	0.00%				
OP		37	MC	0.73	0.54	0.91%				
OP		38	MC	0.71	0.57	0.98%				
OP		39	MC	0.44	0.63	1.37%				
OP		40	MC	0.63	0.52	0.90%				
OP		41	MC	0.51	0.41	0.89%				
OP		42	MC	0.47	0.50	0.85%				
OP		43	MC	0.61	0.49	0.91%				
OP		44	MC	0.82	0.26	0.80%				
OP		45	MC	0.46	0.35	1.07%				
OP		46	MC	0.79	0.48	0.78%				
OP		47	CR	0.42	0.58	0.00%				

Table 7-34 Cont'd
Item Analysis Grade 10 Mathematics

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		48	MC	0.65	0.44	1.34%				
OP		49	MC	0.64	0.53	0.95%				
OP		50	MC	0.38	0.43	0.95%				
OP		51	MC	0.46	0.57	1.07%				
OP		52	MC	0.52	0.51	0.99%		+		
OP		53	MC	0.74	0.50	0.95%				
OP		54	MC	0.68	0.58	1.20%				

Table 7-35
Item Analysis Grade 4 Language Arts*

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.40	0.41	0.23%				
OP		2	MC	0.93	0.36	0.27%				
OP		3	MC	0.89	0.37	0.69%				
OP		4	MC	0.88	0.29	0.55%				
OP		5	MC	0.50	0.25	0.72%				
OP		6	MC	0.83	0.42	0.14%				
OP		7	MC	0.59	0.46	0.11%				
OP		8	MC	0.84	0.48	0.06%				
OP		9	MC	0.71	0.41	0.62%				
OP		10	MC	0.94	0.38	0.39%				
OP		11	MC	0.91	0.42	0.91%				
OP		12	MC	0.94	0.39	0.92%				
OP		13	MC	0.84	0.47	0.80%				
OP		14	MC	0.51	0.44	0.89%				
OP		15	MC	0.56	0.47	2.23%				
OP		16	MC	0.71	0.55	0.45%				
OP		17	MC	0.44	0.35	2.06%				
OP		18	MC	0.66	0.41	0.36%				
OP		19	MC	0.50	0.36	0.73%				
OP		20	MC	0.67	0.48	0.83%				
OP		21	MC	0.74	0.50	0.89%				
OP		22	MC	0.66	0.44	1.84%				
OP		23	MC	0.54	0.40	1.41%				
OP		24	MC	0.61	0.51	0.89%				
OP		25	MC	0.65	0.54	1.34%				

* Operational Writing prompt items are included here. Writing raw score is based on these items, but they do not contribute to the operational (scale) score.

Table 7-35 Cont'd
Item Analysis Grade 4 Language Arts*

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		26	MC	0.53	0.46	2.86%				
OP		27	MC	0.72	0.57	2.08%				
OP		28	MC	0.30	0.32	2.28%				
OP		29	MC	0.74	0.48	2.92%				
OP		30	MC	0.42	0.21	4.47%		+		
FT	A	1A	CR	0.51	0.54	0.33%				
FT	A	1B	CR	0.65	0.36	0.33%				

* Operational Writing prompt items are included here. Writing raw score is based on these items, but they do not contribute to the operational (scale) score.

Table 7-36
Item Analysis Grade 8 Language Arts*

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.89	0.37	0.03%				
OP		2	MC	0.83	0.43	0.40%				
OP		3	MC	0.77	0.47	0.85%				
OP		4	MC	0.86	0.42	0.06%				
OP		5	MC	0.75	0.45	0.10%				
OP		6	MC	0.55	0.38	0.21%				
OP		7	MC	0.89	0.44	0.31%				
OP		8	MC	0.48	0.31	0.12%				
OP		9	MC	0.66	0.53	0.12%				
OP		10	MC	0.89	0.38	3.66%				
OP		11	MC	0.81	0.50	0.12%				
OP		12	MC	0.60	0.52	0.29%				
OP		13	MC	0.88	0.41	0.22%				
OP		14	MC	0.80	0.57	0.31%				
OP		15	MC	0.49	0.41	7.09%			+	
OP		16	MC	0.87	0.41	0.13%				
OP		17	MC	0.73	0.52	0.42%				
OP		18	MC	0.69	0.53	0.34%				
OP		19	MC	0.88	0.49	0.98%				
OP		20	MC	0.83	0.49	0.42%				
OP		21	MC	0.55	0.59	3.59%				
OP		22	MC	0.92	0.48	0.45%				
OP		23	MC	0.63	0.50	1.07%				
OP		24	MC	0.69	0.51	0.91%				
OP		25	MC	0.66	0.35	0.63%				

* Operational Writing prompt items are included here. Writing raw score is based on these items, but they do not contribute to the operational (scale) score.

Table 7-36 Cont'd
Item Analysis Grade 8 Language Arts*

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		26	MC	0.76	0.48	0.89%				
OP		27	MC	0.60	0.52	6.80%			+	
OP		28	MC	0.41	0.09	1.17%	+	+		
OP		29	MC	0.69	0.49	1.32%				
OP		30	MC	0.63	0.51	1.66%				
FT	A	1A**	CR							
FT	A	1B**	CR							

* Operational Writing prompt items are included here. Writing raw score is based on these items, but they do not contribute to the operational (scale) score.
** This item did not converge. See Part 8, for more information.

Table 7-37
Item Analysis Grade 10 Language Arts

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.75	0.46	0.15%				
OP		2	MC	0.84	0.48	0.16%				
OP		3	MC	0.68	0.34	0.08%				
OP		4	MC	0.68	0.34	0.75%				
OP		5	MC	0.77	0.37	0.19%				
OP		6	MC	0.86	0.44	0.16%				
OP		7	MC	0.82	0.40	0.15%				
OP		8	MC	0.60	0.38	0.13%				
OP		9	MC	0.69	0.48	0.19%				
OP		10	MC	0.76	0.49	1.03%				
OP		11	MC	0.61	0.50	0.21%				
OP		12	MC	0.64	0.32	0.24%				
OP		13	MC	0.51	0.26	0.17%				
OP		14	MC	0.57	0.44	0.43%				
OP		15	MC	0.52	0.41	0.31%				
OP		16	MC	0.66	0.41	0.42%				
OP		17	MC	0.86	0.29	0.52%				
OP		18	MC	0.52	0.35	0.40%				
OP		19	MC	0.69	0.52	0.39%				
OP		20	MC	0.58	0.55	0.91%				
OP		21	MC	0.62	0.38	3.67%				
OP		22	MC	0.44	0.39	0.51%				
OP		23	MC	0.30	0.27	1.21%				
OP		24	MC	0.61	0.41	0.68%				
OP		25	MC	0.25	0.19	0.78%		+		+

Table 7-37 Cont'd
Item Analysis Grade 10 Language Arts

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		26	MC	0.80	0.53	0.75%				
OP		27	MC	0.59	0.51	1.01%				
OP		28	MC	0.60	0.49	1.14%				
OP		29	MC	0.71	0.52	1.30%				
OP		30	MC	0.53	0.38	1.38%				
OP		31	CR	0.53	0.66	0.00%				
OP		32	CR	0.67	0.44	0.00%				

Table 7-38
Item Analysis Grade 4 Social Studies

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.88	0.46	0.06%				
OP		2	MC	0.92	0.31	1.00%				
OP		3	MC	0.96	0.38	0.06%				
OP		4	MC	0.90	0.49	0.26%				
OP		5	MC	0.40	0.34	0.56%				
OP		6	MC	0.79	0.49	0.28%				
OP		7	MC	0.49	0.29	1.01%		+		
OP		8	MC	0.74	0.52	1.32%				
OP		9	MC	0.81	0.38	1.95%				
OP		10	MC	0.76	0.36	2.46%				
OP		11	MC	0.86	0.38	2.54%				
OP		12	MC	0.77	0.36	2.18%				
OP		13	MC	0.94	0.34	1.06%				
OP		14	MC	0.96	0.31	1.51%				
OP		15	MC	0.78	0.43	0.20%				
OP		16	MC	0.54	0.35	1.67%				
OP		17	MC	0.71	0.40	4.43%				
OP		18	MC	0.88	0.46	0.78%				
OP		19	MC	0.90	0.42	0.06%				
OP		20	MC	0.96	0.33	0.39%				
OP		21	MC	0.69	0.53	0.25%				
OP		22	MC	0.72	0.51	1.82%				
OP		23	MC	0.87	0.40	0.31%				
OP		24	MC	0.64	0.44	0.64%				
OP		25	MC	0.80	0.45	5.94%			+	
OP		26	MC	0.79	0.49	0.34%				

Table 7-38 Cont'd
Item Analysis Grade 4 Social Studies

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.92	0.47	0.47%				
OP		28	MC	0.94	0.44	0.20%				
OP		29	MC	0.47	0.40	1.06%				
OP		30	MC	0.79	0.45	0.41%				
OP		31	MC	0.62	0.55	0.50%				
OP		32	MC	0.59	0.41	0.65%				
OP		33	MC	0.74	0.48	1.57%				
OP		34	MC	0.95	0.38	0.42%				
OP		35	MC	0.70	0.62	0.45%				
OP		36	MC	0.75	0.57	0.62%				
OP		37	MC	0.92	0.47	0.48%				
OP		38	MC	0.76	0.50	0.84%				

Table 7-39
Item Analysis Grade 8 Social Studies

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.90	0.30	0.01%				
OP		2	MC	0.91	0.42	0.20%				
OP		3	MC	0.78	0.27	1.04%				
OP		4	MC	0.64	0.43	1.20%				
OP		5	MC	0.83	0.35	0.09%				
OP		6	MC	0.97	0.35	0.04%				
OP		7	MC	0.85	0.34	0.07%				
OP		8	MC	0.73	0.42	0.38%				
OP		9	MC	0.96	0.34	0.10%				
OP		10	MC	0.92	0.41	0.70%				
OP		11	MC	0.89	0.47	0.12%				
OP		12	MC	0.87	0.43	0.22%				
OP		13	MC	0.82	0.47	0.52%				
OP		14	MC	0.78	0.44	0.17%				
OP		15	MC	0.27	0.24	1.15%				+
OP		16	MC	0.78	0.59	0.22%				
OP		17	MC	0.67	0.54	0.32%				
OP		18	MC	0.75	0.37	0.73%				
OP		19	MC	0.69	0.57	1.79%				
OP		20	MC	0.48	0.39	0.61%				
OP		21	MC	0.82	0.48	0.71%				
OP		22	MC	0.60	0.47	0.85%				
OP		23	MC	0.76	0.49	0.19%				
OP		24	MC	0.66	0.47	0.32%				
OP		25	MC	0.63	0.56	0.63%				
OP		26	MC	0.65	0.49	0.20%				

Table 7-39 Cont'd
Item Analysis Grade 8 Social Studies

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.68	0.56	0.32%				
OP		28	MC	0.75	0.51	0.47%				
OP		29	MC	0.69	0.45	0.80%				
OP		30	MC	0.63	0.54	0.80%				
OP		31	MC	0.61	0.42	0.74%				
OP		32	MC	0.84	0.52	0.34%				
OP		33	MC	0.79	0.50	0.31%				
OP		34	MC	0.51	0.35	0.82%				
OP		35	MC	0.84	0.52	0.51%				
OP		36	MC	0.68	0.47	0.58%				
OP		37	MC	0.79	0.49	0.50%				
OP		38	MC	0.41	0.43	0.69%				
OP		39	MC	0.61	0.43	0.77%				
OP		40	MC	0.48	0.45	0.93%				

Table 7-40
Item Analysis Grade 10 Social Studies

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.91	0.26	0.13%				
OP		2	MC	0.65	0.50	0.69%				
OP		3	MC	0.63	0.38	0.11%				
OP		4	MC	0.59	0.43	0.38%				
OP		5	MC	0.68	0.52	0.38%				
OP		6	MC	0.38	0.44	0.17%				
OP		7	MC	0.49	0.39	0.60%				
OP		8	MC	0.75	0.30	0.58%				
OP		9	MC	0.93	0.42	0.66%				
OP		10	MC	0.58	0.54	0.86%				
OP		11	MC	0.61	0.47	0.81%				
OP		12	MC	0.87	0.43	0.24%				
OP		13	MC	0.69	0.43	0.28%				
OP		14	MC	0.76	0.40	0.30%				
OP		15	MC	0.51	0.41	0.69%				
OP		16	MC	0.41	0.36	0.83%				
OP		17	MC	0.53	0.44	0.73%				
OP		18	MC	0.67	0.48	0.79%				
OP		19	MC	0.37	0.39	0.47%				
OP		20	MC	0.63	0.55	1.16%				
OP		21	MC	0.89	0.39	0.34%				
OP		22	MC	0.84	0.42	0.59%				
OP		23	MC	0.80	0.48	0.47%				
OP		24	MC	0.75	0.49	1.13%				
OP		25	MC	0.74	0.52	0.82%				
OP		26	MC	0.70	0.47	0.46%				

Table 7-40 Cont'd
Item Analysis Grade 10 Social Studies

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.74	0.48	1.53%				
OP		28	MC	0.62	0.47	0.55%				
OP		29	MC	0.71	0.52	0.35%				
OP		30	MC	0.77	0.49	0.38%				
OP		31	MC	0.80	0.45	0.46%				
OP		32	MC	0.62	0.47	2.12%				
OP		33	MC	0.51	0.31	0.52%				
OP		34	MC	0.42	0.28	0.50%				
OP		35	MC	0.66	0.47	1.13%				
OP		36	MC	0.62	0.53	7.95%			+	
OP		37	MC	0.78	0.47	0.77%				
OP		38	MC	0.76	0.54	0.47%				
OP		39	MC	0.81	0.45	0.47%				
OP		40	MC	0.73	0.58	0.56%				
OP		41	MC	0.63	0.44	0.66%				
OP		42	MC	0.64	0.41	0.44%				
OP		43	MC	0.38	0.48	0.79%		+		
OP		44	MC	0.67	0.56	0.87%				
OP		45	MC	0.75	0.53	0.47%				
OP		46	MC	0.82	0.53	0.63%				
OP		47	MC	0.72	0.49	0.69%				
OP		48	MC	0.28	0.11	1.10%	+	+		+
OP		49	MC	0.61	0.53	0.56%				
OP		50	MC	0.50	0.50	0.58%				

Table 7-41
Item Analysis Grade 4 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.96	0.26	0.16%				
OP		2	MC	0.83	0.31	0.12%				
OP		3	MC	0.58	0.38	0.26%				
OP		4	MC	0.80	0.35	0.76%				
OP		5	MC	0.91	0.25	1.74%				
OP		6	MC	0.96	0.30	0.05%				
OP		7	MC	0.92	0.39	0.30%				
OP		8	MC	0.84	0.30	0.20%				
OP		9	MC	0.78	0.40	0.09%				
OP		10	MC	0.75	0.43	0.82%				
OP		11	MC	0.69	0.47	0.45%				
OP		12	MC	0.82	0.34	0.62%				
OP		13	MC	0.79	0.52	0.53%				
OP		14	MC	0.44	0.26	0.92%		+		
OP		15	MC	0.91	0.38	0.95%				
OP		16	MC	0.79	0.51	0.45%				
OP		17	MC	0.50	0.43	0.82%				
OP		18	MC	0.63	0.44	0.45%				
OP		19	MC	0.48	0.49	0.98%				
OP		20	MC	0.62	0.55	1.11%				
OP		21	MC	0.73	0.53	1.57%				
OP		22	MC	0.82	0.44	0.39%				
OP		23	MC	0.86	0.41	0.50%				
OP		24	MC	0.76	0.48	0.79%				
OP		25	MC	0.82	0.40	0.70%				
OP		26	MC	0.76	0.43	0.59%				

Table 7-41 Cont'd
Item Analysis Grade 4 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.89	0.44	0.56%				
OP		28	MC	0.52	0.44	0.81%				
OP		29	MC	0.67	0.51	1.43%				
OP		30	MC	0.67	0.33	1.34%				
OP		31	MC	0.79	0.41	1.68%				
OP		32	MC	0.53	0.42	1.74%				
OP		33	MC	0.75	0.53	1.95%				
OP		34	MC	0.54	0.25	0.90%				
OP		35	MC	0.49	0.46	2.94%				
OP		36	MC	0.69	0.45	2.18%				
OP		37	MC	0.70	0.33	0.89%				
OP		38	MC	0.51	0.43	0.75%				
OP		39	MC	0.52	0.37	0.84%				
OP		40	MC	0.50	0.51	1.31%				
FT	A	41	MC	0.77	0.48	0.73%				
FT	A	42	MC	0.47	0.36	0.84%				
FT	A	43	MC	0.33	0.15	0.49%	+	+		
FT	A	44	MC	0.81	0.49	0.41%				
FT	A	45	MC	0.65	0.44	0.53%				
FT	A	46	MC	0.75	0.50	0.63%				
FT	A	47	MC	0.80	0.46	0.64%				
FT	A	48	MC	0.65	0.42	0.90%				
FT	A	49	MC	0.72	0.49	1.29%				
FT	A	50	MC	0.67	0.39	1.84%				
FT	B	41	MC	0.52	0.38	0.34%				
FT	B	42	MC	0.59	0.38	0.40%				

Table 7-41 Cont'd
Item Analysis Grade 4 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	B	43	MC	0.81	0.46	0.50%				
FT	B	44	MC	0.65	0.51	0.59%				
FT	B	45	MC	0.64	0.42	0.58%				
FT	B	46	MC	0.91	0.42	0.53%				
FT	B	47	MC	0.84	0.45	1.95%				
FT	B	48	MC	0.58	0.42	3.12%				
FT	B	49	MC	0.77	0.52	1.34%				
FT	B	50	MC	0.72	0.52	1.70%				
FT	C	41	MC	0.65	0.44	0.35%				
FT	C	42	MC	0.66	0.43	0.43%				
FT	C	43	MC	0.47	0.36	0.56%				
FT	C	44	MC	0.42	0.29	0.40%				
FT	C	45	MC	0.76	0.54	0.81%				
FT	C	46	MC	0.54	0.42	1.07%				
FT	C	47	MC	0.57	0.45	0.79%				
FT	C	48	MC	0.97	0.25	1.09%				
FT	C	49	MC	0.44	0.37	1.52%				
FT	C	50	MC	0.52	0.45	2.80%				
FT	D	41	MC	0.74	0.49	0.79%				
FT	D	42	MC	0.82	0.44	0.59%				
FT	D	43	MC	0.59	0.51	0.38%				
FT	D	44	MC	0.93	0.31	0.69%				
FT	D	45	MC	0.83	0.55	0.27%				
FT	D	46	MC	0.50	0.47	0.69%				
FT	D	47	MC	0.66	0.51	1.16%				
FT	D	48	MC	0.71	0.46	1.51%				

Table 7-41 Cont'd
Item Analysis Grade 4 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	D	49	MC	0.71	0.28	2.42%				
FT	D	50	MC	0.65	0.46	1.88%				

Table 7-42
Item Analysis Grade 8 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.86	0.45	0.03%				
OP		2	MC	0.89	0.44	0.07%				
OP		3	MC	0.84	0.35	0.09%				
OP		4	MC	0.82	0.41	0.16%				
OP		5	MC	0.91	0.38	0.04%				
OP		6	MC	0.78	0.30	0.25%				
OP		7	MC	0.92	0.37	0.34%				
OP		8	MC	0.83	0.48	0.23%				
OP		9	MC	0.59	0.45	0.45%				
OP		10	MC	0.82	0.43	4.71%				
OP		11	MC	0.71	0.51	8.63%			+	
OP		12	MC	0.82	0.25	0.36%				
OP		13	MC	0.75	0.46	0.18%				
OP		14	MC	0.87	0.47	0.26%				
OP		15	MC	0.70	0.45	0.29%				
OP		16	MC	0.76	0.49	0.48%				
OP		17	MC	0.70	0.39	0.39%				
OP		18	MC	0.67	0.52	0.71%				
OP		19	MC	0.58	0.41	1.56%				
OP		20	MC	0.49	0.32	2.63%		+		
OP		21	MC	0.77	0.32	0.45%				
OP		22	MC	0.56	0.37	0.55%				
OP		23	MC	0.33	0.35	0.58%				
OP		24	MC	0.64	0.52	2.04%				
OP		25	MC	0.79	0.54	0.31%				
OP		26	MC	0.78	0.50	0.35%				

Table 7-42 Cont'd
Item Analysis Grade 8 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		27	MC	0.45	0.46	0.51%				
OP		28	MC	0.47	0.48	0.34%				
OP		29	MC	0.84	0.47	0.26%				
OP		30	MC	0.62	0.44	0.39%				
OP		31	MC	0.41	0.37	0.71%				
OP		32	MC	0.63	0.48	0.34%				
OP		33	MC	0.53	0.38	0.53%				
OP		34	MC	0.61	0.56	0.48%				
OP		35	MC	0.53	0.33	0.66%				
OP		36	MC	0.43	0.33	1.18%				
OP		37	MC	0.64	0.52	0.58%				
OP		38	MC	0.65	0.42	1.01%				
OP		39	MC	0.72	0.50	0.48%				
OP		40	MC	0.57	0.28	0.83%				
FT	A	41	MC	0.59	0.37	0.33%				
FT	A	42	MC	0.51	0.39	0.38%				
FT	A	43	MC	0.83	0.42	0.31%				
FT	A	44	MC	0.38	0.31	0.47%		+		
FT	A	45	MC	0.47	0.48	0.60%				
FT	A	46	MC	0.19	0.03	0.37%	+	+		+
FT	A	47	MC	0.41	0.37	0.39%				
FT	A	48	MC	0.34	0.43	0.55%				
FT	A	49	MC	0.36	0.40	0.70%				
FT	A	50	MC	0.61	0.49	0.84%				
FT	B	41	MC	0.41	0.42	0.33%				
FT	B	42	MC	0.74	0.50	0.38%				

Table 7-42 Cont'd
Item Analysis Grade 8 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	B	43	MC	0.75	0.29	0.59%				
FT	B	44	MC	0.77	0.46	0.72%				
FT	B	45	MC	0.70	0.49	0.40%				
FT	B	46	MC	0.48	0.17	0.43%		+		
FT	B	47	MC	0.53	0.32	0.54%				
FT	B	48	MC	0.75	0.44	2.34%				
FT	B	49	MC	0.66	0.50	0.79%				
FT	B	50	MC	0.67	0.50	1.06%				
FT	C	41	MC	0.52	0.48	0.27%				
FT	C	42	MC	0.60	0.37	0.30%				
FT	C	43	MC	0.64	0.46	0.30%				
FT	C	44	MC	0.69	0.54	0.52%				
FT	C	45	MC	0.30	0.21	0.62%		+		+
FT	C	46	MC	0.83	0.36	0.36%				
FT	C	47	MC	0.67	0.43	0.42%				
FT	C	48	MC	0.72	0.51	0.50%				
FT	C	49	MC	0.44	0.36	0.79%				
FT	C	50	MC	0.68	0.40	1.02%				
FT	D	41	MC	0.66	0.46	0.26%				
FT	D	42	MC	0.86	0.43	0.27%				
FT	D	43	MC	0.86	0.49	2.46%				
FT	D	44	MC	0.43	0.26	0.33%		+		
FT	D	45	MC	0.68	0.51	0.60%				
FT	D	46	MC	0.72	0.44	0.37%				
FT	D	47	MC	0.69	0.59	0.45%				
FT	D	48	MC	0.65	0.39	0.49%				

Table 7-42 Cont'd
Item Analysis Grade 8 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	D	49	MC	0.52	0.24	0.69%				
FT	D	50	MC	0.86	0.44	0.84%				

Table 7-43
Item Analysis Grade 10 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		1	MC	0.93	0.25	0.62%				
OP		2	MC	0.38	0.25	0.15%				
OP		3	MC	0.82	0.34	0.11%				
OP		4	MC	0.74	0.51	0.20%				
OP		5	MC	0.77	0.47	0.07%				
OP		6	MC	0.93	0.40	0.28%				
OP		7	MC	0.36	0.33	0.26%				
OP		8	MC	0.41	0.31	0.26%				
OP		9	MC	0.44	0.35	0.42%				
OP		10	MC	0.69	0.46	0.59%				
OP		11	MC	0.37	0.31	0.34%				
OP		12	MC	0.85	0.45	0.27%				
OP		13	MC	0.57	0.53	0.21%				
OP		14	MC	0.77	0.37	0.16%				
OP		15	MC	0.67	0.27	2.04%				
OP		16	MC	0.77	0.44	0.17%				
OP		17	MC	0.66	0.39	0.23%				
OP		18	MC	0.70	0.48	0.34%				
OP		19	MC	0.77	0.43	0.27%				
OP		20	MC	0.32	0.22	0.47%		+		
OP		21	MC	0.85	0.29	0.20%				
OP		22	MC	0.46	0.47	0.31%				
OP		23	MC	0.43	0.45	1.17%				
OP		24	MC	0.40	0.32	1.53%				
OP		25	MC	0.89	0.43	0.34%				
OP		36	MC	0.81	0.40	0.54%				

Table 7-43 Cont'd
Item Analysis Grade 10 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
OP		37	MC	0.74	0.57	0.56%				
OP		38	MC	0.26	0.24	0.54%				+
OP		39	MC	0.68	0.51	0.54%				
OP		40	MC	0.56	0.50	2.27%				
OP		41	MC	0.52	0.33	0.60%				
OP		42	MC	0.67	0.47	0.78%				
OP		43	MC	0.45	0.32	0.81%				
OP		44	MC	0.77	0.51	0.78%				
OP		45	MC	0.68	0.53	0.78%				
OP		46	MC	0.72	0.47	0.98%				
OP		47	MC	0.64	0.39	0.63%				
OP		48	MC	0.66	0.51	0.86%				
OP		49	MC	0.75	0.51	0.68%				
OP		50	MC	0.54	0.47	0.74%				
OP		51	MC	0.45	0.46	0.74%				
OP		52	MC	0.46	0.48	0.89%				
OP		53	MC	0.35	0.35	0.99%				
OP		54	MC	0.70	0.44	0.90%				
OP		55	MC	0.53	0.42	1.10%				
OP		56	MC	0.49	0.54	0.82%				
OP		57	MC	0.48	0.36	0.86%				
OP		58	MC	0.44	0.35	0.89%				
OP		59	MC	0.88	0.28	0.99%				
OP		60	MC	0.60	0.45	1.01%				
FT	A	26	MC	0.45	0.44	0.33%				
FT	A	27	MC	0.57	0.52	0.30%				

Table 7-43 Cont'd
Item Analysis Grade 10 Science

Item ID Field			Item Statistic Fields				Flag			
OP/FT	**Form**	**Test Book Item**	**Item Type**	***p*-Value**	**Corr**	**Omit Rate**	**Flag Corr**	**Flag Distracter**	**Flag Omit**	**Flag *p*-Value**
FT	A	28	MC	0.42	0.36	0.40%				
FT	A	29	MC	0.60	0.53	0.32%				
FT	A	30	MC	0.67	0.55	0.40%				
FT	A	31	MC	0.76	0.39	0.51%				
FT	A	32	MC	0.69	0.51	0.46%				
FT	A	33	MC	0.62	0.45	0.34%				
FT	A	34	MC	0.63	0.52	0.39%				
FT	A	35	MC	0.55	0.38	0.37%				

Table 7-44
The Number of Items Flagged*

Content	Grade	OP Items				FT Items			
		Flag Corr	Flag Distracter	Flag Omit	Flag *p*-value	Flag Corr	Flag Distracter	Flag Omit	Flag *p*-value
RD	3		1		1	3	5	0	2
	4		1		1	1	1	0	2
	5		1		1		3	0	2
	6		2			3	5	0	4
	7		2		2	6	4	0	4
	8		2			1		0	1
	10		3						
MA	3				1			0	0
	4				2		2	0	2
	5				3		1	0	0
	6		1		1		7	1	5
	7		3		2	1	3	0	5
	8		2		1	5	4	0	4
	10		3		1				
LA	4		1						
	8	1	1	2					
	10		1		1				
SS	4		1	1					
	8				1				
	10	1	2	1	1				
SC	4		1			1	1		
	8		1	1		1	5		2
	10		1		1				

*Note that the number of FT items flagged reflects the total number flagged across all forms. FT items flagged for a grade and content area are not necessarily from a single form; the table entries reflect the total from across all forms.

