

Oklahoma School Testing Program
Oklahoma Core Curriculum Tests
Grades 3-8
Technical Report
Spring 2009 Administration

September, 2009

Table of Contents

INTRODUCTION.....	1
CHAPTER I. OVERVIEW OF THE OCCT	2
1.1 Skills Assessed by the OCCT	2
1.2 Test Development Procedures	3
1.3 Configuration of the Tests	3
CHAPTER II. STATISTICAL ANALYSIS.....	23
2.1 Data Files for Statistical Analysis.....	23
2.2 Analysis of the Multiple-Choice Tests	23
2.2.1 Classical Item Analyses	23
2.2.2 Differential Item Functioning Analyses.....	26
2.2.3 Item Calibration and Equating	28
2.2.4 Raw Score to Scaled Score Conversion.....	30
2.2.5 Test Score Reliability.....	30
Coefficient Alpha.....	30
Standard Error of Measurement.....	31
Conditional Standard Error of Measurement.....	32
Reliability of Performance-Level Classification Decisions.....	32
2.2.6 Validity	34
2.3 Analysis of the Writing Tests	40
2.3.1 Prompt Scoring Formula.....	40
2.3.2 Statistical Adjustments to Scale the Writing Scores.....	41
Adjustment for Prompt Difficulty and Rater-Year Effects.....	41
Adjustment for Rater-Year Effects	41
A Compound Adjustment	42
2.3.3 Rater Agreement for Operational Writing Prompts.....	44
CHAPTER III. STATE RESULTS.....	46
CHAPTER IV. PERFORMANCE STANDARDS	67
REFERENCES.....	68
APPENDIX A. DATA REVIEW RESULTS.....	70
APPENDIX B. RAW-TO-SCALED SCORE CONVERSION TABLES AND FREQUENCY DISTRIBUTIONS	72

List of Tables

Table 1.3.A	
Number of Operational and Field-test Items by Content Area and Grade	3
Table 1.3.B	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 3 Reading	4
Table 1.3.C	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 4 Reading	5
Table 1.3.D	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 5 Reading	6
Table 1.3.E	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 6 Reading	7
Table 1.3.F	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 7 Reading	8
Table 1.3.G	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 8 Reading	9
Table 1.3.H	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 3 Mathematics	10
Table 1.3.I	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 4 Mathematics	11
Table 1.3.J	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 5 Mathematics	12
Table 1.3.K	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 6 Mathematics	13
Table 1.3.L	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 7 Mathematics	14
Table 1.3.M	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 8 Mathematics	15
Table 1.3.N.1	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 5 Science Process Standards	16
Table 1.3.N.2	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 5 Science Content Standards	17
Table 1.3.O.1	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 8 Science Process Standards	18

Table 1.3.O.2	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 8 Science Content Standards.....	19
Table 1.3.P	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 5 Social Studies.....	20
Table 1.3.Q	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 7 Geography.....	21
Table 1.3.R	
2009 <i>PASS</i> Blueprint and Actual Item Counts: Grade 8 U.S. History	22
Table 2.2.A	
Classical Item and Test Analyses Summary for the Operational Forms	25
Table 2.2.B	
Item Analyses Summary for Field-test Items	26
Table 2.2.C	
Number of Anchor Items by Subject and Grade.....	30
Table 2.2.D	
Conditional Standard Errors of Measurement for Each Achievement Level Cut Score	33
Table 2.2.E	
Estimates of the Reliability of Decisions for Specified Cut Scores.....	34
Table 2.2.G	
Standards Intercorrelation: Grade 4 Reading.....	36
Table 2.2.H	
Standards Intercorrelation: Grade 5 Reading.....	36
Table 2.2.I	
Standards Intercorrelation: Grade 6 Reading.....	36
Table 2.2.J	
Standards Intercorrelation: Grade 7 Reading.....	36
Table 2.2.K	
Standards Intercorrelation: Grade 8 Reading.....	37
Table 2.2.L	
Standards Intercorrelation: Grade 3 Mathematics	37
Table 2.2.M	
Standards Intercorrelation: Grade 4 Mathematics	37
Table 2.2.N	
Standards Intercorrelation: Grade 5 Mathematics	38

Table 2.2.O	
Standards Intercorrelation: Grade 6 Mathematics	38
Table 2.2.P	
Standards Intercorrelation: Grade 7 Mathematics	38
Table 2.2.Q	
Standards Intercorrelation: Grade 8 Mathematics	38
Table 2.2.R	
Standards Intercorrelation: Grade 5 Science.....	39
Table 2.2.S	
Standards Intercorrelation: Grade 8 Science.....	39
Table 2.2.T	
Standards Intercorrelation: Grade 5 Social Studies	39
Table 2.2.U	
Standards Intercorrelation: Grade 7 Geography	39
Table 2.2.V	
Standards Intercorrelation: Grade 8 U. S. History	40
Table 2.3.A	
Weights Assigned to Writing Analytic Traits.....	41
Table 2.3.B	
Scaled Score Ranges for Each Achievement Level.....	42
Table 2.3.C	
Sample Means and Standard Deviations Used for Calculating Constants	43
Table 2.3.D	
Grade 5 Writing Results.....	43
Table 2.3.E	
Grade 8 Writing Results.....	44
Table 2.3.F	
Inter-rater Agreement for Operational Writing Prompts	45
Table 3.1	
Means and Standard Deviations of Students' Raw Scores and Scaled Scores.....	46
Table 3.2	
Percentage of Students Performing within Each Achievement Category in 2007 to 2009	47
Table 3.3	
Subgroup Results: Grade 3 Reading	48

Table 3.4	
Subgroup Results: Grade 4 Reading.....	49
Table 3.5	
Subgroup Results: Grade 5 Reading.....	50
Table 3.6	
Subgroup Results: Grade 6 Reading.....	51
Table 3.7	
Subgroup Results: Grade 7 Reading.....	52
Table 3.8	
Subgroup Results: Grade 8 Reading.....	53
Table 3.9	
Subgroup Results: Grade 3 Mathematics.....	54
Table 3.10	
Subgroup Results: Grade 4 Mathematics.....	55
Table 3.11	
Subgroup Results: Grade 5 Mathematics.....	56
Table 3.12	
Subgroup Results: Grade 6 Mathematics.....	57
Table 3.13	
Subgroup Results: Grade 7 Mathematics.....	58
Table 3.14	
Subgroup Results: Grade 8 Mathematics.....	59
Table 3.15	
Subgroup Results: Grade 5 Science.....	60
Table 3.16	
Subgroup Results: Grade 8 Science.....	61
Table 3.17	
Subgroup Results: Grade 5 Social Studies.....	62
Table 3.18	
Subgroup Results: Grade 7 Geography.....	63
Table 3.19	
Subgroup Results: Grade 8 History.....	64
Table 3.20	
Subgroup Results: Grade 5 Writing.....	65

Table 3.21	
Subgroup Results: Grade 8 Writing.....	66
Table 4.1 Final Scaled Score Ranges for Reading and Mathematics	67
Table A.1	
Data Review Results	71
Table B.1	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 3 Reading	73
Table B.2	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 4 Reading	74
Table B.3	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 5 Reading	75
Table B.4	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 6 Reading	76
Table B.5	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 7 Reading	77
Table B.6	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 8 Reading	78
Table B.7	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 3 Mathematics	79
Table B.8	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 4 Mathematics	80
Table B.9	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 5 Mathematics	81
Table B.10	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 6 Mathematics	82
Table B.11	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 7 Mathematics	83
Table B.12	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 8 Mathematics	84
Table B.13	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 5 Science	85
Table B.14	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 8 Science	86
Table B.15	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 5 Social Studies	87

Table B.16	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 7 Geography	89
Table B.17	
Raw-to-Scaled Score Table and Frequency Distribution: Grade 8 U.S. History.....	90
Table B.18	
Composite Score Frequency Distribution: Grade 5 Writing.....	91
Table B.19	
Composite Score Frequency Distribution: Grade 8 Writing.....	92

INTRODUCTION

The Oklahoma Core Curriculum Test (OCCT) is a component of the Oklahoma School Testing Program (OSTP) administered in Grades 3 through 8. It is a transparent, standard-based, criterion-referenced assessment system designed to monitor student achievement of the Oklahoma *Priority Academic Student Skills (PASS)* adopted by the Oklahoma State Board of Education. Currently, the OCCT includes direct Writing assessments in Grades 5 and 8, and Multiple-Choice (MC) assessments of Reading and Mathematics in Grades 3 through 8, Science in Grades 5 and 8, as well as Social Studies in Grade 5, Geography in Grade 7, and U.S. History, Constitution, and Government in Grade 8.

In 2009, the OCCT was administered during the Spring. The Writing assessments were administered on February 25. The Grade 7 Geography and Grade 8 Reading and Mathematics were administered online during an April 9th to May 1st window. The remaining tests were administered during the MC testing window of April 9 – 24.

This technical report outlines the statistical analyses that were carried out in support of the 2009 OCCT. Chapter I provides an overview of the test content and design. Chapter II details the statistical procedures that were carried out in support of the OCCT. These procedures include preliminary item analyses, differential item functioning analyses, calibration and equating, and various miscellaneous analyses. Chapter III presents statewide test results. Chapter IV describes the performance standard setting process and results. Two appendices are provided. Appendix A presents the data review results. Appendix B presents the raw score to scaled score (RS – SS) conversion tables and frequency distribution by grade.

The technical information provided in this report is intended for use by all interested in how the test is evaluated, how the scores are interpreted, and the subsequent educational decisions based on the test results. 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).

CHAPTER I. OVERVIEW OF THE OCCT

The purpose of the Oklahoma Core Curriculum Test (OCCT) is to fulfill accountability requirements and to provide feedback about student mastery of the knowledge and skills delineated in the Oklahoma *Priority Academic Student Skills (PASS)* standards. In the Spring of 2009, the OCCT assessments were administered to all eligible public school students in Grades 3 through 8. The OCCT includes assessments of Reading and Mathematics in Grades 3 through 8, Writing and Science in Grades 5 and 8, as well as Social Studies in Grade 5, Geography in Grade 7, and U.S. History, Constitution, and Government in Grade 8. All tests were designed to measure the Oklahoma *Priority Academic Student Skills (PASS)* adopted by the Oklahoma State Board of Education. The 2009 administration of the OCCT was the fifteenth for students in Grades 5 and 8 and the fifth for students in Grades 3, 4, and 7 (Geography only). This was the fourth operational administration of the Reading and Mathematics tests in Grades 6 and 7.

Data Recognition Corporation (DRC) worked with the Oklahoma State Department of Education (SDE) to construct OCCT test forms aligned to the *PASS* standards. Each test form included a set of operational items used to produce student test scores, and a set of embedded field-test items. The Writing assessments included one extended constructed-response (CR) item. The Reading, Mathematics, Science, Social Studies, Geography and History assessments were composed of multiple-choice (MC) items only. For each content and grade, there were eight forms consisting of a common set of operational items and a unique set of 10 field-test items. Responses to the operational items were used to produce student scores. Responses to the field-test items were used to evaluate the psychometric properties of these newly developed items for possible inclusion on future forms.

The OCCT is an untimed test. The MC tests in Grades 3 – 5 were administered in either one or two sessions. The Writing tests in Grades 5 and 8 and the MC tests in Grades 6 – 8 were administered in one session. With the exception of Grade 7 Geography and Grade 8 Reading and Mathematics, all assessments were administered as paper-and-pencil tests. The Grade 7 Geography and Grade 8 Reading and Mathematics assessments were delivered primarily online, with paper forms available for accommodated administrations and for make-ups.

In the following sections, more information is provided on the skills assessed by the OCCT, test development procedures, and the configuration of the tests.

1.1 Skills Assessed by the OCCT

The standards assessed by each test can be found at <http://sde.state.ok.us/Curriculum/PASS/default.html>. The OCCT measures all *PASS* standards except for content that cannot be appropriately measured within the limitations of a large scale, Multiple-Choice test. For example, the majority of *PASS* standards (listening, reviewing, etc.) are not measured in the ELA assessment. Standards not measured by the OCCT must be assessed by local school districts.

1.2 Test Development Procedures

The items used in the operational 2009 OCCT were selected from the SDE-owned pool of items. All items selected had previously been reviewed and approved by Oklahoma content, bias, and sensitivity review committees. These operational items had been field tested during previous administrations. The field-test statistics for these operational items indicated that the items were of acceptable quality.

For field-test items embedded in the OCCT, DRC assessment specialists selected field test ready items from SDE's item bank, as well as items newly-developed for 2009. A total of 80 items per content/grade were selected for use as embedded field-test items.

1.3 Configuration of the Tests

Table 1.3.A shows the number of operational and field-test items by content area and grade used in the 2009 operational tests. Also shown is the number of operational items included in the anchor set used for equating the 2009 forms to the previously established reporting scale. For all Multiple-Choice tests, each form contained a core set of operational items common across forms and a unique set of field-test items.

Table 1.3.A
Number of Operational and Field-test Items by Content Area and Grade

Content Area	Grade	Number of Forms	Number of Operational Items per Form ^a	Number of Operational Items in Anchor Set ^a	Number of Field-test Items per Form	Total Number of Items per Form	Total Number of Field-test Items Per Grade
Reading	3-8	8	50	19-20 ^b	10	60	80
Writing	5, 8	1	1	0	0	1	0
Mathematics	3-8	8	45	18-20 ^c	10	55	80
Science	5, 8	8	45	16,19	10	55	80
Social Studies	5	8	60	20	10	70	80
Geography	7	8	45	18	10	55	80
U.S. History	8	8	45	19	10	55	80

^a Operational item counts include anchor items.

^b Anchor counts for Reading tests were 20, 20, 19, 20, 20, and 19 in Grades 3 through 8 respectively.

^c Anchor counts for Mathematics tests were 20, 18, 18, 19, 18, and 20 in Grades 3 through 8 respectively.

Tables 1.3.B through 1.3.R provide information drawn from the official 2009 test *PASS* blueprints. These tables show the number of items by content standard specified in the blueprints and the number of items that appeared on the 2009 operational assessments.

Table 1.3.B
2009 *PASS* Blueprint and Actual Item Counts: Grade 3 Reading

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Vocabulary	12	12	24%
Words in Context (2.1)	2-4	3	
Affixes, Roots, and Stems (2.2)	2-4	3	
Synonyms, Antonyms, and Homonyms (2.3)	2-4	3	
Using Resource Materials (2.4)	2-4	3	
Comprehension/Critical Literacy	24	24	48%
Literal Understanding (4.1)	5	6	
Inferences and Interpretation (4.2)	7	7	
Summary and Generalization (4.3)	6	5	
Analysis and Evaluation (4.4)	6	6	
Literature	8	8	16%
Literary Elements (5.2)	4	4	
Figurative Language/Sound Devices (5.3)	4	4	
Research and Information	6	6	12%
Accessing Information (6.1)	6	6	
Total Test	50	50	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.C
2009 PASS Blueprint and Actual Item Counts: Grade 4 Reading

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Vocabulary	12	12	24%
Words in Context (1.1)	4	4	
Affixes, Roots, and Stems (1.2)	4	4	
Synonyms, Antonyms, and Homonyms (1.3)	4	4	
Comprehension/Critical Literacy	23	23	46%
Literal Understanding (3.1)	4	4	
Inferences and Interpretation (3.2)	6	6	
Summary and Generalization (3.3)	7	7	
Analysis and Evaluation (3.4)	6	6	
Literature	9	9	18%
Literary Elements (4.2)	5	6	
Figurative Language/Sound Devices (4.3)	4	3	
Research and Information	6	6	12%
Accessing Information (5.1)	6	6	
Total Test	50	50	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.D
2009 PASS Blueprint and Actual Item Counts: Grade 5 Reading

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Vocabulary	12	12	24%
Words in Context (1.1)	4	4	
Affixes, Roots, and Stems (1.2)	4	4	
Synonyms, Antonyms, and Homonyms (1.3)	4	4	
Comprehension/Critical Literacy	20	20	40%
Literal Understanding (3.1)	4	4	
Inferences and Interpretation (3.2)	4-6	6	
Summary and Generalization (3.3)	4-6	4	
Analysis and Evaluation (3.4)	4-6	6	
Literature	12	12	24%
Literary Genre (4.1)	4	4	
Literary Elements (4.2)	4	4	
Figurative Language/Sound Devices (4.3)	4	4	
Research and Information	6	6	12%
Accessing Information (5.1)	2-4	4	
Interpreting Information (5.2)	2-4	2	
<i>Total Test</i>	<i>50</i>	<i>50</i>	<i>100%</i>

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.E
2009 PASS Blueprint and Actual Item Counts: Grade 6 Reading

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Vocabulary	8	7	16%
Words in Context (1.1)	4	4	
Word Origins (1.2)	4	3	
Comprehension/Critical Literacy	20	20	40%
Literal Understanding (3.1)	4	4	
Inferences and Interpretation (3.2)	4-6	6	
Summary and Generalization (3.3)	4-6	5	
Analysis and Evaluation (3.4)	4-6	5	
Literature	14	15	28%
Literary Genres (4.1)	4	3	
Literary Elements (4.2)	4-6	6	
Figurative Language/Sound Devices (4.3)	4-6	6	
Research and Information	8	8	16%
Accessing Information (5.1)	4	4	
Interpreting Information (5.2)	4	4	
Total Test	50	50	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.F
2009 PASS Blueprint and Actual Item Counts: Grade 7 Reading

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Vocabulary	10	10	20%
Words in Context (1.1)	3-4	4	
Word Origins (1.2)	3-4	3	
Idioms and Comparisons (1.3)	3-4	3	
Comprehension/Critical Literacy	20	20	40%
Literal Understanding (3.1)	4	4	
Inferences and Interpretation (3.2)	4-6	5	
Summary and Generalization (3.3)	4-6	6	
Analysis and Evaluation (3.4)	4-6	5	
Literature	12	12	24%
Literary Genres (4.1)	4	4	
Literary Elements (4.2)	4	4	
Figurative Language/Sound Devices (4.3)	4	4	
Research and Information	8	8	16%
Accessing Information (5.1)	4	4	
Interpreting Information (5.2)	4	4	
Total Test	50	50	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.G
2009 PASS Blueprint and Actual Item Counts: Grade 8 Reading

PASS Standards & Objectives	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Vocabulary	6	6	12%
Words in Context (1.1)	2	3	
Word Origins (1.2)	2	1	
Idioms and Comparisons (1.3)	2	2	
Comprehension/Critical Literacy	21	21	42%
Literal Understanding (3.1)	4	4	
Inferences and Interpretation (3.2)	4-6	5	
Summary and Generalization (3.3)	5-7	5	
Analysis and Evaluation (3.4)	6-8	7	
Literature	15	15	30%
Literary Genre (4.1)	4	4	
Literary Elements (4.2)	5-7	7	
Figurative Language/Sound Devices (4.3)	4-6	4	
Research and Information	8	8	16%
Accessing Information (5.1)	4	4	
Interpreting Information (5.2)	4	4	
Total Test	50	50	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.H
2009 PASS Blueprint and Actual Item Counts: Grade 3 Mathematics

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Patterns and Algebraic Reasoning	8	8	18%
Algebra Patterns (1.1)	4	4	
Problem Solving (1.2)	4	4	
Number Sense	7	7	16%
Place Value (2.1)	3-4	3	
Whole Numbers and Fractions (2.2)	3-4	4	
Number Operations and Computation	12	12	27%
Estimation (3.1)	4	4	
Multiplication (3.2)	4	4	
Money Problems (3.3)	4	4	
Geometry and Measurement	12	12	27%
Spatial Reasoning (4.1)	4	4	
Measurement (4.2)	4	4	
Time and Temperature (4.4)	4	4	
Data Analysis and Probability	6	6	13%
Data Analysis (5.1)	2-4	3	
Probability (5.2)	2-4	3	
Total Test	45	45	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.I
2009 PASS Blueprint and Actual Item Counts: Grade 4 Mathematics

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Patterns and Algebraic Reasoning	8	8	18%
Algebra Patterns (1.1)	4	4	
Functions (1.2)	4	4	
Number Sense	10	10	22%
Place Value (2.1)	4	4	
Whole Numbers and Decimals (2.2)	2-4	3	
Fractions (2.3)	2-4	3	
Number Operations and Computation	11	11	24%
Multiplication (3.1)	2-4	3	
Division (3.2)	2-4	3	
Estimation (3.3)	4-5	5	
Geometry and Measurement	10	10	22%
Lines and Angles (4.1)	2-4	3	
Spatial Reasoning (4.3)	2-4	3	
Measurement (4.4)	4	4	
Data Analysis and Probability	6	6	13%
Data Analysis (5.1)	6	6	
Total Test	45	45	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.J
2009 PASS Blueprint and Actual Item Counts: Grade 5 Mathematics

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Patterns and Algebraic Reasoning	8	8	18%
Algebra Patterns (1.1)	4	4	
Problem Solving (1.2)	4	4	
Number Sense	8	8	18%
Fractions/Decimals/Percents (2.1)	4	4	
Number Theory (2.2)	4	4	
Number Operations and Computation	8	8	18%
Estimation (3.1)	4	4	
Whole Numbers/Decimals/Fractions (3.2)	4	4	
Geometry and Measurement	12	12	27%
Geometric Figure Properties (4.1)	4	4	
Perimeter/Area (4.2)	4	4	
Convert Measurements (4.5)	4	4	
Data Analysis and Probability	9	9	20%
Data Analysis (5.1)	5	5	
Probability (5.2)	4	4	
Total Test	45	45	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.K
2009 PASS Blueprint and Actual Item Counts: Grade 6 Mathematics

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Algebraic Reasoning	10	10	22%
Patterns (1.1)	5	5	
Order of Operations (1.2)	5	5	
Number Sense	13	13	29%
Multiply/Divide Fractions (2.1)	2-3	2	
Decimals (2.2)	2-3	3	
Estimation (2.3)	4	4	
Expressions (2.5)	4	4	
Geometry	6	6	13%
Angles (3.1)	2-4	3	
Congruent and Similar Figures (3.2)	2-4	3	
Measurement	7	7	16%
Compare/Convert Units (4.2)	3-4	3	
Estimate Measurements (4.3)	3-4	4	
Data Analysis and Statistics	9	9	20%
Collect/Organize/Interpret Data (5.1)	2-3	3	
Construct/Interpret Graphs (5.2)	2-3	2	
Median/Mode (5.3)	4	4	
Total Test	45	45	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.L
2009 PASS Blueprint and Actual Item Counts: Grade 7 Mathematics

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Algebraic Reasoning	8	8	18%
Properties (1.1)	4	4	
Linear Equations (1.2)	4	4	
Number Sense	12	12	27%
Integers (2.1)	4	4	
Ratio/Proportion/Percent (2.2)	4	4	
Exponents (2.3)	4	4	
Geometry	9	9	20%
Geometric Figures (3.1)	2-3	3	
Angles (3.2)	2-3	2	
Coordinate System (3.3)	4	4	
Measurement	7	7	16%
Area and Perimeter (4.1)	2-4	3	
Customary/Metric Measurements (4.2)	2-4	4	
Data Analysis and Probability	9	9	20%
Outcomes/Simple Probability (5.1)	4	4	
Probability with Or, And, or Not (5.2)	2-3	3	
Combinations/Permutations (5.3)	2-3	2	
Total Test	45	45	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.M
2009 PASS Blueprint and Actual Item Counts: Grade 8 Mathematics

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Algebraic Reasoning	9	9	20%
Equations (1.1)	5	5	
Inequalities (1.2)	4	4	
Number Sense	8	8	18%
Rational Numbers/Proportions (2.1)	4	4	
Exponents (2.2)	4	4	
Geometry	8	8	18%
Classify Solids (3.1)	4	4	
Pythagorean Theorem (3.2)	4	4	
Measurement	12	12	27%
Estimate Surface Area/Volume (4.1)	4	4	
Similar Figures (4.2)	4	4	
Formulas (4.3)	4	4	
Data Analysis and Statistics	8	8	18%
Data Representation (5.1)	4	4	
Central Tendency (5.2)	4	4	
Total Test	45	45	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.N.1
2009 PASS Blueprint and Actual Item Counts: Grade 5 Science Process Standards

<i>PASS</i> Process Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Observe and Measure	10	10	22%
SI Metric (P1.1)	5	5	
Similar/different characteristics (P1.2)	5	5	
Classify	10	10	22%
Observable properties (P2.1)	5	5	
Serial order (P2.2)	5	5	
Experiment	11	11	24%
Experimental design (P3.2)	7	7	
Hazards/practice safety (P3.4)	4	4	
Interpret and Communicate	14	14	31%
Data tables/line/bar/trend and circle graphs (P4.2)	6	6	
Prediction based on data (P4.3)	4	4	
Explanations based on data (P4.4)	4	4	
Total Test	45	45	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.N.2
2009 PASS Blueprint and Actual Item Counts: Grade 5 Science Content Standards

<i>PASS</i> Content Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i>	Actual Number of Items on 2009 Test	Ideal Percentage of Items
Properties of Matter and Energy	18	18	44%
Matter has physical properties (1.1)	6	6	
Physical properties can be measured (1.2)	6	6	
Energy can be transferred (1.3)	6	6	
Organisms and Environments	12	12	29%
Dependence upon community (2.1)	6	6	
Individual organism and species survival (2.2)	6	6	
Structures of the Earth and the Solar System	11	11	27%
Weather patterns (3.2)	6	6	
Earth as a planet (3.3)	5	5	
<i>Total Test</i>	<i>41*</i>	<i>41*</i>	<i>100%**</i>

* Safety items are not included within the content blueprint

** The ideal percents are based on the total number of items on a test that are matched to the content standards and do not include items added for safety.

Table 1.3.O.1
2009 PASS Blueprint and Actual Item Counts: Grade 8 Science Process Standards

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Observe and Measure	8	8	18%
Qualitative/quantitative observations/changes (P1.1)	4	4	
SI (metrics) units/appropriate tools (P1.2 and P1.3)	4	4	
Classify	8	8	18%
Classification system (P2.1)	4	4	
Properties ordered (P2.2)	4	4	
Experiment	16	16	36%
Experimental design (P3.2)	6	6	
Identify variables (P3.3)	6	6	
Hazards/practice safety (P3.6)	4	4	
Interpret and Communicate	13	13	29%
Data tables/line/bar/trend and circle graphs (P4.2)	7	7	
Explanations/prediction (P4.3)	6	6	
Total Test	45	45	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.O.2
2009 PASS Blueprint and Actual Item Counts: Grade 8 Science Content Standards

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i>	Actual Number of Items on 2009 Test	Ideal Percentage of Items
Properties and Chemical Changes in Matter	7-8	8	19%
Chemical reactions (1.1)	3-4	4	
Conservation of matter (1.2)	3-4	4	
Motion and Forces	8	8	20%
Motion of an object (2.1)	4	4	
Object subjected to a force (2.2)	4	4	
Diversity and Adaptations of Organisms	9	9	22%
Classification (3.1)	5	5	
Internal and external structures (3.2)	4	4	
Structures/Forces of the Earth/Solar System	8	8	20%
Landforms result from constructive and destructive forces (4.1)	4	4	
Rock cycle (4.2)	4	4	
Earth's History	7-8	8	19%
Catastrophic events (5.1)	3-4	4	
Fossil evidence (5.2)	3-4	4	
<i>Total Test</i>	<i>41*</i>	<i>41</i>	<i>100%**</i>

* Safety items are not included within the content blueprint

** The ideal percents are based on the total number of items on a test that are matched to the content standards and do not include items added for safety.

Table 1.3.P
2009 PASS Blueprint and Actual Item Counts: Grade 5 Social Studies

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Early Exploration	8	8	13%
Expeditions (2.1)	4	4	
Native American Reaction (2.2)	4	4	
Colonial America	12	12	20%
Settlements and Migration (3.1)	4	4	
Colonial Life (3.2)	4	4	
Individuals and Groups (3.3)	4	4	
American Revolution	12	12	20%
Causes and Results (4.1)	4	4	
Declaration of Independence (4.3)	4	4	
Individuals (4.4)	4	4	
Early Federal Period	8	8	13%
Constitutional Provisions (5.2)	4	4	
Ratification and Rights (5.3)	4	4	
Geographic Skills	20	20	33%
Maps/Charts/Graphs Usage (7.1)	7	7	
Human/Environment Interaction (7.2)	5	5	
Historical Places (7.3)	4	4	
Westward Movement (7.4)	4	4	
Total Test	60	60	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.Q
2009 PASS Blueprint and Actual Item Counts: Grade 7 Geography

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Geographic Tools	4	4	9%
Map Concepts (1.2)	4	4	
Regions	12	12	27%
Regional Characteristics (2.1)	4	4	
Conflict/Cooperation (2.2)	4	4	
Locations (2.4)	4	4	
Physical Systems	8	8	18%
Climate/Weather (3.2)	4	4	
Natural Disasters (3.3)	4	4	
Human Systems	8	8	18%
World Cultures (4.1)	4	4	
Population Issues (4.5)	4	4	
Human/Environment Interaction	8	8	18%
Natural Resources (5.1)	4	4	
Human Modification (5.2)	4	4	
Geography Skills	5	5	11%
Maps/Charts/Graphs (6.1)	5	5	
Total Test	45	45	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

Table 1.3.R
2009 PASS Blueprint and Actual Item Counts: Grade 8 U.S. History

<i>PASS</i> Standards & Objectives	Ideal Number of Items for Alignment to <i>PASS</i> *	Actual Number of Items on 2009 Test	Ideal Percentage of Items **
Social Studies Process Skills (1.0)	6	6	13%
Causes of the American Revolution (3.0)	5	5	11%
Results of the American Revolution (4.0)	5	5	11%
Governing Documents/Early Federal Period (5.0)	6	6	13%
Northern/Southern Economic Growth (6.0)	4	4	9%
Jacksonian Era (7.0)	4	4	9%
Cultural Growth and Reform (8.0)	4	4	9%
Westward Movement (9.0)	6	6	13%
Eve of War (10.0)	5	5	11%
<i>Total Test</i>	45	45	100%

* A minimum of 4 items is required to report results for an objective, and a minimum of 6 items is required to report a standard. While the actual numbers of items on the test may not match the blueprint exactly, each future test will move toward closer alignment with the ideal blueprint.