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 9-1
Scale Score Descriptive Statistics based on Census Data

Content	Grade	N Count	Mean	SD	Skewness	Kurtosis	Min	Max	LOSS	HOSS
Reading	3	58187	458.53	38.93	-1.09	4.30	270	640	270	640
	4	57788	477.33	45.17	-0.80	2.27	280	650	280	650
	5	59019	484.76	46.95	-0.95	2.61	290	690	290	690
	6	60066	503.10	49.52	-0.89	2.46	300	730	300	730
	7	62610	512.86	47.77	-0.77	2.05	310	780	310	780
	8	64080	526.73	52.38	-0.74	2.32	330	790	330	790
	10	69693	539.36	62.01	-0.47	0.89	350	820	350	820
Mathematics	3	58719	434.33	46.81	-0.16	1.25	220	630	220	630
	4	58446	466.31	43.12	-0.45	1.78	240	650	240	650
	5	59415	489.39	44.09	-0.26	1.43	270	680	270	680
	6	60156	513.18	45.46	-0.21	1.65	310	700	310	700
	7	62855	535.01	43.51	-0.50	1.90	330	710	330	710
	8	64353	542.58	48.36	-0.47	1.07	350	730	350	730
	10	69956	563.40	49.26	-0.41	0.90	410	750	410	750
Language Arts	4	57763	295.79	29.04	-0.41	4.83	140	420	140	420
	8	64046	394.95	38.60	-0.04	2.03	250	520	250	520
	10	69242	448.68	38.73	-0.47	1.04	290	630	290	630
Social Studies	4	58032	298.00	27.54	0.62	3.62	170	400	170	400
	8	63999	397.14	36.80	-0.10	2.56	230	530	230	530
	10	69258	447.66	45.73	-0.94	3.88	240	620	240	620
Science	4	58091	299.59	31.23	0.20	3.01	170	440	170	440
	8	64085	398.09	40.87	-0.33	2.85	230	560	230	560
	10	69351	451.06	43.08	-0.83	3.23	240	610	240	610

Table 9-2
Scale Score Descriptive Statistics based on 14 Districts

Content	Grade	N Count	Mean	SD	Skewness	Kurtosis	Min	Max	LOSS	HOSS
Reading	3	6451	457.30	41.50	-1.01	4.02	270	640	270	640
	4	6424	477.44	48.71	-0.80	1.93	280	639	280	650
	5	6547	485.79	51.84	-0.88	2.22	290	673	290	690
	6	6575	502.60	54.51	-0.83	2.12	300	730	300	730
	7	6769	510.44	51.78	-0.71	1.79	310	780	310	780
	8	6883	527.66	56.86	-0.63	1.88	330	790	330	790
	10	7586	535.86	68.91	-0.42	0.56	350	820	350	820
Mathematics	3	6500	434.01	47.28	-0.04	0.92	220	630	220	630
	4	6491	468.47	46.31	-0.35	1.67	240	650	240	650
	5	6580	490.88	46.53	-0.14	1.26	270	680	270	680
	6	6609	515.83	47.31	-0.11	1.45	310	700	310	700
	7	6795	535.07	45.83	-0.53	1.81	330	710	330	710
	8	6903	545.40	51.17	-0.38	0.92	350	730	350	730
	10	7670	562.24	54.19	-0.33	0.69	410	750	410	750
Language Arts	4	6400	295.37	32.20	-0.77	5.39	140	420	140	420
	8	6853	396.23	41.62	-0.01	1.64	250	520	250	520
	10	7449	447.82	41.81	-0.40	0.90	290	630	290	630
Social Studies	4	6417	298.01	28.93	0.54	3.38	170	400	170	400
	8	6864	397.56	40.29	-0.06	2.34	230	530	230	530
	10	7449	447.82	50.03	-0.80	3.29	240	620	240	620
Science	4	6424	298.71	33.97	0.35	2.99	170	440	170	440
	8	6862	396.94	43.60	-0.21	2.20	230	560	230	560
	10	7453	448.57	46.82	-0.71	2.83	240	610	240	610

Table 9-3
Scale Score Descriptive Statistics by Gender

Content	Grade	Male					Female				
		N Count	Mean	SD	Min	Max	N Count	Mean	SD	Min	Max
Reading	3	29665	454.52	40.63	270	640	28521	462.70	36.62	270	640
	4	29426	474.43	46.14	280	650	28362	480.34	43.94	280	650
	5	30228	480.58	48.00	290	690	28791	489.15	45.41	290	690
	6	30695	498.18	51.22	300	730	29370	508.24	47.14	300	730
	7	31896	509.98	49.45	310	780	30713	515.85	45.76	310	780
	8	32944	522.02	54.61	330	790	31135	531.72	49.43	330	790
	10	35673	533.53	63.36	350	820	34020	545.48	59.95	350	820
Mathematics	3	29996	435.17	47.20	220	630	28722	433.44	46.38	220	630
	4	29832	468.08	43.67	240	650	28614	464.46	42.46	240	650
	5	30483	489.88	44.90	270	680	28932	488.89	43.22	270	680
	6	30776	513.72	47.20	310	700	29379	512.62	43.55	310	700
	7	32015	535.64	45.10	330	710	30840	534.36	41.77	330	710
	8	33103	543.79	49.73	350	730	31249	541.31	46.83	350	730
	10	35806	564.49	51.43	410	750	34150	562.25	46.85	410	750
Language Arts	4	29412	292.27	29.60	140	420	28351	299.44	27.98	140	420
	8	32905	389.31	39.75	250	520	31139	400.91	36.41	250	520
	10	35411	442.12	40.10	290	630	33831	455.53	35.98	290	630
Social Studies	4	29597	297.72	28.37	170	400	28435	298.31	26.64	170	400
	8	32894	397.11	38.82	230	530	31103	397.18	34.54	230	530
	10	35420	447.22	48.83	240	620	33838	448.12	42.25	240	620
Science	4	29638	300.62	32.41	170	440	28453	298.53	29.92	170	440
	8	32950	398.96	43.64	230	560	31134	397.18	37.69	230	560
	10	35476	454.11	46.55	240	610	33875	447.86	38.85	240	610

Table 9-4
Scale Score Descriptive Statistics for Reading by Ethnicity

Content	Ethnicity	Grade	N Count	Mean	SD	Min	Max
Reading	W	3	44380	464.54	36.02	270	640
		4	44396	484.51	41.91	280	650
		5	45513	492.62	42.68	290	690
		6	46385	511.94	44.80	300	730
		7	48781	521.68	43.04	310	780
		8	50211	534.90	48.77	330	790
		10	56715	549.50	56.52	350	820
	AA	3	6387	433.33	43.32	270	589
		4	6244	446.03	48.41	280	593
		5	6247	449.89	51.04	290	623
		6	6470	463.58	53.99	300	619
		7	6794	471.74	50.36	310	626
		8	6711	487.37	54.40	330	664
		10	6020	481.56	64.18	350	705
	H	3	4621	439.75	38.37	270	589
		4	4325	455.38	44.14	280	634
		5	4280	460.33	49.33	290	690
		6	4229	475.96	50.11	300	645
		7	3935	486.75	48.28	310	673
		8	3920	500.16	52.28	330	669
		10	3461	500.90	63.95	350	717
	A	3	1969	453.76	40.50	270	589
		4	2013	469.40	47.75	280	617
		5	2138	474.92	51.26	290	644
		6	2126	491.34	50.59	300	722
		7	2143	497.84	50.44	310	642
		8	2246	517.48	51.57	330	718
		10	2462	511.56	66.41	350	820
	AI	3	828	446.84	37.28	270	558
		4	810	462.32	42.59	280	564
		5	841	468.28	42.21	290	587
		6	854	486.06	47.28	300	679
		7	956	495.78	44.87	310	673
		8	990	505.57	52.51	330	676
		10	1033	514.87	57.37	350	757

Table 9-5
Scale Score Descriptive Statistics for Mathematics by Ethnicity

Content	Ethnicity	Grade	N Count	Mean	SD	Min	Max
Mathematics	**W**	3	44493	442.53	43.40	220	630
		4	44530	473.90	39.39	240	650
		5	45619	496.89	40.92	270	680
		6	46268	521.17	42.13	310	700
		7	48783	543.06	39.49	330	710
		8	50255	550.91	44.64	350	730
		10	56736	571.94	44.55	410	750
	AA	3	6420	395.83	48.32	220	630
		4	6294	429.09	46.69	240	650
		5	6272	452.00	44.20	270	680
		6	6469	472.80	44.92	310	700
		7	6787	493.45	44.75	330	710
		8	6707	498.62	47.85	350	730
		10	6016	509.33	49.11	410	720
	H	3	4915	412.60	41.57	220	570
		4	4691	446.85	39.35	240	591
		5	4481	467.80	40.48	270	634
		6	4373	490.88	40.03	310	700
		7	4115	513.04	39.25	330	641
		8	4102	517.31	44.44	350	730
		10	3634	532.73	46.98	410	723
	A	3	2054	435.20	48.83	220	630
		4	2117	465.23	45.20	240	650
		5	2200	490.09	46.37	270	680
		6	2192	515.53	46.17	310	700
		7	2219	533.03	43.94	330	710
		8	2294	542.70	48.53	350	730
		10	2536	554.24	51.22	410	750
	AI	3	835	418.71	41.45	220	551
		4	814	453.79	36.69	245	592
		5	843	474.83	37.81	338	593
		6	852	494.29	41.31	310	624
		7	951	518.55	40.45	330	655
		8	993	522.09	45.10	350	668
		10	1032	539.47	44.15	410	661

Table 9-6
Scale Score Descriptive Statistics for Language Arts by Ethnicity

Content	Ethnicity	Grade	N Count	Mean	SD	Min	Max
Language Arts	W	4	44443	299.60	27.67	140	420
		8	50285	400.34	37.15	250	520
		10	56583	454.10	36.24	290	630
	AA	4	6233	277.89	30.17	140	420
		8	6631	369.03	37.37	250	520
		10	5870	416.22	40.44	290	549
	H	4	4306	284.61	28.08	140	420
		8	3891	377.64	35.05	250	520
		10	3380	426.53	39.00	290	547
	A	4	1975	294.68	30.58	140	420
		8	2246	388.14	38.55	250	520
		10	2386	438.87	37.69	290	584
	AI	4	806	286.47	25.58	140	420
		8	990	378.37	34.71	250	520
		10	1022	430.66	36.31	290	545

Table 9-7
Scale Score Descriptive Statistics for Social Studies by Ethnicity

Content	Ethnicity	Grade	N Count	Mean	SD	Min	Max
Social Studies	W	4	44610	302.20	26.49	170	400
		8	50231	403.26	34.29	230	530
		10	56615	454.42	41.56	240	620
	AA	4	6265	277.96	26.09	170	400
		8	6628	365.67	36.83	230	530
		10	5858	404.01	54.01	240	620
	H	4	4354	287.08	23.46	170	400
		8	3901	378.46	34.50	230	530
		10	3374	423.81	46.76	240	620
	A	4	1990	294.73	29.03	170	400
		8	2246	392.12	36.08	230	530
		10	2386	436.75	41.93	240	620
	AI	4	813	288.97	23.82	170	400
		8	990	382.47	31.38	230	530
		10	1024	427.67	44.55	240	569

Table 9-8
Scale Score Descriptive Statistics for Science by Ethnicity

Content	Ethnicity	Grade	N Count	Mean	SD	Min	Max
Science	W	4	44629	305.22	29.08	170	440
		8	50297	405.43	37.36	230	560
		10	56653	458.51	38.25	240	610
	AA	4	6285	272.65	30.98	170	440
		8	6639	361.67	41.59	230	560
		10	5880	402.76	48.53	240	602
	H	4	4369	284.62	27.01	170	440
		8	3904	375.40	39.10	230	560
		10	3395	424.96	42.64	240	602
	A	4	1993	296.08	31.71	170	440
		8	2254	387.97	41.86	230	560
		10	2395	437.93	42.64	240	610
	AI	4	815	288.28	27.83	170	440
		8	989	382.14	37.85	230	509
		10	1026	433.44	38.39	240	545

Table 9-9
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	20546	441.84	40.59	270	589	37641	467.64	34.77	270	640
	4	19652	456.36	46.20	280	634	38136	488.14	40.60	280	650
	5	19736	461.46	48.76	290	690	39283	496.47	41.32	290	690
	6	19947	477.96	51.63	300	730	40119	515.60	43.32	300	730
	7	19967	487.41	49.49	310	673	42643	524.77	41.93	310	780
	8	20020	500.79	54.02	330	790	44060	538.52	47.12	330	790
	10	17307	503.02	63.60	350	820	52386	551.37	56.56	350	820
Mathematics	3	20970	413.01	45.15	220	630	37749	446.17	43.39	220	630
	4	20192	446.17	43.53	240	650	38254	476.94	38.91	240	650
	5	20050	467.69	42.77	270	680	39365	500.45	40.50	270	680
	6	20143	489.95	43.87	310	700	40013	524.88	41.59	310	700
	7	20190	512.27	43.45	330	710	42665	545.77	39.17	330	710
	8	20238	517.09	47.73	350	730	44115	554.28	43.95	350	730
	10	17508	533.94	49.49	410	750	52448	573.23	45.08	410	750
Language Arts	4	19618	284.13	28.26	140	420	38145	301.79	27.58	140	420
	8	19927	377.17	36.57	250	520	44119	402.98	36.76	250	520
	10	17003	427.53	38.73	290	587	52239	455.56	36.15	290	630
Social Studies	4	19792	286.25	25.19	170	400	38240	304.09	26.73	170	400
	8	19941	378.27	35.51	230	530	44058	405.68	34.10	230	530
	10	17010	422.31	49.49	240	620	52248	455.91	41.20	240	620
Science	4	19831	284.98	30.03	170	440	38260	307.17	29.08	170	440
	8	19963	377.15	41.22	230	560	44122	407.57	37.00	230	560
	10	17061	426.41	45.68	240	610	52290	459.10	38.95	240	610

Table 9-10
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	7094	425.06	48.48	270	589	51093	463.18	34.97	270	640
	4	7262	435.68	54.73	280	616	50526	483.32	40.22	280	650
	5	7698	437.25	57.67	290	644	51321	491.89	40.58	290	690
	6	7654	448.99	59.64	300	668	52412	511.00	42.44	300	730
	7	8204	461.09	53.69	310	637	54406	520.66	41.55	310	780
	8	8541	468.94	59.76	330	691	55539	535.62	44.99	330	790
	10	8629	471.05	61.00	350	759	61064	549.02	55.76	350	820
Mathematics	3	7302	407.77	48.50	220	630	51417	438.10	45.32	220	630
	4	7490	435.86	49.83	240	650	50956	470.78	40.14	240	650
	5	7862	454.58	48.12	270	680	51553	494.70	40.92	270	680
	6	7692	471.24	49.27	310	700	52464	519.33	41.45	310	700
	7	8229	490.45	47.70	330	710	54626	541.72	38.61	330	710
	8	8587	492.54	50.27	350	665	55766	550.29	43.18	350	730
	10	8654	510.47	48.35	410	750	61302	570.87	44.58	410	750
Language Arts	4	7326	277.41	31.03	140	420	50437	298.46	27.74	140	420
	8	8519	358.16	38.86	250	520	55527	400.60	35.31	250	520
	10	8501	406.42	35.68	290	564	60741	454.59	35.31	290	630
Social Studies	4	7582	284.41	26.93	170	400	50450	300.05	27.05	170	400
	8	8580	364.06	38.17	230	530	55419	402.26	33.81	230	530
	10	8544	404.36	52.44	240	620	60714	453.75	41.20	240	620
Science	4	7601	282.87	32.57	170	440	50490	302.11	30.23	170	440
	8	8598	362.09	45.15	230	560	55487	403.67	37.16	230	560
	10	8577	412.05	48.32	240	610	60774	456.56	39.28	240	610

Table 9-11
Scale Score Descriptive Statistics by English Language Proficiency

Content	Grade	Proficient					Not Proficient				
		N Count	Mean	SD	Min	Max	N Count	Mean	SD	Min	Max
Reading	3	54562	460.19	38.44	270	640	3625	433.58	37.77	270	589
	4	54417	479.21	44.59	280	650	3371	446.97	43.54	280	634
	5	55630	486.97	45.95	290	690	3389	448.62	48.42	290	622
	6	56985	505.32	48.60	300	730	3081	461.99	48.50	300	626
	7	59768	514.72	46.94	310	780	2842	473.56	48.15	310	646
	8	61528	528.47	51.75	330	790	2552	484.97	50.18	330	646
	10	67519	541.61	60.77	350	820	2174	469.69	59.51	350	820
Mathematics	3	54725	435.95	46.84	220	630	3994	412.09	40.16	220	630
	4	54603	467.79	42.92	240	650	3843	445.26	40.35	240	650
	5	55757	491.02	43.81	270	680	3658	464.65	40.83	270	680
	6	56848	514.62	45.38	310	700	3308	488.52	39.29	310	700
	7	59759	536.32	43.24	330	710	3096	509.74	40.76	330	650
	8	61561	543.93	48.09	350	730	2792	512.90	44.69	350	730
	10	67536	564.84	48.70	410	750	2420	523.22	47.75	410	723
Language Arts	4	54434	296.64	28.82	140	420	3328	281.93	29.18	140	420
	8	61509	396.08	38.41	250	520	2535	367.73	32.78	250	520
	10	67178	449.73	38.35	290	630	2063	414.34	35.09	290	530
Social Studies	4	54653	298.86	27.59	170	400	3378	284.21	22.70	170	400
	8	61462	398.20	36.60	230	530	2535	371.51	32.23	230	530
	10	67196	448.78	45.32	240	620	2061	411.03	43.96	240	620
Science	4	54695	300.69	31.13	170	440	3396	281.89	27.27	170	440
	8	61533	399.44	40.44	230	560	2552	365.54	37.51	230	560
	10	67269	452.23	42.59	240	610	2082	413.22	41.72	240	541

Table 9-12
The Number of Students and Percents at LOSS and HOSS

Content	Grade	LOSS	N	Percent	HOSS	N	Percent
Reading	3	**270**	524	0.90	**640**	14	0.02
	4	**280**	374	0.65	**650**	5	0.01
	5	**290**	627	1.06	**690**	7	0.01
	6	**300**	537	0.89	**730**	5	0.01
	7	**310**	465	0.74	**780**	6	0.01
	8	**330**	882	1.38	**790**	18	0.03
	10	**350**	946	1.36	**820**	37	0.05
Mathematics	3	**220**	88	0.15	**630**	129	0.22
	4	**240**	57	0.10	**650**	65	0.11
	5	**270**	66	0.11	**680**	65	0.11
	6	**310**	157	0.26	**700**	89	0.15
	7	**330**	207	0.33	**710**	41	0.07
	8	**350**	239	0.37	**730**	46	0.07
	10	**410**	851	1.22	**750**	77	0.11
Language Arts	4	**140**	257	0.44	**420**	314	0.54
	8	**250**	472	0.74	**520**	899	1.40
	10	**290**	293	0.42	**630**	5	0.01
Social Studies	4	**170**	113	0.19	**400**	1230	2.12
	8	**230**	257	0.40	**530**	642	1.00
	10	**240**	646	0.93	**620**	216	0.31
Science	4	**170**	128	0.22	**440**	387	0.67
	8	**230**	458	0.71	**560**	365	0.57
	10	**240**	356	0.51	**610**	48	0.07

Table 9-13
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	58187	1	Determines Meaning	11	0	11	6.98	0.63	2.27	63.40	17.61
	58187	2	Understands Text	17	0	17	12.91	0.76	3.72	75.83	20.78
	58187	3	Analyzes Text	22	1	25	16.92	0.68	5.10	67.89	19.53
	58187	4	Evaluates/Extends Text	4	1	7	4.29	0.61	1.51	61.71	16.53
4	57788	1	Determines Meaning	10	0	10	7.37	0.74	2.21	73.49	19.78
	57788	2	Understands Text	18	0	18	12.99	0.72	3.69	71.70	19.10
	57788	3	Analyzes Text	19	1	22	12.91	0.59	4.35	58.96	18.41
	57788	4	Evaluates/Extends Text	6	1	9	4.40	0.49	1.86	49.48	16.39
5	59019	1	Determines Meaning	12	0	12	8.15	0.68	2.49	67.93	17.89
	59019	2	Understands Text	16	0	16	12.26	0.77	3.10	76.30	18.15
	59019	3	Analyzes Text	17	1	20	11.62	0.58	3.66	58.58	16.59
	59019	4	Evaluates/Extends Text	9	1	12	6.97	0.58	2.68	58.97	19.82
6	60066	1	Determines Meaning	11	0	11	7.77	0.71	2.32	70.44	18.49
	60066	2	Understands Text	14	0	14	10.35	0.74	2.61	73.71	16.27
	60066	3	Analyzes Text	18	1	21	12.70	0.60	3.63	60.94	15.58
	60066	4	Evaluates/Extends Text	11	1	14	8.61	0.61	2.67	62.01	16.73
7	62610	1	Determines Meaning	11	0	11	7.27	0.66	2.21	65.90	17.14
	62610	2	Understands Text	15	0	15	10.46	0.70	2.93	69.53	17.48
	62610	3	Analyzes Text	18	1	21	12.59	0.60	4.23	60.02	18.57
	62610	4	Evaluates/Extends Text	10	1	13	7.47	0.57	2.62	57.92	17.38
8	64080	1	Determines Meaning	11	0	11	7.97	0.72	2.32	72.15	18.38
	64080	2	Understands Text	13	0	13	9.34	0.72	2.68	71.41	18.78
	64080	3	Analyzes Text	17	1	20	11.59	0.58	4.01	58.67	18.36
	64080	4	Evaluates/Extends Text	13	1	16	10.46	0.65	3.23	65.83	18.29

Table 9-13 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	69693	1	Determines Meaning	7	0	7	5.13	0.73	1.67	73.16	20.45
	69693	2	Understands Text	6	0	6	3.98	0.66	1.51	65.68	19.94
	69693	3	Analyzes Text	23	1	26	16.08	0.62	4.85	62.33	17.60
	69693	4	Evaluates/Extends Text	14	1	17	11.02	0.65	3.51	65.00	18.96

Table 9-14
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	58719	A	Mathematical Processes	0	4	8	3.08	0.39	2.25	38.32	24.54
	58719	B	Number Operations	11	1	12	8.76	0.73	2.54	73.25	18.74
	58719	C	Geometry	9	1	10	7.77	0.78	1.81	77.86	15.05
	58719	D	Measurement	9	0	9	6.96	0.77	1.86	76.42	16.90
	58719	E	Statistics/Probability	8	1	9	7.02	0.78	1.84	77.30	17.64
	58719	F	Algebraic Relationships	8	1	9	6.54	0.73	1.82	73.70	17.18
4	58446	A	Mathematical Processes	0	4	8	2.55	0.32	2.01	34.14	20.16
	58446	B	Number Operations	11	0	11	9.51	0.86	1.83	85.86	14.53
	58446	C	Geometry	9	1	10	7.64	0.76	1.61	76.35	12.62
	58446	D	Measurement	8	1	9	6.28	0.70	1.94	69.93	18.73
	58446	E	Statistics/Probability	8	1	9	6.30	0.70	2.11	68.32	19.94
	58446	F	Algebraic Relationships	9	1	10	7.74	0.77	2.04	77.75	17.91
5	59415	A	Mathematical Processes	0	4	8	3.44	0.43	1.92	43.59	20.31
	59415	B	Number Operations	12	0	12	9.52	0.79	2.13	78.85	15.13
	59415	C	Geometry	9	1	10	6.79	0.68	1.77	68.09	13.23
	59415	D	Measurement	9	1	10	6.96	0.70	1.77	70.24	13.88
	59415	E	Statistics/Probability	10	1	11	7.18	0.65	2.68	64.52	21.23
	59415	F	Algebraic Relationships	10	1	11	7.23	0.66	2.38	66.51	18.22

Table 9-14 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	60156	A	Mathematical Processes	0	4	8	3.38	0.42	2.19	43.48	24.00
	60156	B	Number Operations	11	0	11	7.20	0.65	2.58	65.91	21.19
	60156	C	Geometry	8	1	9	6.79	0.75	1.97	73.65	16.24
	60156	D	Measurement	9	1	10	5.89	0.59	2.29	59.32	19.43
	60156	E	Statistics/Probability	11	1	12	7.96	0.66	2.85	62.84	20.91
	60156	F	Algebraic Relationships	10	1	11	7.90	0.72	2.31	71.76	18.91
7	62855	A	Mathematical Processes	0	4	8	3.77	0.47	2.08	46.93	21.94
	62855	B	Number Operations	12	0	12	7.53	0.63	2.60	63.67	18.77
	62855	C	Geometry	10	2	12	7.21	0.60	2.57	60.71	17.82
	62855	D	Measurement	10	0	10	5.14	0.51	2.32	52.47	19.87
	62855	E	Statistics/Probability	8	1	9	6.15	0.68	2.00	66.14	19.17
	62855	F	Algebraic Relationships	10	1	11	8.96	0.81	1.94	81.30	15.01
8	64353	A	Mathematical Processes	0	4	8	4.50	0.56	2.29	55.98	25.32
	64353	B	Number Operations	9	0	9	4.46	0.50	2.24	49.34	20.90
	64353	C	Geometry	8	1	9	4.67	0.52	2.22	53.44	20.82
	64353	D	Measurement	11	1	12	6.28	0.52	2.40	52.64	16.87
	64353	E	Statistics/Probability	8	1	9	5.19	0.58	2.15	56.63	20.02
	64353	F	Algebraic Relationships	14	1	15	11.46	0.76	2.66	75.91	15.84
10	69956	A	Mathematical Processes	7	1	9	5.60	0.62	2.06	61.78	19.18
	69956	B	Number Operations	7	0	7	4.87	0.70	1.62	69.16	18.54
	69956	C	Geometry	8	1	10	5.22	0.52	2.54	53.58	21.60
	69956	D	Measurement	9	1	11	6.05	0.55	3.00	56.15	24.17
	69956	E	Statistics/Probability	9	0	9	5.55	0.62	2.06	61.25	18.58
	69956	F	Algebraic Relationships	10	1	12	6.91	0.58	2.69	58.65	19.98

Table 9-15
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	57763	B	Writing	19	0	19	13.43	0.71	3.39	70.84	16.94
	57763	D	Language	5	0	5	3.41	0.68	1.17	67.52	16.75
	57763	F	Research and Inquiry	6	0	6	3.40	0.57	1.55	57.36	20.18
8	64046	B	Writing	16	0	16	12.49	0.78	2.87	78.07	16.91
	64046	D	Language	8	0	8	5.35	0.67	1.97	67.30	21.84
	64046	F	Research and Inquiry	6	0	6	3.76	0.63	1.49	62.57	18.17
10	69225	B	Writing	15	2	24	14.77	0.62	3.58	61.51	14.10
	69225	D	Language	9	0	9	5.92	0.66	2.21	66.36	21.29
	69225	F	Research and Inquiry	6	0	6	3.76	0.63	1.48	62.76	17.86

Table 9-16
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	58032	A	Geography	9	0	9	7.03	0.78	1.74	78.12	16.73
	58032	B	History	8	0	8	6.28	0.78	1.54	78.10	16.02
	58032	C	Political Science	7	0	7	5.15	0.74	1.42	74.06	14.97
	58032	D	Economics	7	0	7	5.93	0.85	1.09	84.81	12.41
	58032	E	Behavioral Science	7	0	7	5.33	0.76	1.71	76.70	20.73
8	63999	A	Geography	10	0	10	7.48	0.75	1.87	74.92	15.45
	63999	B	History	13	0	13	9.54	0.73	2.64	73.79	18.31
	63999	C	Political Science	6	0	6	4.19	0.70	1.51	69.44	20.03
	63999	D	Economics	6	0	6	4.67	0.78	1.20	77.48	15.90
	63999	E	Behavioral Science	5	0	5	3.06	0.61	1.43	62.74	22.63
10	69258	A	Geography	10	0	10	7.30	0.73	2.04	72.00	17.11
	69258	B	History	12	0	12	7.50	0.63	2.81	63.49	21.09
	69258	C	Political Science	12	0	12	7.25	0.60	2.54	60.56	18.63
	69258	D	Economics	8	0	8	5.75	0.72	1.88	71.18	20.06
	69258	E	Behavioral Science	8	0	8	5.19	0.65	2.04	65.34	21.44

Table 9-17
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	58091	A	Science Connections	5	0	5	4.31	0.86	1.00	85.12	15.68
	58091	B	Nature of Science	3	0	3	2.40	0.80	0.71	78.96	15.30
	58091	C	Science Inquiry	6	0	6	4.13	0.69	1.46	69.35	19.46
	58091	D	Physical Science	6	0	6	4.30	0.72	1.16	73.28	13.37
	58091	E	Earth and Space	6	0	6	4.02	0.67	1.49	67.01	18.66
	58091	F	Life and Environment	6	0	6	4.13	0.69	1.45	67.89	18.49
	58091	G	Science Applications	3	0	3	2.00	0.67	0.84	67.44	16.47
	58091	H	Personal/Social Perspectives	5	0	5	3.60	0.72	1.29	72.50	20.52
8	64085	A	Science Connections	3	0	3	2.50	0.83	0.75	82.29	18.53
	64085	B	Nature of Science	3	0	3	1.94	0.65	0.85	65.68	17.33
	64085	C	Science Inquiry	7	0	7	5.17	0.74	1.58	74.09	19.03
	64085	D	Physical Science	6	0	6	3.22	0.54	1.48	54.44	18.60
	64085	E	Earth and Space	6	0	6	3.58	0.60	1.55	59.48	19.46
	64085	F	Life and Environment	6	0	6	4.32	0.72	1.35	72.29	16.56
	64085	G	Science Applications	5	0	5	3.89	0.78	1.17	77.39	18.25
	64085	H	Personal/Social Perspectives	4	0	4	2.98	0.75	1.04	74.19	18.06
10	69351	A	Science Connections	5	0	5	3.03	0.61	1.27	60.65	17.96
	69351	B	Nature of Science	5	0	5	3.47	0.69	1.16	70.16	17.39
	69351	C	Science Inquiry	10	0	10	7.65	0.77	1.98	76.27	16.90
	69351	D	Physical Science	7	0	7	3.47	0.50	1.66	49.45	16.71
	69351	E	Earth and Space	6	0	6	3.99	0.66	1.35	66.44	16.76
	69351	F	Life and Environment	7	0	7	3.61	0.52	1.58	53.81	17.04
	69351	G	Science Applications	5	0	5	3.16	0.63	1.38	63.09	21.64
	69351	H	Personal/Social Perspectives	5	0	5	2.95	0.59	1.45	59.58	22.05

Table 9-18
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

Table 9-19
Percentage of Students in Each Performance Level by Sub-Group (Reading)

Grade	Proficiency Level	Examinees		Gender		Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
3	M	2640	4.54	3.10	5.92	2.91	12.46	8.72	4.77	6.40	4.17	10.10	19.02	2.53	8.80	2.21
	B	7896	13.57	11.52	15.54	10.26	26.74	24.32	17.47	19.93	12.61	28.08	31.04	11.14	22.33	8.79
	P	21293	36.59	36.23	36.94	34.81	41.27	44.08	41.19	43.72	35.92	46.76	32.41	37.18	42.19	33.54
	A	26358	45.30	49.15	41.60	52.02	19.52	22.87	36.57	29.95	47.31	15.06	17.54	49.15	26.68	55.46
Total		58187	100.00	28521	29665	44380	6387	4621	1969	828	54562	3625	7094	51093	20546	37641
4	M	2340	4.05	3.19	4.87	2.58	11.64	7.61	5.12	4.57	3.71	9.55	18.66	1.95	8.10	1.96
	B	7444	12.88	11.89	13.83	9.56	27.32	22.66	17.78	19.26	11.95	27.86	30.80	10.31	22.03	8.17
	P	23528	40.71	40.66	40.76	39.08	44.63	48.99	42.87	50.25	40.21	48.86	35.44	41.47	46.73	37.62
	A	24476	42.36	44.25	40.53	48.78	16.40	20.74	34.23	25.93	44.13	13.73	15.09	46.27	23.14	52.26
Total		57788	100.00	28362	29426	44396	6244	4325	2013	810	54417	3371	7262	50526	19652	38136
5	M	2708	4.59	3.66	5.47	2.78	13.49	9.86	5.99	5.83	4.11	12.48	21.49	2.05	9.44	2.15
	B	6167	10.45	9.06	11.77	7.44	23.85	19.00	15.58	17.00	9.58	24.79	27.07	7.96	18.96	6.17
	P	24917	42.22	40.98	43.40	40.43	47.78	49.63	44.95	53.03	41.70	50.72	39.14	42.68	49.31	38.66
	A	25227	42.74	46.30	39.35	49.34	14.87	21.52	33.49	24.14	44.62	12.01	12.30	47.31	22.28	53.02
Total		59019	100.00	28791	30228	45513	6247	4280	2138	841	55630	3389	7698	51321	19736	39283

Table 9-19 Cont'd
Percentage of Students in Each Performance Level by Sub-Group (Reading)

Grade	Proficiency Level	Examinees		Gender		Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
6	M	2879	4.79	3.32	6.21	2.81	14.78	10.12	5.69	8.08	4.30	13.86	24.22	1.96	10.01	2.20
	B	5573	9.28	8.11	10.39	6.18	22.86	18.78	15.19	13.00	8.46	24.34	25.74	6.87	17.25	5.32
	P	24974	41.58	40.98	42.15	39.49	47.76	49.80	47.74	52.11	41.05	51.25	39.13	41.94	49.51	37.64
	A	26640	44.35	47.59	41.25	51.52	14.61	21.31	31.37	26.81	46.18	10.55	10.91	49.23	23.24	54.85
Total		60066	100.00	29370	30695	46385	6470	4229	2126	854	56985	3081	7654	52412	19947	40119
7	M	3238	5.17	3.95	6.35	2.78	18.06	10.80	7.42	7.64	4.69	15.27	24.62	2.24	11.39	2.26
	B	5936	9.48	8.60	10.33	6.33	24.20	18.45	15.59	14.75	8.82	23.43	26.86	6.86	17.88	5.55
	P	25325	40.45	40.88	40.03	38.87	43.58	49.02	46.52	49.69	40.02	49.44	38.10	40.80	47.18	37.30
	A	28111	44.90	46.57	43.29	52.02	14.16	21.73	30.47	27.93	46.47	11.86	10.41	50.10	23.54	54.90
Total		62610	100.00	30713	31896	48781	6794	3935	2143	956	59768	2842	8204	54406	19967	42643
8	M	3432	5.36	3.62	6.99	3.45	15.53	11.38	5.57	8.79	4.95	15.20	25.88	2.20	11.21	2.69
	B	6055	9.45	8.46	10.38	6.86	22.40	17.58	11.84	15.56	8.88	23.08	27.53	6.67	17.10	5.97
	P	26961	42.07	42.36	41.80	40.16	47.58	49.39	51.16	52.22	41.67	51.72	37.06	42.85	48.80	39.02
	A	27632	43.12	45.55	40.82	49.53	14.50	21.66	31.43	23.43	44.50	9.99	9.54	48.28	22.89	52.32
Total		64080	100.00	31135	32944	50211	6711	3920	2246	990	61528	2552	8541	55539	20020	44060

Table 9-19 Cont'd
Percentage of Students in Each Performance Level by Sub-Group (Reading)

Grade	Proficiency Level	Examinees		Gender		Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
10	M	6647	9.54	7.38	11.59	5.83	32.03	23.40	18.36	14.62	8.61	38.22	39.21	5.35	21.67	5.53
	B	9864	14.15	13.07	15.19	11.38	28.27	25.37	23.60	23.81	13.61	31.05	29.08	12.04	23.87	10.94
	P	23101	33.15	33.45	32.85	33.75	27.89	30.83	33.51	37.66	33.43	24.43	23.57	34.50	33.58	33.01
	A	30081	43.16	46.09	40.37	49.04	11.81	20.40	24.53	23.91	44.35	6.30	8.15	48.11	20.88	50.52
Total		69693	100.00	34020	35673	56715	6020	3461	2462	1033	67519	2174	8629	61064	17307	52386