** Percents are approximations and may result in a sum other than 100 due to rounding.

CHAPTER II. STATISTICAL ANALYSIS

This section provides an overview of the research and statistical analyses carried out for the 2009 administration of the OCCT. Following the administration of the OCCT, student demographic and item response data were transmitted to DRC's Psychometric Services (PS) department. PS staff is responsible for analyzing the OCCT test data and producing the scoring tables used for reporting.

The analyses of the test data can be broken down into several components: 1) classical item analyses; 2) differential item functioning (DIF) analyses; 3) reliability analyses; 4) calibration and equating; 5) production of scoring tables; and 6) validity analyses. In the following sections, the analysis procedures for each component are described. Separate sections are provided for the Multiple-Choice tests and the Writing tests.

2.1 Data Files for Statistical Analysis

Preliminary item and DIF analyses and the final calibration/equating for the Multiple-Choice tests were conducted using early-return sample data consisting of approximately 50% of the examinee population. The final item and DIF analyses were conducted using data files that contained 100% of the student data. For the analysis of the Writing tests, the rater-year effect analysis was conducted based on a sample ($n = 510$) which was randomly drawn from 2007 and rescored in 2009. All other analyses on writing were conducted using the final population data file.

2.2 Analysis of the Multiple-Choice Tests

2.2.1 Classical Item Analyses

Classical item analyses were conducted using DRC's software, iTEMs (DRC, 2009). The analyses involved computing a set of statistics based on classical test theory for every item in each form. Each statistic is designed to provide empirical information about the characteristics of each item. The statistics estimated for OCCT items are described below.

Classical Item Difficulty (“ p -value”):

This statistic indicates the proportion of examinees in the sample that answered the item correctly. Desired p -values generally fall within the range of 0.30 to 0.90. Occasionally, items that fall outside this range can be justified for inclusion in an assessment based upon other quality indicators (e.g., adequate point-biserial), the educational importance of the item's content, or to better measure students with very high or low achievement.

Item Discrimination (“point-biserial”):

This statistic describes the relationship between performance on the specific item and performance on the entire form. Estimated as the correlation between the item score and the total test score, it indicates the extent to which test takers with high test scores tend to answer the item correctly, and those with low scores tend to answer incorrectly. The point-biserial correlation for item i is given by

$$pbis_i = \frac{\mu_+ - \mu_x}{\sigma_x} \sqrt{\frac{p_i}{1 - p_i}},$$

where μ_+ is the mean score for those students who answered item i correctly, and μ_x and σ_x are the mean score and standard deviation for the test form, and p_i is the item difficulty for item i .

Items with negative or low correlations can indicate problems with the item (e.g., incorrect key, multiple correct answers or unusually complex content), or can indicate that students have not been taught the content.

Examination of Empirical Item Response Curves (EIRC):

iTEMS provides graphical displays of student performance on each item. In the MC item plots, the x-axis represents the criterion score level (the total number-correct score) and y-axis represents the percentage of examinees choosing the response option. Each response option is plotted, representing the percentage of examinees that chose that particular option by ability level. One would expect the curve for the correct option to increase as ability level increases. These graphs were reviewed by DRC psychometricians.

Percentage of Students Omitting or not Reaching an Item:

This statistic is useful for identifying issues related to testing time and item/test layout. Testing time issues do not exist for the OCCT's as they are untimed. However, if the omit percentage is greater than 5% for a single item, this could be an indication of an item/test layout problem. For example, students might accidentally skip an item that follows a lengthy stem.

For the OCCT operational and field-test analyses, a series of flags were created in order to automatically identify items with performance characteristics that are at times considered unusual. The following flagging criteria were applied to all items tested in Spring 2009:

- P -value less than .30 or greater than .90;
- The percentage choosing an “incorrect” option is equal to or greater than the percentage choosing the correct option;
- The percentage of students selecting any of the “incorrect” options is larger than 30%;
- Point-biserial correlation is less than .30 for the correct answer;
- Any of the “incorrect” answer options (distractors) with a positive point-biserial;
- Percentage of test takers omitting the item is greater than 5%.

After the operational administration, the early-return data were used to conduct preliminary analyses to verify the accuracy of the scoring keys and to obtain an early indication of how items were functioning. Content specialists examined all flagged items to ensure that the items were correctly keyed.

Upon receipt of the complete student data, the items were further scrutinized during a final round of classical item analysis. After content specialist’s review and verification, item statistics were prepared for uploading to the item bank.

Summary statistics describing the difficulty and discrimination of items comprising the operational forms are given in Table 2.2.A. Results are combined across test forms for a given grade and content area because the operational items were the same for all test forms. Differential item functioning (DIF) flags and reliability indices (alpha, SEM, and stratified alpha) are also provided. These statistics are described in the sections that follow.

Table 2.2.A
Classical Item and Test Analyses Summary for the Operational Forms

Content Area/Grade	Grade	N Items	Alpha	SEM	Stratified Alpha	p-value		Pt-Biserial		Flagged Item Count	
						Mean	SD	Mean	SD	Stats ^a	DIF
Reading	3	50	0.89	2.79	0.89	0.72	0.15	0.40	0.08	12	1
	4	50	0.89	2.72	0.89	0.75	0.12	0.41	0.08	8	2
	5	50	0.89	2.69	0.89	0.75	0.15	0.41	0.07	12	1
	6	50	0.90	2.76	0.90	0.74	0.12	0.42	0.08	7	5
	7	50	0.89	2.78	0.89	0.73	0.13	0.40	0.07	12	1
	8	50	0.87	2.76	0.87	0.74	0.15	0.38	0.07	13	2
Mathematics	3	45	0.91	2.66	0.91	0.72	0.10	0.44	0.08	1	0
	4	45	0.88	2.74	0.88	0.70	0.12	0.40	0.07	5	0
	5	45	0.89	2.80	0.89	0.66	0.14	0.41	0.08	8	0
	6	45	0.89	2.79	0.90	0.67	0.12	0.42	0.07	3	0
	7	45	0.87	2.88	0.87	0.64	0.13	0.39	0.06	6	1
	8	45	0.88	2.85	0.88	0.66	0.11	0.40	0.08	8	0
Science	5	45	0.88	2.77	0.88	0.67	0.15	0.41	0.07	9	2
	8	45	0.88	2.72	0.88	0.68	0.16	0.40	0.07	9	1
Social Studies	5	60	0.91	3.35	0.91	0.63	0.13	0.39	0.07	11	1
Geography	7	45	0.88	2.87	0.88	0.65	0.12	0.40	0.07	6	0
U.S. History	8	45	0.89	2.86	0.89	0.64	0.13	0.42	0.09	9	2

^aClassical item statistics flagged using the criteria from the bulleted points above.

As shown in Table 2.2.A, the mean item difficulties of the tests across content/grade ranged from 0.63 to 0.75, and the mean point-biserial correlations ranged from 0.38 to 0.44. The internal consistency reliability estimates (coefficient alpha) of all tests were high, ranging from 0.87 to 0.91. The stratified alpha coefficients were almost identical to the alpha (after rounding). The SEMs ranged from 2.66 to 3.35.

Table 2.2.A also shows that a small number of operational test items were flagged for out-of-range statistics and/or C-category DIF. Items flagged for out-of-range statistics were scrutinized by content experts to verify the accuracy of the items in the test books, to verify keys, and to judge whether items were performing as expected. All items flagged for out-of-range statistics were found to be accurate, correctly keyed, and performing in a satisfactory manner with respect to content. DIF results for the operational items are discussed in the next section.

The results of the classical item analysis for the field-test items are presented in Table 2.2.B. Field-test items with extreme difficulty values, low point-biserials or poorly functioning distractors, and/or DIF were flagged for review by content experts. The number of items flagged for poor statistics ranged from 13 to 49 per content/grade. A very small number of items were

flagged for DIF. All flagged items were evaluated by DRC Test Development content specialists, DRC Psychometric Services experts, and the SDE at a Data Review Meeting held on July 20th in Maple Grove, and individual decisions were made about each item. Items rejected during Data Review will not be eligible as operational items in the future test administrations. Items accepted with revisions will be returned to the item bank and will be revised and re-field tested as necessary. The results of the Data Review are presented in Appendix A.

Table 2.2.B
Item Analyses Summary for Field-test Items

Content Area	Grade	Sample Size	N Items	p-value		Pt-Biserial		Flagged Item Count	
		Range		Mean	SD	Mean	SD	Stats ^a	DIF
Reading	3	5366-5515	80	0.62	0.18	0.40	0.12	22	0
	4	5230-5352	80	0.61	0.17	0.37	0.11	24	0
	5	5035-5372	80	0.67	0.16	0.43	0.10	13	1
	6	5037-5120	80	0.59	0.15	0.38	0.10	21	2
	7	4951-5023	80	0.59	0.17	0.38	0.11	21	2
	8	4912-5597	80	0.67	0.15	0.37	0.09	24	3
Mathematics	3	5469-5567	80	0.71	0.17	0.36	0.12	24	0
	4	5296-5406	80	0.68	0.19	0.35	0.07	33	4
	5	5096-5415	80	0.57	0.19	0.35	0.10	36	3
	6	5085-5153	80	0.56	0.16	0.34	0.10	26	0
	7	4973-5039	80	0.52	0.19	0.31	0.10	47	3
	8	4861-6098	80	0.50	0.18	0.31	0.11	43	0
Science	5	5195-5454	80	0.59	0.16	0.32	0.10	34	0
	8	5053-5160	80	0.49	0.20	0.26	0.14	49	3
Social Studies	5	5657-5748	80	0.49	0.15	0.33	0.11	36	0
Geography	7	5046-8054	80	0.57	0.18	0.35	0.11	27	1
History	8	5345-5449	80	0.47	0.14	0.36	0.12	30	1

^aOut-of-range classical item statistics

2.2.2 Differential Item Functioning Analyses

One of the goals of test development is to assemble a set of items that provides an estimate of a student’s ability that is as fair and accurate as possible for all groups within the population. DIF statistics are used to identify items that groups of students with the same underlying level of ability have different probabilities of answering correctly. If the item is differentially more difficult for an identifiable subgroup when conditioned on ability, the item may be measuring something different from the intended construct. However, it is important to recognize that DIF-flagged items might be related to actual differences in relevant knowledge or skills (item impact) or a statistical Type I error (a “false positive”). As a result, DIF statistics are used to identify potential sources of item bias. Subsequent review by content experts and bias/sensitivity committees is required to determine the source and meaning of performance differences.

For the OCCT, the Mantel-Haenszel (MH) procedure (Mantel & Haenszel, 1959; Holland & Thayer, 1988) was used to estimate DIF statistics for subgroups of interest defined by the SDE for NCLB accountability. Comparison groups were based on gender (female versus male), ethnicity (Hispanic versus White, American Indian versus White, African American versus White, Asian versus White, Pacific Islander versus White), and economic status (students who are economically disadvantaged as indicated by participation in a free and reduced-price school lunch program versus students who are not economically disadvantaged). Items with statistically significant differences in performance were flagged for possible biased or unfair content that was

undetected in earlier fairness and bias reviews. DIF analyses results were not considered as valid when the sample size for either the reference group (i.e., male, White, not economically disadvantaged) or focal group (i.e., female, Hispanic, American Indian, African American, Asian, Pacific Islander, economically disadvantaged) was less than 300 *and* the sample size for the two groups combined was less than 700.

The MH procedure is one of the more commonly used methods to detect DIF. This method uses contingency tables to compare the probability of success on each item for the studied groups of interest after matching on overall ability (i.e., total test score). The common odds ratio is estimated across all categories of matched examinee ability. The resulting estimate is interpreted as the relative likelihood of success on a particular item for members of two groups when matched on ability. As such, the common odds ratio provides an estimated effect size where a value of unity indicates equal odds, and thus no DIF (Dorans & Holland, 1993).

The common odds ratio (α) is estimated as $\hat{\alpha}_{MH} = \frac{\sum_{s=1}^S R_{rs} W_{fs} / N_{ts}}{\sum_{s=1}^S R_{fs} W_{rs} / N_{ts}}$, where

R_{rs} = the number of examinees in the reference group who answer the item correctly,

W_{fs} = the number of examinees in the focal group who answer the item incorrectly,

R_{fs} = the number of examinees in the focal group who answer the item correctly,

W_{rs} = the number of examinees in the reference group who answer the item incorrectly,

N_{ts} = the total number of examinees.

The odds ratio takes on values from 0 to infinity and is interpreted as the average factor by which the odds that an examinee of the reference group will answer an item correctly exceed that of a member of the comparable focal group. The statistical test is $H_0: \alpha = 1$, where α is a common odds ratio assumed equal for all matched score categories $s = 1$ to S . Values less than unity indicate DIF in favor of the focal group, a value of unity indicates the null condition, and a value greater than one indicates DIF in favor of the reference group. The associated MH χ^2 is distributed as a chi-square random variable with 1 degree of freedom. As an index of magnitude, the odds ratio is frequently transformed to a delta scale given by $MH\ D-DIF = -2.35 \ln(\hat{\alpha}_{MH})$ where negative values indicate DIF in favor of the reference group and positive values favor the focal group.