Table 9-20
Percentage of Students in Each Performance Level by Sub-Group (Mathematics)

Grade	Proficiency Level	Examinees		Gender		Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
3	M	9730	16.57	16.93	16.23	11.01	44.84	28.69	16.94	23.23	15.67	28.92	34.77	13.99	29.50	9.39
	B	5439	9.26	9.46	9.08	8.01	13.21	14.22	10.03	14.37	8.90	14.30	13.08	8.72	12.75	7.33
	P	22746	38.74	39.19	38.31	39.55	30.62	41.24	39.00	42.51	38.55	41.34	35.52	39.19	39.37	38.39
	A	20804	35.43	34.43	36.39	41.43	11.32	15.85	34.03	19.88	36.89	15.45	16.63	38.10	18.38	44.90
Total		58719	100	28722	29996	44493	6420	4915	2054	835	54725	3994	7302	51417	20970	37749
4	M	7670	13.12	13.94	12.34	8.45	38.88	22.23	13.13	17.20	12.41	23.26	34.94	9.92	24.65	7.04
	B	5373	9.19	9.80	8.61	7.59	15.22	15.50	10.06	11.67	8.73	15.79	13.18	8.61	13.42	6.96
	P	24832	42.49	43.14	41.86	42.82	35.67	46.07	44.40	51.23	42.22	46.24	36.18	43.41	43.49	41.96
	A	20571	35.20	33.12	37.18	41.14	10.23	16.20	32.40	19.90	36.64	14.70	15.70	38.06	18.44	44.04
Total		58446	100	28614	29832	44530	6294	4691	2117	814	54603	3843	7490	50956	20192	38254
5	M	8274	13.93	14.00	13.86	9.00	40.23	25.73	14.41	20.64	12.99	28.16	38.97	10.11	26.56	7.49
	B	6474	10.90	10.87	10.92	9.21	18.03	17.25	11.55	13.64	10.45	17.77	16.59	10.03	16.26	8.16
	P	22906	38.55	39.78	37.39	39.16	31.97	40.50	38.55	44.13	38.50	39.37	31.35	39.65	39.06	38.29
	A	21761	36.63	35.36	37.83	42.62	9.77	16.51	35.50	21.59	38.06	14.71	13.09	40.21	18.11	46.05
Total		59415	100	28932	30483	45619	6272	4481	2200	843	55757	3658	7862	51553	20050	39365

Table 9-20 Cont'd
Percentage of Students in Each Performance Level by Sub-Group (Mathematics)

Grade	Proficiency Level	Examinees		Gender		Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
6	M	7193	11.96	11.31	12.57	7.18	38.77	21.95	10.13	21.24	11.28	23.64	41.48	7.63	24.02	5.89
	B	6862	11.41	11.56	11.27	9.11	20.65	20.38	12.64	16.90	10.86	20.77	17.86	10.46	18.07	8.05
	P	26281	43.69	45.77	41.70	45.09	32.97	44.02	44.94	44.13	43.70	43.41	31.97	45.41	42.87	44.10
	A	19820	32.95	31.36	34.47	38.63	7.61	13.65	32.30	17.72	34.16	12.18	8.68	36.51	15.05	41.96
Total		60156	100	29379	30776	46268	6469	4373	2192	852	56848	3308	7692	52464	20143	40013
7	M	5902	9.39	9.03	9.74	5.17	34.08	17.28	9.78	14.72	8.87	19.38	36.94	5.24	19.72	4.50
	B	6957	11.07	11.11	11.03	8.31	23.40	20.17	14.42	17.56	10.52	21.61	22.47	9.35	18.43	7.58
	P	29471	46.89	48.72	45.12	48.16	36.22	49.31	44.84	51.84	46.87	47.25	33.90	48.84	47.19	46.74
	A	20525	32.66	31.13	34.12	38.36	6.31	13.24	30.96	15.88	33.74	11.76	6.68	36.57	14.66	41.17
Total		64353	100	31249	33103	50255	6707	4102	2294	993	61561	2792	8587	55766	20238	44115
8	M	6473	10.06	9.88	10.23	6.14	32.95	19.19	9.98	16.21	9.52	21.99	38.85	5.63	20.65	5.20
	B	9260	14.39	14.97	13.84	11.40	28.78	24.91	15.13	23.26	13.87	25.86	26.85	12.47	23.49	10.22
	P	31342	48.70	50.08	47.41	50.82	33.56	46.88	49.61	49.35	48.87	44.95	29.99	51.59	45.04	50.38
	A	17278	26.85	25.07	28.52	31.64	4.71	9.02	25.28	11.18	27.74	7.20	4.31	30.32	10.82	34.20
Total		64353	100	31249	33103	50255	6707	4102	2294	993	61561	2792	8587	55766	20238	44115

Table 9-20 Cont'd
Percentage of Students in Each Performance Level by Sub-Group (Mathematics)

Grade	Proficiency Level	Examinees		Gender		Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
10	M	10334	14.77	14.32	15.20	9.29	52.33	32.03	19.12	25.58	13.89	39.46	51.69	9.56	31.67	9.13
	B	9489	13.56	14.43	12.74	11.71	21.38	23.31	18.89	22.48	13.19	23.97	21.52	12.44	20.46	11.26
	P	31995	45.74	47.31	44.24	48.97	22.86	36.35	42.35	42.54	46.26	31.24	23.58	48.86	38.50	48.15
	A	18138	25.93	23.94	27.82	30.02	3.44	8.31	19.64	9.40	26.67	5.33	3.21	29.13	9.37	31.45
Total		69956	100	34150	35806	56736	6016	3634	2536	1032	67536	2420	8654	61302	17508	52448

Table 9-21
Percentage of Students in Each Performance Level by Sub-Group (Language Arts)

Grade	Proficiency Level	Examinees		Gender		Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
4	M	2673	4.63	3.16	6.05	3.11	12.74	8.29	4.61	5.96	4.36	8.98	13.62	3.32	8.79	2.49
	B	9980	17.28	14.61	19.85	14.18	33.08	24.66	17.37	26.30	16.69	26.83	31.50	15.21	26.34	12.62
	P	26259	45.46	45.03	45.88	45.33	42.26	49.63	48.00	49.13	45.15	50.54	42.75	45.85	47.55	44.39
	A	18851	32.64	37.21	28.23	37.38	11.92	17.42	30.03	18.61	33.80	13.64	12.12	35.61	17.32	40.51
Total		57763	100	28351	29412	44443	6233	4306	1975	806	54434	3328	7326	50437	19618	38145
8	M	8507	13.28	8.87	17.46	9.71	32.11	23.03	16.43	23.03	12.59	30.06	45.32	8.37	24.58	8.18
	B	15334	23.94	21.93	25.84	21.13	34.55	34.59	32.68	33.94	23.18	42.41	32.68	22.60	32.93	19.88
	P	24892	38.87	41.18	36.68	41.20	26.84	33.49	34.02	33.03	39.48	24.02	18.02	42.06	32.34	41.81
	A	15313	23.91	28.02	20.02	27.95	6.50	8.89	16.87	10.00	24.75	3.51	3.98	26.97	10.15	30.13
Total		64046	100	31139	32905	50285	6631	3891	2246	990	61509	2535	8519	55527	19927	44119
10	M	5442	7.86	4.30	11.26	5.48	24.14	16.63	9.14	14.19	7.41	22.35	33.25	4.31	16.75	4.97
	B	13151	18.99	16.00	21.85	15.90	35.28	32.34	28.16	31.12	18.33	40.67	40.38	16.00	30.88	15.12
	P	38664	55.84	58.56	53.24	58.61	37.70	45.95	52.10	48.34	56.45	35.87	25.14	60.14	46.59	58.85
	A	11985	17.31	21.14	13.65	20.02	2.88	5.09	10.60	6.36	17.81	1.11	1.22	19.56	5.78	21.06
Total		69242	100	33831	35411	56583	5870	3380	2386	1022	67178	2063	8501	60741	17003	52239

Table 9-22
Percentage of Students in Each Performance Level by Sub-Group (Social Studies)

Grade	Proficiency Level	Examinees		Gender		Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
4	M	899	1.55	1.26	1.83	0.76	6.56	2.37	1.66	1.35	1.49	2.55	4.60	1.09	3.26	0.66
	B	3096	5.34	5.04	5.62	3.48	15.37	9.00	5.78	9.23	5.07	9.68	11.43	4.42	10.13	2.85
	P	14851	25.59	25.53	25.65	21.09	43.86	39.11	33.87	38.87	24.39	44.97	38.21	23.69	37.88	19.23
	A	39186	67.53	68.18	66.90	74.66	34.21	49.52	58.69	50.55	69.05	42.81	45.75	70.80	48.72	77.26
Total		58032	100	28435	29597	44610	6265	4354	1990	813	54653	3378	7582	50450	19792	38240
8	M	2231	3.49	2.91	4.03	1.80	13.67	7.38	3.38	5.56	3.25	9.19	15.73	1.59	7.77	1.55
	B	7843	12.26	11.69	12.79	8.46	33.00	22.53	14.38	20.61	11.67	26.51	31.74	9.24	23.33	7.24
	P	25793	40.30	41.92	38.77	39.16	40.04	48.40	48.58	49.49	39.84	51.40	40.37	40.29	46.10	37.68
	A	28132	43.96	43.48	44.41	50.58	13.29	21.69	33.66	24.34	45.24	12.90	12.16	48.88	22.80	53.53
Total		63999	100	31103	32894	50231	6628	3901	2246	990	61462	2535	8580	55419	19941	44058
10	M	10353	14.95	13.36	16.47	10.41	46.21	30.29	19.11	26.86	14.22	38.77	47.44	10.38	31.03	9.71
	B	4647	6.71	6.71	6.71	5.73	11.37	11.20	10.27	11.04	6.48	14.31	12.24	5.93	10.50	5.48
	P	21976	31.73	33.97	29.59	31.38	29.53	34.80	39.44	35.55	31.60	36.10	27.41	32.34	34.64	30.78
	A	32282	46.61	45.96	47.24	52.48	12.89	23.71	31.18	26.56	47.71	10.82	12.91	51.35	23.83	54.03
Total		69258	100	33838	35420	56615	5858	3374	2386	1024	67196	2061	8544	60714	17010	52248

Table 9-23
Percentage of Students in Each Performance Level by Sub-Group (Science)

Grade	Proficiency Level	Examinees		Gender		Ethnicity					ELP		Disability		SES	
		N	%	Female	Male	White	African-American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
4	M	2824	4.86	4.78	4.94	2.54	18.63	8.67	4.67	5.77	4.58	9.42	12.81	3.66	10.14	2.12
	B	9561	16.46	17.12	15.82	12.11	36.93	27.95	19.67	27.48	15.59	30.48	28.50	14.65	27.49	10.74
	P	32659	56.22	56.93	55.54	58.47	39.92	56.31	57.25	55.83	56.30	54.98	48.82	57.33	52.74	58.02
	A	13047	22.46	21.17	23.70	26.88	4.52	7.07	18.41	10.92	23.54	5.12	9.87	24.36	9.63	29.11
Total		58091	100	28453	29638	44629	6285	4369	1993	815	54695	3396	7601	50490	19831	38260
8	M	5745	8.97	8.17	9.72	5.06	30.83	18.72	12.60	14.05	8.29	25.31	31.83	5.42	19.13	4.37
	B	9721	15.17	15.76	14.61	11.71	30.58	26.77	22.94	24.17	14.39	33.97	27.89	13.20	25.09	10.68
	P	30133	47.02	49.70	44.49	49.04	33.05	45.06	46.01	48.33	47.49	35.78	32.62	49.25	43.40	48.66
	A	18486	28.85	26.37	31.19	34.19	5.54	9.45	18.46	13.45	29.84	4.94	7.65	32.13	12.39	36.29
Total		64085	100	31134	32950	50297	6639	3904	2254	989	61533	2552	8598	55487	19963	44122
10	M	10059	14.50	14.59	14.42	8.91	53.71	32.22	21.75	23.39	13.67	41.50	44.21	10.31	31.65	8.91
	B	8179	11.79	13.71	9.96	10.10	19.86	19.12	18.58	19.20	11.46	22.43	19.25	10.74	17.91	9.80
	P	24856	35.84	38.79	33.03	37.46	20.02	34.85	36.53	38.69	36.04	29.35	26.77	37.12	33.32	36.66
	A	26257	37.86	32.91	42.59	43.54	6.41	13.81	23.13	18.71	38.82	6.72	9.77	41.83	17.12	44.63
Total		69351	100	33875	35476	56653	5880	3395	2395	1026	67269	2082	8577	60774	17061	52290

Table 10-1
Reliability for Total Group and Subgroups

Content	Grade	Total	Gender		Ethnicity					ELP		Disability		SES	
			Female	Male	White	African American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
Reading	3	0.93	0.93	0.93	0.92	0.92	0.92	0.93	0.92	0.93	0.91	0.93	0.92	0.93	0.92
	4	0.91	0.91	0.92	0.90	0.90	0.90	0.92	0.90	0.91	0.89	0.92	0.90	0.91	0.90
	5	0.91	0.91	0.91	0.90	0.90	0.90	0.92	0.89	0.91	0.89	0.90	0.90	0.90	0.90
	6	0.90	0.89	0.90	0.88	0.89	0.89	0.89	0.88	0.89	0.86	0.89	0.87	0.89	0.87
	7	0.90	0.90	0.91	0.89	0.89	0.89	0.91	0.89	0.90	0.88	0.89	0.89	0.90	0.89
	8	0.91	0.90	0.91	0.90	0.90	0.90	0.90	0.90	0.91	0.88	0.89	0.89	0.90	0.89
	10	0.91	0.90	0.91	0.89	0.89	0.90	0.91	0.89	0.90	0.87	0.88	0.89	0.90	0.89
Mathematics	3	0.91	0.90	0.91	0.89	0.91	0.89	0.91	0.89	0.91	0.89	0.91	0.90	0.90	0.89
	4	0.90	0.90	0.90	0.88	0.90	0.88	0.90	0.87	0.90	0.88	0.91	0.89	0.90	0.88
	5	0.90	0.90	0.90	0.89	0.89	0.88	0.90	0.87	0.90	0.88	0.90	0.89	0.89	0.89
	6	0.92	0.92	0.93	0.91	0.90	0.90	0.92	0.91	0.92	0.90	0.91	0.91	0.91	0.91
	7	0.91	0.91	0.92	0.90	0.88	0.88	0.91	0.89	0.91	0.88	0.89	0.90	0.90	0.90
	8	0.91	0.91	0.92	0.91	0.86	0.88	0.92	0.89	0.91	0.87	0.86	0.91	0.89	0.91
	10	0.92	0.92	0.93	0.92	0.87	0.90	0.92	0.90	0.92	0.88	0.87	0.92	0.91	0.92
Language Arts	4	0.82	0.81	0.82	0.81	0.80	0.79	0.82	0.79	0.82	0.78	0.79	0.81	0.80	0.80
	8	0.84	0.83	0.85	0.83	0.82	0.82	0.83	0.82	0.84	0.77	0.81	0.82	0.83	0.83
	10	0.83	0.81	0.83	0.81	0.79	0.80	0.80	0.80	0.82	0.72	0.75	0.80	0.81	0.81
Social Studies	4	0.87	0.86	0.87	0.84	0.87	0.85	0.86	0.84	0.87	0.84	0.87	0.86	0.86	0.84
	8	0.88	0.87	0.89	0.87	0.86	0.86	0.87	0.85	0.88	0.83	0.86	0.86	0.87	0.86
	10	0.91	0.90	0.92	0.90	0.88	0.89	0.89	0.89	0.91	0.84	0.88	0.90	0.89	0.90
Science	4	0.86	0.86	0.86	0.83	0.84	0.83	0.85	0.84	0.86	0.82	0.86	0.85	0.85	0.83
	8	0.87	0.86	0.88	0.85	0.83	0.84	0.86	0.84	0.87	0.80	0.84	0.85	0.85	0.84
	10	0.89	0.87	0.90	0.87	0.84	0.86	0.87	0.85	0.89	0.82	0.85	0.88	0.87	0.87

Table 10-2
Standard Error of Measurement for Total Group and Subgroups

Content	Grade	Total	Gender		Ethnicity					ELP		Disability		SES	
			Female	Male	White	African American	Hispanic	Asian	American Indian	Proficient	Not Proficient	Disabled	Not Disabled	Economically Disadvantaged	Not Economically Disadvantaged
Reading	3	3.02	2.85	3.14	2.98	3.40	3.26	3.11	3.20	2.97	3.34	3.40	2.96	3.14	2.85
	4	3.21	3.14	3.07	3.16	3.43	3.33	3.17	3.22	3.17	3.37	3.29	3.09	3.25	3.05
	5	3.19	3.11	3.23	3.11	3.35	3.38	3.18	3.25	3.13	3.36	3.49	3.04	3.37	3.00
	6	3.10	3.12	3.17	3.07	3.41	3.30	3.34	3.33	3.19	3.49	3.48	3.11	3.36	3.08
	7	3.34	3.25	3.25	3.20	3.39	3.46	3.31	3.38	3.29	3.41	3.38	3.16	3.35	3.12
	8	3.26	3.29	3.35	3.20	3.44	3.47	3.39	3.43	3.22	3.51	3.55	3.18	3.50	3.24
	10	3.10	3.16	3.17	3.14	3.39	3.38	3.24	3.26	3.21	3.35	3.30	3.10	3.33	3.13
Mathematics	3	2.94	3.08	2.95	2.95	3.17	3.11	2.91	3.09	2.93	3.04	3.16	2.97	3.18	2.88
	4	2.89	2.89	2.89	2.88	3.05	3.04	2.93	2.98	2.87	3.07	3.08	2.84	2.97	2.80
	5	3.21	3.17	3.24	3.13	3.23	3.31	3.25	3.31	3.18	3.28	3.30	3.16	3.29	3.06
	6	3.33	3.25	3.18	3.26	3.49	3.46	3.32	3.40	3.31	3.42	3.51	3.26	3.45	3.18
	7	3.32	3.25	3.19	3.24	3.40	3.42	3.35	3.41	3.30	3.43	3.36	3.23	3.33	3.21
	8	3.45	3.40	3.30	3.29	3.51	3.42	3.30	3.41	3.44	3.45	3.49	3.26	3.46	3.26
	10	3.35	3.28	3.20	3.16	3.32	3.26	3.35	3.28	3.33	3.37	3.30	3.17	3.25	3.18
Language Arts	4	2.21	2.18	2.25	2.15	2.36	2.32	2.22	2.32	2.19	2.35	2.41	2.18	2.33	2.16
	8	2.18	2.06	2.21	2.10	2.40	2.31	2.21	2.30	2.16	2.41	2.45	2.07	2.30	2.03
	10	2.30	2.28	2.39	2.30	2.46	2.45	2.43	2.44	2.35	2.55	2.51	2.29	2.43	2.28
Social Studies	4	2.17	2.20	2.21	2.14	2.51	2.41	2.30	2.40	2.15	2.45	2.47	2.14	2.45	2.06
	8	2.46	2.47	2.42	2.31	2.76	2.71	2.48	2.66	2.43	2.77	2.78	2.40	2.65	2.33
	10	2.88	2.89	2.83	2.83	3.14	3.13	2.98	3.10	2.85	3.23	3.17	2.81	3.17	2.81
Science	4	2.48	2.46	2.49	2.44	2.78	2.66	2.55	2.61	2.45	2.70	2.72	2.46	2.67	2.39
	8	2.54	2.54	2.52	2.46	2.80	2.74	2.66	2.71	2.51	2.84	2.87	2.52	2.77	2.49
	10	2.93	3.01	2.90	2.92	3.10	3.08	3.12	3.12	2.90	3.14	3.19	2.87	3.13	2.94

Table 10-3
Cronbach Alpha Reliability Coefficients for Content Standards

Content Area	Grade	Content Standard								
		A/1	B/2	C/3	D/4	E	F	G	H	Total
		Alpha	Alpha	Alpha	Alpha	Alpha	Alpha	Alpha	Alpha	Alpha
Reading	3	0.66	0.84	0.85	0.51					0.93
	4	0.71	0.79	0.79	0.51					0.91
	5	0.68	0.78	0.73	0.69					0.91
	6	0.69	0.69	0.71	0.67					0.90
	7	0.64	0.72	0.77	0.67					0.90
	8	0.68	0.72	0.75	0.72					0.91
	10	0.64	0.57	0.79	0.76					0.91
Mathematics	3	0.64	0.74	0.59	0.65	0.66	0.61			0.91
	4	0.52	0.70	0.51	0.65	0.69	0.69			0.90
	5	0.60	0.66	0.51	0.54	0.74	0.68			0.90
	6	0.62	0.73	0.63	0.67	0.76	0.71			0.92
	7	0.58	0.71	0.66	0.66	0.68	0.66			0.91
	8	0.67	0.66	0.68	0.62	0.66	0.71			0.91
	10	0.59	0.58	0.70	0.76	0.62	0.73			0.92
Language Arts	4		0.74		0.43		0.54			0.82
	8		0.75		0.65		0.48			0.84
	10		0.66		0.68		0.46			0.83
Social Studies	4	0.62	0.59	0.48	0.45	0.67				0.87
	8	0.62	0.72	0.59	0.48	0.57				0.88
	10	0.63	0.73	0.69	0.65	0.66				0.91
Science	4	0.54	0.26	0.53	0.34	0.51	0.51	0.25	0.56	0.86
	8	0.44	0.27	0.61	0.46	0.51	0.44	0.49	0.43	0.87
	10	0.42	0.43	0.65	0.46	0.48	0.49	0.56	0.56	0.89

Table 10-4
Standard Error of Measurement per Content Standard

Content Area	Grade	SEM Per Content Standard								
		A/1	B/2	C/3	D/4	E	F	G	H	Total
Reading	3	1.32	1.49	1.98	1.06					3.02
	4	1.19	1.69	1.99	1.30					3.21
	5	1.41	1.45	1.90	1.49					3.19
	6	1.29	1.45	1.95	1.53					3.10
	7	1.33	1.55	2.03	1.51					3.34
	8	1.31	1.42	2.01	1.71					3.26
	10	1.00	0.99	2.22	1.72					3.10
Mathematics	3	1.35	1.30	1.16	1.10	1.07	1.14			2.94
	4	1.39	1.00	1.13	1.15	1.17	1.14			2.89
	5	1.21	1.24	1.24	1.20	1.37	1.35			3.21
	6	1.35	1.34	1.26	1.32	1.40	1.24			3.33
	7	1.35	1.40	1.50	1.35	1.13	1.13			3.32
	8	1.32	1.31	1.26	1.48	1.25	1.43			3.45
	10	1.32	1.05	1.39	1.47	1.27	1.40			3.35
Language Arts	4		1.73		0.88		1.05			2.21
	8		1.44		1.17		1.07			2.18
	10		1.64		1.25		1.09			2.30
Social Studies	4	1.07	0.99	1.02	0.81	0.98				2.17
	8	1.15	1.40	0.97	0.87	0.94				2.46
	10	1.24	1.46	1.41	1.11	1.19				2.88
Science	4	0.68	0.61	1.00	0.94	1.04	1.02	0.73	0.86	2.48
	8	0.56	0.73	0.99	1.09	1.09	1.01	0.84	0.79	2.54
	10	0.97	0.88	1.17	1.22	0.97	1.13	0.92	0.96	2.93

Table 10-5
Classification Consistency and Classification Accuracy for Reading Grade 3

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.03	0.01	0.00	0.00	0.04
Basic	0.01	0.11	0.03	0.00	0.15
Proficient	0.00	0.03	0.29	0.06	0.37
Advanced	0.00	0.00	0.05	0.39	0.44
Sum	0.04	0.15	0.36	0.44	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.97	0.94	0.89	0.81
Probability of Chance	0.92	0.69	0.51	0.35
Kappa (k)	0.65	0.82	0.79	0.71
Classification Accuracy	0.98	0.96	0.93	0.87

Table 10-6
Classification Consistency and Classification Accuracy for Reading Grade 4

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.03	0.02	0.00	0.00	0.05
Basic	0.02	0.10	0.03	0.00	0.15
Proficient	0.00	0.03	0.31	0.06	0.40
Advanced	0.00	0.00	0.06	0.36	0.41
Sum	0.05	0.15	0.39	0.42	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.97	0.94	0.88	0.79
Probability of Chance	0.91	0.69	0.51	0.35
Kappa (k)	0.59	0.81	0.76	0.68
Classification Accuracy	0.98	0.95	0.92	0.85

Table 10-7
Classification Consistency and Classification Accuracy for Reading Grade 5

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.04	0.02	0.00	0.00	0.07
Basic	0.02	0.10	0.04	0.00	0.16
Proficient	0.00	0.03	0.31	0.05	0.39
Advanced	0.00	0.00	0.05	0.33	0.38
Sum	0.07	0.16	0.40	0.38	

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	0.88	0.66	0.53	0.33
Kappa (k)	0.63	0.79	0.79	0.68
Classification Accuracy	0.97	0.95	0.93	0.85

Table 10-8
Classification Consistency and Classification Accuracy for Reading Grade 6

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.04	0.02	0.00	0.00	0.06
Basic	0.02	0.07	0.03	0.00	0.12
Proficient	0.00	0.03	0.30	0.06	0.39
Advanced	0.00	0.00	0.06	0.36	0.42
Sum	0.06	0.12	0.39	0.42	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.96	0.93	0.88	0.78
Probability of Chance	0.89	0.70	0.51	0.35
Kappa (k)	0.65	0.77	0.76	0.66
Classification Accuracy	0.97	0.95	0.92	0.84

Table 10-9
Classification Consistency and Classification Accuracy for Reading Grade 7

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.05	0.02	0.00	0.00	0.07
Basic	0.02	0.07	0.03	0.00	0.12
Proficient	0.00	0.03	0.32	0.06	0.40
Advanced	0.00	0.00	0.05	0.35	0.41
Sum	0.07	0.12	0.40	0.41	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.96	0.93	0.89	0.79
Probability of Chance	0.87	0.69	0.52	0.35
Kappa (k)	0.68	0.79	0.77	0.67
Classification Accuracy	0.97	0.95	0.92	0.84

Table 10-10
Classification Consistency and Classification Accuracy for Reading Grade 8

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.05	0.02	0.00	0.00	0.07
Basic	0.02	0.07	0.03	0.00	0.12
Proficient	0.00	0.03	0.31	0.05	0.40
Advanced	0.00	0.00	0.05	0.36	0.41
Sum	0.07	0.12	0.40	0.41	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.95	0.94	0.89	0.79
Probability of Chance	0.87	0.69	0.52	0.35
Kappa (k)	0.66	0.79	0.78	0.67
Classification Accuracy	0.97	0.95	0.93	0.85

Table 10-11
Classification Consistency and Classification Accuracy for Reading Grade 10

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.09	0.03	0.00	0.00	0.12
Basic	0.03	0.11	0.04	0.00	0.18
Proficient	0.00	0.05	0.22	0.05	0.31
Advanced	0.00	0.00	0.05	0.35	0.40
Sum	0.12	0.18	0.31	0.39	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.94	0.91	0.90	0.76
Probability of Chance	0.79	0.58	0.52	0.30
Kappa (k)	0.73	0.78	0.79	0.65
Classification Accuracy	0.96	0.94	0.93	0.83

Table 10-12
Classification Consistency and Classification Accuracy for Mathematics Grade 3

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.14	0.03	0.01	0.00	0.17
Basic	0.02	0.03	0.03	0.00	0.09
Proficient	0.01	0.03	0.27	0.06	0.37
Advanced	0.00	0.00	0.06	0.32	0.37
Sum	0.17	0.09	0.37	0.37	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.93	0.92	0.89	0.76
Probability of Chance	0.72	0.62	0.53	0.31
Kappa (k)	0.76	0.78	0.76	0.65
Classification Accuracy	0.95	0.94	0.92	0.82

Table 10-13
Classification Consistency and Classification Accuracy for Mathematics Grade 4

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.15	0.03	0.01	0.00	0.19
Basic	0.03	0.04	0.03	0.00	0.10
Proficient	0.01	0.04	0.29	0.05	0.38
Advanced	0.00	0.00	0.05	0.27	0.33
Sum	0.19	0.10	0.38	0.33	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.93	0.91	0.89	0.75
Probability of Chance	0.70	0.59	0.56	0.30
Kappa (k)	0.76	0.79	0.76	0.64
Classification Accuracy	0.95	0.94	0.93	0.81

Table 10-14
Classification Consistency and Classification Accuracy for Mathematics Grade 5

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.17	0.03	0.01	0.00	0.21
Basic	0.03	0.05	0.04	0.00	0.12
Proficient	0.01	0.04	0.26	0.05	0.35
Advanced	0.00	0.00	0.05	0.27	0.32
Sum	0.21	0.12	0.35	0.32	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.92	0.91	0.90	0.74
Probability of Chance	0.67	0.56	0.56	0.29
Kappa (k)	0.76	0.79	0.77	0.64
Classification Accuracy	0.94	0.93	0.93	0.81

Table 10-15
Classification Consistency and Classification Accuracy for Mathematics Grade 6

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.10	0.02	0.00	0.00	0.13
Basic	0.02	0.05	0.03	0.00	0.11
Proficient	0.00	0.03	0.31	0.05	0.39
Advanced	0.00	0.00	0.05	0.33	0.37
Sum	0.13	0.11	0.39	0.37	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.95	0.93	0.91	0.79
Probability of Chance	0.78	0.64	0.53	0.32
Kappa (k)	0.76	0.80	0.80	0.69
Classification Accuracy	0.96	0.95	0.93	0.85

Table 10-16
Classification Consistency and Classification Accuracy for Mathematics Grade 7

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.10	0.03	0.00	0.00	0.13
Basic	0.03	0.06	0.04	0.00	0.13
Proficient	0.00	0.04	0.35	0.04	0.43
Advanced	0.00	0.00	0.04	0.26	0.30
Sum	0.13	0.13	0.44	0.30	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.94	0.91	0.91	0.77
Probability of Chance	0.77	0.61	0.58	0.32
Kappa (k)	0.72	0.77	0.79	0.66
Classification Accuracy	0.95	0.94	0.94	0.83

Table 10-17
Classification Consistency and Classification Accuracy for Mathematics Grade 8

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.11	0.04	0.00	0.00	0.15
Basic	0.04	0.08	0.05	0.00	0.16
Proficient	0.00	0.04	0.36	0.04	0.44
Advanced	0.00	0.00	0.03	0.22	0.25
Sum	0.15	0.15	0.44	0.25	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.92	0.90	0.93	0.76
Probability of Chance	0.74	0.57	0.63	0.30
Kappa (k)	0.69	0.77	0.82	0.66
Classification Accuracy	0.94	0.93	0.95	0.83

Table 10-18
Classification Consistency and Classification Accuracy for Mathematics Grade 10

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.16	0.04	0.00	0.00	0.20
Basic	0.04	0.07	0.04	0.00	0.15
Proficient	0.01	0.04	0.32	0.03	0.40
Advanced	0.00	0.00	0.03	0.21	0.25
Sum	0.20	0.15	0.40	0.24	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.92	0.90	0.94	0.77
Probability of Chance	0.68	0.54	0.63	0.29
Kappa (k)	0.74	0.79	0.83	0.67
Classification Accuracy	0.94	0.93	0.95	0.83

Table 10-19
Classification Consistency and Classification Accuracy for Language Arts Grade 4

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.05	0.03	0.00	0.00	0.08
Basic	0.03	0.11	0.06	0.00	0.19
Proficient	0.00	0.06	0.29	0.07	0.42
Advanced	0.00	0.00	0.07	0.24	0.31
Sum	0.08	0.19	0.42	0.31	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.94	0.88	0.87	0.70
Probability of Chance	0.85	0.60	0.57	0.31
Kappa (k)	0.61	0.70	0.69	0.56
Classification Accuracy	0.96	0.91	0.90	0.78