A classification scheme puts items into three DIF categories on the basis of a combination of statistical significance and magnitude (absolute value) of MH D-DIF (Zwick and Ercikan, 1989):

A-items or negligible DIF: MH D-DIF is not statistically different from 0 (at the .05 level) or its absolute value is less than 1 delta unit;

B-items or intermediate DIF: MH D-DIF is statistically different from 0 (at the .05 level) and its absolute value is at least 1 but less than 1.5 or an absolute value of at least 1 but not significantly greater than 1 (at the .05 level);

C-items or large DIF: MH D-DIF is statistically different from 1 (at the .05 level) and its absolute value is at least 1.5.

Items classified as B+ or C+ tend to be easier for members of the focal group than for members of the reference group whose total scores on the test are like those of the focal group. Items classified as B- or C- tend to be harder for members of the focal group than for members of the reference group whose total scores on the test are like those of the focal group.

Items classified in category C were sent to test development staff for review. They were asked to consider any identifiable characteristics that may have contributed to the differential item functioning. The items were then submitted to the SDE for further review.

Table 2.2.A shows that a small number of operational items were flagged for C DIF. These items were reviewed by DRC's content experts. Recommendations were made by DRC on whether to remove an item with C DIF from scoring or not. SDE content experts further reviewed these items and made the final decision. As a result, no items were dropped in the 2009 administration.

DIF analysis was also conducted on the field tests. Items with C DIF were flagged and reviewed by SDE and DRC's content experts at the data review meeting. Appendix A reports the items rejected due to DIF and/or other poor statistics.

2.2.3 Item Calibration and Equating

The purpose of item calibration and equating is to create a common scale for expressing the difficulty estimates of all the items across forms within a test. The scale is initially defined so that the examinees used in the calibration will have a mean score of 0 and a standard deviation of 1. It should be noted that the metric of this scale is often referred to as the "theta" metric. This scale is not used for reporting purposes because its values typically range from -3.0 to +3.0, which is a scale that is not easily understood. Therefore, following calibration and equating, the scale is usually transformed to a reporting scale that can be understood more easily by students, teachers, and other stakeholders.

The three-parameter logistic (3PL) model was used to calibrate the OCCT test items. The 3PL model expresses the probability that a person with ability θ will respond correctly to item j as a function of item and ability parameters:

$$P(U_j = 1 | \theta) = P_j(\theta) = c_j + \frac{1 - c_j}{1 + e^{-1.7a(\theta - b)}},$$

where:

- U_j is the response to item j , 1 if correct and 0 if incorrect;
- a_j is the slope parameter of item j , characterizing its discriminating power;
- b_j is the threshold parameter of item j , characterizing its difficulty; and
- c_j is the lower asymptote parameter of item j , reflecting the chance that students with very low proficiency will select the correct answer, sometimes called the "pseudo-guessing" level.

The parameters estimated for the 3PL model were discrimination (*a*), difficulty (*b*), and the pseudo-guessing level (*c*). All item response theory (IRT) based analyses were conducted using PARSCALE (Muraki and Bock, 2003).

For each operational test, items were calibrated separately by content and grade. The calibrations were examined to assess the quality of the parameter estimates and model-data fit. Items were flagged for:

- *a*-parameters less than 0.3 or greater than 2.3
- *b*-parameters less than -3.5 or greater than 3.5
- *c*-parameters greater than 0.35 for 4-option items
- Not calibrated due to biserial correlations less than 0.10
- Data model fit is “bad.” (This criterion varies depending on the response n-count.)

Flagged items were reviewed to determine whether they should be excluded from scoring. No items in the 2009 OCCT were excluded from scoring because of IRT results.

After the final set of item parameter estimates were established, the scales for the 2009 operational tests were linked to the reporting scale using the test characteristic curve (TCC) method described by Stocking and Lord (1983). The Stocking and Lord procedure involves finding a linear transformation that will minimize the sum of squared differences between two TCCs generated from two sets of anchor item parameters.

Embedded in the 2009 OCCT were sets of anchor items that had served as operational items in the 2008 OCCT. These items were positioned so their sequences were very similar to that in the prior year. The sets were chosen for being both content and statistically representative of the entire test to ensure an accurate equating result. The anchor set is mostly unique for each testing cycle, though some items may be used as anchors for multiple administrations. Repeated use of an item creates the risk of overexposure and is avoided in practice. Table 2.2.C summarizes the number of anchor items per test.

The parameters for the 2008 items were expressed on the reporting scale. These 2008 item parameters served as a reference item set and were used with their 2009 counterparts and the Stocking and Lord procedure to find transformation constants. These constants were used to transform the 2009 item parameters so that they were expressed on the reporting scale. Once this was done, the transformed parameters were used to generate raw score to scaled score conversion tables.

Table 2.2.C
Number of Anchor Items by Subject and Grade

Subject	Grade	Number of Operational Items per Form ^a	Number of Operational Items in Anchor Set ^a
Reading	3-8	50	20, 20, 19, 20, 20, and 19 respectively
Writing	5, 8	1	0
Mathematics	3-8	45	20, 18, 18, 19, 18, and 20 respectively
Science	5, 8	45	16,19
Social Studies	5	60	20
Geography	7	45	18
U.S. History	8	45	19

^a Operational item counts include anchor items.

Field-test items were placed on the operational scale in a similar fashion. For each content/grade, field-test items were calibrated with the operational items. Resulting field-test parameters were placed on the OCCT reporting scale using the operational items as the anchor set in the Stocking and Lord procedure.

2.2.4 Raw Score to Scaled Score Conversion

Since 2005, the OCCT scaled scores have been produced using a number-correct scoring procedure that is based on IRT. This procedure produces maximum-likelihood trait estimates for each obtainable raw score, except for raw scores at chance or below-chance levels and the perfect raw score. It is conventional to assign scaled scores to at and below-chance level raw scores and perfect raw scores using a rational, but not necessarily maximum likelihood, procedure. These values are called the lowest obtainable scaled score (LOSS) and the highest obtainable scaled score (HOSS). The LOSS and HOSS values assigned to all OCCT operational tests were 400 and 990, respectively.

For all MC tests, the OCCT score scale uses a three-digit integer that spans a range from 400 (LOSS) to 990 (HOSS). The Proficient cut score for reading and Mathematics and the Satisfactory cut score for science and social studies is 700 for all tests. The raw-score to scaled-score conversion tables are provided in Appendix B.

2.2.5 Test Score Reliability

Test score reliability focuses on the extent to which differences in test scores reflect true differences in the knowledge, ability, or skills being tested rather than random fluctuations. The variance in the distributions of test scores, essentially the differences among individuals, is partly due to real differences in the knowledge, skills, or ability being tested (called true score variance) and partly due to random factors that cause variability in examinee performance (called error variance). The number used to describe reliability is an estimate of the proportion of true score variance to total variance. Several different ways of estimating this proportion exist.

Coefficient Alpha

When the goal is to estimate the precision of a set of test scores from a single administration, a measure of internal consistency (sensitive to random errors associated with item content sampling) is frequently used to estimate reliability. For the OCCT, the measure of internal

consistency called coefficient alpha (Cronbach, 1951) was used to estimate the reliability of the test scores. The formula for coefficient alpha is given by

$$\rho_{xx'} \geq \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2}\right),$$

where k is the number of items on the test, $\sum \sigma_i^2$ is item score variance summed over all items, and σ_x^2 is observed-score variance.

Internal consistency measures apply only to the test form being analyzed. They do not take into account form-to-form variation due to equating limitations, nor are they sensitive to day-to-day variation due, for example, to state of health or testing environment. Reliability coefficients may range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely it would be for individuals to obtain very similar scores over replicated testing (e.g., using the same number of items, sampling same content domain(s), etc.). The internal consistency of the multiple-choice test scores are reported in Table 2.2.A for all examinees and in Tables 3.3 – 3.21 by demographic subgroup.

When a test contains different components (e.g., content standards), the stratified alpha coefficient can provide a more accurate estimate of the overall test reliability (Qualls, 1995). The stratified alpha coefficient is calculated by

$$\text{Strat } \alpha \rho_{xx'} = 1 - \frac{\sum \sigma_{x_j}^2 (1 - \alpha \rho_{x_j x_j'})}{\sigma_{Total}^2},$$

where, σ_{Total}^2 is the variance of the total test scores; $\sigma_{x_j}^2$ is the variance of scores for each test component (i.e., content standards in this case); and $\alpha \rho_{x_j x_j'}$ is the coefficient alpha reliability for scores from content standard J . The stratified alpha coefficients for the multiple-choice test scores are reported in Table 2.2.A for all examinees and in Tables 3.3-3.21 by demographic subgroup.

Standard Error of Measurement

The standard error of measurement (SEM) is the standard deviation of the errors of measurement that is associated with the test scores of a specific group of test takers. In Classical Test theory (CTT), an overall SEM can be estimated as a function of the standard deviation of observed scores and test reliability coefficient:

$$SEM = s_x \sqrt{1 - r_{xx'}},$$

where SEM is standard error of measurement, s_x is standard deviation of observed scores, and $r_{xx'}$ is a coefficient of reliability.

The SEM is particularly useful in determining the confidence interval (CI) that captures an examinee's true score. Assuming that measurement error is normally distributed, it can be said that upon infinite testing replications, approximately 95 percent of the CIs of ± 1.96 SEM around the observed score would contain an examinee's true score (Crocker & Algina, 1986).

For example, if an examinee's observed score on a given test equals 15 points, and SEM equals 1.92, one can be 95% confident that the examinee's true score lies between 11 and 19 points (15 ± 3.76 rounded to the nearest integer). Table 2.2.A provides the SEM for each multiple-choice test.

Conditional Standard Error of Measurement

From the IRT framework, a standard error of measurement can be estimated for each measured ability. Thus, it is often referred to as a conditional standard error of measurement (CSEM). The expected a posteriori estimation of CSEM proposed by Kolen, Zeng, and Hanson (1996) was used for the OCCT. The calculation of CSEM can be expressed as:

$$CSEM(S_x | \theta) = \sqrt{\left[\sum_{X=0}^{MaxX} S_x^2 p(X | \theta) \right] - \left[\sum_{X=0}^{MaxX} S_x \cdot p(X | \theta) \right]^2}$$

where S_x is the scaled score for a particular number correct score X ; θ is the IRT ability scaled value conditioned on; and $p(X | \theta)$ is the probability function that is computed using a recursive algorithm given by Thissen, Pommerich, Billeaud, and Williams (1995). For the operational OCCT, CSEMs were provided for each obtainable scaled score (see Appendix B).

Reliability of Performance-Level Classification Decisions

Student performance on the OCCT is classified into one of four achievement levels using cut scores adopted by the SDE. Table 2.2.D provides the cut score for each achievement level and the CSEM associated with each cut score in 2009.

Table 2.2.D
Conditional Standard Errors of Measurement for Each Achievement Level Cut Score

Content Area	Grade	Limited Knowledge		Satisfactory		Advanced	
		Scale Score	CSEM	Scale Score	CSEM	Scale Score	CSEM
Reading	3	649	24	700	23	891	54
	4	658	20	700	20	845	60
	5	641	23	700	23	830	51
	6	647	22	700	22	828	52
	7	668	21	700	21	802	38
	8	655	25	700	25	833	49
Mathematics	3	636	27	700	21	798	39
	4	639	31	700	26	816	39
	5	642	33	700	23	767	23
	6	660	23	700	20	754	21
	7	667	29	700	26	766	25
	8	662	28	700	23	771	26
Science	5	638	53	700	25	814	24
	8	647	40	700	23	829	26
Social Studies	5	645	32	700	22	786	20
Geography	7	595	70	700	30	847	33
History	8	622	47	700	23	821	35
Writing ^a	5	26	na	36	na	54	na
	8	25	na	36	na	54	na

^aWriting cut scores are in the composite score metric.

The reliability of 2009 achievement-level classification decisions was assessed using the computer program BB-CLASS (Brennan, 2004), which provides two statistics that describe the reliability of classifications based on test scores (Livingston & Lewis, 1995). More specifically, information from an administration of one form is used to estimate the following:

Decision Accuracy, which describes the extent to which performance-level classification decisions based on the administered test form would agree with the decisions that would be made on the basis of a perfectly reliable test (i.e., meaning if it was possible to know each examinee's true score). Decision accuracy answers the question: How does the actual classification of test takers, based on their single-form scores, agree with the classification that would be made on the basis of their true scores, if their true scores were somehow known?

Decision Consistency, which describes the extent to which classification decisions based on the administered test form would agree with the decisions made if a parallel alternate form had been administered. Decision consistency answers the question: What is the agreement between the classifications based on two non-overlapping, equally difficult forms of the test?

For each performance level and test, true scores and single-form scores on forms parallel to the one actually given are estimated following the Livingston and Lewis (1995) method. The decision accuracy is estimated using an estimated joint distribution of reported performance level classifications on the current form of the exam and the performance-level classifications based on the true score. Decision consistency is estimated using an estimated joint distribution of

reported performance-level classifications on the current form of the exam and performance-level classifications on the parallel alternate form.

In each case, the proportion of performance-level classifications with exact agreement is the sum of the entries in the diagonal of the contingency table representing the joint distribution.

Reliability of classification at each performance-level cut score is estimated by collapsing the joint distribution at the passing score boundary into a 2-by-2 table and summing the two entries.

Table 2.2.E provides the results for decision accuracy and consistency analyses that were conducted at the Limited Knowledge, Proficient/Satisfactory, and Advanced cut scores, and for the four performance levels (total). It should be noted that decision accuracy and consistency indices for the four performance levels should be lower than those for each cut, as shown in Table 2.2.D. This is not surprising since classification using four levels would allow more opportunity to change achievement levels. Hence there would be more classification errors in the four achievement levels, resulting in lower consistency indices.

For the OCCT, a *PASS*ing score is one that meets or exceeds the Proficient/Satisfactory cut score. Across all tests, the decision accuracy of the Proficient/Satisfactory cut scores ranged from 0.89–0.94 and decision consistency ranged from 0.85–0.92. These results indicate that at least 89% students meeting or exceeding the Proficient/Satisfactory cut score would receive the same *PASS*/fail classification if their true scores were known. If a parallel test were administered, at least 85% or more of students meeting or exceeding the Proficient/Satisfactory cut score would be classified in the same way.

Table 2.2.E
Estimates of the Reliability of Decisions for Specified Cut Scores ^a

Content Area	Grade	Decision Accuracy				Decision Consistency			
		Limited Knowledge	Proficient/Satisfactory	Advanced	Total	Limited Knowledge	Proficient/Satisfactory	Advanced	Total
Reading	3	0.95	0.91	0.96	0.82	0.92	0.87	0.95	0.76
	4	0.94	0.91	0.94	0.80	0.92	0.87	0.92	0.72
	5	0.95	0.91	0.91	0.78	0.93	0.87	0.90	0.71
	6	0.95	0.91	0.92	0.78	0.93	0.88	0.90	0.71
	7	0.94	0.91	0.89	0.76	0.92	0.88	0.86	0.67
	8	0.94	0.90	0.91	0.76	0.92	0.86	0.89	0.69
Mathematics	3	0.95	0.92	0.91	0.79	0.93	0.89	0.88	0.71
	4	0.94	0.91	0.92	0.77	0.92	0.87	0.89	0.69
	5	0.94	0.91	0.91	0.76	0.91	0.87	0.88	0.67
	6	0.93	0.91	0.91	0.76	0.90	0.87	0.87	0.67
	7	0.91	0.89	0.90	0.72	0.88	0.85	0.87	0.64
	8	0.92	0.90	0.92	0.75	0.89	0.86	0.88	0.66
Science	5	0.98	0.94	0.91	0.82	0.96	0.91	0.87	0.75
	8	0.98	0.94	0.92	0.84	0.97	0.92	0.89	0.78
Social Studies	5	0.94	0.92	0.93	0.79	0.92	0.88	0.90	0.71
Geography	7	0.98	0.93	0.91	0.82	0.97	0.90	0.88	0.75
History	8	0.95	0.91	0.94	0.81	0.93	0.88	0.92	0.73

a: The analysis was based on the final data files with students who took the standard OCCT.

2.2.6 Validity

As noted in the *Standards for Educational and Psychological Testing*, “validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of the tests” (AERA, APA, & NCME, 1999, P.9). Content representativeness

considerations, item bias (i.e., DIF) analysis, and correlations among content standards are often used as sources of validity evidence.

Each test's blueprint specifies the proportion of items that should be devoted to any given content unit. The blueprint is used as a guide by test developers when assembling a test from a pool of candidate items that are classified by content unit. Validity evidence related to test content is bolstered to the extent that the numbers of items allocated to each *PASS* Standard/Objective reflect what is specified by the test blueprints. Tables 1.2 to 1.18 in Chapter 1 provide content validity evidence by standard for the 2009 OCCT.

Differential item functioning with respect to gender, ethnicity, and economic status helps address construct-irrelevant variance, which represents an important threat to the validity of achievement tests. As noted in the section of Differential Item Functioning Analyses, field-test items are screened and reviewed for DIF by SDE content specialists. Only items approved by SDE are eligible for operational use. DIF analyses were also conducted on the operational items. After SDE and DRC's content specialists' review, no item was dropped from the operational tests. The number of operational and field-test items with C DIF were reported in Tables 2.2.A and 2.2.B.

Intercorrelations among standards provide evidence of convergent test validity. The analyses were performed by summing the obtained raw score points for each standard and then correlating the subtotals associated with each standard. Standards with low point totals, e.g., less than five, usually have markedly attenuated coefficients, meaning that they will be spuriously low in magnitude. Tables 1.3.B to 1.3.R list the numbers of items associated with each standard. The correlations among standards are reported in the left corner of Tables 2.2.F through 2.2.V.

The correlations corrected for attenuation are reported in the right corner of Tables 2.2.E through 2.2.U. Correcting for attenuation adjusts the correlation between the two measures to account for the unreliability of both. Although the theoretical upper bound for a correlation is 1.0, disattenuated correlations can be greater. This is often seen in practice when the correlations are relatively high and the reliabilities relatively low. However, two underlying factors should be noted. The first is that sample statistics are being used to estimate population parameters. The second, and likely more prevailing issue, is that something akin to a "design misspecification" occurs. The internal consistency reliability indices used for the OCCT likely do not capture all the sources of random error in the test scores, and, as such, might over estimate reliability. One might also postulate potential negative biases (e.g., lack of item homogeneity due to multidimensional content standards). Thus, it is possible that any given tabled disattenuated correlation may be too high, or too low, depending on which bias prevails. Also note that the correlations between standards and total test are spuriously inflated given they have items in common.

Given that none of these tests have perfect reliabilities (equal to one), the disattenuated correlations are somewhat higher than the correlations. Disattenuated correlations less than 1.0 suggest that the different strands are measuring slightly different aspects of the constructs. Values around 1.0 suggest that the same or very similar constructs are being measured.

Table 2.2.F
Standards Intercorrelation: Grade 3 Reading

	Reading	Vocabulary	Comprehension /Critical Literacy	Literature	Research and Information
Reading	--	1.09	1.11	1.15	1.09
Vocabulary	0.84	--	0.94	0.98	0.96
Comprehension/Critical Literacy	0.94	0.69	--	0.96	0.95
Literature	0.73	0.54	0.58	--	0.96
Research and Information	0.78	0.60	0.65	0.49	--

Table 2.2.G
Standards Intercorrelation: Grade 4 Reading

	Reading	Vocabulary	Comprehension /Critical Literacy	Literature	Research and Information
Reading	--	1.10	1.10	1.14	1.10
Vocabulary	0.85	--	0.97	0.98	0.95
Comprehension/Critical Literacy	0.95	0.72	--	1.01	0.94
Literature	0.82	0.61	0.70	--	0.96
Research and Information	0.70	0.52	0.57	0.49	--

Table 2.2.H
Standards Intercorrelation: Grade 5 Reading

	Reading	Vocabulary	Comprehension /Critical Literacy	Literature	Research and Information
Reading	--	1.07	1.11	1.13	1.06
Vocabulary	0.85	--	0.91	0.95	0.92
Comprehension/Critical Literacy	0.91	0.67	--	1.00	0.89
Literature	0.87	0.65	0.71	--	0.95
Research and Information	0.73	0.57	0.57	0.56	--

Table 2.2.I
Standards Intercorrelation: Grade 6 Reading

	Reading	Vocabulary	Comprehension /Critical Literacy	Literature	Research and Information
Reading	--	1.13	1.10	1.12	1.11
Vocabulary	0.73	--	0.98	1.01	0.98
Comprehension/Critical Literacy	0.93	0.61	--	0.99	0.95
Literature	0.89	0.58	0.75	--	0.99
Research and Information	0.79	0.51	0.64	0.63	--

Table 2.2.J
Standards Intercorrelation: Grade 7 Reading

	Reading	Vocabulary	Comprehension /Critical Literacy	Literature	Research and Information
Reading	--	1.10	1.11	1.15	1.12
Vocabulary	0.80	--	0.96	0.98	0.94
Comprehension/Critical Literacy	0.93	0.66	--	0.99	1.00
Literature	0.83	0.58	0.68	--	1.00
Research and Information	0.81	0.55	0.67	0.59	--

Table 2.2.K
Standards Intercorrelation: Grade 8 Reading

	Reading	Vocabulary	Comprehension /Critical Literacy	Literature	Research and Information
Reading	--	1.15	1.13	1.14	1.15
Vocabulary	0.66	--	0.95	0.99	0.96
Comprehension/Critical Literacy	0.92	0.50	--	0.98	1.02
Literature	0.88	0.50	0.70	--	0.99
Research and Information	0.80	0.44	0.66	0.61	--

Table 2.2.L
Standards Intercorrelation: Grade 3 Mathematics

	Mathematics	Patterns and Algebraic Reasoning	Number Sense	Number Operations and Computation	Geometry and Measurement	Data Analysis and Probability
Mathematics	--	1.06	1.07	1.06	1.10	1.05
Patterns and Algebraic Reasoning	0.83	--	0.94	0.90	0.94	0.92
Number Sense	0.80	0.61	--	0.90	0.98	0.95
Number Operations and Computation	0.88	0.64	0.62	--	0.92	0.90
Geometry and Measurement	0.86	0.64	0.64	0.66	--	0.95
Data Analysis and Probability	0.81	0.61	0.61	0.64	0.64	--

Table 2.2.M
Standards Intercorrelation: Grade 4 Mathematics

	Mathematics	Patterns and Algebraic Reasoning	Number Sense	Number Operations and Computation	Geometry and Measurement	Data Analysis and Probability
Mathematics	--	1.11	1.13	1.07	1.11	1.08
Patterns and Algebraic Reasoning	0.79	--	0.94	0.94	0.90	0.97
Number Sense	0.80	0.54	--	0.90	0.93	0.94
Number Operations and Computation	0.88	0.62	0.59	--	0.87	0.91
Geometry and Measurement	0.77	0.50	0.52	0.57	--	0.91
Data Analysis and Probability	0.76	0.55	0.53	0.60	0.51	--

Table 2.2.N
Standards Intercorrelation: Grade 5 Mathematics

	Mathematics	Patterns and Algebraic Reasoning	Number Sense	Number Operations and Computation	Geometry and Measurement	Data Analysis and Probability
Mathematics	--	1.10	1.11	1.08	1.12	1.07
Patterns and Algebraic Reasoning	0.80	--	0.98	0.92	0.95	0.94
Number Sense	0.82	0.59	--	0.96	0.97	0.92
Number Operations and Computation	0.83	0.58	0.61	--	0.96	0.86
Geometry and Measurement	0.86	0.59	0.62	0.63	--	0.93
Data Analysis and Probability	0.79	0.57	0.57	0.56	0.59	--

Table 2.2.O
Standards Intercorrelation: Grade 6 Mathematics

	Mathematics	Algebraic Reasoning	Number Sense	Geometry	Measurement	Data Analysis and Statistics
Mathematics	--	1.02	1.11	0.98	1.05	1.08
Algebraic Reasoning	0.82	--	0.89	0.79	0.84	0.86
Number Sense	0.88	0.64	--	0.80	0.95	0.91
Geometry	0.69	0.50	0.50	--	0.79	0.83
Measurement	0.81	0.58	0.65	0.48	--	0.88
Data Analysis and Statistics	0.80	0.57	0.60	0.48	0.56	--

Table 2.2.P
Standards Intercorrelation: Grade 7 Mathematics

	Mathematics	Algebraic Reasoning	Number Sense	Geometry	Measurement	Data Analysis and Probability
Mathematics	--	1.12	1.13	1.08	1.13	1.12
Algebraic Reasoning	0.78	--	0.96	0.89	0.95	0.93
Number Sense	0.86	0.58	--	0.90	0.99	0.95
Geometry	0.80	0.53	0.59	--	0.91	0.88
Measurement	0.73	0.49	0.56	0.50	--	0.91
Data Analysis and Probability	0.79	0.52	0.58	0.53	0.47	--

Table 2.2.Q
Standards Intercorrelation: Grade 8 Mathematics

	Mathematics	Algebraic Reasoning	Number Sense	Geometry	Measurement	Data Analysis and Statistics
Mathematics	--	1.07	1.14	1.08	1.10	1.01
Algebraic Reasoning	0.80	--	0.93	0.83	0.91	0.85
Number Sense	0.78	0.54	--	0.90	0.99	0.89
Geometry	0.73	0.48	0.47	--	0.92	0.78
Measurement	0.89	0.62	0.62	0.57	--	0.86
Data Analysis and Statistics	0.77	0.55	0.53	0.46	0.60	--

Table 2.2.R
Standards Intercorrelation: Grade 5 Science

	Science	Observe and Measure	Classify	Experiment	Interpret and Communicate
Science	--	1.12	1.12	1.13	1.13
Observe and Measure	0.84	--	0.99	0.98	0.97
Classify	0.86	0.65	--	0.98	1.00
Experiment	0.84	0.62	0.63	--	0.98
Interpret and Communicate	0.89	0.65	0.68	0.64	--

Table 2.2.S
Standards Intercorrelation: Grade 8 Science

	Science	Observe and Measure	Classify	Experiment	Interpret and Communicate
Science	--	1.14	1.14	1.13	1.14
Observe and Measure	0.78	--	1.02	0.98	1.01
Classify	0.82	0.57	--	0.98	1.04
Experiment	0.89	0.60	0.64	--	0.98
Interpret and Communicate	0.89	0.62	0.67	0.69	--

Table 2.2.T
Standards Intercorrelation: Grade 5 Social Studies

	Social Studies	Early Exploration	Colonial America	American Revolution	Early Federal Period	Geographic Skills
Social Studies	--	1.09	1.10	1.08	1.07	1.06
Early Exploration	0.81	--	0.99	0.96	0.95	0.94
Colonial America	0.84	0.62	--	0.99	0.96	0.92
American Revolution	0.84	0.61	0.65	--	0.95	0.88
Early Federal Period	0.78	0.57	0.59	0.60	--	0.91
Geographic Skills	0.88	0.64	0.65	0.63	0.61	--

Table 2.2.U
Standards Intercorrelation: Grade 7 Geography

	Geography	Geographic Tools	Regions	Physical Systems	Human Systems	Human /Environment Interaction	Geography Skills
Geography	--	1.11	1.12	1.13	1.15	1.09	1.07
Geographic Tools	0.61	--	0.93	0.92	0.97	0.89	1.02
Regions	0.88	0.45	--	0.99	1.00	0.97	0.91
Physical Systems	0.78	0.39	0.61	--	0.97	0.98	0.93
Human Systems	0.80	0.42	0.62	0.53	--	0.97	0.94
Human/Environment Interaction	0.82	0.41	0.65	0.58	0.57	--	0.87
Geography Skills	0.68	0.40	0.52	0.46	0.47	0.47	--

Table 2.2.V
Standards Intercorrelation: Grade 8 U. S. History

	History	HA	HB	HC	HD	HE	HF	HG	HH	HI
History	--	1.11	1.08	1.10	1.15	1.07	1.09	1.20	1.11	1.12
HA	0.75	--	0.96	0.96	0.98	0.92	0.92	1.06	1.02	1.00
HB	0.80	0.54	--	0.99	0.99	0.88	0.95	1.01	0.97	0.99
HC	0.76	0.51	0.57	--	0.98	0.95	0.91	1.04	0.98	1.00
HD	0.72	0.46	0.51	0.47	--	0.95	1.00	1.03	0.97	1.00
HE	0.59	0.39	0.40	0.41	0.36	--	0.86	0.91	0.95	0.94
HF	0.67	0.43	0.49	0.43	0.43	0.33	--	0.97	0.96	0.97
HG	0.63	0.42	0.44	0.42	0.38	0.29	0.35	--	1.03	1.07
HH	0.78	0.55	0.57	0.54	0.48	0.41	0.47	0.43	--	1.00
HI	0.76	0.52	0.56	0.53	0.48	0.40	0.46	0.43	0.54	--

HA – Social Studies Process Skills
 HB – Causes of American Revolution
 HC – Results of American Revolution
 HD – Governing Documents/Early Federal Period
 HE – Northern/Southern Economic Growth
 HF – Jacksonian Era
 HG – Cultural Growth and Reform
 HH – Westward Movement
 HI – Eve of War

2.3 Analysis of the Writing Tests

The administration of the Spring 2009 Writing assessment took place on February 25. Students at Grades 5 and 8 were given one operational writing prompt. The Grade 5 operational prompt was field-test prompt #9 in 2007; the Grade 8 operational prompt was field-test prompt #3 in 2007. The following sections describe the statistical analyses conducted to place the 2009 operational writing prompts on the scale established in 2006.