Table 10-20
Classification Consistency and Classification Accuracy for Language Arts Grade 8

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.11	0.03	0.00	0.00	0.14
Basic	0.03	0.10	0.05	0.00	0.19
Proficient	0.00	0.05	0.20	0.07	0.32
Advanced	0.00	0.00	0.07	0.28	0.35
Sum	0.14	0.19	0.32	0.35	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.93	0.89	0.86	0.69
Probability of Chance	0.76	0.56	0.55	0.28
Kappa (k)	0.73	0.75	0.70	0.57
Classification Accuracy	0.95	0.92	0.91	0.78

Table 10-21
Classification Consistency and Classification Accuracy for Language Arts Grade 10

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.08	0.03	0.00	0.00	0.11
Basic	0.03	0.12	0.06	0.00	0.21
Proficient	0.00	0.06	0.37	0.05	0.49
Advanced	0.00	0.00	0.05	0.14	0.19
Sum	0.11	0.21	0.49	0.19	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.93	0.87	0.89	0.70
Probability of Chance	0.80	0.56	0.69	0.33
Kappa (k)	0.64	0.70	0.65	0.55
Classification Accuracy	0.95	0.91	0.93	0.78

Table 10-22
Classification Consistency and Classification Accuracy for Social Studies Grade 4

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.02	0.01	0.00	0.00	0.03
Basic	0.01	0.05	0.03	0.00	0.09
Proficient	0.00	0.03	0.19	0.06	0.28
Advanced	0.00	0.00	0.06	0.54	0.60
Sum	0.03	0.09	0.28	0.60	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.98	0.94	0.88	0.80
Probability of Chance	0.94	0.79	0.52	0.45
Kappa (k)	0.61	0.73	0.76	0.65
Classification Accuracy	0.98	0.96	0.92	0.86

Table 10-23
Classification Consistency and Classification Accuracy for Social Studies Grade 8

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.05	0.02	0.00	0.00	0.08
Basic	0.02	0.09	0.04	0.00	0.15
Proficient	0.00	0.04	0.29	0.06	0.38
Advanced	0.00	0.00	0.06	0.33	0.39
Sum	0.08	0.15	0.38	0.39	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.95	0.92	0.89	0.76
Probability of Chance	0.86	0.65	0.53	0.33
Kappa (k)	0.66	0.76	0.76	0.64
Classification Accuracy	0.97	0.94	0.92	0.82

Table 10-24
Classification Consistency and Classification Accuracy for Social Studies Grade 10

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.17	0.03	0.02	0.00	0.21
Basic	0.03	0.02	0.03	0.00	0.08
Proficient	0.01	0.03	0.22	0.05	0.30
Advanced	0.00	0.00	0.05	0.36	0.41
Sum	0.21	0.07	0.31	0.41	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.92	0.91	0.91	0.77
Probability of Chance	0.67	0.59	0.52	0.31
Kappa (k)	0.75	0.78	0.81	0.66
Classification Accuracy	0.94	0.94	0.94	0.83

Table 10-25
Classification Consistency and Classification Accuracy for Science Grade 4

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.04	0.03	0.00	0.00	0.07
Basic	0.02	0.12	0.05	0.00	0.20
Proficient	0.00	0.05	0.39	0.06	0.50
Advanced	0.00	0.00	0.06	0.17	0.23
Sum	0.07	0.20	0.50	0.23	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.95	0.89	0.89	0.73
Probability of Chance	0.87	0.60	0.65	0.35
Kappa (k)	0.61	0.73	0.68	0.59
Classification Accuracy	0.96	0.92	0.92	0.81

Table 10-26
Classification Consistency and Classification Accuracy for Science Grade 8

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.11	0.03	0.00	0.00	0.15
Basic	0.04	0.09	0.05	0.00	0.17
Proficient	0.00	0.05	0.30	0.05	0.41
Advanced	0.00	0.00	0.06	0.22	0.27
Sum	0.15	0.17	0.41	0.27	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.92	0.89	0.89	0.71
Probability of Chance	0.75	0.57	0.60	0.29
Kappa (k)	0.69	0.75	0.73	0.59
Classification Accuracy	0.94	0.92	0.93	0.79

Table 10-27
Classification Consistency and Classification Accuracy for Science Grade 10

Contingency Table with All Cut Scores

	Minimal	Basic	Proficient	Advanced	Sum
Minimal	0.17	0.04	0.01	0.00	0.21
Basic	0.03	0.05	0.05	0.00	0.13
Proficient	0.01	0.04	0.23	0.05	0.33
Advanced	0.00	0.00	0.05	0.28	0.33
Sum	0.21	0.13	0.33	0.33	

Indexes for Classification Consistency and Classification Accuracy

	Cut 1	Cut 2	Cut 3	All Cuts
Classification Consistency (P)	0.91	0.90	0.90	0.73
Probability of Chance	0.66	0.55	0.56	0.28
Kappa (k)	0.74	0.77	0.77	0.62
Classification Accuracy	0.94	0.92	0.93	0.79

Table 10-28
Inter-Rater Reliability Reading

				Percentage Absolute Difference								Frequency			
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra. Corr.	Weighted Kappa	Mean	N	0	1	2	3
3	A	31	3	71	23	1	5	0.89	0.78	1.33	1514	236	586	652	40
3	B	31	3	73	22	1	4	0.89	0.77	1.36	1500	204	601	653	42
3	C	31	3	72	23	1	5	0.89	0.77	1.33	1524	217	624	644	39
3	D	31	3	73	22	1	4	0.90	0.80	1.34	1452	236	532	640	44
3	A	48	3	68	26	1	5	0.88	0.75	1.43	1514	215	493	754	52
3	B	48	3	67	27	2	5	0.88	0.75	1.40	1500	240	471	739	50
3	C	48	3	67	27	1	5	0.88	0.77	1.36	1524	274	472	727	51
3	D	48	3	65	30	0	4	0.88	0.75	1.43	1452	220	436	742	54
3	A	63	3	66	24	2	8	0.90	0.79	0.90	682	234	312	107	29
3	B	63	3	66	27	2	5	0.89	0.78	0.91	732	266	299	133	34
3	C	67	3	71	17	2	10	0.95	0.89	1.20	772	203	262	259	48
3	D	67	3	72	20	1	7	0.94	0.88	1.24	722	185	254	209	74
4	A	14	3	66	26	2	5	0.87	0.75	1.07	1528	377	732	355	64
4	B	14	3	71	22	2	5	0.88	0.76	1.07	1476	344	737	348	47
4	C	14	3	69	24	2	5	0.89	0.77	1.16	1408	293	661	385	69
4	D	14	3	70	24	2	5	0.89	0.78	1.13	1504	326	721	388	69
4	A	33	3	82	11	1	6	0.95	0.89	0.66	1528	798	498	186	46
4	B	33	3	86	9	1	4	0.96	0.92	0.70	1476	754	459	219	44
4	C	33	3	84	11	1	4	0.96	0.92	0.74	1408	688	450	225	45
4	D	33	3	84	9	1	5	0.96	0.91	0.73	1504	723	507	238	36
4	A	62	3	64	25	3	8	0.86	0.71	0.82	1302	514	525	250	13
4	B	62	3	71	21	2	6	0.90	0.80	1.17	1294	283	543	430	38
4	C	66	3	67	28	2	3	0.87	0.73	0.89	706	254	290	150	12
4	D	66	3	67	25	1	6	0.89	0.78	0.88	712	266	287	141	18
5	A	13	3	65	28	2	5	0.87	0.73	1.46	1638	197	618	700	123
5	B	13	3	63	30	4	4	0.86	0.71	1.46	1484	176	580	598	130
5	C	13	3	65	29	2	4	0.87	0.75	1.51	1460	163	525	631	141
5	D	13	3	64	31	1	4	0.86	0.72	1.55	1546	147	545	704	150

Table 10-28 Cont'd
Inter-Rater Reliability Reading

				Percentage Absolute Difference								Frequency			
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra. Corr.	Weighted Kappa	Mean	N	0	1	2	3
5	A	48	3	63	31	1	4	0.86	0.72	1.33	1638	228	728	589	93
5	B	48	3	66	30	1	4	0.87	0.73	1.31	1484	212	681	513	78
5	C	48	3	64	31	1	4	0.86	0.71	1.35	1460	190	655	531	84
5	D	48	3	66	29	1	3	0.86	0.72	1.40	1546	171	690	583	102
5	A	71	3	69	26	1	4	0.92	0.83	1.11	644	210	193	199	42
5	B	71	3	68	26	1	5	0.92	0.83	1.21	638	188	182	216	52
5	C	71	3	63	28	4	5	0.84	0.68	0.95	668	241	233	180	14
5	D	71	3	62	30	3	4	0.86	0.72	0.97	620	225	210	167	18
6	A	19	3	61	32	2	5	0.81	0.61	1.11	1518	285	804	402	27
6	B	19	3	60	35	3	2	0.78	0.56	1.23	1632	215	877	483	57
6	C	19	3	61	34	2	3	0.81	0.61	1.19	1536	235	814	450	37
6	D	19	3	57	36	3	4	0.77	0.54	1.20	1486	221	790	439	36
6	A	27	3	70	23	2	5	0.89	0.78	0.83	1518	630	548	304	36
6	B	27	3	68	27	3	2	0.86	0.72	0.87	1632	637	625	323	47
6	C	27	3	66	27	2	4	0.87	0.74	0.81	1536	655	551	294	36
6	D	27	3	65	28	2	5	0.87	0.74	0.90	1486	551	570	327	38
6	A	63	3	66	26	0	8	0.91	0.82	1.22	632	137	262	189	44
6	B	63	3	60	34	2	4	0.87	0.73	1.27	714	156	251	266	41
6	C	67	3	64	25	2	9	0.91	0.81	0.91	1328	522	486	235	85
6	D	67	3	64	26	1	8	0.92	0.83	0.99	1204	446	431	222	105
7	A	37	3	69	23	3	5	0.92	0.85	1.42	1624	466	251	670	237
7	B	37	3	71	22	3	3	0.92	0.84	1.45	1646	446	248	717	235
7	C	37	3	71	24	2	4	0.93	0.86	1.53	1638	401	230	744	263
7	D	37	3	72	22	3	3	0.93	0.85	1.47	1550	405	236	685	224
7	A	56	3	68	22	3	6	0.87	0.75	0.73	1624	799	520	257	48
7	B	56	3	67	26	2	5	0.87	0.74	0.76	1646	762	554	299	31
7	C	56	3	69	26	2	4	0.89	0.79	0.78	1638	769	521	286	62
7	D	56	3	69	25	2	4	0.88	0.76	0.78	1550	687	560	255	48

Table 10-28 Cont'd
Inter-Rater Reliability Reading

				Percentage Absolute Difference								Frequency			
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra. Corr.	Weighted Kappa	Mean	N	0	1	2	3
7	A	72	3	65	27	1	7	0.88	0.77	0.95	1134	361	495	252	26
7	B	72	3	68	28	0	4	0.91	0.81	1.16	648	159	269	175	45
7	C	72	3	60	32	1	7	0.88	0.75	0.99	1170	361	497	271	41
7	D	72	3	74	23	0	3	0.91	0.82	1.14	598	140	265	160	33
8	A	31	3	64	29	1	6	0.91	0.82	1.14	1594	518	460	488	128
8	B	31	3	65	28	2	5	0.90	0.80	1.10	1756	599	478	578	101
8	C	31	3	63	30	2	4	0.89	0.79	1.16	1714	563	446	580	125
8	D	31	3	66	28	2	4	0.90	0.81	1.10	1706	584	500	498	124
8	A	56	3	63	30	2	5	0.90	0.79	1.58	1594	274	349	745	226
8	B	56	3	66	29	1	4	0.91	0.82	1.60	1756	300	381	806	269
8	C	56	3	66	28	2	4	0.90	0.80	1.60	1714	271	379	833	231
8	D	56	3	65	30	2	4	0.89	0.79	1.57	1706	281	397	798	230
8	A	63	3	64	23	2	11	0.89	0.78	1.12	1338	322	574	401	41
8	B	63	3	68	24	1	7	0.92	0.84	1.47	1314	245	347	587	135
8	C	67	3	69	19	1	11	0.89	0.78	0.92	1340	383	715	207	35
8	D	67	3	65	23	3	9	0.91	0.82	1.13	1286	422	386	363	115
10	A	31	3	62	30	1	6	0.89	0.78	1.74	6612	629	1,841	2,785	1,357
10	A	52	3	61	30	2	7	0.88	0.76	1.35	6612	1,066	2,949	1,787	810

Table 10-29
Inter-Rater Reliability Mathematics

				Percentage Absolute Difference								Frequency		
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra Corr.	Weighted Kappa	Mean	N	0	1	2
3	A	19A	1	94	2	1	3	0.97	0.94	0.66	1514	510	1,004	0
3	B	19A	1	93	2	1	4	0.98	0.96	0.64	1500	539	961	0
3	C	19A	1	95	1	0	3	0.99	0.97	0.64	1524	544	980	0
3	D	19A	1	93	2	0	4	0.97	0.95	0.67	1452	487	965	0
3	A	19B	2	78	14	3	4	0.89	0.78	0.69	1514	675	630	209
3	B	19B	2	77	15	2	6	0.89	0.79	0.69	1500	681	603	216
3	C	19B	2	78	16	2	5	0.88	0.77	0.68	1524	674	662	188
3	D	19B	2	73	18	3	6	0.85	0.71	0.72	1452	617	626	209
3	A	26A	1	96	1	0	3	0.99	0.98	0.71	1514	437	1,077	0
3	B	26A	1	95	2	0	3	0.98	0.96	0.71	1500	442	1,058	0
3	C	26A	1	94	2	0	3	0.98	0.95	0.71	1524	443	1,081	0
3	D	26A	1	96	1	0	3	0.99	0.98	0.74	1452	383	1,069	0
3	A	26B	2	75	14	5	6	0.91	0.81	0.64	1514	943	178	393
3	B	26B	2	77	14	5	4	0.90	0.80	0.62	1500	952	169	379
3	C	26B	2	77	14	5	4	0.90	0.80	0.63	1524	959	171	394
3	D	26B	2	76	15	4	5	0.91	0.81	0.67	1452	880	178	394
3	A	42A	1	95	2	0	3	0.98	0.96	0.68	1514	486	1,028	0
3	B	42A	1	95	1	0	3	0.98	0.97	0.69	1500	471	1,029	0
3	C	42A	1	96	2	0	2	0.98	0.96	0.68	1524	494	1,030	0
3	D	42A	1	96	1	0	3	0.99	0.97	0.72	1452	413	1,039	0
3	A	42B	2	80	12	3	4	0.93	0.87	1.04	1514	614	229	671
3	B	42B	2	80	12	3	5	0.94	0.88	1.03	1500	624	213	663
3	C	42B	2	81	12	2	4	0.94	0.88	0.97	1524	682	203	639
3	D	42B	2	80	12	5	3	0.91	0.83	1.11	1452	551	192	709
3	A	52A	1	92	1	0	6	0.99	0.98	0.54	1274	585	689	0
3	B	52A	1	94	1	0	5	0.99	0.99	0.46	1296	697	599	0
3	A	52B	2	80	8	2	11	0.96	0.93	0.78	1274	668	220	386
3	B	52B	2	78	14	3	5	0.93	0.86	0.71	1296	728	217	351

Table 10-29 Cont'd
Inter-Rater Reliability Mathematics

				Percentage Absolute Difference								Frequency		
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra Corr.	Weighted Kappa	Mean	N	0	1	2
3	C	58A	1	91	3	1	5	0.97	0.94	0.36	1316	841	475	0
3	D	58A	1	95	0	0	5	0.99	0.98	0.89	1242	132	1,110	0
3	C	58B	2	74	17	1	8	0.92	0.83	0.75	1316	548	554	214
3	D	58B	2	87	7	0	5	0.97	0.94	1.30	1242	328	210	704
3	A	8A	1	96	1	0	3	0.99	0.98	0.49	1514	775	739	0
3	B	8A	1	95	2	0	3	0.98	0.97	0.51	1500	734	766	0
3	C	8A	1	95	2	0	3	0.98	0.96	0.51	1524	755	769	0
3	D	8A	1	94	1	0	4	0.99	0.98	0.51	1452	719	733	0
3	A	8B	2	74	18	2	6	0.90	0.79	0.69	1514	716	545	253
3	B	8B	2	75	18	2	5	0.89	0.78	0.70	1500	686	574	240
3	C	8B	2	74	19	2	5	0.89	0.78	0.71	1524	706	557	261
3	D	8B	2	74	19	1	5	0.90	0.80	0.77	1452	606	574	272
4	A	12A	1	95	1	0	4	0.99	0.97	0.60	1528	616	912	0
4	B	12A	1	97	1	0	2	0.99	0.98	0.64	1476	529	947	0
4	C	12A	1	97	1	0	2	0.99	0.99	0.65	1408	499	909	0
4	D	12A	1	96	1	0	3	0.99	0.97	0.64	1504	535	969	0
4	A	12B	2	83	10	2	5	0.95	0.90	0.77	1528	730	425	373
4	B	12B	2	85	9	2	3	0.95	0.89	0.84	1476	650	406	420
4	C	12B	2	87	9	2	3	0.96	0.91	0.91	1408	592	346	470
4	D	12B	2	86	11	0	3	0.96	0.91	0.91	1504	599	435	470
4	A	18A	1	95	2	0	3	0.98	0.96	0.48	1528	802	726	0
4	B	18A	1	97	1	0	2	1.00	0.99	0.51	1476	727	749	0
4	C	18A	1	96	1	0	2	0.99	0.98	0.51	1408	692	716	0
4	D	18A	1	97	1	0	3	1.00	0.99	0.51	1504	734	770	0
4	A	18B	2	84	8	5	3	0.90	0.81	0.50	1528	1,099	90	339
4	B	18B	2	84	9	5	3	0.90	0.81	0.57	1476	1,008	101	367

Table 10-29 Cont'd
Inter-Rater Reliability Mathematics

				Percentage Absolute Difference								Frequency		
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra Corr.	Weighted Kappa	Mean	N	0	1	2
4	C	18B	2	85	6	5	3	0.91	0.82	0.54	1408	995	72	341
4	D	18B	2	84	8	5	3	0.91	0.81	0.56	1504	1,041	90	373
4	A	26A	1	96	1	0	3	0.99	0.98	0.45	1528	840	688	0
4	B	26A	1	97	1	0	2	0.99	0.98	0.45	1476	808	668	0
4	C	26A	1	96	1	0	2	0.99	0.97	0.46	1408	768	640	0
4	D	26A	1	96	1	0	3	0.99	0.98	0.47	1504	804	700	0
4	A	26B	2	76	17	2	5	0.86	0.72	0.48	1528	893	536	99
4	B	26B	2	75	21	1	4	0.87	0.74	0.55	1476	785	568	123
4	C	26B	2	73	20	1	5	0.85	0.71	0.51	1408	791	520	97
4	D	26B	2	75	19	1	4	0.85	0.71	0.52	1504	840	548	116
4	A	47A	1	97	1	0	2	0.99	0.98	0.75	1528	387	1,141	0
4	B	47A	1	97	1	0	2	1.00	0.99	0.75	1476	374	1,102	0
4	C	47A	1	97	1	0	2	1.00	0.98	0.76	1408	343	1,065	0
4	D	47A	1	97	1	0	2	0.99	0.99	0.78	1504	334	1,170	0
4	A	47B	2	80	15	2	4	0.91	0.82	0.52	1528	998	264	266
4	B	47B	2	78	17	2	3	0.91	0.82	0.58	1476	899	293	284
4	C	47B	2	81	13	3	3	0.92	0.83	0.58	1408	879	245	284
4	D	47B	2	78	17	2	3	0.90	0.80	0.58	1504	929	282	293
4	A	52A	1	95	1	0	4	0.99	0.99	0.56	1302	573	729	0
4	B	52A	1	97	1	0	3	1.00	0.99	0.52	1294	624	670	0
4	C	52A*	1	0	0	0	0	0.00	0.00	0.00	1260	1,260	0	0
4	A	52B	2	84	9	1	6	0.96	0.92	1.00	1302	517	274	511
4	B	52B	2	76	19	2	4	0.91	0.82	0.70	1294	639	406	249
4	C	52B*	2	0	0	0	0	0.00	0.00	0.00	1260	1,260	0	0
4	D	58A	1	95	1	0	4	1.00	0.99	0.58	1312	546	766	0
4	D	58B	2	75	19	1	5	0.92	0.83	0.85	1312	490	528	294
5	A	10A	1	96	2	0	2	0.98	0.97	0.51	1638	807	831	0

*This item was suppressed during range finding, so there are no inter-rater reliability statistics to report for this item.

Table 10-29 Cont'd
Inter-Rater Reliability Mathematics*

				Percentage Absolute Difference								Frequency		
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra Corr.	Weighted Kappa	Mean	N	0	1	2
5	B	10A	1	96	2	0	3	0.98	0.97	0.51	1484	726	758	0
5	C	10A	1	96	2	0	2	0.99	0.97	0.53	1460	692	768	0
5	D	10A	1	95	2	0	2	0.98	0.96	0.56	1546	677	869	0
5	A	10B	2	77	20	1	2	0.89	0.78	1.01	1638	458	714	466
5	B	10B	2	74	21	2	3	0.88	0.75	1.05	1484	364	687	433
5	C	10B	2	78	19	1	3	0.91	0.81	1.07	1460	362	632	466
5	D	10B	2	76	20	1	3	0.90	0.79	1.07	1546	384	675	487
5	A	19A	1	97	0	0	3	1.00	1.00	0.83	1638	275	1,363	0
5	B	19A	1	97	0	0	3	1.00	0.99	0.81	1484	276	1,208	0
5	C	19A	1	96	1	0	3	0.99	0.97	0.84	1460	241	1,219	0
5	D	19A	1	97	1	0	2	0.99	0.98	0.83	1546	269	1,277	0
5	A	19B	2	84	10	1	5	0.93	0.86	0.43	1638	1,127	315	196
5	B	19B	2	83	10	1	5	0.93	0.86	0.43	1484	1,019	291	174
5	C	19B	2	84	12	1	4	0.94	0.87	0.48	1460	961	302	197
5	D	19B	2	83	13	1	4	0.93	0.85	0.43	1546	1,054	321	171
5	A	29A	1	96	1	0	3	0.98	0.96	0.21	1638	1,302	336	0
5	B	29A	1	96	1	0	3	0.99	0.97	0.23	1484	1,149	335	0
5	C	29A	1	96	1	0	3	0.98	0.97	0.20	1460	1,168	292	0
5	D	29A	1	96	1	0	2	0.98	0.97	0.23	1546	1,185	361	0
5	A	29B	2	85	10	0	4	0.93	0.86	1.00	1638	317	1,008	313
5	B	29B	2	85	11	1	4	0.92	0.84	1.05	1484	250	905	329
5	C	29B	2	85	11	0	4	0.92	0.84	1.03	1460	243	926	291
5	D	29B	2	88	8	1	3	0.93	0.86	1.09	1546	226	951	369
5	A	41A	1	96	0	0	3	1.00	0.99	0.74	1638	430	1,208	0
5	B	41A	1	97	1	0	2	0.98	0.97	0.77	1484	348	1,136	0
5	C	41A	1	96	1	0	3	0.99	0.98	0.76	1460	352	1,108	0

Table 10-29 Cont'd
Inter-Rater Reliability Mathematics

				Percentage Absolute Difference								Frequency		
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra Corr.	Weighted Kappa	Mean	N	0	1	2
5	D	41A	1	97	1	0	2	0.98	0.97	0.78	1546	341	1,205	0
5	A	41B	2	72	24	1	3	0.89	0.77	0.75	1638	721	608	309
5	B	41B	2	70	25	1	4	0.88	0.77	0.81	1484	601	569	314
5	C	41B	2	72	24	1	4	0.89	0.78	0.82	1460	583	558	319
5	D	41B	2	72	23	2	2	0.89	0.77	0.85	1546	621	544	381
5	A	57A	1	95	2	0	3	0.98	0.96	0.38	1284	797	487	0
5	B	57A	1	95	1	0	4	0.99	0.98	0.67	1236	413	823	0
5	C	57A	1	94	1	0	4	0.99	0.98	0.46	1270	690	580	0
5	A	57B	2	84	10	1	4	0.95	0.90	0.83	1284	543	412	329
5	B	57B	2	78	13	4	6	0.91	0.82	0.59	1236	805	131	300
5	C	57B	2	86	7	1	6	0.98	0.95	0.99	1270	570	141	559
5	D	63A	1	94	1	0	5	0.99	0.98	0.49	1228	633	595	0
5	D	63B	2	82	10	1	7	0.96	0.91	0.67	1228	738	164	326
6	A	11A	1	94	2	0	4	0.98	0.97	0.48	1518	793	725	0
6	B	11A	1	95	2	0	2	0.98	0.97	0.47	1632	866	766	0
6	C	11A	1	94	2	1	3	0.98	0.97	0.46	1536	830	706	0
6	D	11A	1	93	2	1	4	0.98	0.96	0.48	1486	770	716	0
6	A	11B	2	82	11	3	5	0.95	0.89	1.02	1518	631	228	659
6	B	11B	2	84	11	2	3	0.95	0.89	1.09	1632	609	263	760
6	C	11B	2	82	11	2	5	0.95	0.90	1.02	1536	657	197	682
6	D	11B	2	83	10	2	5	0.96	0.92	1.09	1486	574	208	704
6	A	22A	1	97	1	0	3	0.99	0.98	0.84	1518	248	1,270	0
6	B	22A	1	98	1	0	2	0.99	0.98	0.84	1632	261	1,371	0
6	C	22A	1	96	1	0	3	0.99	0.98	0.87	1536	208	1,328	0
6	D	22A	1	97	0	0	3	0.99	0.99	0.86	1486	204	1,282	0
6	A	22B	2	81	13	2	4	0.90	0.79	0.55	1518	816	565	137

Table 10-29 Cont'd
Inter-Rater Reliability Mathematics

				Percentage Absolute Difference								Frequency		
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra Corr.	Weighted Kappa	Mean	N	0	1	2
6	B	22B	2	84	12	2	3	0.90	0.79	0.57	1632	866	610	156
6	C	22B	2	80	14	1	4	0.89	0.78	0.57	1536	791	618	127
6	D	22B	2	79	15	2	4	0.90	0.80	0.58	1486	782	552	152
6	A	33A	1	95	1	0	3	0.99	0.97	0.65	1518	538	980	0
6	B	33A	1	97	1	0	2	1.00	0.99	0.65	1632	569	1,063	0
6	C	33A	1	96	1	0	3	0.99	0.98	0.66	1536	524	1,012	0
6	D	33A	1	95	2	0	3	0.98	0.97	0.65	1486	520	966	0
6	A	33B	2	82	13	1	4	0.94	0.89	0.92	1518	544	557	417
6	B	33B	2	84	13	1	2	0.94	0.89	0.93	1632	568	606	458
6	C	33B	2	84	12	1	4	0.95	0.90	0.95	1536	508	597	431
6	D	33B	2	82	13	1	4	0.94	0.88	0.93	1486	506	577	403
6	A	46A	1	95	1	1	4	0.99	0.98	0.21	1518	1,195	323	0
6	B	46A	1	97	0	0	2	0.99	0.99	0.23	1632	1,251	381	0
6	C	46A	1	96	1	0	3	0.99	0.97	0.23	1536	1,186	350	0
6	D	46A	1	96	0	0	4	1.00	1.00	0.25	1486	1,111	375	0
6	A	46B	2	88	6	1	4	0.97	0.94	0.74	1518	762	397	359
6	B	46B	2	90	6	1	3	0.97	0.94	0.79	1632	749	474	409
6	C	46B	2	89	5	2	4	0.97	0.93	0.79	1536	704	455	377
6	D	46B	2	89	6	1	4	0.97	0.95	0.80	1486	690	402	394
6	A	56A	1	94	1	0	5	0.99	0.98	0.19	1264	1,021	243	0
6	B	56A	1	94	1	0	5	0.99	0.98	0.40	1348	805	543	0
6	A	56B	2	84	9	0	6	0.95	0.89	0.59	1264	676	429	159
6	B	56B	2	71	21	2	7	0.89	0.78	0.69	1348	656	450	242
6	C	63A	1	92	1	0	6	0.99	0.97	0.36	1328	846	482	0
6	D	63A	1	93	1	0	6	0.99	0.99	0.58	1204	504	700	0
6	C	63B	2	81	11	1	8	0.96	0.92	0.75	1328	730	202	396

Table 10-29 Cont'd
Inter-Rater Reliability Mathematics

				Percentage Absolute Difference								Frequency		
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra Corr.	Weighted Kappa	Mean	N	0	1	2
6	D	63B	2	83	8	1	8	0.97	0.93	0.77	1204	689	98	417
7	A	19A	1	95	1	0	4	0.99	0.98	0.49	1624	821	803	0
7	B	19A	1	96	1	0	3	1.00	0.99	0.51	1646	807	839	0
7	C	19A	1	96	1	0	3	0.99	0.98	0.51	1638	805	833	0
7	D	19A	1	95	2	0	3	0.98	0.97	0.51	1550	754	796	0
7	A	19B	2	87	5	2	6	0.97	0.93	0.52	1624	951	498	175
7	B	19B	2	90	4	1	4	0.97	0.93	0.50	1646	982	506	158
7	C	19B	2	90	5	1	5	0.96	0.92	0.56	1638	889	589	160
7	D	19B	2	87	6	1	5	0.96	0.93	0.56	1550	863	509	178
7	A	31A	1	95	1	0	4	1.00	0.99	0.50	1624	809	815	0
7	B	31A	1	96	1	0	3	0.99	0.99	0.53	1646	770	876	0
7	C	31A	1	96	1	0	3	0.99	0.98	0.56	1638	724	914	0
7	D	31A	1	96	1	0	3	0.99	0.98	0.54	1550	712	838	0
7	A	31B	2	90	3	1	5	0.99	0.97	1.14	1624	626	148	850
7	B	31B	2	90	4	1	5	0.99	0.97	1.18	1646	624	103	919
7	C	31B	2	91	3	1	4	0.98	0.97	1.25	1638	556	120	962
7	D	31B	2	91	4	1	5	0.99	0.97	1.23	1550	543	104	903
7	A	47A	1	93	2	0	4	0.97	0.95	0.38	1624	1,014	610	0
7	B	47A	1	93	4	0	3	0.97	0.93	0.41	1646	969	677	0
7	C	47A	1	94	3	0	3	0.97	0.93	0.41	1638	975	663	0
7	D	47A	1	93	4	0	3	0.96	0.92	0.40	1550	931	619	0
7	A	47B	2	79	15	1	6	0.89	0.79	0.64	1624	696	817	111
7	B	47B	2	78	17	0	4	0.88	0.75	0.70	1646	615	905	126
7	C	47B	2	78	18	1	4	0.86	0.72	0.69	1638	604	946	88
7	D	47B	2	78	17	1	4	0.87	0.73	0.70	1550	571	880	99
7	B	56A*	1	0	0	0	0	0.00	0.00	0.00	1320	1,320	0	0

*This item was suppressed during range finding, so there are no inter-rater reliability statistics to report for this item.