2.3.1 Prompt Scoring Formula

The writing score is a weighted composite of five analytic scores that focus on specific domains of writing skills. These skills are listed in Table 2.3.A. Each student’s response to a prompt is read by two independent raters; the raters’ scores for each domain are averaged. The domain scores range from 1 (the lowest score) to 4 (the highest score).

The raw writing score is calculated as a weighted composite of the average of two independent ratings for each of the five analytic traits:

$$\text{Raw Composite Score (RCS)} = 15*(0.30*ID + 0.25*OUC + 0.15*WC + 0.15*SP + 0.15*GUM)$$

Table 2.3.A
Weights Assigned to Writing Analytic Traits

Writing Analytic Traits	Weight
Ideas and Development (ID)	30%
Organization, Unity, and Coherence (OUC)	25%
Word Choice (WC)	15%
Sentences and Paragraphs (SP)	15%
Grammar, Usage, and Mechanics (GUM)	15%

2.3.2 Statistical Adjustments to Scale the Writing Scores

The baseline for each grade’s operational writing scale was 2006. To place the 2009 operational prompts on the 2006 scale, transformation constants were obtained to adjust RCS scores for prompt difficulty and for rater-year effects relative to a target distribution. All calculations were performed on the RCS prior to rounding. For reporting, the scaled composite scores (SCS) were then rounded to the nearest integer between 15 and 60.

Adjustment for Prompt Difficulty and Rater-Year Effects

For each of the 2007 field-test prompts, ETS provided a set of unique transformation constants to adjust for both prompt difficulty and rater-year effects. Based on ETS’ report, *OCCT Writing: Scaling the 2007 Field-Test Prompts* (ETS, 2007), the following equation was used to adjust the 2009 raw composite scores ($RCS_{Op09,grade}$):

$$SCS_{1Op09,grade} = B_{07, grade} \times RCS_{Op09,grade} + A_{07, grade} .$$

Where $SCS_{1Op09,grade}$ represents the scaled composite score after adjusting for the 2007 prompt difficulty and rater-year effects; $A_{07, grade}$ and $B_{07, grade}$ are the additive and multiplicative constants (Grade 5: $A_{07, g5} = -1.006108$ $B_{07, g5} = 1.015014$; Grade 8: $A_{07, g8} = 0.213224$ $B_{07, g8} = 0.977976$).

Adjustment for Rater-Year Effects

In 2009, DRC performed a rater drift study similar to the one conducted by ETS in 2007 to adjust for the rater-year effects. DRC’s Performance Assessment Services (PAS) staff randomly pulled 510 student responses from 2007 for each grade’s prompt and distributed these into the current administration scoring throughout the entire scoring timeframe. The student responses were pulled by lithocode and were only the valid scored responses (i.e., no condition codes such as off-topic present). 2009 scorers then rescored these papers. The lithocodes randomly pulled by PAS were provided to EIS for generating the data files for Psychometric Services (PS) department.

The 09 rater-year effect constants, $C_{09, grade}$ and $D_{09, grade}$, were determined by using the means (M) and standard deviations (S) of the 2007 raw composite scores and the 2009 rescored raw composite scores ($RS09$) as calculated below for each grade:

$$C_{09, grade} = M_{07, grade} - M_{RS09, grade} \times D_{09, grade}$$

$$D_{09, grade} = S_{07, grade} / S_{RS09, grade}$$

The formula for the 09 rater-year effects adjusted score is:

$$SCS_{2_{Op09,grade}} = D_{09, grade} \times RCS_{Op09,grade} + C_{09, grade}$$

Once the transformation constants are applied to the 2009 rescored raw composite scores, the mean and standard deviation of the adjusted 2009 scores should be the same as the 2007 mean and standard deviation.

A Compound Adjustment

Following the calculation of the 2009 transformation constants, compound adjustments were made to the 2009 operational raw composite scores. The generic formula for producing the final 2009 scaled composite score ($SCS_{Op09,grade}$) is:

$$SCS_{Op09,grade} = B_{07, grade} \times (D_{09, grade} \times RCS_{Op09,grade} + C_{09, grade}) + A_{07, grade}$$

To simplify the calculation, transformation constants for each of the Grades 5 and 8 were calculated as below:

$$E_{09, grade} = B_{07, grade} \times C_{09, grade} + A_{07, grade}$$

$$F_{09, grade} = D_{09, grade} \times B_{07, grade}$$

The following formula was used to calculate the final scaled composite scores. The calculated values are rounded to the nearest whole integer. Resulting values outside of the 15-60 range are set to the nearest bound.

$$SCS_{Op09,grade} = F_{09, grade} \times RCS_{Op09,grade} + E_{09, grade}$$

The scaled composite score will be converted to the performance level using Table 2.3.B.

Table 2.3.B
Scaled Score Ranges for Each Achievement Level

GRADE 5 SCALED COMPOSITE SCORE	GRADE 8 SCALED COMPOSITE SCORE	Performance Level
54 – 60	54 – 60	Advanced
36 – 53	36 – 53	Satisfactory
26 – 35	25 – 35	Limited Knowledge
15 – 25	15 – 24	Unsatisfactory
Unscorable	Unscorable	Unsatisfactory

Summary statistics for the scaling analysis of the operational writing prompts are provided in Tables 2.3.C to 2.3.E. Table 2.3.C provides the sample means and standard deviations used to calculate the transformation constants for each grade. The results indicate that sampled students in both grades had lower 2009 prompt scores. Because the responses scored were the same across the two years, this indicates that the raters were more strict in 2009.

Table 2.3.C
Sample Means and Standard Deviations Used for Calculating Constants

Grade	Statistic	Raters	
		2007	2009
5	N	505	505
	MIN	15	15
	MAX	60	60
	MEAN*	43.75	39.43
	STD*	9.39	9.38
8	N	508	508
	MIN	15	15
	MAX	60	60
	MEAN*	44.30	41.39
	STD*	9.00	9.30

*Tabled values are rounded for display purposes. Transformations were performed without rounding.

Tables 2.3.D and 2.3.E provide the resulting score distribution statistics with no adjustment, only the ETS adjustment, and the compound DRC and ETS adjustment. Transformation constants are provided at the bottom of the Tables. The 2006 and 2007 score distributions—as reported in *OCCT Writing: Scaling the 2007 Field-Test Prompts (ETS, 2007)*—are also provided for comparison. Relative to no adjustment and the ETS only adjustment, the DRC and ETS compound adjustment led to higher mean scores at both grades.

Table 2.3.D
Grade 5 Writing Results

		2009 No Adjustment	2009 ETS Only	2009 DRC & ETS	2008 Scores	2007 Scores
Statistic	N	43665	43665	43665	41988	42649
	MIN	15	15	19	17	15
	MAX	60	60	60	60	60
	MEAN	40.79	40.34	44.57	44.01	43.25
	STD	8.67	8.81	8.54	8.91	9.00
Perf Level %	PL 1 U	4.8	5.5	3.5	4.6	3.2
	PL 2 L	23.1	23.2	13.7	15.0	16.2
	PL 3 S	65.5	65.3	69.0	67.4	67.5
	PL 4 A	6.6	6.0	13.8	13.0	13.1
Constants						
DRC	C	Additive	4.259475			
	D	Multiplicative	1.001316			
ETS	A	Additive	-1.006108			
	B	Multiplicative	1.015014			
Combined	E	Additive	3.317319			
	F	Multiplicative	1.016350			

**Table 2.3.E
Grade 8 Writing Results**

		2009 No Adjustment	2009 ETS Only	2009 DRC & ETS	2008 Scores	2007 Scores
Statistic	N	40962	40962	40962	42271	43589
	MIN	15	15	19	18	15
	MAX	60	59	60	60	59
	MEAN	43.71	42.86	45.73	45.50	44.08
	STD	7.97	7.75	7.42	7.04	8.18
Perf Level %	PL 1 U	2.6	2.8	2.0	1.8	2.1
	PL 2 L	12.8	13.4	8.8	8.7	11.4
	PL 3 S	74.3	75.3	75.2	78.0	74.6
	PL 4 A	10.3	8.4	14.1	11.6	11.9
Constants						
DRC	C	Additive	4.217337			
	D	Multiplicative	0.968307			
ETS	A	Additive	0.213224			
	B	Multiplicative	0.977976			
Combined	E	Additive	4.337679			
	F	Multiplicative	0.946981			

2.3.3 Rater Agreement for Operational Writing Prompts

As stated earlier, student responses were rated by two independent raters, and the score for each domain was the average of the two ratings. The average of the two ratings was used for the calculation of the final composite score.

Consistency between the two ratings was evaluated with the following statistics:

- Percentage of exact agreement between raters
- Percentage of adjacent agreement between raters
- Correlation between ratings 1 and 2

Table 2.3.F provides a summary of the rater-agreement analysis for the Grade 5 and Grade 8 operational prompts. Included are the mean and standard deviation of assigned ratings, the percentage of exact and adjacent ratings, and the correlation between ratings. In Grade 5 writing, the exact agreement rate ranged from 70% to 76%, and the sum of exact plus adjacent agreement rates ranged from 99% to 100%. For Grade 8 Writing, the exact agreement rate ranged from 68% to 72%, and the sum of the exact plus adjacent agreement rates was 100%. The correlations between ratings ranged from 0.70 to 0.73 in Grade 5 and 0.59 to 0.65 in Grade 8. In general, the raters were fairly consistent in each domain.

Table 2.3.F
Inter-rater Agreement for Operational Writing Prompts

Grade	Domain ^a	N	Rating 1		Rating 2		Percent Agreement			Corr ^b
			Mean	SD	Mean	SD	Exact	Adjacent	Exact+ Adjacent	
5	ID	44543	2.76	0.66	2.76	0.67	76	24	100	0.72
	OUC	44543	2.71	0.67	2.71	0.68	75	25	100	0.71
	WC	44543	2.72	0.67	2.72	0.68	73	26	99	0.70
	SP	44543	2.64	0.76	2.64	0.76	70	29	99	0.73
	GUM	44543	2.68	0.76	2.68	0.76	70	29	99	0.72
8	ID	42290	2.94	0.65	2.94	0.65	69	31	100	0.64
	OUC	42290	2.94	0.66	2.94	0.66	69	31	100	0.64
	WC	42290	3.00	0.59	2.99	0.59	72	28	100	0.59
	SP	42290	2.82	0.68	2.82	0.68	68	32	100	0.65
	GUM	42290	2.81	0.63	2.81	0.64	69	31	100	0.61

^a ID=Ideas and Development; OUC=Organization, Unity, and Coherence; WC=Word Choice; SP=Sentences and Grammar; GUM=Grammar, Usage, and Mechanics

^b Pearson correlation between first and second ratings

CHAPTER III. STATE RESULTS

In this section, performance on the OCCT is summarized for the participating Oklahoma student population and for demographic subgroups. All reported results are based on valid scores on the 2009 forms in the final student data received by July 17, 2009. These data differ from the analysis data in several ways: corrections were made to student and school information, invalidations and missing data issues were resolved, and all students who took the standard, equivalent, and braille OCCTs were included. Thus, final counts of examinees by test differ somewhat from samples used for item and test analysis.

As described in Chapter II, prior to the release of student reports, raw scores were converted to a reporting scale metric. Raw scores on the Multiple-Choice tests were converted to scaled scores using the conversion tables provided in Appendix B. For the Writing tests, analytic scores were converted to composite scores using the formulas provided in the previous section. Achievement level scores were assigned as well using the SDE-established OCCT cut scores.

The means and standard deviations of students' raw scores and scaled or composite scores are shown in Table 3.1. Table 3.2 provides the percentage of students in each achievement category in 2006, 2007 and 2008. Tables 3.3 to 3.21 provide test results by demographic subgroups. Tables B.1 through B.19 provide the raw score, scaled score, CSEM, achievement level, and frequency distributions for each OCCT test.

Table 3.1
Means and Standard Deviations of Students' Raw Scores and Scaled Scores

Content Area	Grade	Valid N	Raw Scores				Scaled Scores			
			Mean	SD	Min	Max	Mean	SD	Min	Max
Reading	3	43703	36.1	8.3	0	50	727.9	80.9	400	990
	4	42754	37.6	8.4	0	50	718.3	75.0	400	990
	5	42563	37.3	8.2	0	50	722.3	78.8	400	990
	6	41235	36.9	8.6	0	50	718.9	78.0	400	990
	7	40375	36.6	8.2	0	50	729.6	72.1	400	990
	8	40380	36.8	7.8	1	50	727.5	80.1	400	990
Mathematics	3	44278	32.4	8.8	4	45	728.1	89.0	400	990
	4	43335	31.5	7.9	0	45	729.7	89.1	400	990
	5	42866	29.8	8.4	0	45	722.1	83.5	400	990
	6	41548	30.3	8.5	0	45	719.3	74.9	400	990
	7	40595	29.0	8.0	1	45	716.5	82.5	400	990
	8	40390	29.9	8.3	0	45	716.2	85.8	400	990
Science	5	43365	30.2	8.1	0	45	768.0	73.0	400	990
	8	41287	30.8	7.8	0	45	767.3	67.2	400	990
Social Studies	5	46184	37.8	10.9	0	60	726.4	76.8	400	990
Geography	7	43750	29.4	8.3	1	45	768.9	88.3	400	990
History	8	43806	28.7	8.7	1	45	726.8	83.1	400	990
Writing	5	44543	40.7	8.7	15	60	44.6	8.5	19	60
	8	42290	43.6	7.9	15	60	45.7	7.4	19	60

^a Mean writing composite scores are reported.

Table 3.2 presents the percentage of students scoring in each of the four achievement levels for all students for the current year and the past two years. It shows that for all grades and subject

areas, the percentage of students scoring at or above the Satisfactory achievement level increased from 2007 to 2009.