Table 10-29 Cont'd
Inter-Rater Reliability Mathematics

				Percentage Absolute Difference								Frequency		
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra Corr.	Weighted Kappa	Mean	N	0	1	2
7	C	56A	1	90	2	0	8	0.97	0.94	0.69	1408	439	969	0
7	D	56A	1	87	3	1	8	0.93	0.87	0.14	1276	1,100	176	0
7	B	56B*	2	0	0	0	0	0.00	0.00	0.00	1320	1,320	0	0
7	C	56B	2	76	13	2	9	0.94	0.88	1.21	1408	460	195	753
7	D	56B	2	81	7	1	11	0.88	0.76	0.14	1276	1,143	88	45
7	A	63A	1	85	5	0	9	0.95	0.89	0.51	1346	660	686	0
7	A	63B	2	83	9	0	8	0.92	0.83	0.36	1346	897	414	35
7	A	7A	1	94	0	1	5	1.00	0.98	0.88	1624	203	1,421	0
7	B	7A	1	94	1	0	4	0.97	0.93	0.89	1646	180	1,466	0
7	C	7A	1	95	1	1	3	0.98	0.94	0.91	1638	144	1,494	0
7	D	7A	1	95	1	0	4	0.99	0.96	0.90	1550	150	1,400	0
7	A	7B	2	85	9	2	3	0.95	0.90	1.31	1624	453	218	953
7	B	7B	2	81	12	2	4	0.93	0.86	1.34	1646	447	196	1,003
7	C	7B	2	82	13	2	3	0.93	0.86	1.33	1638	428	244	966
7	D	7B	2	84	10	2	3	0.94	0.88	1.36	1550	413	164	973
8	A	20A	1	96	1	0	4	1.00	0.98	0.74	1594	421	1,173	0
8	B	20A	1	97	0	0	2	1.00	0.99	0.74	1756	450	1,306	0
8	C	20A	1	97	0	0	3	0.99	0.99	0.78	1714	376	1,338	0
8	D	20A	1	97	1	0	3	0.99	0.98	0.77	1706	400	1,306	0
8	A	20B	2	79	11	6	4	0.89	0.78	1.34	1594	451	148	995
8	B	20B	2	84	9	4	3	0.93	0.85	1.34	1756	491	173	1,092
8	C	20B	2	82	11	4	3	0.91	0.83	1.36	1714	472	159	1,083
8	D	20B	2	83	11	4	3	0.92	0.84	1.36	1706	462	171	1,073
8	A	32A	1	82	9	1	8	0.89	0.78	0.30	1594	1,116	478	0
8	B	32A	1	85	8	1	7	0.91	0.81	0.28	1756	1,262	494	0
8	C	32A	1	83	8	0	9	0.91	0.82	0.32	1714	1,166	548	0

*This item was suppressed during range finding, so there are no inter-rater reliability statistics to report for this item.

Table 10-29 Cont'd
Inter-Rater Reliability Mathematics

				Percentage Absolute Difference								Frequency		
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra Corr.	Weighted Kappa	Mean	N	0	1	2
8	D	32A	1	85	7	0	7	0.92	0.83	0.29	1706	1,219	487	0
8	A	32B	2	80	11	1	8	0.94	0.88	0.73	1594	701	625	268
8	B	32B	2	82	10	1	7	0.94	0.88	0.69	1756	771	766	219
8	C	32B	2	81	10	1	8	0.95	0.89	0.72	1714	731	727	256
8	D	32B	2	83	10	1	6	0.95	0.89	0.68	1706	758	729	219
8	A	44A	1	94	1	0	5	0.99	0.99	0.44	1594	893	701	0
8	B	44A	1	94	1	0	5	1.00	0.99	0.45	1756	971	785	0
8	C	44A	1	94	1	0	5	0.99	0.98	0.45	1714	935	779	0
8	D	44A	1	94	1	0	5	0.99	0.99	0.45	1706	936	770	0
8	A	44B	2	79	13	1	6	0.95	0.90	0.95	1594	651	380	563
8	B	44B	2	82	12	1	6	0.96	0.91	0.99	1756	669	435	652
8	C	44B	2	83	10	1	6	0.96	0.92	1.02	1714	630	427	657
8	D	44B	2	82	11	1	6	0.95	0.91	0.97	1706	670	412	624
8	A	57A	1	93	1	0	6	0.99	0.98	0.70	734	221	513	0
8	B	57A	1	98	1	0	2	1.00	0.99	0.75	768	190	578	0
8	C	57A	1	86	1	0	13	0.99	0.98	0.18	1340	1,093	247	0
8	A	57B	2	84	9	1	6	0.93	0.86	0.88	734	214	396	124
8	B	57B	2	86	12	1	2	0.93	0.86	0.95	768	205	395	168
8	C	57B	2	71	14	1	14	0.94	0.87	0.59	1340	769	347	224
8	D	65A	1	90	1	1	8	0.99	0.97	0.30	1286	896	390	0
8	D	65B	2	82	7	1	10	0.97	0.95	0.72	1286	739	168	379
8	A	9A	1	96	1	0	3	0.97	0.95	0.84	1594	263	1,331	0
8	B	9A	1	97	1	0	2	0.98	0.96	0.88	1756	213	1,543	0
8	C	9A	1	96	2	0	3	0.96	0.93	0.88	1714	213	1,501	0
8	D	9A	1	96	1	0	2	0.98	0.95	0.86	1706	244	1,462	0
8	A	9B	2	78	16	2	4	0.92	0.83	1.40	1594	352	254	988
8	B	9B	2	78	18	1	3	0.91	0.82	1.46	1756	312	322	1,122

Table 10-29 Cont'd
Inter-Rater Reliability Mathematics

				Percentage Absolute Difference								Frequency		
Grade	Form	Item No.	Max Score	Perfect	Adjacent	Discrepant	Codes	Intra Corr.	Weighted Kappa	Mean	N	0	1	2
8	C	9B	2	77	18	1	4	0.92	0.83	1.45	1714	325	299	1,090
8	D	9B	2	77	18	2	3	0.91	0.81	1.42	1706	355	273	1,078
10	A	16	2	88	3	0	9	0.98	0.96	0.44	6612	3,934	2,476	202
10	A	30	2	78	9	4	10	0.93	0.87	0.75	6612	3,747	772	2,093
10	A	36	2	82	10	1	8	0.97	0.93	0.78	6612	3,267	1,515	1,830
10	A	47	2	76	15	1	8	0.93	0.86	0.79	6612	2,983	2,064	1,565

Table 10-30
Inter-Rater Reliability Language Arts*

				Percentage Absolute Difference								Frequency						
Grade	Form	Item No.	Max Score	P	A	D	Codes	Intra. Corr.	Weighted Kappa	Mean	N	0	1	2	3	4	5	6
4	E	1A	6	59	33	3	5	0.89	0.78	2.98	122372	6646	2558	20754	54441	33084	4284	605
4	E	1B	3	90	7	0	3	0.89	0.78	1.89	122372	3813	6986	110604	969	0	0	0
8	E	1A	6	59	35	3	4	0.90	0.80	3.27	137856	5728	1576	16804	51290	50867	9950	1011
8	E	1B	3	92	5	0	3	0.92	0.84	1.94	137856	4101	2407	128542	2806	0	0	0
10	B	1A	6	60	33	2	5	0.92	0.84	3.19	144334	7256	2417	19237	54618	49756	10016	1034
10	B	1B	3	86	9	0	5	0.90	0.81	1.94	144334	6890	2711	127085	7648	0	0	0

*P is percent perfect agreement, A is percent adjacent agreement, and D is percent discrepant.

Table 11-1
Items Flagged for DIF, By Gender

Content	Grade	Form	Test Book Number	Item Type	Female			Male			SMD	Delta	LH Flag Female	LH Flag Male	Flag MH
					D+	D-	Z	D+	D1	Z					
RD	6	A	15	MC	0.00	-0.04	-21.07	0.04	0.00	22.45		-1.94			-C
RD	6	A	63	CR	0.12	-0.08	5.19	0.00	-0.12	-6.47	0.26	.		-CC	CC
RD	7	A	16	MC	0.00	-0.07	-28.32	0.06	0.00	23.04		-1.78			-C
RD	8	A	5	MC	0.00	-0.13	-49.74	0.11	0.00	45.31		-3.07	-C	CC	-C
RD	8	A	56	CR	0.13	0.00	30.26	0.00	-0.13	-32.37	0.29	.	CC	-CC	CC
RD	8	B	63	CR	0.10	0.00	5.51	0.01	-0.09	-4.96	0.23	.	CC		BB
RD	10	A	5	MC	0.00	-0.13	-49.74	0.11	0.00	45.31		-3.07	-C	CC	-C
RD	10	A	56	CR	0.13	0.00	30.26	0.00	-0.13	-32.37	0.29	.	CC	-CC	CC
RD	10	B	63	CR	0.10	0.00	5.51	0.01	-0.09	-4.96	0.23	.	CC		BB
MA	3	D	50	MC	0.00	-0.02	-5.08	0.02	0.00	6.08		-1.52			-C
MA	4	A	43	MC	0.00	-0.04	-16.99	0.04	0.00	18.78		-1.56			-C
MA	6	A	15	MC	0.00	-0.05	-19.36	0.04	-0.01	16.78		-1.57			-C
MA	7	A	60	MC	0.00	-0.09	-16.87	0.07	0.00	13.76		-2.31			-C
LA	4	A	1A	CR	0.13	0.00	24.80	0.07	-0.15	-27.79	0.33	.	CC	-CC	CC
LA	8	A	1A	CR	0.11	-0.02	21.18	0.19	-0.15	-29.94	0.31	.		-CC	CC
LA	10	A	1A	CR	0.11	-0.02	21.18	0.19	-0.15	-29.94	0.31	.		-CC	CC
SC	8	A	11	MC	0.00	-0.05	-18.38	0.05	0.00	21.73		-1.81			-C
SC	10	A	11	MC	0.00	-0.05	-18.38	0.05	0.00	21.73		-1.81			-C

Table 11-2
Items Flagged for DIF, By Ethnicity, African American

Content	Grade	Form	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
RD	3	A	61	MC	0.01	-0.11	-8.53		-1.71	-C	-C
RD	4	A	17	MC	0.00	-0.05	-9.37		-1.50		-C
RD	5	C	61	MC	0.00	-0.09	-12.67		-2.10		-C
RD	7	A	23	MC	0.00	-0.05	-11.79		-1.61		-C
RD	7	A	37	CR	0.01	-0.22	-18.31	-0.22	.	-CC	-BB
RD	8	A	6	MC	0.00	-0.14	-24.76		-2.09	-C	-C
RD	8	A	31	CR	0.13	0.00	12.30	0.21	.	CC	BB
RD	10	A	31	CR	0.00	-0.13	-13.33	-0.04	.	-CC	
RD	10	A	40	MC	0.03	-0.01	-0.61		-1.82		-C
MA	5	A	47	MC	0.00	-0.05	-12.75		-1.57		-C
MA	5	B	61	MC	0.00	-0.08	-7.62		-2.11		-C
MA	6	A	29	MC	0.00	-0.07	-13.66		-1.73		-C
MA	6	A	32	MC	0.00	-0.09	-16.74		-1.57		-C
MA	7	A	16	MC	0.00	-0.09	-16.48		-1.59		-C
MA	7	B	61	MC	0.00	-0.06	-5.65		-1.58		-C
MA	7	C	56	CR	0.16	-0.11	3.56	0.16	.	CC	
MA	8	A	35	MC	0.00	-0.07	-14.73		-1.73		-C
MA	8	A	63	MC	0.00	-0.06	-6.12		-1.70		-C
MA	8	B	58	MC	0.01	-0.06	-4.06		-1.67		-C
LA	10	A	1A	CR	0.18	-0.05	11.04	-0.01	.	CC	
SS	8	A	14	MC	0.00	-0.10	-17.98		-1.52		-C
SS	8	A	30	MC	0.08	0.00	14.85		1.54		C
SC	4	A	4	MC	0.00	-0.09	-17.30		-1.52		-C
SC	4	B	46	MC	0.04	-0.05	-3.36		-1.72		-C

Table 11-3
Items Flagged for DIF, By Ethnicity, Hispanic

Content	Grade	Form	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
RD	4	A	17	MC	0.00	-0.05	-8.18		-1.52		-C
RD	6	D	67	CR	0.20	-0.09	3.85	0.07	.	CC	
RD	7	A	23	MC	0.00	-0.07	-9.85		-1.73		-C
RD	7	A	40	MC	0.00	-0.12	-16.29		-1.76	-C	-C
RD	8	A	6	MC	0.00	-0.10	-13.98		-1.65		-C
RD	8	A	31	CR	0.12	0.00	9.16	0.20	.	CC	BB
RD	10	A	40	MC	0.02	-0.02	-0.85		-1.71		-C
MA	3	A	42	CR	0.00	-0.10	-9.03	-0.13	.	-CC	
MA	3	A	54	MC	0.00	-0.06	-6.62		-1.93		-C
MA	5	A	49	MC	0.00	-0.02	-5.07		-1.7		-C
MA	5	B	61	MC	0.04	-0.05	-4.39		-1.52		-C
LA	10	A	6	MC	0.01	-0.06	-7.70		-1.67		-C
LA	10	A	1A	CR	0.22	-0.08	10.64	0.03	.	CC	
SS	4	A	35	MC	0.00	-0.07	-8.18		-1.57		-C
SC	4	B	46	MC	0.01	-0.03	-1.02		-1.5		-C

Table 11-4
Items Flagged for DIF, By Ethnicity, Asian

Content	Grade	Form	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
RD	3	A	61	MC	0.00	-0.12	-5.71		-1.84	-C	-C
RD	3	C	67	CR	0.21	-0.08	2.66	0.12	.	CC	
RD	4	A	17	MC	0.00	-0.06	-5.75		-1.57		-C
RD	4	B	62	CR	0.20	-0.03	3.44	0.30	.	CC	CC
RD	5	A	48	CR	0.24	-0.13	14.62	0.32	.	CC	CC
RD	5	A	68	MC	0.13	0.00	8.89		1.67	C	C
RD	5	A	71	CR	0.17	-0.28	2.91	0.13	.	CC	
RD	6	A	19	CR	0.14	-0.02	9.15	0.21	.	CC	BB
RD	6	A	27	CR	0.13	0.00	8.18	0.17	.	CC	BB
RD	6	A	46	MC	0.00	-0.09	-11.74		-2.02		-C
RD	6	A	47	MC	0.00	-0.09	-12.80		-2.07		-C
RD	6	A	63	CR	0.19	0.00	4.49	0.29	.	CC	CC
RD	6	A	64	MC	0.02	-0.11	-6.45		-1.67		-C
RD	6	D	67	CR	0.20	-0.08	3.39	0.11	.	CC	
RD	7	A	23	MC	0.00	-0.07	-7.23		-1.87		-C
RD	7	A	26	MC	0.03	0.00	4.50		2.04		C
RD	7	A	40	MC	0.00	-0.12	-12.53		-1.86	-C	-C
RD	7	A	42	MC	0.04	-0.01	5.48		1.78		C
RD	7	A	56	CR	0.12	0.00	8.05	0.17	.	CC	
RD	7	B	72	CR	0.24	-0.02	3.51	0.19	.	CC	BB
RD	8	A	5	MC	0.00	-0.11	-11.38		-1.61	-C	-C
RD	8	A	20	MC	0.00	-0.05	-6.19		-1.92		-C
RD	8	A	31	CR	0.28	0.00	15.63	0.34	.	CC	CC
RD	8	A	50	MC	0.00	-0.07	-9.52		-1.62		-C
RD	8	A	56	CR	0.20	-0.02	10.57	0.22	.	CC	BB
RD	8	A	57	MC	0.06	-0.01	5.04		1.7		C
RD	8	A	63	CR	0.23	-0.09	4.46	0.28	.	CC	CC
RD	8	A	66	MC	0.00	-0.07	-5.15		-1.91		-C
RD	8	B	61	MC	0.00	-0.13	-6.54		-1.52	-C	-C

Table 11-4 Cont'd
Items Flagged for DIF, By Ethnicity, Asian

Content	Grade	Form	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
RD	8	B	63	CR	0.17	-0.01	3.25	0.14	.	CC	
RD	8	C	65	MC	0.01	-0.13	-5.41		-1.65		-C
RD	8	C	67	CR	0.22	-0.05	3.97	0.22	.	CC	BB
RD	8	D	59	MC	0.00	-0.12	-7.14		-1.96	-C	-C
RD	8	D	67	CR	0.24	-0.06	3.30	0.20	.	CC	BB
RD	10	A	34	MC	0.01	-0.08	-8.04		-1.67		-C
RD	10	A	40	MC	0.00	-0.04	-4.03		-2.9		-C
RD	10	A	52	CR	0.15	-0.03	10.16	0.25	.	CC	BB
MA	3	A	42	CR	0.00	-0.11	-6.93	-0.14	.	-CC	
MA	3	A	43	MC	0.01	-0.06	-6.21		-2.03		-C
MA	3	A	54	MC	0.00	-0.07	-5.26		-2.23		-C
MA	3	D	56	MC	0.08	-0.10	-6.25		-1.76		-C
MA	5	A	24	MC	0.00	-0.10	-10.60		-1.62		-C
MA	5	A	33	MC	0.00	-0.05	-8.20		-1.58		-C
MA	5	A	34	MC	0.01	-0.05	-5.51		-1.64		-C
MA	5	A	49	MC	0.00	-0.02	-4.43		-2.13		-C
MA	5	B	61	MC	0.01	-0.12	-6.58		-2.62		-C
MA	5	D	63	CR	0.03	-0.11	-2.61	-0.26	.		-CC
MA	6	A	8	MC	0.09	-0.02	8.59		1.58		C
MA	6	A	11	CR	0.02	-0.14	-6.17	-0.14	.	-CC	
MA	6	A	53	MC	0.00	-0.09	-11.54		-2.03		-C
MA	6	D	63	CR	0.04	-0.15	-2.63	-0.23	.	-CC	-BB
MA	7	A	3	MC	0.00	-0.03	-7.55		-1.56		-C
MA	7	A	19	CR	0.12	-0.03	8.55	0.18	.	CC	BB
MA	7	A	52	MC	0.00	-0.11	-10.97		-1.44	-C	-B
MA	7	A	63	CR	0.20	-0.02	3.72	0.27	.	CC	CC
MA	8	A	57	CR	0.04	-0.14	-2.98	-0.21	.	-CC	-BB
MA	8	D	61	MC	0.01	-0.14	-5.68		-1.45	-C	-B
MA	10	A	21	MC	0.00	-0.11	-13.54		-2.14	-C	-C

Table 11-4 Cont'd
Items Flagged for DIF, By Ethnicity, Asian

Content	Grade	Form	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
LA	8	A	14	MC	0.00	-0.09	-9.89		-1.63		-C
LA	8	A	18	MC	0.00	-0.11	-12.40		-1.58	-C	-C
LA	8	A	1A	CR	0.11	0.00	6.36	0.12	.	CC	
LA	10	A	6	MC	0.00	-0.10	-11.90		-2.46		-C
LA	10	A	11	MC	0.00	-0.11	-11.89		-1.6	-C	-C
LA	10	A	24	MC	0.00	-0.13	-12.91		-1.35	-C	-B
LA	10	A	1A	CR	0.26	0.00	14.93	0.16	.	CC	
SS	4	A	35	MC	0.01	-0.09	-6.34		-1.85		-C
SS	10	A	2	MC	0.00	-0.10	-10.79		-1.65	-C	-C
SS	10	A	12	MC	0.00	-0.11	-15.50		-2.15	-C	-C
SC	4	A	1	MC	0.02	0.00	4.01		1.58		C
SC	4	B	46	MC	0.01	-0.08	-4.74		-2.4		-C
SC	8	A	17	MC	0.00	-0.19	-20.10		-2.34	-C	-C
SC	8	A	45	MC	0.05	-0.13	-5.60		-1.79	-C	-C
SC	10	A	26	MC	0.05	-0.13	-11.50		-1.52	-C	-C

Table 11-5
Items Flagged for DIF, By Ethnicity, American Indian

Content	Grade	Form	Test Book Number	Item Type	D+	D-	Z	SMD	Delta	LH Flag	MH Flag
RD	3	C	67	CR	0.12	-0.34	-1.54	-0.30	.		-CC
RD	6	A	19	CR	0.00	-0.11	-4.91	-0.11	.	-CC	
RD	6	C	60	MC	0.15	-0.05	3.84		1.43	C	B
RD	8	A	56	CR	0.01	-0.11	-4.10	-0.07	.	-CC	
RD	10	A	31	CR	0.00	-0.16	-7.16	-0.06	.	-CC	
RD	10	A	52	CR	0.01	-0.13	-5.39	-0.04	.	-CC	
MA	3	A	42	CR	0.05	-0.14	-3.98	-0.12	.	-CC	
MA	3	A	52	CR	0.58	-0.48	2.80	0.44	.	CC	CC
MA	3	B	52	CR	0.64	-0.52	-2.04	-0.34	.		-CC
MA	4	B	52	CR	0.11	-0.33	-0.71	-0.32	.		-CC
MA	5	A	57	CR	1.56	1.87		-0.50	.		-CC
MA	6	A	56	CR	0.55	-0.13	-1.07	-0.26	.		-CC
MA	6	B	56	CR	0.27	-0.20	-2.09	-0.35	.		-CC
MA	7	A	63	CR	1.25	1.37		-0.33	.		-CC
MA	7	C	56	CR	0.39	-0.33	-1.96	-0.26	.		-CC
MA	10	A	47	CR	0.00	-0.13	-6.01	-0.07	.	-CC	
LA	8	A	1A	CR	0.24	-0.16	-5.03	-0.18	.	-CC	-BB
SS	10	A	4	MC	0.11	0.00	7.03		0.49	C	

Table 11-6
Items Flagged for DIF, By English Language Proficiency

Content	Grade	Form	Test Book Number	Item Type	Not Proficient			Proficient			SMD	Delta	LH Flag Not Proficient	LH Flag Proficient	MH Flag
					D+	D-	Z	D+	D1	Z					
RD	4	A	17	MC	0	-0.08	-11.76	0.01	0	7.18		-1.55			-C
RD	4	B	62	CR	0.19	-0.04	3.65	0.06	-0.06	-0.93	0.15	.	CC		
RD	4	C	66	CR	0.17	-0.07	3.31	0.05	-0.04	-1.7	0.14	.	CC		
RD	5	A	48	CR	0.24	-0.15	14.15	0.05	-0.07	-3.58	0.24	.	CC		BB
RD	5	A	67	MC	0	-0.1	-8.74	0.01	-0.01	1.28		-1.59			-C
RD	6	A	3	MC	0	-0.11	-12.29	0.01	0	2.21		-1.41	-C		-B
RD	6	A	4	MC	0	-0.12	-14.19	0.01	0	6.08		-1.78	-C		-C
RD	6	A	46	MC	0.01	-0.11	-12.58	0.01	0	6.05		-1.59			-C
RD	6	A	63	CR	0.17	-0.05	4.16	0.05	-0.06	-2.57	0.14	.	CC		
RD	6	A	64	MC	0	-0.11	-9.31	0.01	-0.01	2.4		-1.56	-C		-C
RD	6	D	67	CR	0.21	-0.02	3.54	0.06	-0.05	-1.78	0.13	.	CC		
RD	7	A	23	MC	0	-0.09	-12.61	0.01	0	6.09		-1.53			-C
RD	7	A	40	MC	0	-0.15	-17.27	0.02	-0.01	3		-1.66	-C		-C
RD	8	A	6	MC	0	-0.11	-11.65	0.01	-0.01	1.98		-1.35	-C		-B
RD	8	A	31	CR	0.25	0	15.75	0.02	-0.03	-5.17	0.3	.	CC		CC
RD	8	A	41	MC	0	-0.1	-11.26	0.01	0	3.67		-1.41	-C		-B
RD	8	A	56	CR	0.16	-0.01	8.64	0.02	-0.02	-3.34	0.17	.	CC		
RD	8	A	66	MC	0	-0.06	-6.43	0.01	0	4.2		-1.58			-C
RD	8	B	63	CR	0.17	-0.06	3.26	0.04	-0.06	-0.51	0.15	.	CC		
RD	8	D	59	MC	0	-0.11	-5.71	0.01	-0.01	1.58		-1.53	-C		-C
RD	10	A	6	MC	0	-0.11	-11.65	0.01	-0.01	1.98		-1.35	-C		-B
RD	10	A	31	CR	0.25	0	15.75	0.02	-0.03	-5.17	0.3	.	CC		CC
RD	10	A	41	MC	0	-0.1	-11.26	0.01	0	3.67		-1.41	-C		-B
RD	10	A	56	CR	0.16	-0.01	8.64	0.02	-0.02	-3.34	0.17	.	CC		
RD	10	A	66	MC	0	-0.06	-6.43	0.01	0	4.2		-1.58			-C
RD	10	B	63	CR	0.17	-0.06	3.26	0.04	-0.06	-0.51	0.15	.	CC		
RD	10	D	59	MC	0	-0.11	-5.71	0.01	-0.01	1.58		-1.53	-C		-C
MA	3	A	42	CR	0	-0.13	-10.06	0.02	-0.02	-0.85	-0.15	.	-CC		
MA	3	A	54	MC	0	-0.08	-8.14	0.01	0	4.5		-1.84			-C

Table 11-6 Cont'd
Items Flagged for DIF, By English Language Proficiency

Content	Grade	Form	Test Book Number	Item Type	Not Disabled			Disabled			SMD	Delta	LH Flag Not Proficient	LH Flag Proficient	MH Flag
					D+	D-	Z	D+							
MA	5	A	34	MC	0	-0.07	-9.5	0.03	-0.02	-0.65		-1.52			-C
MA	5	A	49	MC	0	-0.03	-6.24	0.01	0	2.97		-1.54			-C
MA	5	D	63	CR	0	-0.1	-2.77	0.08	-0.03	0.52	-0.18	.	-CC		-BB
MA	6	A	53	MC	0	-0.11	-16.37	0.01	0	5.54		-1.76	-C		-C
MA	7	A	63	CR	0.13	-0.07	4.26	0.06	-0.06	-1.31	0.23	.	CC		BB
MA	8	A	44	CR	0.13	-0.01	7.77	0.04	-0.04	-7.17	0.12	.	CC		
MA	8	D	61	MC	0	-0.14	-6.87	0.04	-0.03	1.17		-1.61	-C		-C
MA	10	A	44	CR	0.13	-0.01	7.77	0.04	-0.04	-7.17	0.12	.	CC		
MA	10	D	61	MC	0	-0.14	-6.87	0.04	-0.03	1.17		-1.61	-C		-C
LA	4	A	11	MC	0.01	0	10.07	0	-0.06	-8.41		-1.78			-C
LA	4	A	26	MC	0.01	-0.01	0.16	0.04	-0.13	-13.11		-1.27		-CC	-B
LA	4	A	1A	CR	0.03	-0.03	-3.43	0.1	0	5.82	0.05	.		CC	
SS	4	A	35	MC	0.02	-0.01	13.57	0.01	-0.12	-13.73		-2.12		-CC	-C
SC	4	A	1	MC	0	0	1.76	0.03	0	7.38		1.76			C
SC	4	A	14	MC	0	-0.01	-5.12	0.11	0	12.79		1.33		CC	B
SC	8	A	12	MC	0.02	-0.02	-1.97	0.1	0	10.82		1.55			C
SC	8	A	17	MC	0.01	0	2.4	0	-0.13	-13.38		-1.38		-CC	-B
SC	8	A	45	MC	0.01	-0.01	-0.12	0.09	-0.11	-5.33		-1.58			-C
SC	10	A	12	MC	0.02	-0.02	-1.97	0.1	0	10.82		1.55			C
SC	10	A	17	MC	0.01	0	2.4	0	-0.13	-13.38		-1.38		-CC	-B
SC	10	A	45	MC	0.01	-0.01	-0.12	0.09	-0.11	-5.33		-1.58			-C

Table 11-7
Items Flagged for DIF, By Disability Status

Content	Grade	Form	Test Book Number	Item Type	Not Disabled			Disabled			SMD	Delta	LH Flag Not Disabled	LH Flag Disabled	MH Flag
					D+	D-	Z	D+	D1	Z					
RD	8	A	56	CR	0.03	-0.03	3.84	0	-0.13	-14.79	-0.17	.		-CC	-BB
RD	10	A	56	CR	0.03	-0.03	3.84	0	-0.13	-14.79	-0.17	.		-CC	-BB
MA	3	A	5	MC	0.02	0	4.16	0	-0.04	-6.76		-1.55			-C
MA	3	A	52	CR	0.04	-0.05	-1.04	0.13	-0.19	3.97	0.11	.		CC	
MA	3	B	52	CR	0.02	-0.05	-2.02	0.21	-0.08	3.63	0.2	.		CC	BB
MA	3	C	52	MC	0.01	0	2.81	0.01	-0.04	-3.1		-1.55			-C
MA	3	D	58	CR	0.02	-0.05	-0.88	0.21	-0.14	3.49	0.21	.		CC	BB
MA	5	C	55	MC	0.01	0	3.52	0	-0.06	-6.37		-1.63			-C
MA	5	D	58	MC	0.01	0	1.49	0	-0.02	-0.91		-1.52			-C
LA	4	A	11	MC	0.01	0	11.25	0.03	-0.03	-8.5		-1.68			-C
LA	4	A	1A	CR	0.04	-0.04	6.47	0.1	-0.29	-24.37	-0.47	.		-CC	-CC
LA	4	A	1B	CR	0.02	-0.01	8.69	0	-0.07	-23.01	-0.59	.			-CC
LA	8	A	1A	CR	0.04	-0.04	-1.02	0.24	-0.21	-16.03	-0.4	.		-CC	-CC
LA	10	A	1A	CR	0.04	-0.04	-1.02	0.24	-0.21	-16.03	-0.4	.		-CC	-CC
LA	10	A	1B	CR							-0.36	.			-CC
SS	4	A	4	MC	0.02	0	15.02	0	-0.04	-10.22		-1.91			-C
SC	4	A	15	MC	0.02	0	4.55	0	-0.04	-9.45		-1.57			-C

Table 11-8
Correlations among Reading Objectives

Grade	CS	1	2	3
3	2	0.72		
	3	0.74	0.84	
	4	0.58	0.65	0.68
4	2	0.71		
	3	0.73	0.75	
	4	0.59	0.61	0.65
5	2	0.71		
	3	0.70	0.75	
	4	0.68	0.73	0.73
6	2	0.66		
	3	0.68	0.68	
	4	0.68	0.66	0.70
7	2	0.66		
	3	0.69	0.74	
	4	0.63	0.67	0.73
8	2	0.67		
	3	0.69	0.73	
	4	0.68	0.71	0.75
10	2	0.59		
	3	0.71	0.66	
	4	0.69	0.66	0.78

Table 11-9
Correlations among Mathematics Objectives

Grade	CS	A	B	C	D	E
3	B	0.62				
	C	0.53	0.58			
	D	0.50	0.58	0.54		
	E	0.62	0.63	0.60	0.58	
	F	0.59	0.65	0.54	0.53	0.58
4	B	0.45				
	C	0.45	0.50			
	D	0.54	0.59	0.55		
	E	0.52	0.55	0.51	0.60	
	F	0.56	0.66	0.54	0.64	0.59
5	B	0.56				
	C	0.50	0.49			
	D	0.55	0.54	0.49		
	E	0.60	0.63	0.54	0.57	
	F	0.59	0.64	0.52	0.56	0.64
6	B	0.64				
	C	0.49	0.49			
	D	0.65	0.67	0.47		
	E	0.68	0.67	0.52	0.64	
	F	0.65	0.71	0.48	0.65	0.65
7	B	0.60				
	C	0.62	0.62			
	D	0.57	0.65	0.61		
	E	0.71	0.62	0.60	0.56	
	F	0.58	0.61	0.56	0.52	0.60
8	B	0.58				
	C	0.66	0.63			
	D	0.58	0.62	0.61		
	E	0.62	0.64	0.62	0.60	
	F	0.65	0.61	0.61	0.57	0.62
10	B	0.61				
	C	0.64	0.61			
	D	0.66	0.62	0.73		
	E	0.63	0.59	0.63	0.65	
	F	0.67	0.65	0.70	0.71	0.66

Table 11-10
Correlations among Language Arts Objectives

Grade	CS	B	D
4	D	0.53	
	F	0.57	0.42
8	D	0.69	
	F	0.52	0.50
10	D	0.66	
	F	0.53	0.54

Table 11-11
Correlations among Social Studies Objectives

Grade	CS	A	B	C	D
4	B	0.61			
	C	0.50	0.50		
	D	0.55	0.55	0.46	
	E	0.61	0.60	0.51	0.55
8	B	0.64			
	C	0.54	0.62		
	D	0.53	0.60	0.51	
	E	0.52	0.63	0.56	0.51
10	B	0.63			
	C	0.62	0.70		
	D	0.61	0.66	0.64	
	E	0.60	0.68	0.67	0.63

Table 11-12
Correlations among Science Objectives

Grade	CS	A	B	C	D	E	F	G
	B	0.37						
	C	0.48	0.37					
	D	0.34	0.29	0.41				
4	E	0.43	0.34	0.49	0.41			
	F	0.40	0.33	0.49	0.41	0.52		
	G	0.30	0.25	0.34	0.30	0.35	0.35	
	H	0.43	0.37	0.52	0.43	0.50	0.52	0.38
	B	0.33						
	C	0.49	0.39					
	D	0.36	0.33	0.49				
8	E	0.40	0.36	0.51	0.50			
	F	0.41	0.34	0.51	0.44	0.46		
	G	0.46	0.38	0.54	0.42	0.46	0.47	
	H	0.42	0.34	0.50	0.42	0.44	0.44	0.48
	B	0.46						
	C	0.53	0.54					
	D	0.44	0.43	0.50				
10	E	0.47	0.48	0.54	0.44			
	F	0.44	0.44	0.51	0.43	0.45		
	G	0.49	0.49	0.57	0.47	0.52	0.49	
	H	0.47	0.47	0.56	0.45	0.49	0.48	0.56

Table 11-13
Factor Analysis

Content Area	Grade	First Eigenvalue	Second Eigenvalue	Ratio of First Two Eigenvalues
Reading	3	16.25	1.27	12.83
	4	12.97	0.92	14.16
	5	11.96	1.00	11.98
	6	10.36	0.81	12.77
	7	11.30	0.99	11.42
	8	12.19	0.97	12.62
	10	12.03	1.02	11.75
Mathematics	3	12.15	1.59	7.66
	4	10.66	1.35	7.90
	5	11.34	1.53	7.43
	6	17.16	3.93	4.36
	7	14.27	3.44	4.15
	8	14.10	1.57	8.97
	10	14.00	1.48	9.44
Language Arts	4	5.36	0.76	7.08
	8	6.28	0.53	11.83
	10	6.47	0.93	7.00
Social Studies	4	7.65	0.96	7.96
	8	8.50	0.84	10.14
	10	11.14	0.91	12.25
Science	4	6.82	0.73	9.40
	8	7.40	0.73	10.07
	10	9.22	0.84	10.91

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%

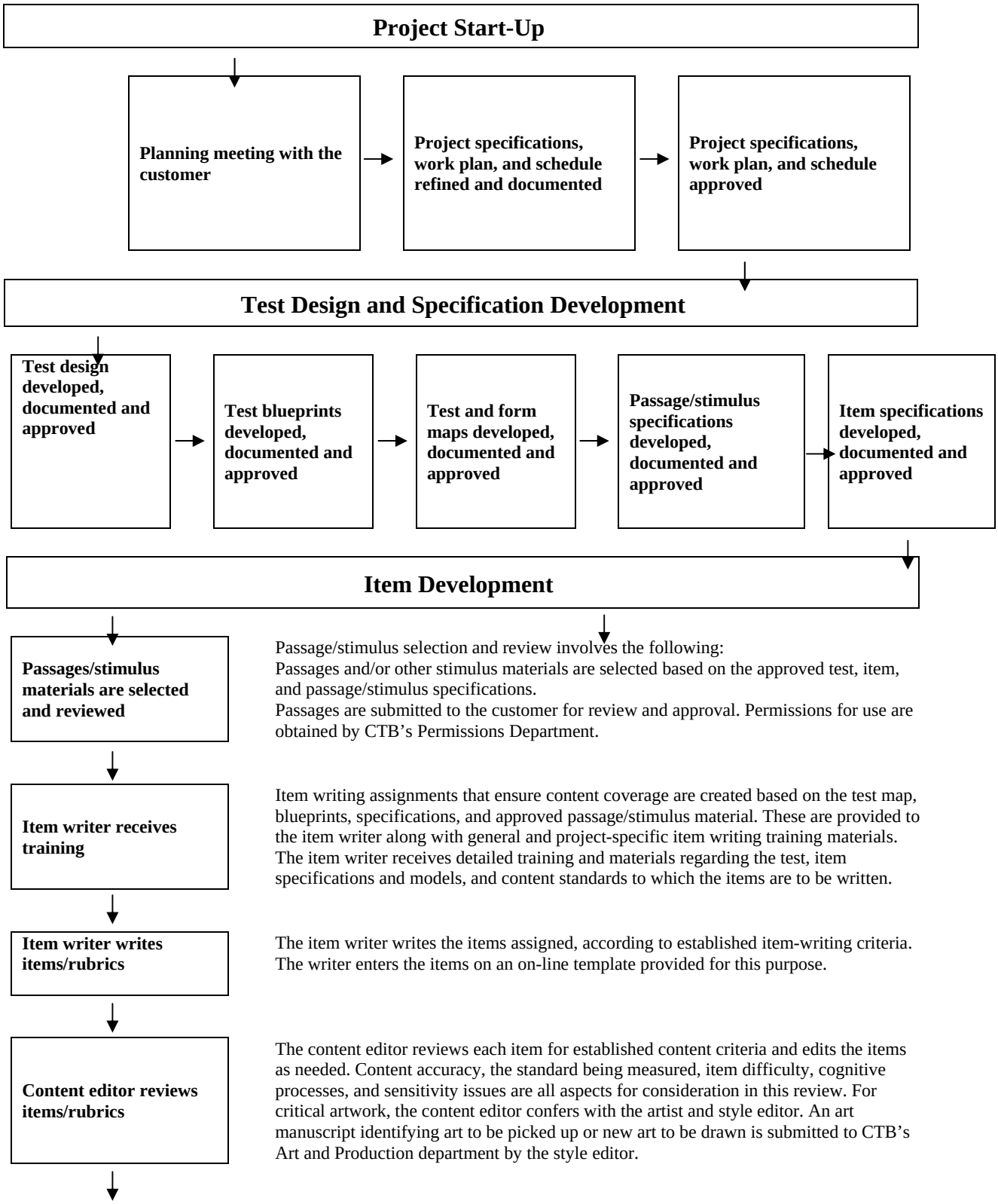
Table 12-5
Cut scores and Associated Impact Data for WKCE-CRT Social Studies