Table 3.2
Percentage of Students Performing within Each Achievement Category in 2007 to 2009

Content Area	Grade	Unsatisfactory			Limited Knowledge			Satisfactory			Advanced			Satisfactory or Advanced		
		2007	2008	2009	2007	2008	2009	2007	2008	2009	2007	2008	2009	2007	2008	2009
Reading	3	3	2	14	10	11	19	82	83	63	5	4	4	87	87	67
	4	3	1	17	7	6	20	86	88	57	4	4	6	90	92	63
	5	7	4	13	13	12	22	71	74	56	10	10	9	81	84	65
	6	9	7	15	14	12	20	71	72	56	7	9	8	78	81	64
	7	8	5	17	15	17	13	66	64	54	11	13	16	77	78	70
	8	7	4	15	14	14	18	70	73	57	9	9	10	79	82	67
Mathematics	3	3	2	12	21	20	22	60	62	45	15	16	21	75	78	66
	4	2	2	13	16	15	20	63	64	50	19	19	17	82	83	67
	5	3	2	13	14	11	22	57	59	35	26	27	30	83	87	65
	6	10	6	18	13	13	19	52	48	31	24	32	32	76	80	63
	7	12	7	22	14	15	15	52	51	35	22	26	28	74	77	63
	8	8	4	21	14	14	18	54	57	36	23	24	25	77	81	61
Science	5	5	3	4	13	12	12	58	58	58	24	27	26	82	84	84
	8	7	4	3	10	9	11	72	75	69	11	12	17	83	87	86
Social Studies	5	15	12	14	20	20	19	47	46	46	18	22	22	65	68	68
Geography	7	4	3	3	17	17	16	66	67	63	14	13	18	80	80	81
History	8	9	9	9	26	25	23	55	56	58	10	10	10	65	66	68

Tables 3.3 to 3.21 present the scaled score and achievement level results for each test by population subgroups. Ethnic category membership is based on identifying one ethnicity; those identifying more than one or none are classified as Other. Economically disadvantaged is based on participation in Free and Reduced-Price Lunch.

Table 3.3
Subgroup Results: Grade 3 Reading

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		43703	36.1	8.3	0.89	0.89	727.9	80.9	13.7	19.2	63.4	3.7
Gender	Female	21753	35.6	8.5	0.89	0.89	722.8	81.9	15.3	19.7	61.7	3.3
	Male	21950	36.7	8.0	0.88	0.88	732.9	79.5	12.1	18.6	65.2	4.1
Ethnicity	African American	4643	33.0	8.9	0.89	0.89	697.9	81.8	23.0	24.4	51.1	1.5
	American Indian	8075	35.8	8.0	0.88	0.88	724.1	77.1	13.6	21.0	62.3	3.1
	Hispanic	5052	33.0	8.6	0.88	0.88	697.9	77.7	23.0	24.4	51.1	1.4
	Asian	813	37.7	8.2	0.89	0.90	744.6	84.9	10.6	14.9	68.8	5.8
	White	70	36.0	8.8	0.90	0.90	727.9	87.0	14.3	15.7	64.3	5.7
	Pacific Islander	24537	37.5	7.8	0.88	0.88	740.7	79.0	10.1	16.6	68.6	4.7
	Other	513	34.9	8.4	0.89	0.89	715.5	81.4	16.2	23.8	57.3	2.7
IEP	No	42508	36.4	8.1	0.88	0.88	730.2	79.4	12.7	19.1	64.5	3.8
	Yes	1195	26.9	9.4	0.89	0.89	643.6	85.8	49.5	23.6	26.6	0.3
ELL	No	42812	36.3	8.2	0.88	0.89	729.3	80.3	13.1	19.0	64.1	3.8
	Yes	891	28.5	8.9	0.88	0.88	658.5	77.5	41.8	26.3	31.8	0.2
FLS	No	18415	38.6	7.3	0.87	0.87	752.5	77.7	7.3	14.2	72.4	6.1
	Yes	25288	34.3	8.5	0.88	0.88	710.0	78.4	18.3	22.8	56.9	2.0

Table 3.4
Subgroup Results: Grade 4 Reading

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		42754	37.6	8.4	0.89	0.89	718.3	75.0	17.3	19.5	57.0	6.2
Gender	Female	21349	37.0	8.7	0.90	0.90	713.5	76.6	19.3	20.1	54.7	5.8
	Male	21405	38.2	8.1	0.89	0.89	723.0	73.0	15.3	18.9	59.2	6.5
Ethnicity	African American	4434	34.2	9.0	0.89	0.89	689.3	71.9	29.0	24.7	44.0	2.3
	American Indian	7973	36.9	8.3	0.89	0.89	711.5	72.1	18.7	21.9	54.7	4.7
	Hispanic	4683	34.3	9.0	0.89	0.89	689.4	71.4	28.4	24.3	44.8	2.5
	Asian	753	39.5	7.9	0.89	0.89	738.5	80.6	12.6	16.2	61.1	10.1
	Pacific Islander	80	36.2	8.6	0.89	0.89	708.1	74.4	21.3	25.0	46.3	7.5
	White	24229	39.0	7.8	0.89	0.89	731.0	73.6	12.7	16.9	62.4	8.0
	Other	602	37.0	8.5	0.89	0.89	712.0	72.2	19.3	21.6	54.2	5.0
IEP	No	41314	37.9	8.1	0.89	0.89	721.1	73.2	16.0	19.4	58.2	6.4
	Yes	1440	27.8	9.9	0.90	0.90	638.3	80.9	56.2	21.7	21.5	0.6
ELL	No	42165	37.7	8.3	0.89	0.89	719.4	74.4	16.7	19.5	57.5	6.3
	Yes	589	27.9	9.3	0.88	0.88	641.3	73.6	58.2	20.7	20.4	0.7
FLS	No	18704	40.2	7.2	0.88	0.88	741.7	72.9	9.3	14.8	65.9	10.0
	Yes	24050	35.6	8.7	0.89	0.89	700.0	71.4	23.5	23.2	50.1	3.2

Table 3.5
Subgroup Results: Grade 5 Reading

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		42563	37.3	8.2	0.89	0.89	722.3	78.8	12.9	21.7	56.2	9.2
Gender	Female	21310	36.7	8.5	0.90	0.90	715.8	80.6	15.3	22.3	54.3	8.2
	Male	21253	38.0	7.7	0.88	0.88	728.8	76.4	10.6	21.0	58.1	10.2
Ethnicity	African American	4526	34.6	8.6	0.89	0.89	695.7	77.3	20.4	27.7	47.8	4.1
	American Indian	7995	36.4	8.3	0.89	0.89	712.6	77.8	15.1	24.2	53.7	7.0
	Hispanic	4460	34.9	8.6	0.89	0.89	698.9	77.0	19.9	27.1	48.0	5.0
	Asian	787	39.8	7.8	0.90	0.90	749.8	83.0	8.3	16.9	56.2	18.7
	Pacific Islander	86	34.0	9.8	0.92	0.92	692.0	97.8	27.9	20.9	46.5	4.7
	White	24063	38.6	7.7	0.88	0.89	734.2	76.9	9.6	18.8	60.1	11.4
	Other	646	36.8	7.9	0.88	0.88	715.6	74.2	13.9	22.6	56.8	6.7
IEP	No	40907	37.7	7.9	0.89	0.89	725.7	76.8	11.5	21.3	57.6	9.5
	Yes	1656	27.9	9.3	0.89	0.89	637.5	80.2	47.4	29.4	22.5	0.7
ELL	No	42130	37.4	8.1	0.89	0.89	723.1	78.4	12.6	21.6	56.5	9.3
	Yes	433	28.6	9.2	0.89	0.89	645.2	78.0	45.3	30.5	22.6	1.6
FLS	No	18926	39.6	7.2	0.88	0.88	745.0	75.8	7.1	16.3	62.4	14.2
	Yes	23637	35.5	8.4	0.89	0.89	704.1	76.5	17.6	26.0	51.2	5.2

Table 3.6
Subgroup Results: Grade 6 Reading

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		41235	36.9	8.6	0.90	0.90	718.9	78.0	15.0	20.3	56.2	8.4
Gender	Female	20366	36.1	9.1	0.91	0.91	712.1	81.4	17.9	20.9	53.4	7.7
	Male	20869	37.7	8.0	0.88	0.88	725.6	73.9	12.2	19.7	58.9	9.2
Ethnicity	African American	4153	33.5	9.2	0.90	0.90	688.2	77.9	25.2	26.6	44.6	3.6
	American Indian	7664	36.1	8.4	0.89	0.89	710.5	74.0	16.7	23.3	53.8	6.2
	Hispanic	4186	33.9	9.0	0.90	0.90	691.5	76.8	23.4	26.4	46.5	3.6
	Asian	787	39.5	7.9	0.89	0.89	744.3	78.4	9.0	15.9	60.5	14.6
	Pacific Islander	78	35.1	10.5	0.93	0.93	702.6	88.6	21.8	16.7	55.1	6.4
	White	23559	38.3	8.1	0.89	0.89	731.4	76.2	11.4	17.2	60.7	10.7
	Other	808	36.3	8.6	0.89	0.89	713.7	78.1	15.6	23.8	53.0	7.7
IEP	No	39773	37.2	8.4	0.89	0.89	721.7	76.4	13.9	20.1	57.3	8.7
	Yes	1462	28.1	9.8	0.90	0.90	643.8	82.2	46.3	27.1	25.9	0.8
ELL	No	40873	37.0	8.5	0.90	0.90	719.8	77.4	14.7	20.3	56.5	8.5
	Yes	362	26.0	10.0	0.90	0.90	625.9	87.4	56.4	22.9	19.9	0.8
FLS	No	19553	39.3	7.6	0.88	0.88	740.8	73.9	8.7	15.1	63.6	12.6
	Yes	21682	34.8	8.9	0.89	0.89	699.2	76.3	20.7	25.1	49.5	4.7

Table 3.7
Subgroup Results: Grade 7 Reading

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		40375	36.6	8.2	0.89	0.89	729.6	72.1	16.9	13.3	53.7	16.2
Gender	Female	20295	35.9	8.6	0.89	0.89	723.5	73.7	19.4	13.6	52.5	14.4
	Male	20080	37.3	7.7	0.88	0.88	735.9	69.9	14.2	13.0	54.8	17.9
Ethnicity	African American	4043	32.9	8.6	0.88	0.88	698.2	69.1	29.2	19.4	44.8	6.6
	American Indian	7375	35.8	8.3	0.88	0.88	721.9	70.0	19.3	14.9	52.6	13.2
	Hispanic	4031	33.5	8.7	0.88	0.88	703.0	70.4	27.8	16.8	47.0	8.4
	Asian	773	39.5	7.2	0.87	0.87	757.2	72.1	8.9	8.7	55.0	27.4
	Pacific Islander	60	33.7	8.4	0.87	0.87	702.9	63.9	20.0	20.0	51.7	8.3
	White	23279	38.0	7.7	0.88	0.88	741.4	70.3	12.3	11.3	56.5	19.9
	Other	814	36.3	7.9	0.87	0.87	725.8	67.5	16.6	13.5	57.0	12.9
IEP	No	38993	36.9	8.0	0.88	0.88	732.0	70.8	15.7	13.1	54.5	16.7
	Yes	1382	28.5	9.3	0.89	0.89	663.2	74.7	48.3	19.2	30.2	2.2
ELL	No	40080	36.7	8.2	0.88	0.88	730.2	71.7	16.5	13.3	53.9	16.3
	Yes	295	26.5	9.4	0.89	0.89	649.4	80.4	59.3	18.0	19.0	3.7
FLS	No	20062	38.8	7.3	0.87	0.87	749.0	69.6	10.0	10.0	57.1	22.9
	Yes	20313	34.4	8.5	0.88	0.88	710.5	69.3	23.6	16.6	50.3	9.6

Table 3.8
Subgroup Results: Grade 8 Reading

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		40380	36.8	7.8	0.87	0.87	727.5	80.1	14.9	17.9	57.1	10.1
Gender	Female	20051	36.3	7.9	0.88	0.88	722.4	80.3	16.2	18.5	56.6	8.7
	Male	20329	37.3	7.6	0.87	0.87	732.5	79.6	13.6	17.3	57.7	11.5
Ethnicity	African American	4034	33.3	8.4	0.88	0.88	692.3	79.5	27.0	23.8	45.3	3.9
	American Indian	7482	36.3	7.6	0.86	0.86	721.5	76.1	15.4	19.9	56.7	8.1
	Hispanic	3925	32.7	8.8	0.89	0.89	687.6	83.7	30.8	22.0	43.1	4.1
	Asian	732	38.3	8.3	0.90	0.90	746.3	91.3	13.0	11.9	57.0	18.2
	Pacific Islander	93	33.4	9.4	0.90	0.90	694.2	95.6	28.0	18.3	46.2	7.5
	White	23349	38.2	7.0	0.85	0.86	741.9	75.9	9.9	15.8	61.7	12.6
	Other	765	36.6	7.5	0.86	0.86	723.8	75.5	14.8	18.0	58.3	8.9
IEP	No	39359	37.0	7.7	0.87	0.87	729.3	79.3	14.1	17.7	57.9	10.3
	Yes	1021	29.4	8.4	0.86	0.86	656.3	77.2	43.1	26.7	28.7	1.5
ELL	No	40165	36.9	7.7	0.87	0.87	728.2	79.5	14.5	17.9	57.4	10.2
	Yes	215	22.6	7.7	0.82	0.82	594.8	76.6	79.5	12.6	7.4	0.5
FLS	No	20819	38.9	6.8	0.85	0.85	749.3	76.0	8.3	13.9	63.2	14.6
	Yes	19561	34.5	8.1	0.87	0.87	704.3	77.7	21.9	22.1	50.7	5.3

Table 3.9
Subgroup Results: Grade 3 Mathematics

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		44278	32.4	8.8	0.91	0.91	728.1	89.0	12.0	21.6	45.5	20.9
Gender	Female	22179	32.9	8.7	0.91	0.91	732.6	89.4	11.0	20.7	46.0	22.3
	Male	22099	32.0	8.9	0.91	0.91	723.6	88.2	13.1	22.6	44.9	19.4
Ethnicity	African American	4682	28.2	9.5	0.91	0.91	687.2	91.4	24.1	27.9	37.3	10.8
	American Indian	8191	32.1	8.6	0.90	0.90	724.8	85.1	12.0	23.1	45.9	19.0
	Hispanic	5158	30.0	9.0	0.90	0.91	705.1	86.6	17.4	26.8	42.1	13.6
	Asian	844	35.4	8.2	0.91	0.91	759.8	93.4	7.2	14.2	45.7	32.8
	Pacific Islander	71	30.7	9.6	0.92	0.92	715.2	93.5	19.7	26.8	36.6	16.9
	White	24810	33.7	8.3	0.90	0.90	740.9	86.6	8.8	19.0	47.6	24.6
	Other	522	31.4	8.8	0.90	0.90	717.9	84.3	13.2	26.1	43.7	17.0
IEP	No	42606	32.7	8.7	0.91	0.91	730.3	88.2	11.4	21.3	46.0	21.4
	Yes	1672	26.4	9.2	0.90	0.90	670.9	87.9	29.7	31.3	32.5	6.6
ELL	No	43028	32.5	8.7	0.91	0.91	729.2	88.6	11.7	21.4	45.7	21.1
	Yes	1250	28.5	9.6	0.91	0.92	690.8	94.1	23.8	28.2	36.2	11.8
FLS	No	18551	34.8	7.9	0.90	0.90	751.8	85.6	6.9	16.3	48.0	28.7
	Yes	25727	30.7	9.0	0.91	0.91	711.0	87.4	15.7	25.5	43.7	15.2