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

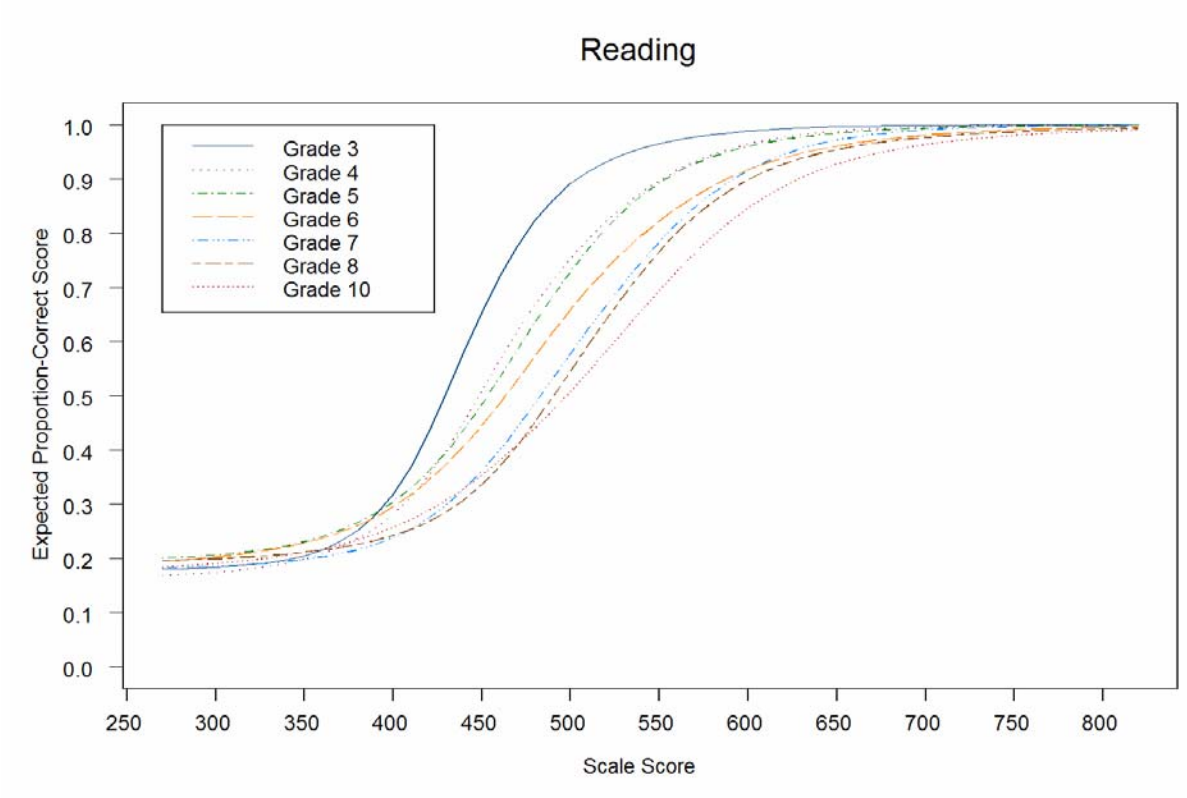


Figure 8-2
TCC Curve for Mathematics Grades 3-8, 10

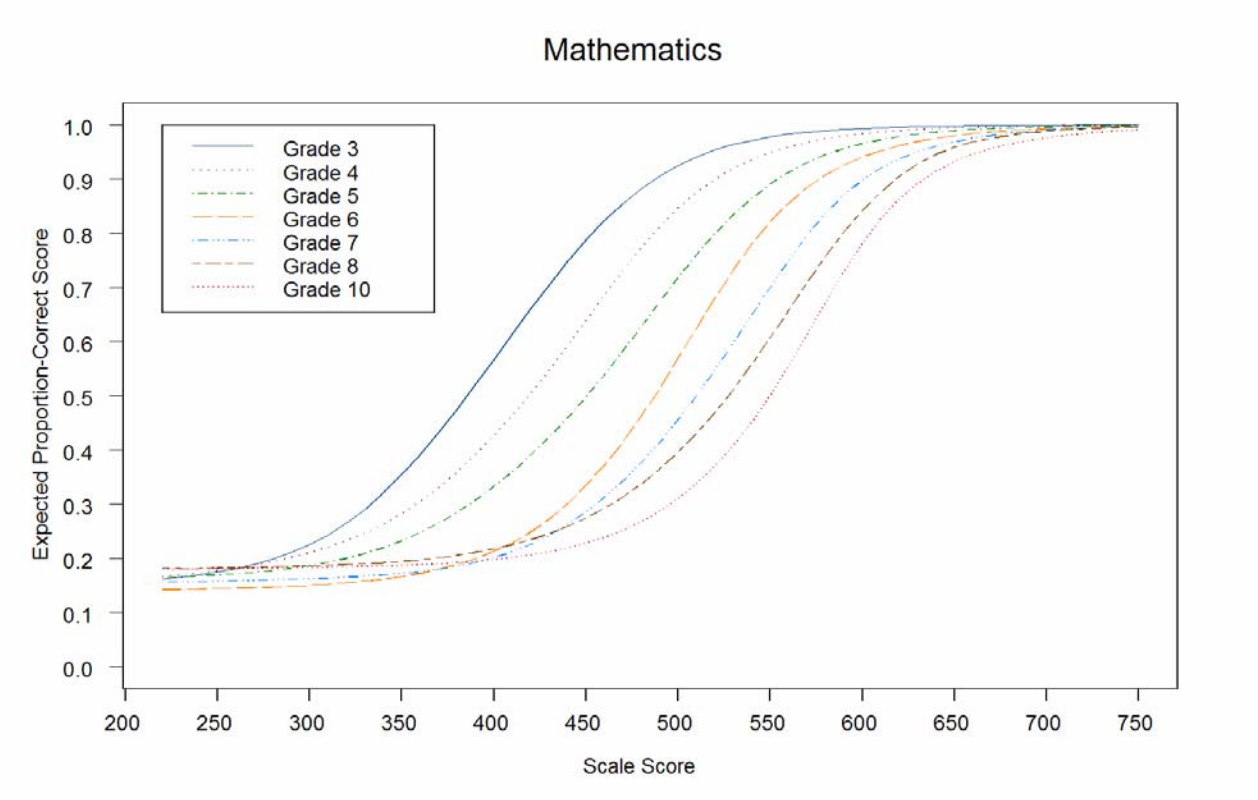


Figure 8-3
TCC Curve for Language Arts Grades 4, 8, 10

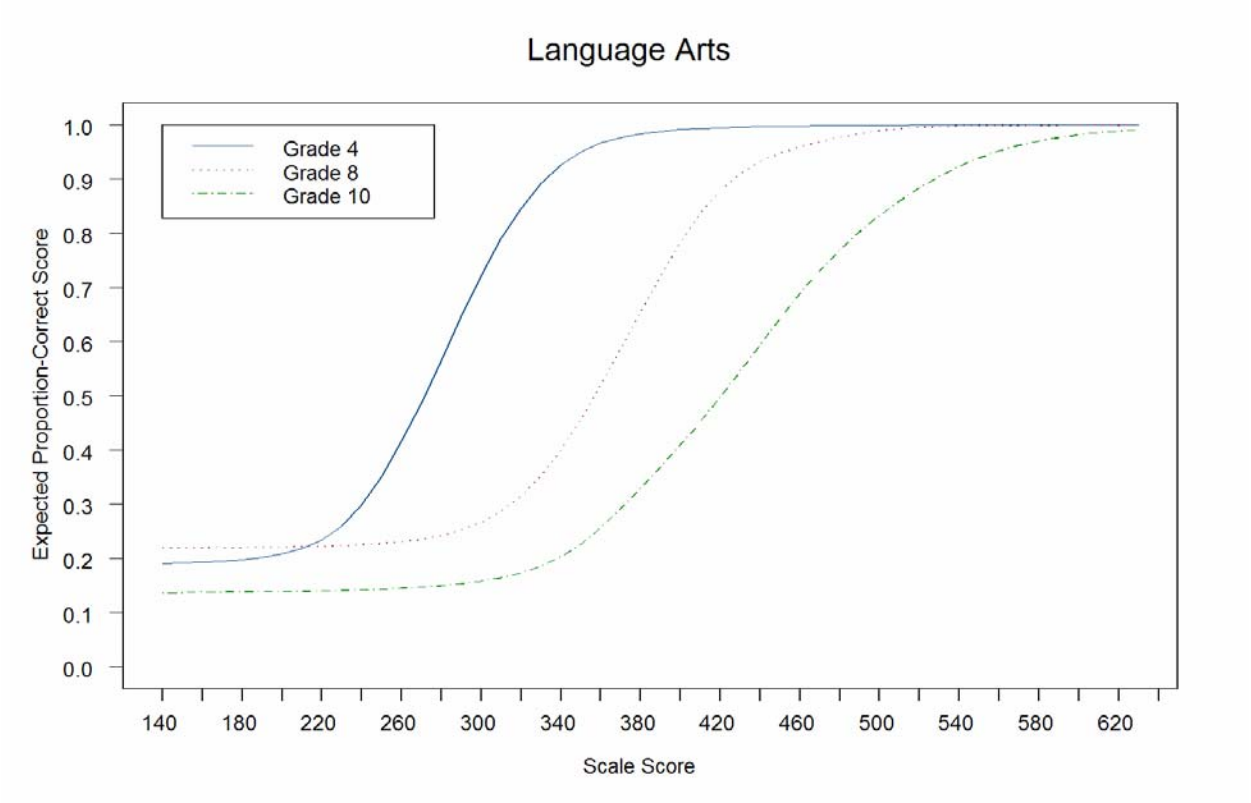


Figure 8-4
TCC Curve for Social Studies Grades 4, 8, 10

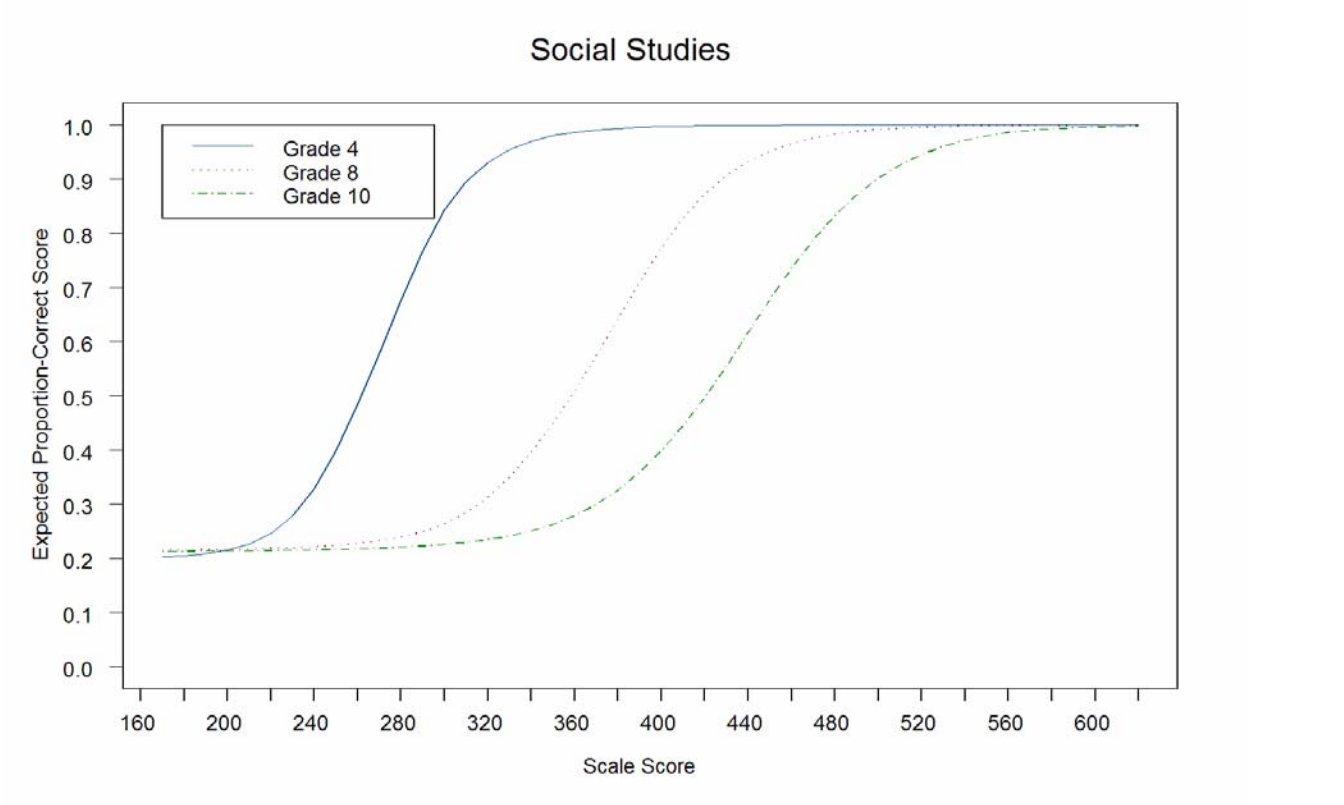


Figure 8-5
TCC Curve for Science Grades 4, 8, 10

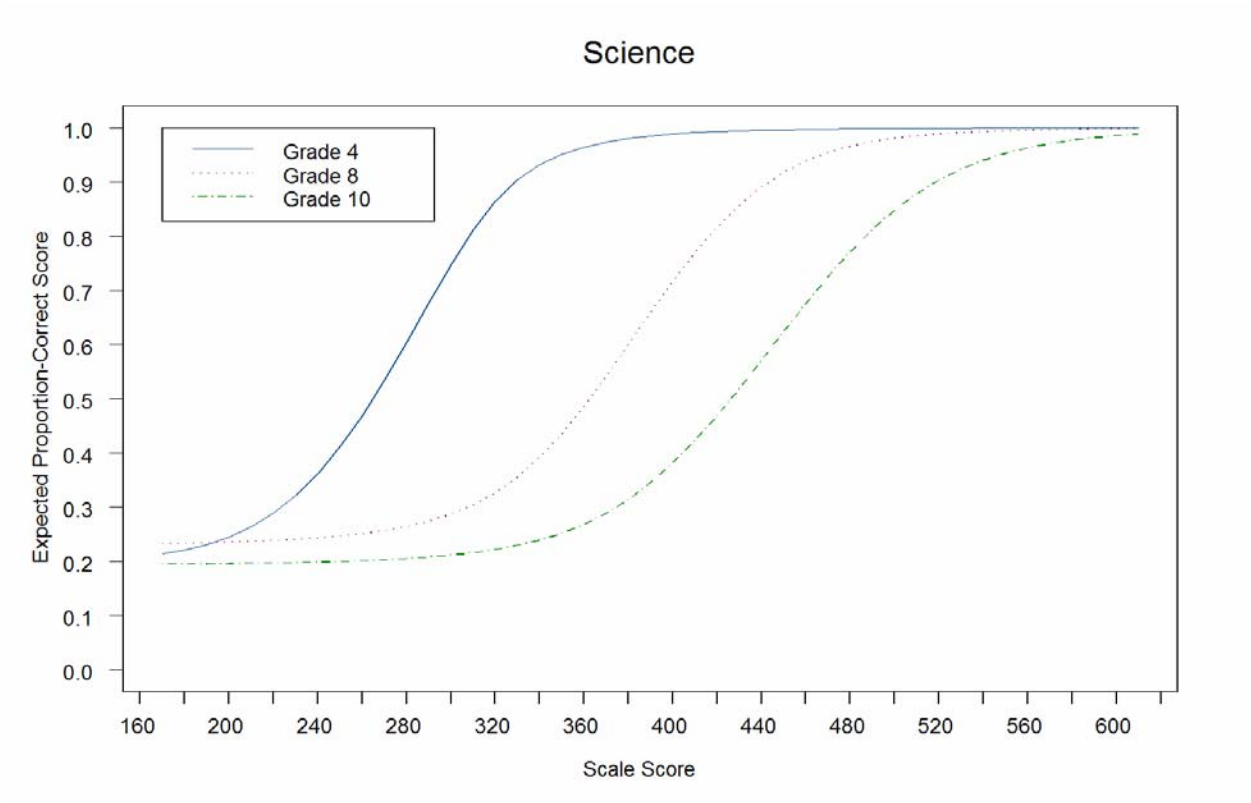


Figure 8-6
SEM Curves, Reading Grades 3-6

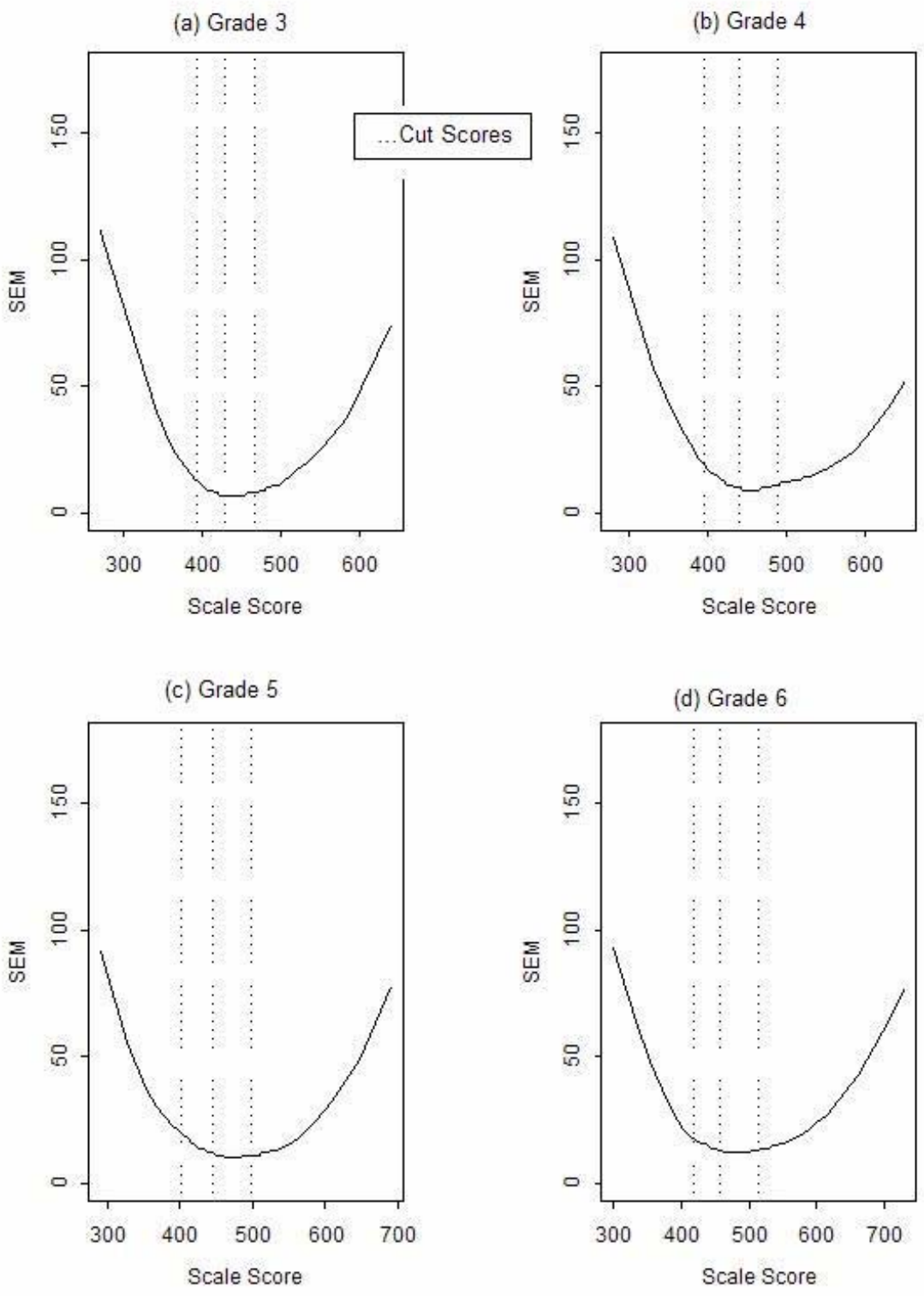


Figure 8-6 Cont'd
SEM Curves, Reading Grades 7, 8, 10

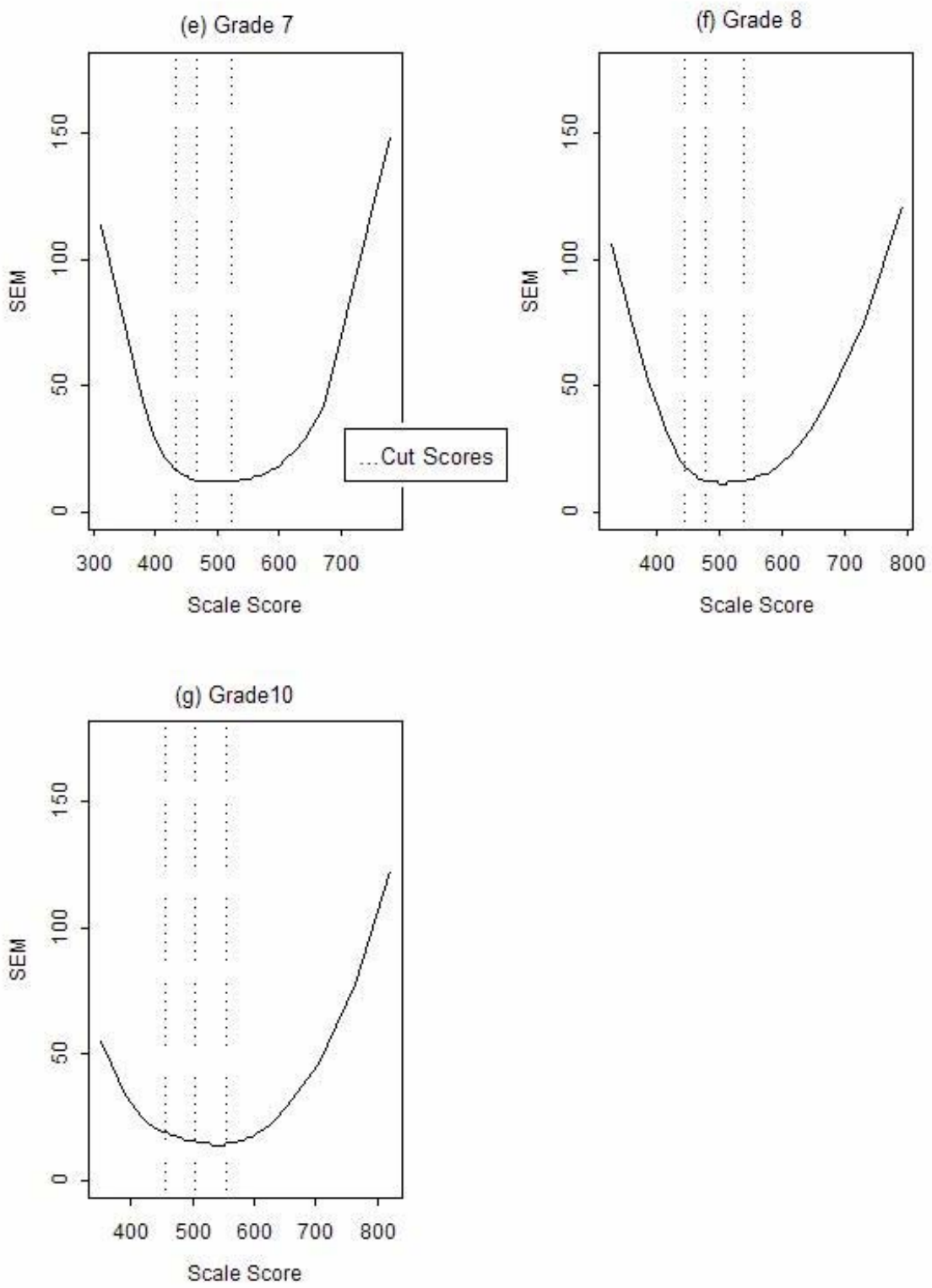


Figure 8-7
SEM Curves, Mathematics Grades 3-6

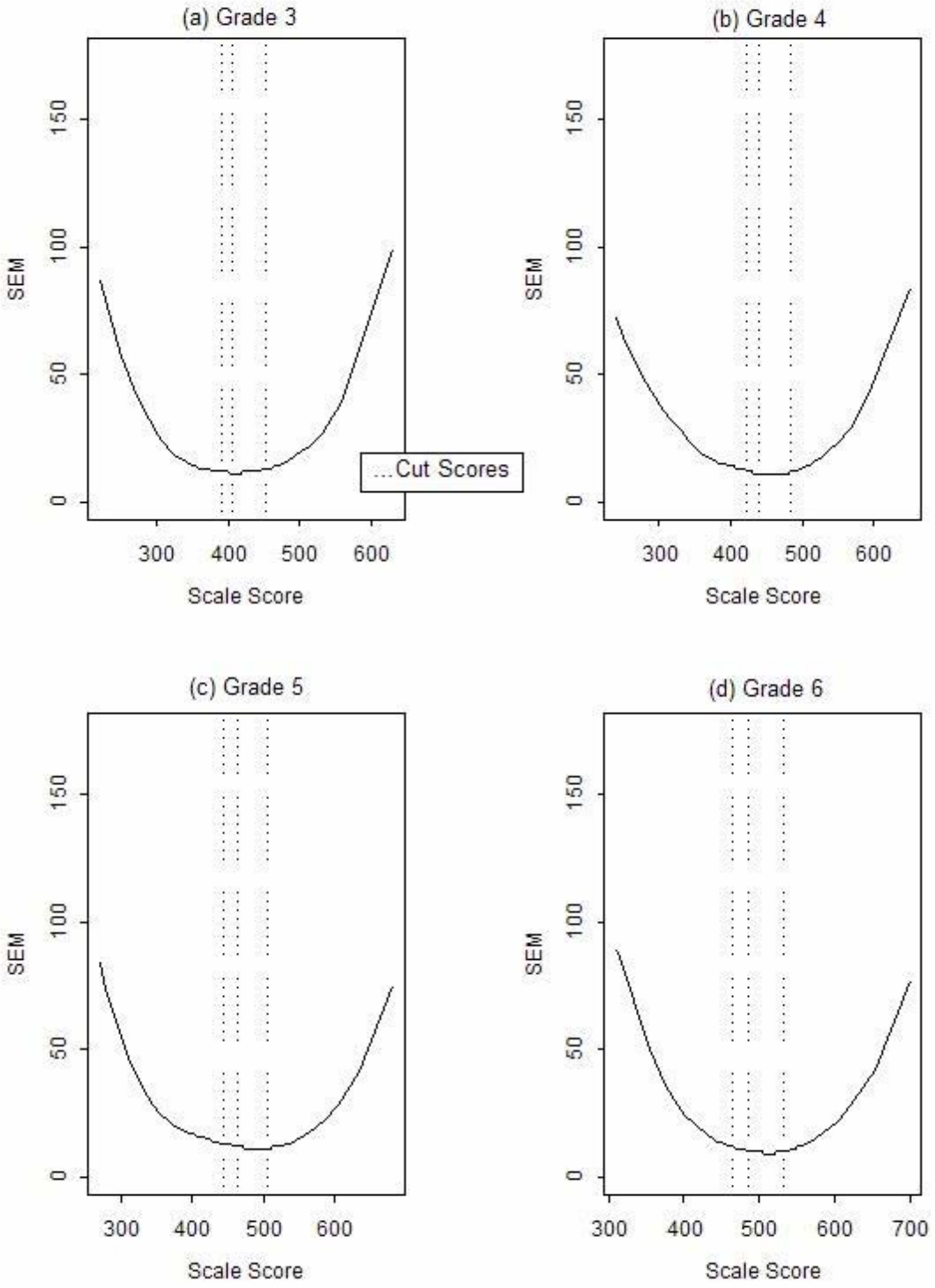


Figure 8-7 Cont'd
SEM Curves, Mathematics Grades 7, 8, 10

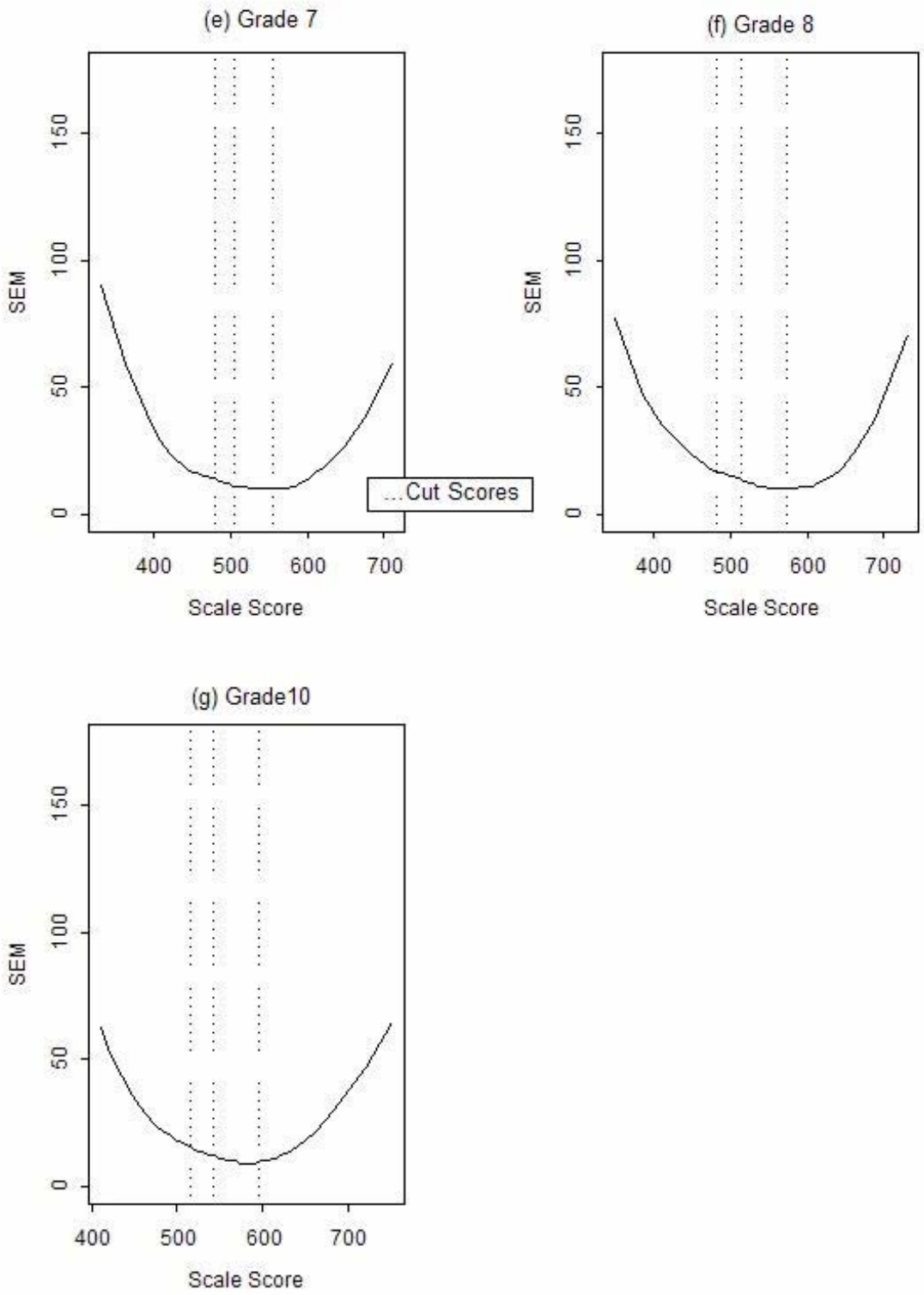


Figure 8-8
SEM Curves, Language Arts Grades 4, 8, 10

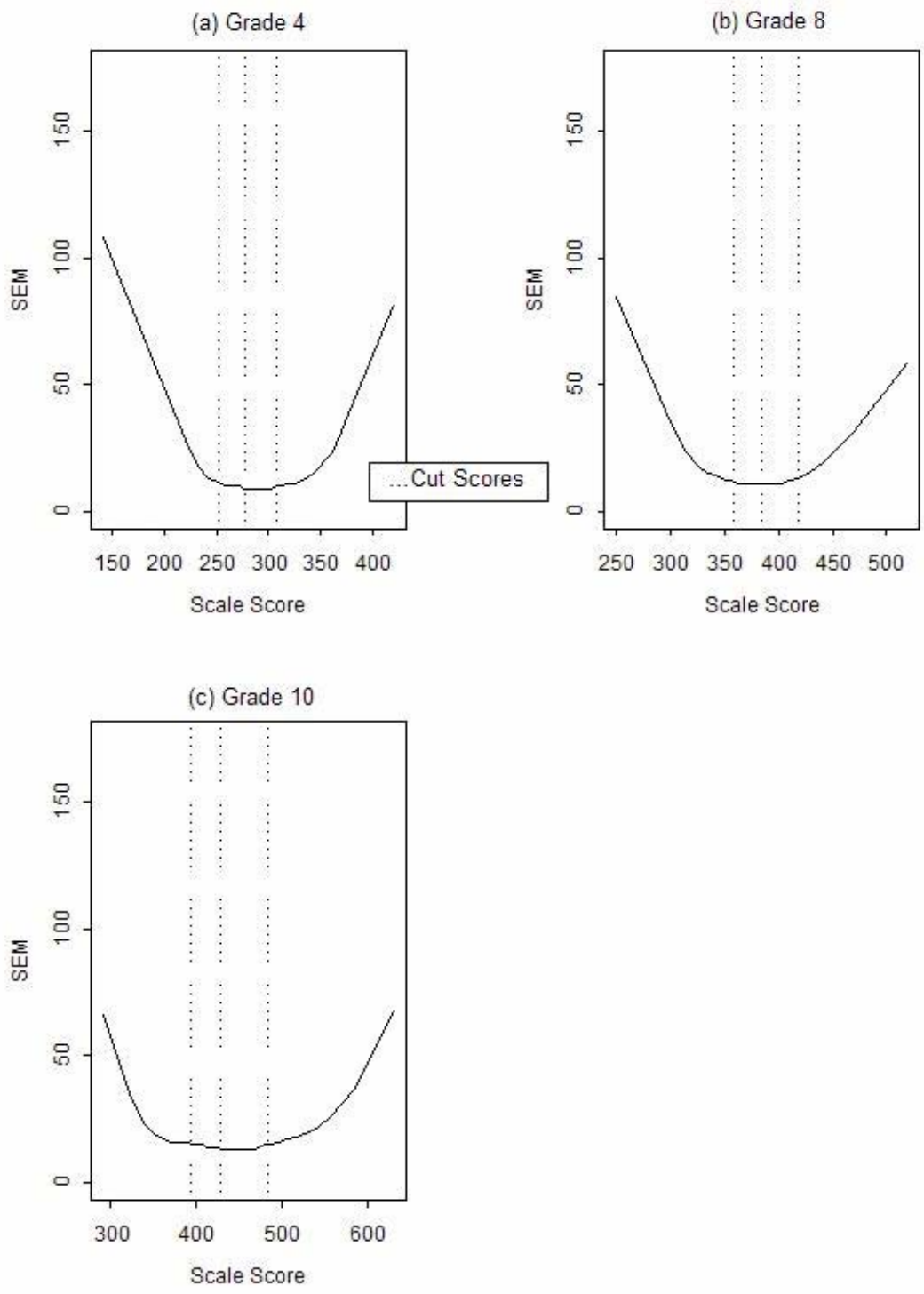


Figure 8-9
SEM Curves, Social Studies Grades 4, 8, 10

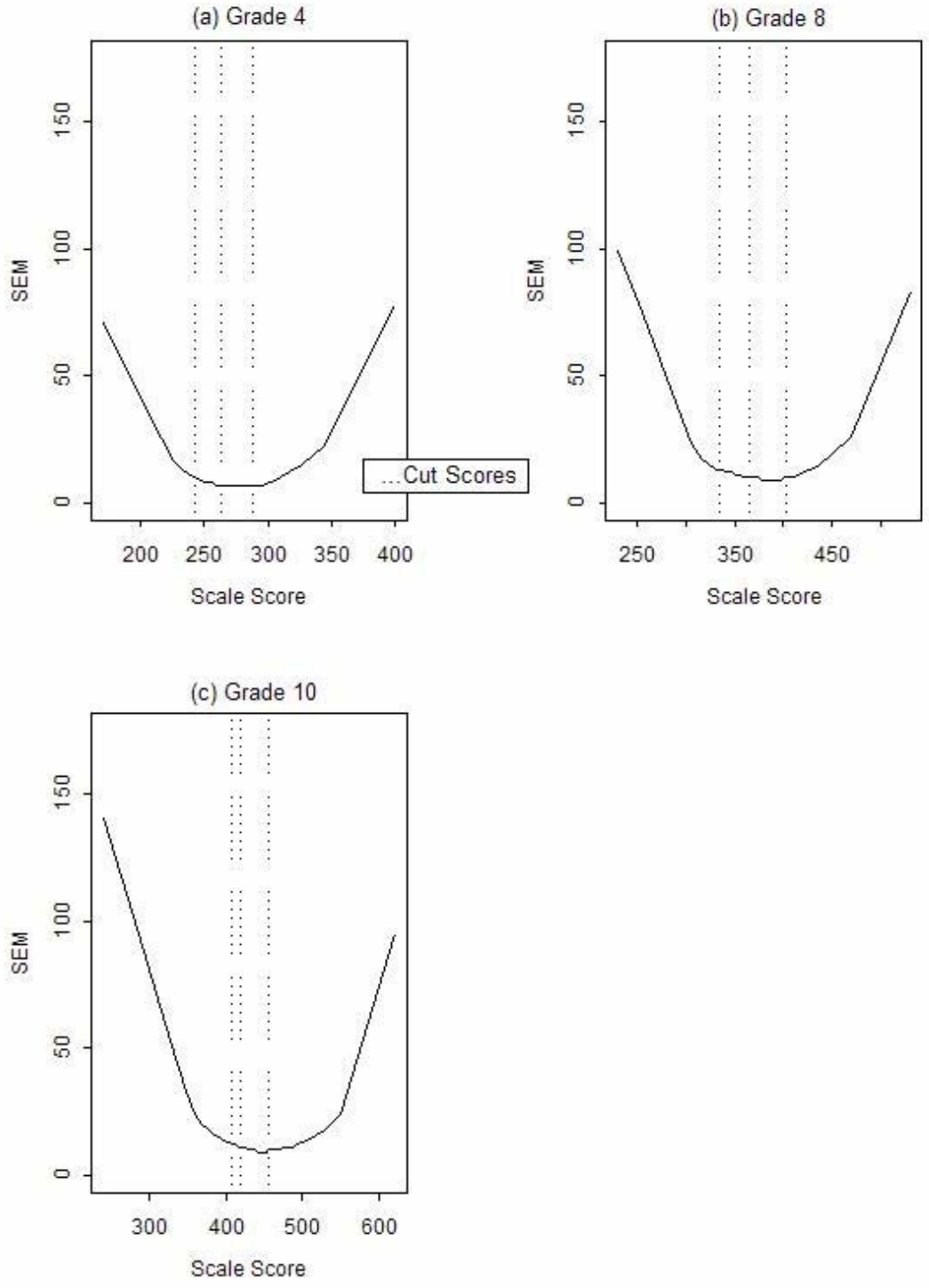


Figure 8-10
SEM Curves, Science Grades 4, 8, 10

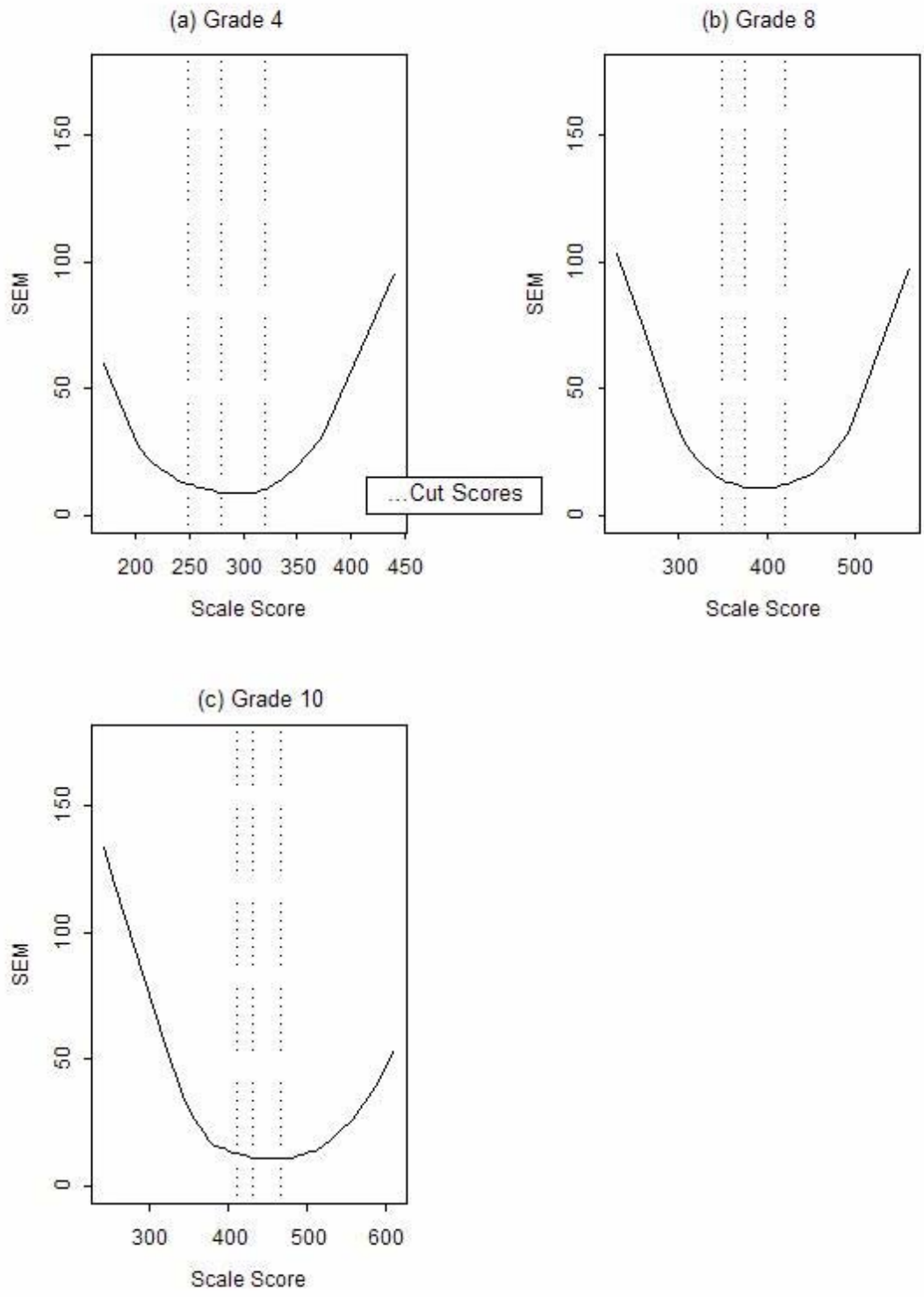


Figure 11-1
Reading Indices for Classification Consistency and Classification Accuracy

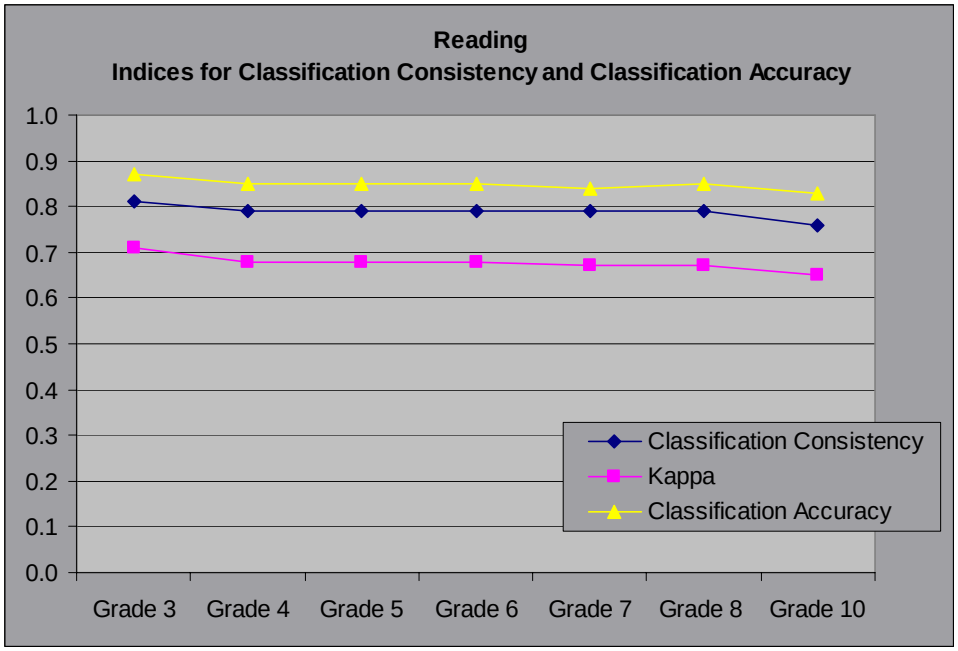


Figure 11-2
Mathematics Indices for Classification Consistency and Classification Accuracy

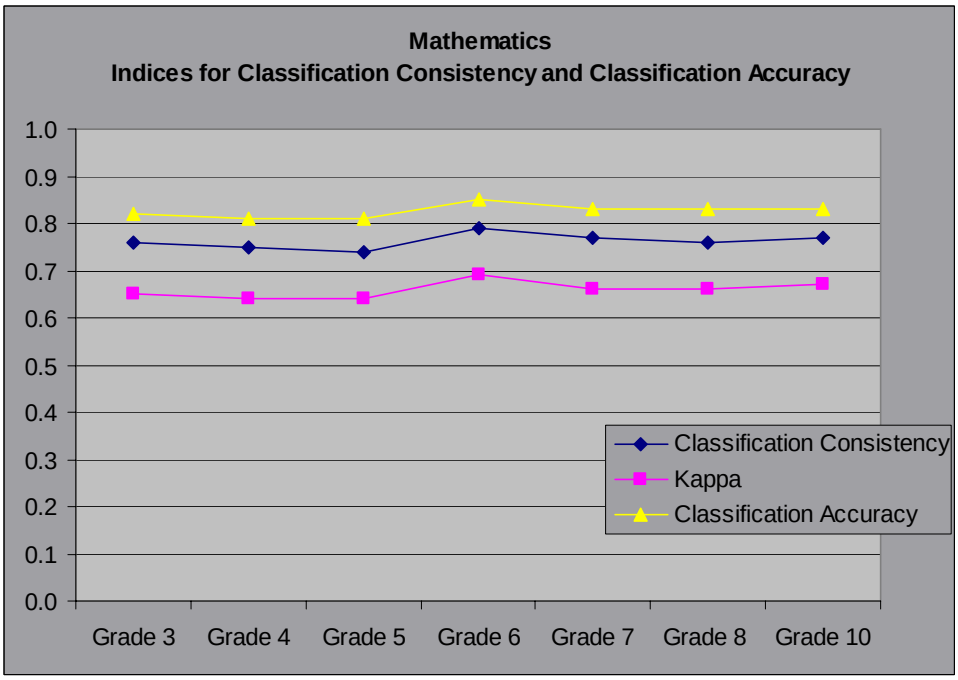


Figure 11-3
Language Arts Indices for Classification Consistency and Classification Accuracy

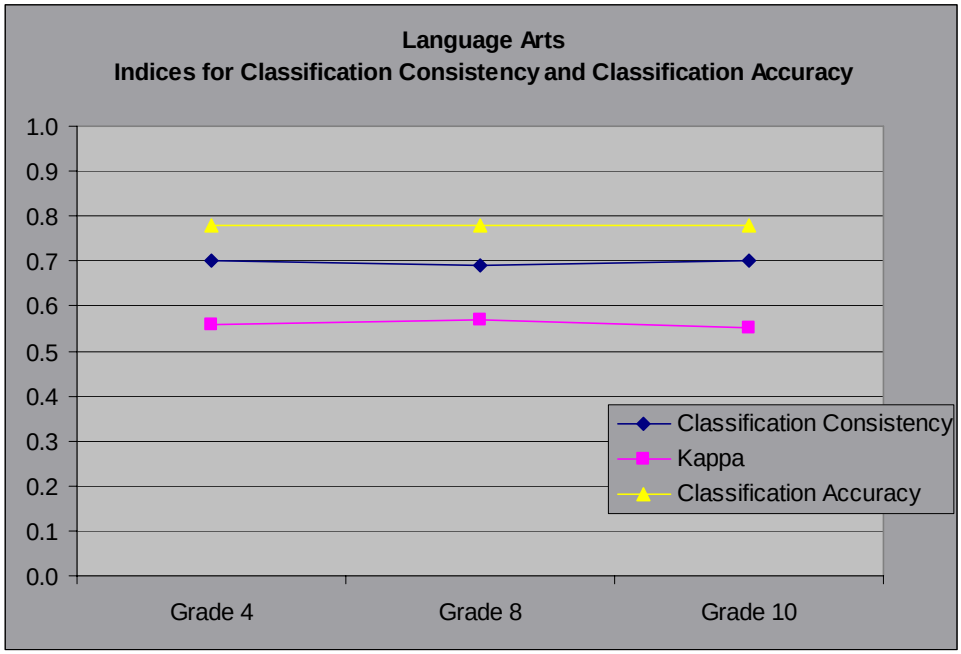


Figure 11-4
Social Studies Indices for Classification Consistency and Classification Accuracy

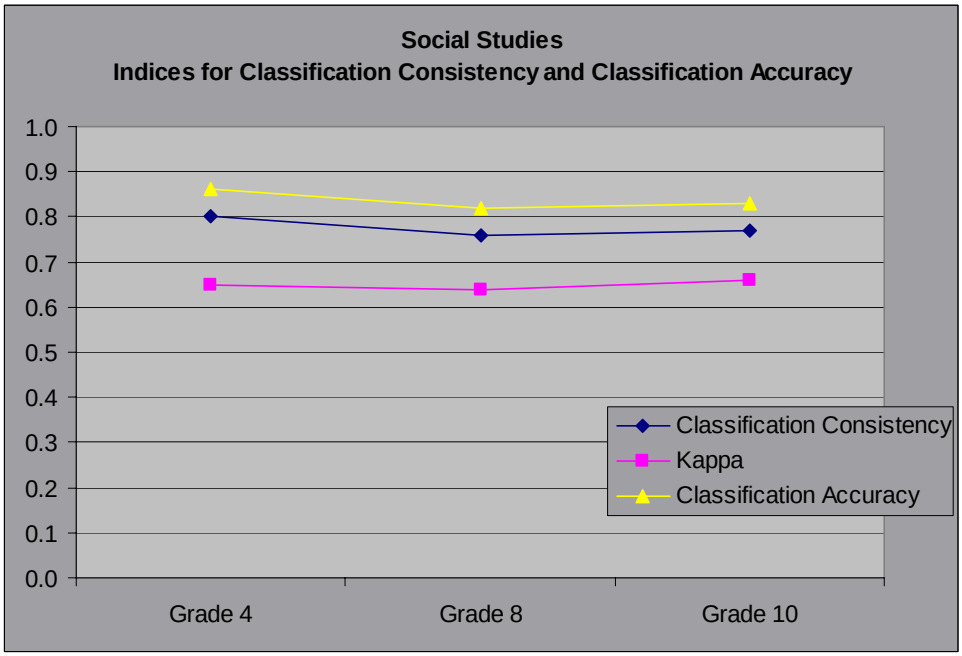


Figure 11-5
Science Indices for Classification Consistency and Classification Accuracy

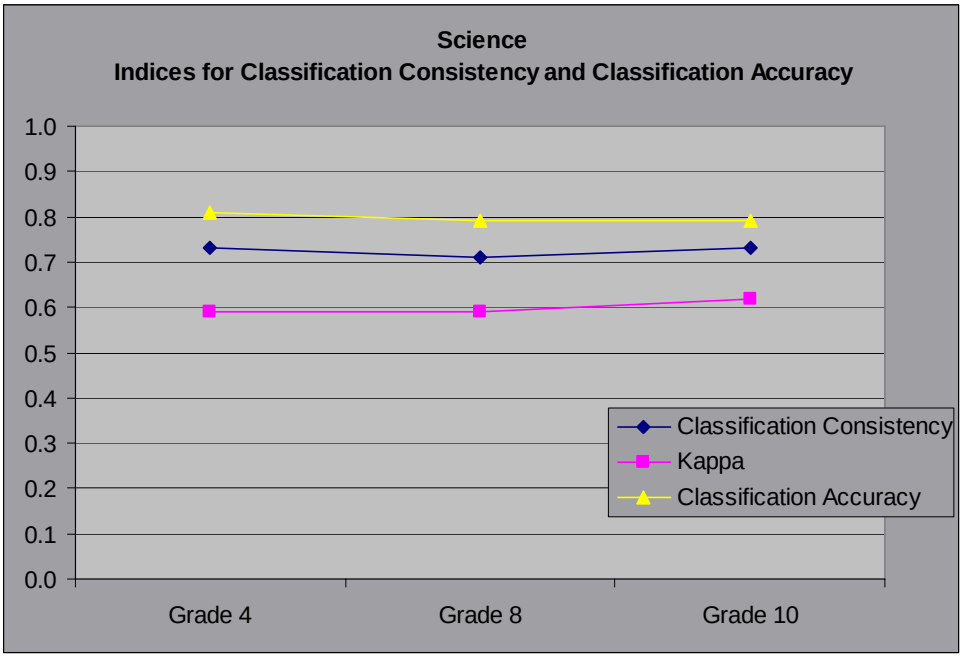


Figure 12-1
Cut Scores for Reading

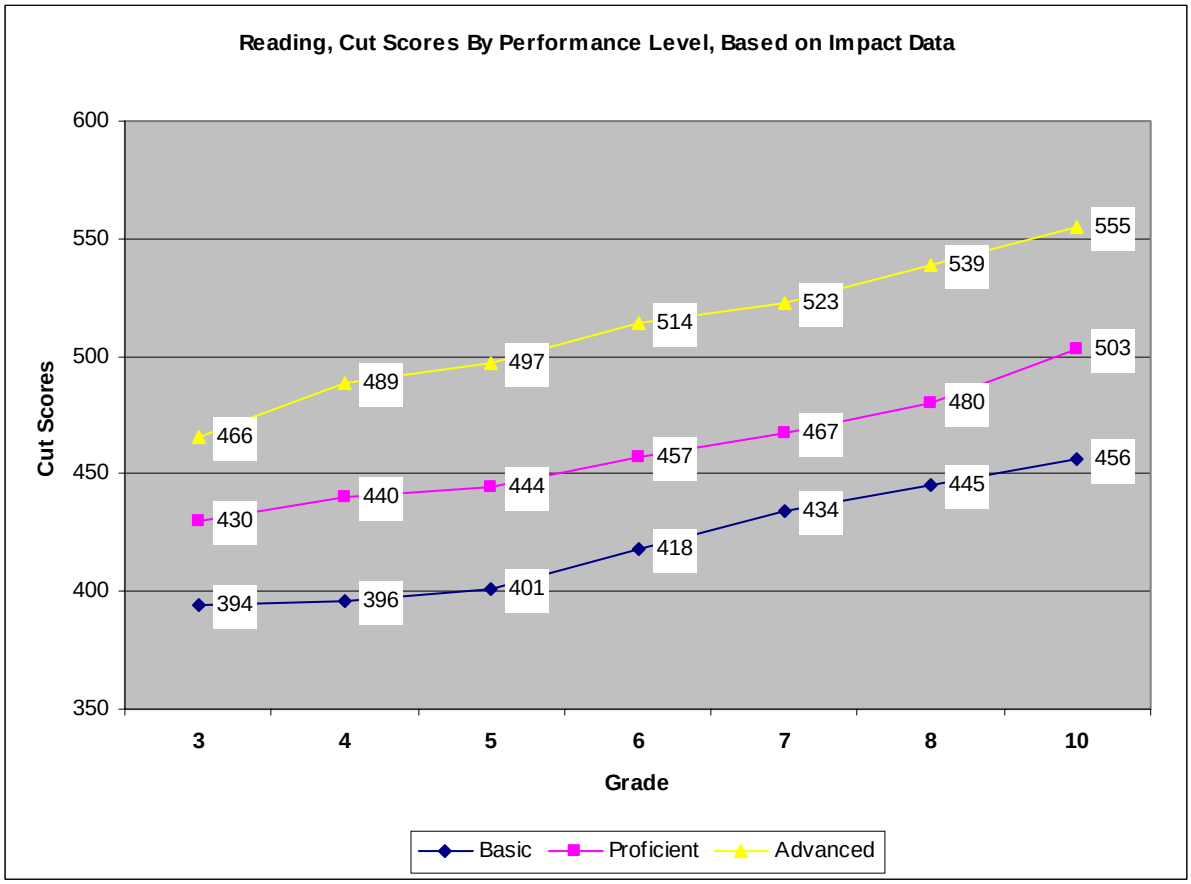


Figure 12-2
Cut Scores for Mathematics

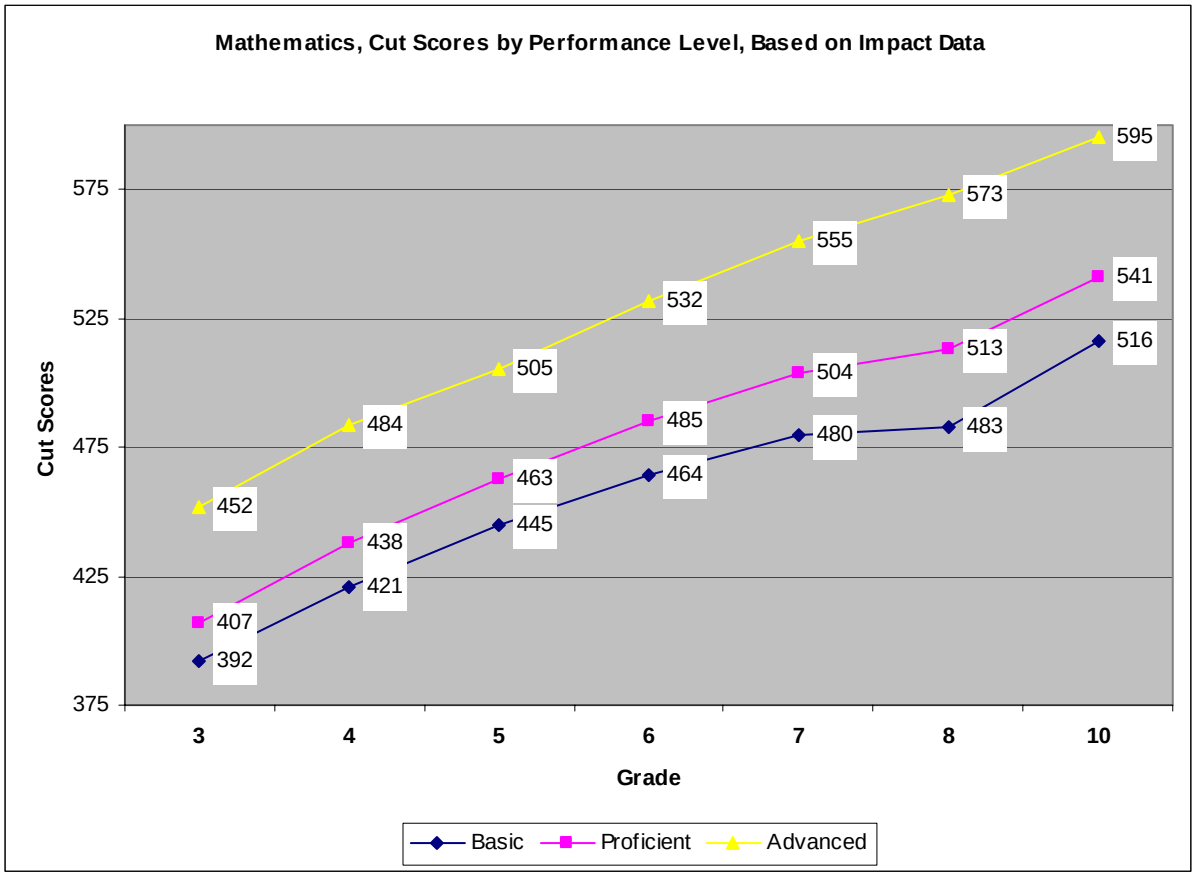


Figure 12-3
Cut Scores for Language Arts

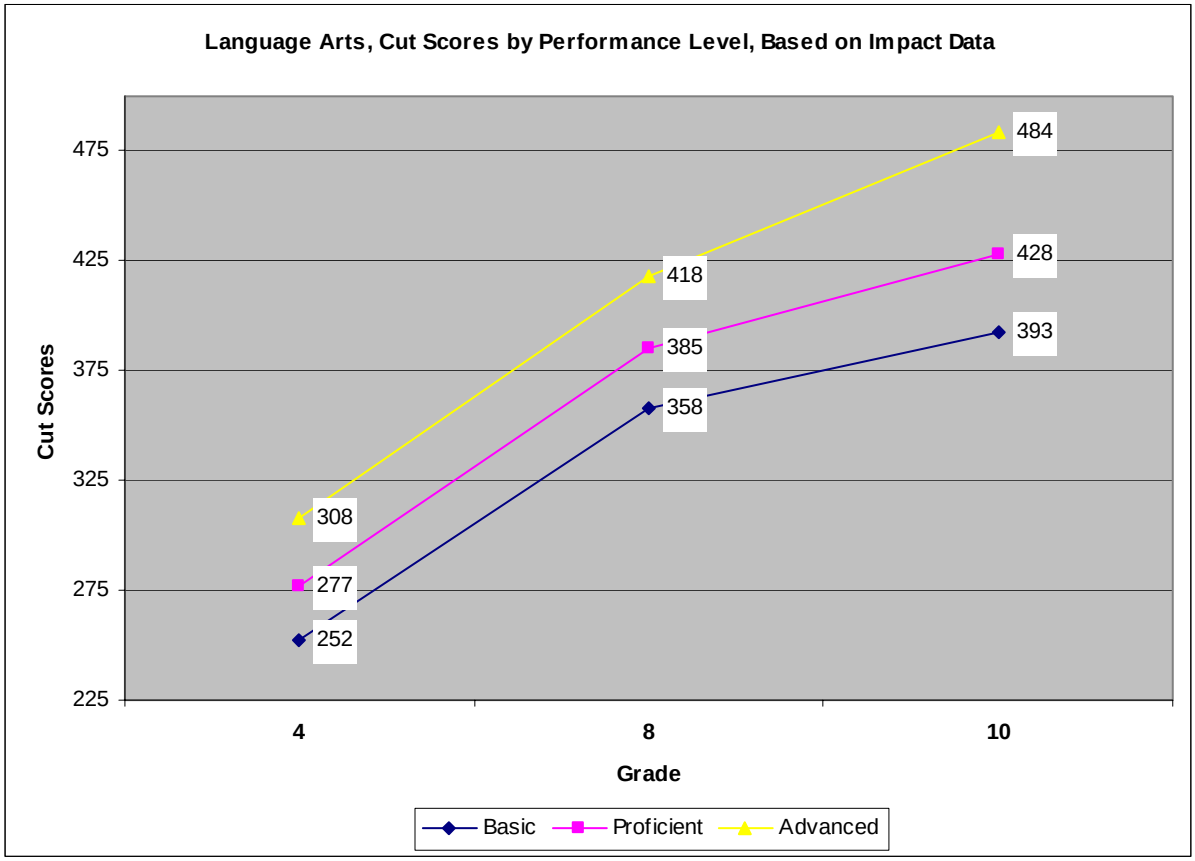


Figure 12-4
Cut Scores for Social Studies

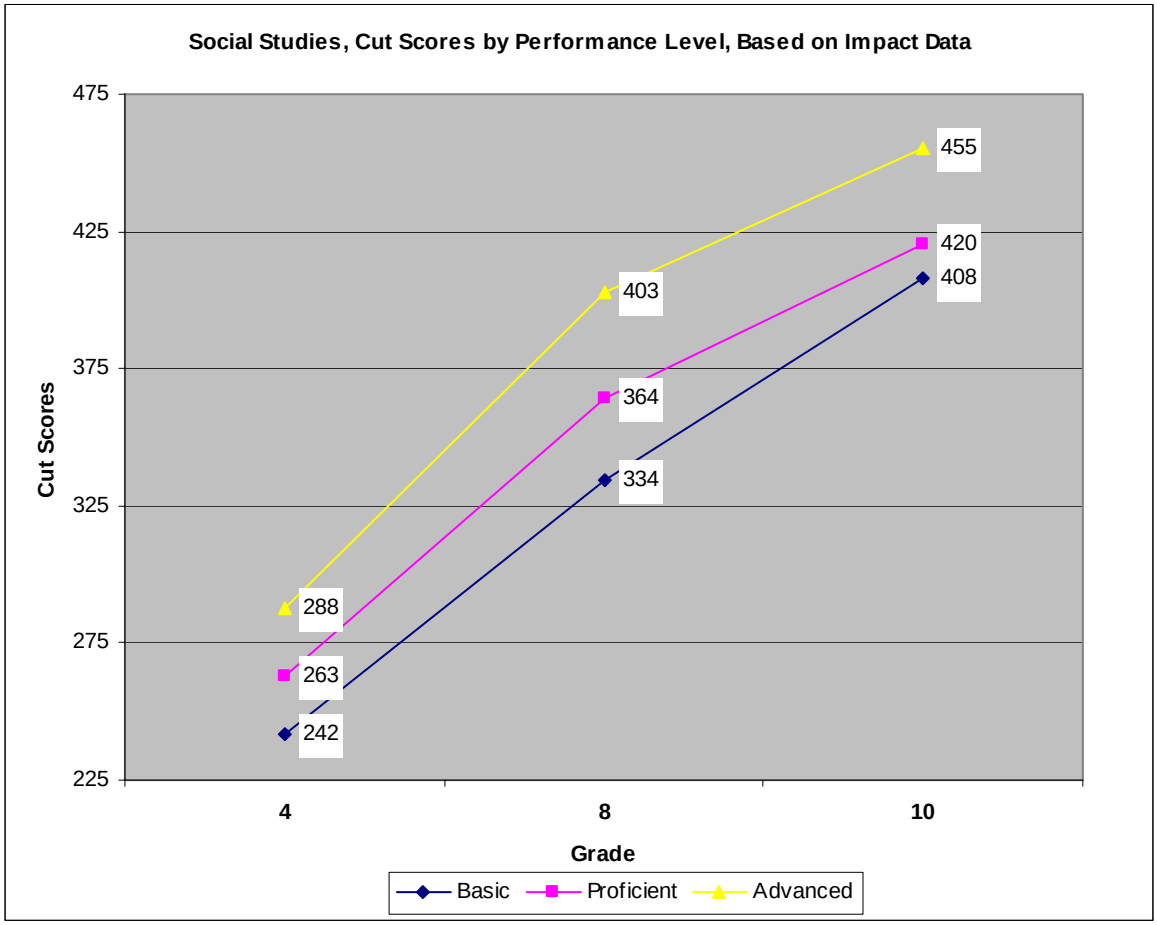


Figure 12-5
Cut Scores for Science

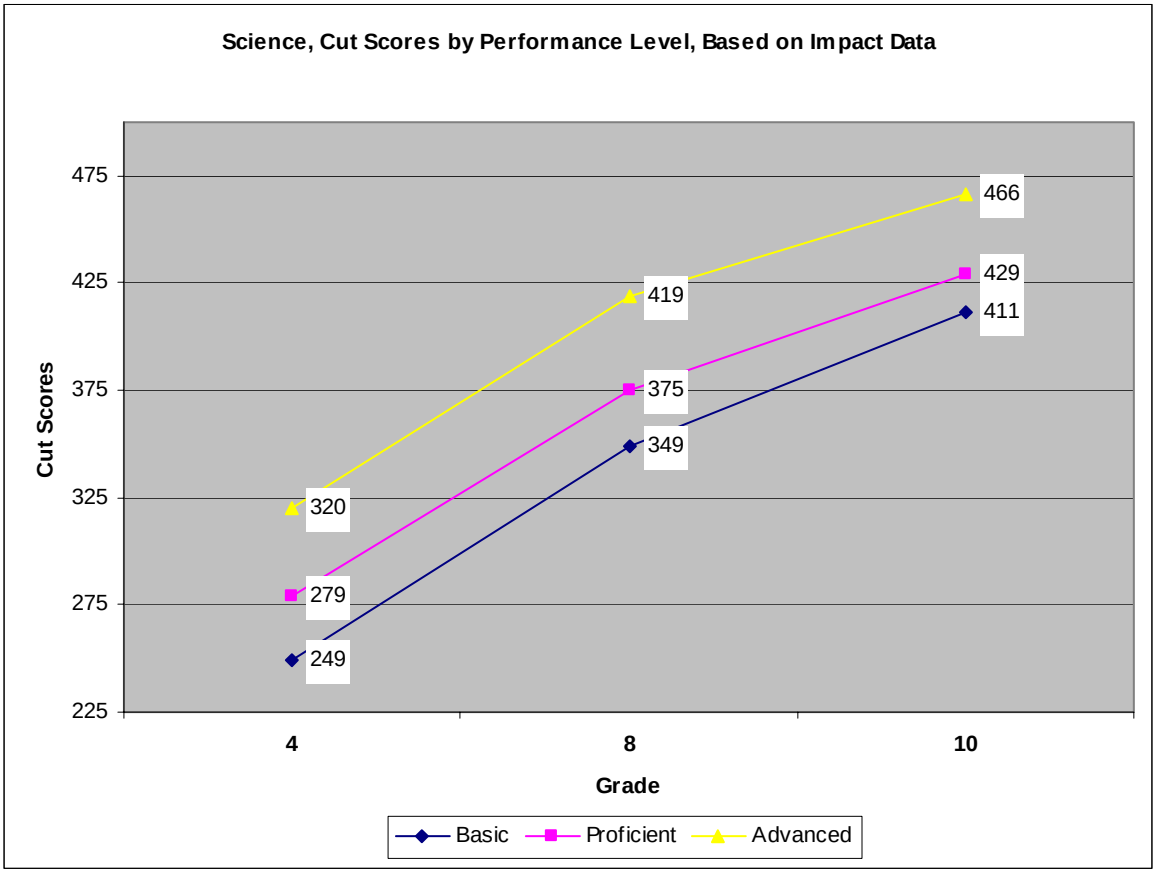


Figure 12-6
Percent of Students for Reading

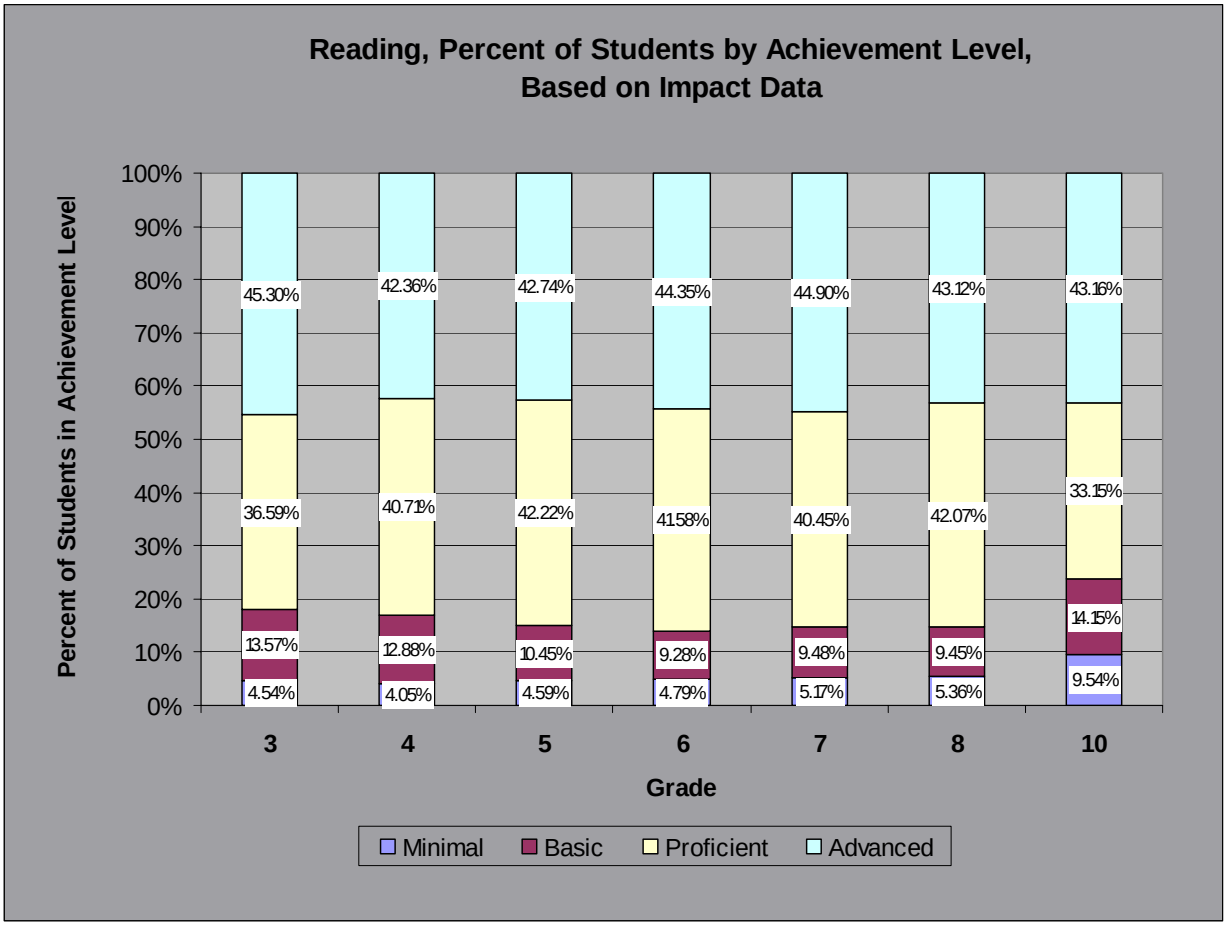


Figure 12-7
Percent of Students for Mathematics

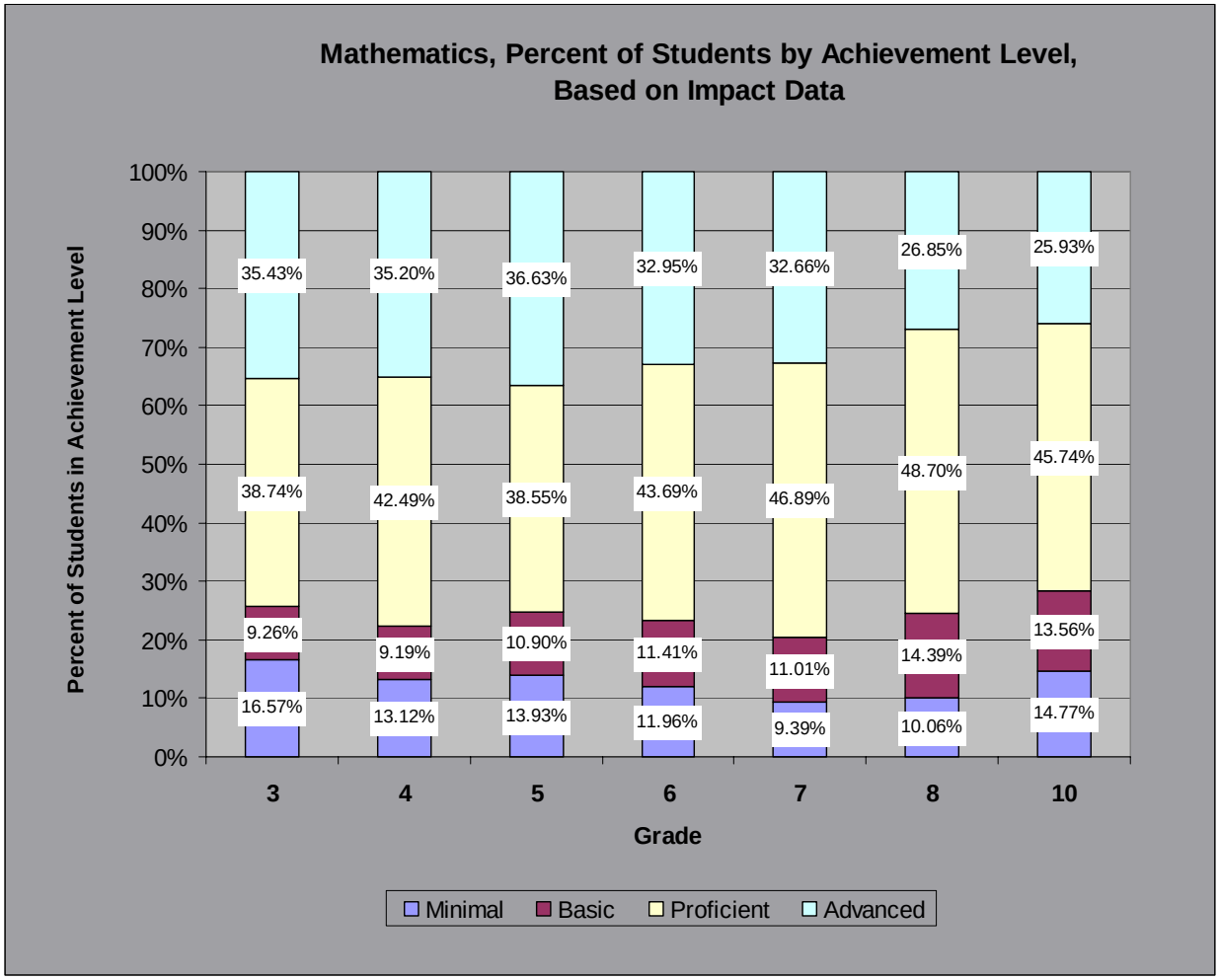


Figure 12-8
Percent of Students for Language Arts

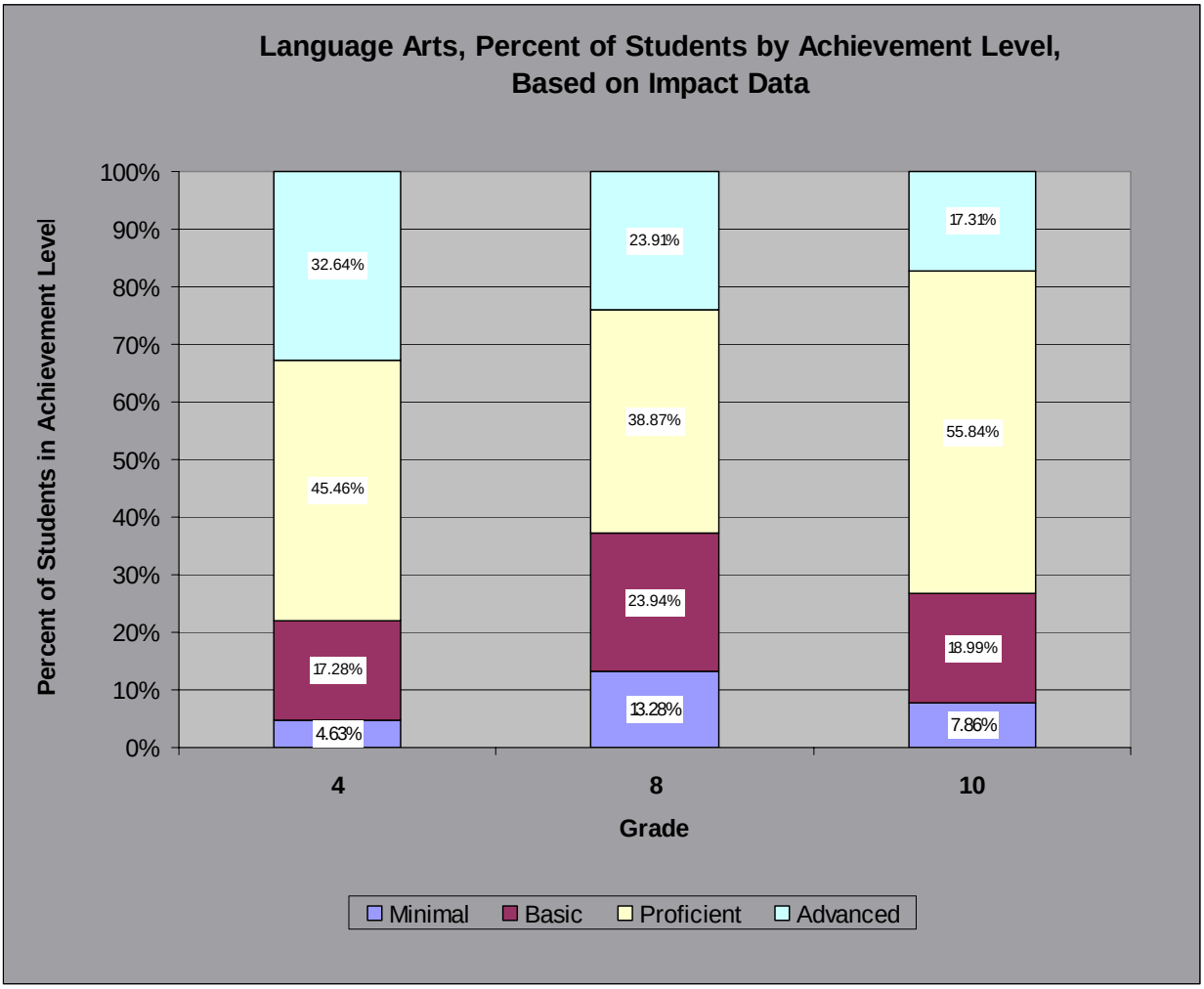


Figure 12-9
Percent of Students for Social Studies

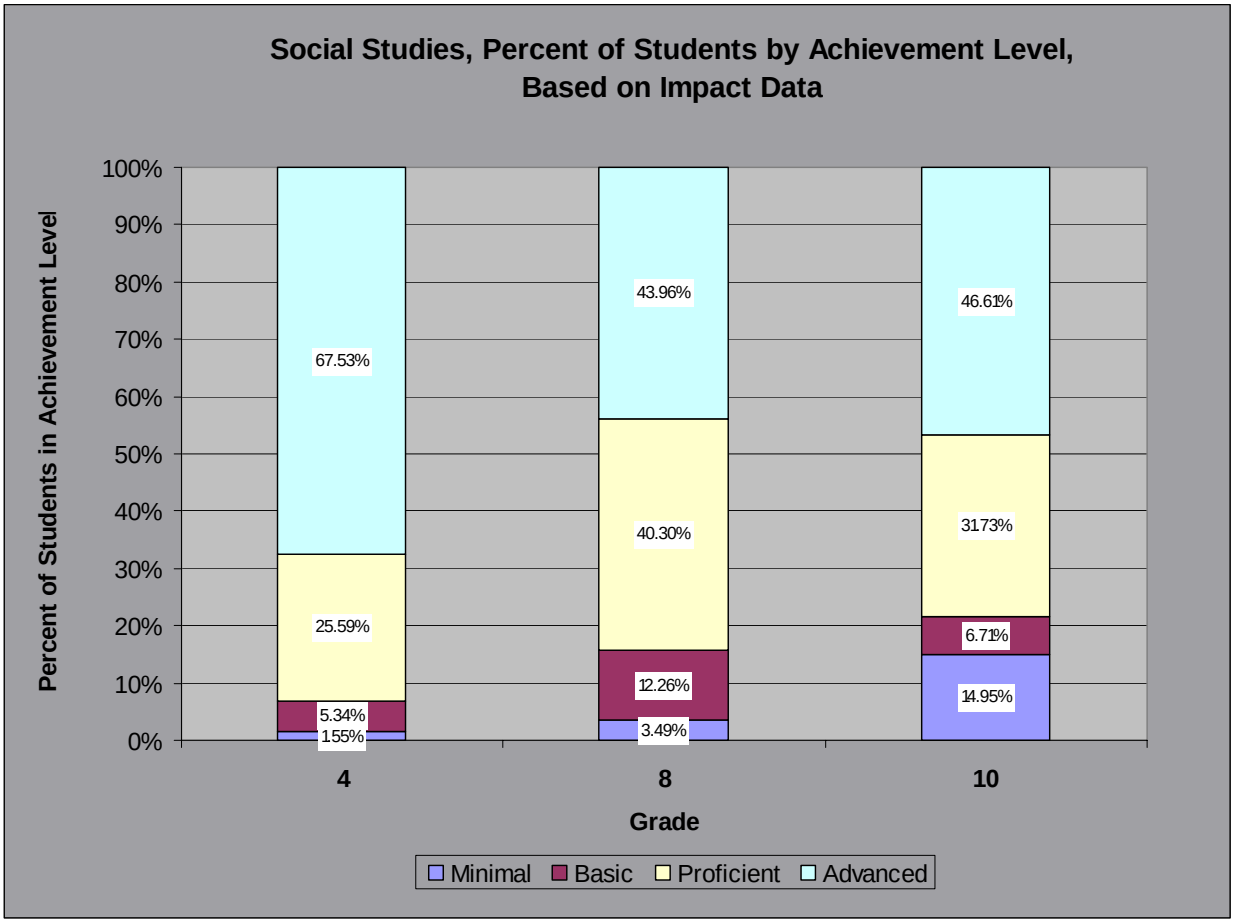
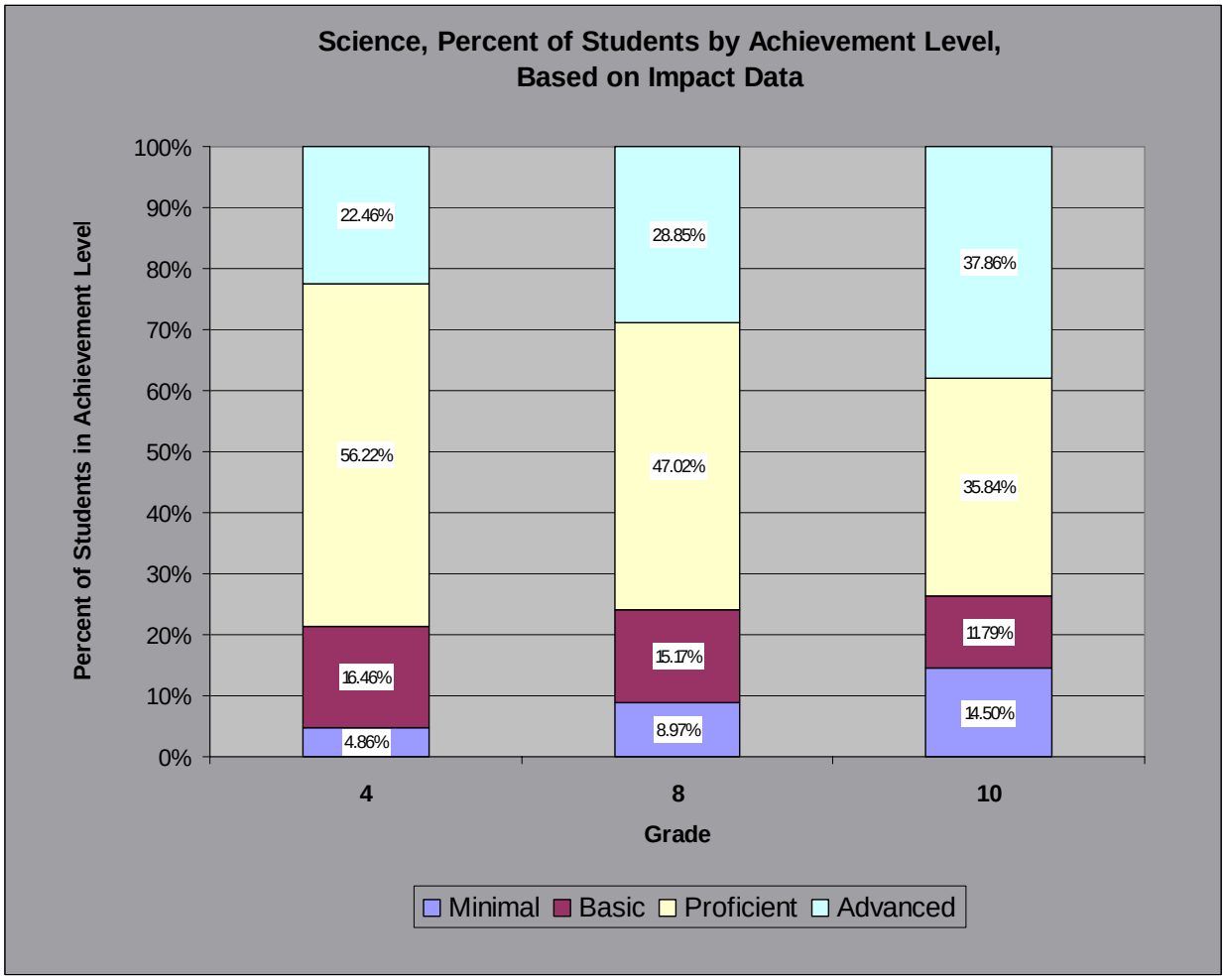


Figure 12-10
Percent of Students for Science



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 tow 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.

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., and Novick. M. R. (1973). Toward an integration of theory and

method for criterion-referenced tests. *Journal of Educational Measurement*, *10*, 159-96.

Kolen, M and Kim, D. (2004). Personnel Correspondence

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.

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.

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., and Algina, J. (1974). Reliability of criterion referenced tests: A decision theoretic formulation. *Journal of Educational Measurement*, *11*, 263-268.

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.

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.