Table 3.10
Subgroup Results: Grade 4 Mathematics

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		43335	31.5	7.9	0.88	0.88	729.7	89.1	12.7	20.1	50.2	17.0
Gender	Female	21775	31.9	8.0	0.89	0.89	734.3	91.2	12.2	18.9	50.2	18.8
	Male	21560	31.1	7.9	0.88	0.88	725.1	86.7	13.2	21.3	50.3	15.2
Ethnicity	African American	4492	28.0	8.4	0.88	0.88	691.5	91.3	24.0	26.1	41.4	8.5
	American Indian	8080	30.9	7.8	0.87	0.87	721.7	84.5	13.5	22.1	50.7	13.8
	Hispanic	4824	29.0	8.2	0.88	0.88	702.0	89.6	20.0	24.2	46.2	9.7
	Asian	787	35.0	7.6	0.89	0.89	773.3	96.9	6.9	11.1	47.3	34.8
	Pacific Islander	85	29.4	9.5	0.91	0.92	704.9	112.3	21.2	21.2	41.2	16.5
	White	24459	32.8	7.5	0.87	0.88	743.7	86.0	9.1	17.7	52.7	20.5
	Other	608	31.0	7.6	0.87	0.87	724.3	83.4	11.8	23.4	50.0	14.8
IEP	No	41416	31.8	7.8	0.88	0.88	733.2	87.6	11.6	19.7	51.1	17.6
	Yes	1919	24.7	8.3	0.87	0.87	655.9	90.8	36.6	29.2	31.0	3.2
ELL	No	42451	31.6	7.9	0.88	0.88	731.0	88.6	12.3	20.0	50.5	17.2
	Yes	884	25.9	8.5	0.88	0.88	668.7	91.7	33.7	25.0	36.4	4.9
FLS	No	18842	33.8	7.2	0.87	0.87	755.8	85.0	6.8	15.0	53.5	24.7
	Yes	24493	29.7	8.0	0.88	0.88	709.7	87.0	17.2	24.0	47.8	11.0

Table 3.11
Subgroup Results: Grade 5 Mathematics

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		42866	29.8	8.4	0.89	0.89	722.1	83.5	13.1	22.2	35.2	29.6
Gender	Female	21567	30.1	8.6	0.90	0.90	724.7	86.7	13.3	20.7	34.1	31.9
	Male	21299	29.5	8.2	0.88	0.88	719.4	80.1	12.8	23.7	36.2	27.3
Ethnicity	African American	4546	26.8	8.5	0.88	0.88	691.9	85.3	22.3	26.8	32.8	18.1
	American Indian	8040	28.8	8.3	0.88	0.88	711.9	80.4	15.4	24.4	36.1	24.1
	Hispanic	4580	27.9	8.6	0.89	0.89	703.5	84.5	18.2	26.4	32.3	23.1
	Asian	807	34.4	8.1	0.90	0.91	770.5	89.9	5.5	14.4	26.5	53.7
	Pacific Islander	89	26.9	9.7	0.91	0.91	693.9	95.8	27.0	24.7	25.8	22.5
	White	24155	31.0	8.2	0.89	0.89	733.4	81.2	9.8	20.0	36.0	34.2
	Other	649	29.1	7.7	0.86	0.86	714.3	72.5	12.8	23.4	40.5	23.3
IEP	No	40891	30.2	8.3	0.89	0.89	725.5	81.8	11.9	21.6	35.8	30.7
	Yes	1975	22.6	7.9	0.85	0.85	650.3	87.7	37.5	33.2	22.7	6.6
ELL	No	42209	29.9	8.4	0.89	0.89	723.0	83.1	12.7	22.0	35.4	29.9
	Yes	657	23.8	8.8	0.89	0.89	662.4	93.1	35.9	30.0	21.0	13.1
FLS	No	18991	32.2	8.0	0.89	0.89	745.2	80.6	7.5	16.9	35.5	40.1
	Yes	23875	27.9	8.3	0.88	0.88	703.7	81.3	17.5	26.3	34.9	21.2

Table 3.12
Subgroup Results: Grade 6 Mathematics

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		41548	30.3	8.5	0.89	0.90	719.3	74.9	17.6	18.6	31.4	32.3
Gender	Female	20652	30.4	8.8	0.90	0.90	720.3	78.2	18.4	17.3	30.6	33.7
	Male	20896	30.2	8.3	0.89	0.89	718.3	71.5	16.9	19.9	32.3	31.0
Ethnicity	African American	4164	26.5	8.6	0.88	0.89	686.7	74.4	29.9	24.6	28.1	17.5
	American Indian	7708	29.3	8.4	0.88	0.89	710.5	71.1	19.4	21.3	32.2	27.1
	Hispanic	4266	28.0	8.6	0.89	0.89	699.4	73.5	25.7	21.7	30.3	22.4
	Asian	814	35.1	7.6	0.89	0.90	765.6	80.1	5.9	12.2	25.3	56.6
	Pacific Islander	83	28.9	10.1	0.92	0.93	706.3	84.4	31.3	12.0	20.5	36.1
	White	23708	31.6	8.2	0.89	0.89	730.1	73.1	13.7	16.4	32.1	37.8
	Other	805	29.6	8.4	0.89	0.89	713.9	73.1	20.2	17.8	36.0	26.0
IEP	No	39755	30.6	8.4	0.89	0.89	722.0	73.7	16.4	18.4	31.9	33.4
	Yes	1793	23.1	8.2	0.86	0.87	659.8	77.1	46.3	24.0	21.2	8.5
ELL	No	41121	30.4	8.5	0.89	0.89	719.9	74.5	17.3	18.6	31.6	32.5
	Yes	427	22.7	8.9	0.89	0.89	656.1	87.9	50.8	21.5	17.3	10.3
FLS	No	19634	32.6	8.0	0.89	0.89	739.4	73.3	10.9	14.1	31.9	43.1
	Yes	21914	28.2	8.4	0.88	0.89	701.3	71.7	23.7	22.7	31.1	22.6

Table 3.13
Subgroup Results: Grade 7 Mathematics

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		40595	29.0	8.0	0.87	0.87	716.5	82.5	22.2	15.0	35.0	27.8
Gender	Female	20517	29.1	8.1	0.87	0.88	717.4	84.2	21.9	14.6	35.4	28.2
	Male	20078	28.9	7.9	0.87	0.87	715.6	80.7	22.7	15.3	34.5	27.5
Ethnicity	African American	4071	25.5	8.0	0.86	0.86	681.0	83.8	37.4	18.0	29.4	15.1
	American Indian	7411	27.9	7.7	0.86	0.86	705.9	77.5	26.0	16.1	35.7	22.2
	Hispanic	4105	26.4	8.0	0.86	0.86	690.1	84.3	32.4	18.0	31.6	18.1
	Asian	800	34.1	7.3	0.88	0.88	773.3	85.0	7.9	8.3	29.0	54.9
	Pacific Islander	63	27.0	8.0	0.86	0.87	693.4	88.2	25.4	17.5	36.5	20.6
	White	23334	30.2	7.7	0.87	0.87	729.2	79.5	17.1	13.7	36.4	32.8
	Other	811	28.1	7.7	0.86	0.86	707.8	78.3	23.8	15.5	37.2	23.4
IEP	No	39002	29.2	7.9	0.87	0.87	719.2	81.3	21.0	14.8	35.4	28.7
	Yes	1593	22.6	7.5	0.83	0.83	651.4	84.6	51.9	18.3	23.4	6.5
ELL	No	40242	29.0	8.0	0.87	0.87	717.2	82.1	22.0	14.9	35.1	28.0
	Yes	353	22.1	8.3	0.86	0.87	645.1	94.0	54.1	17.3	19.3	9.3
FLS	No	20123	31.2	7.5	0.86	0.87	739.0	79.1	14.0	11.9	36.4	37.6
	Yes	20472	26.8	7.8	0.85	0.86	694.5	79.8	30.3	17.9	33.5	18.3

Table 3.14
Subgroup Results: Grade 8 Mathematics

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		40390	29.9	8.3	0.88	0.88	716.2	85.8	21.1	17.5	35.9	25.4
Gender	Female	20134	29.9	8.5	0.89	0.89	716.5	89.0	21.8	17.4	34.3	26.5
	Male	20256	29.8	8.1	0.88	0.88	715.9	82.5	20.5	17.6	37.5	24.3
Ethnicity	African American	4039	26.7	8.3	0.87	0.87	683.2	86.6	32.7	21.0	32.6	13.7
	American Indian	7438	28.8	8.1	0.87	0.87	705.3	81.1	24.0	19.7	36.3	20.1
	Hispanic	3985	27.2	8.3	0.87	0.87	689.5	84.8	31.3	20.4	32.8	15.5
	Asian	751	34.6	8.1	0.90	0.91	767.8	95.8	10.5	7.6	32.8	49.1
	Pacific Islander	94	27.9	8.5	0.88	0.88	697.3	89.2	30.9	21.3	27.7	20.2
	White	23316	31.1	8.1	0.88	0.88	728.9	83.3	16.7	16.0	37.0	30.3
	Other	767	28.4	8.3	0.87	0.88	700.7	87.6	24.5	18.8	38.2	18.5
IEP	No	39168	30.1	8.2	0.88	0.88	718.3	84.8	20.2	17.4	36.4	26.0
	Yes	1222	23.1	8.0	0.85	0.86	647.8	90.5	50.7	22.1	20.3	7.0
ELL	No	40117	29.9	8.3	0.88	0.88	716.7	85.5	20.9	17.5	36.1	25.5
	Yes	273	23.0	8.8	0.88	0.88	647.0	100.4	53.1	20.5	16.5	9.9
FLS	No	20780	32.0	7.9	0.88	0.88	737.8	83.5	14.0	14.5	37.2	34.3
	Yes	19610	27.6	8.1	0.87	0.87	693.4	82.2	28.6	20.8	34.6	16.0

Table 3.15
Subgroup Results: Grade 5 Science

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		43365	30.2	8.1	0.88	0.88	768.0	73.0	3.9	12.3	58.1	25.7
Gender	Female	21804	30.7	8.3	0.89	0.89	771.6	75.7	4.1	12.0	55.5	28.4
	Male	21561	29.8	7.9	0.88	0.88	764.4	70.0	3.8	12.6	60.7	22.9
Ethnicity	African American	4653	25.8	8.1	0.87	0.87	728.8	72.8	9.4	22.0	58.7	10.0
	American Indian	8153	29.6	7.9	0.88	0.88	762.3	70.0	4.3	13.1	60.5	22.0
	Hispanic	4628	27.1	8.0	0.87	0.87	741.2	71.1	6.4	19.8	59.7	14.1
	Asian	807	32.7	7.8	0.89	0.89	791.2	73.5	1.5	10.2	50.4	37.9
	Pacific Islander	88	27.2	9.0	0.90	0.90	741.2	84.9	8.0	17.0	60.2	14.8
	White	24380	31.8	7.7	0.88	0.88	782.0	69.9	2.4	8.8	56.9	31.9
	Other	656	29.9	7.4	0.85	0.86	763.7	66.4	2.9	11.0	66.8	19.4
IEP	No	41047	30.6	8.0	0.88	0.88	771.3	71.3	3.4	11.5	58.3	26.8
	Yes	2318	23.8	8.0	0.86	0.86	710.3	78.8	13.7	26.9	53.6	5.8
ELL	No	42817	30.3	8.1	0.88	0.88	768.8	72.7	3.8	12.0	58.3	25.9
	Yes	548	23.2	7.9	0.86	0.86	708.9	73.6	13.0	35.4	44.9	6.8
FLS	No	19112	32.9	7.3	0.87	0.87	791.2	68.0	1.7	6.8	54.5	37.0
	Yes	24253	28.2	8.1	0.87	0.87	749.7	71.5	5.7	16.6	60.9	16.7

Table 3.16
Subgroup Results: Grade 8 Science

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		41287	30.8	7.8	0.88	0.88	767.3	67.2	2.9	11.0	68.7	17.3
Gender	Female	20671	31.0	8.1	0.89	0.89	769.2	70.5	3.4	11.1	66.3	19.2
	Male	20616	30.5	7.6	0.87	0.87	765.4	63.6	2.5	10.9	71.2	15.5
Ethnicity	African American	4200	26.5	8.0	0.87	0.87	732.5	66.7	7.1	21.1	65.3	6.5
	American Indian	7650	29.9	7.6	0.87	0.87	759.8	63.0	3.2	11.8	71.8	13.2
	Hispanic	4060	27.4	7.9	0.87	0.87	739.6	66.0	5.6	18.9	67.6	8.0
	Asian	748	33.1	8.1	0.90	0.90	788.4	75.5	2.7	8.0	59.2	30.1
	Pacific Islander	98	27.3	8.7	0.89	0.89	739.1	70.3	8.2	19.4	63.3	9.2
	White	23720	32.3	7.3	0.87	0.87	780.3	64.4	1.7	7.6	68.8	21.9
	Other	811	30.0	7.6	0.87	0.87	760.5	63.6	3.3	12.2	71.1	13.3
IEP	No	39322	31.2	7.6	0.87	0.87	770.6	65.2	2.4	9.9	69.6	18.1
	Yes	1965	22.7	7.7	0.85	0.85	701.0	72.2	14.5	32.9	50.6	2.0
ELL	No	40985	30.8	7.8	0.88	0.88	767.9	66.8	2.8	10.8	69.0	17.4
	Yes	302	20.5	7.5	0.84	0.84	684.0	71.2	22.2	37.7	37.4	2.6
FLS	No	21041	33.0	7.2	0.87	0.87	786.2	64.0	1.4	6.4	67.2	24.9
	Yes	20246	28.4	7.8	0.87	0.87	747.6	64.7	4.5	15.7	70.3	9.4

Table 3.17
Subgroup Results: Grade 5 Social Studies

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		46184	37.8	10.9	0.91	0.91	726.4	76.8	13.6	18.6	46.1	21.8
Gender	Female	23603	38.4	11.3	0.91	0.92	730.0	81.1	14.1	16.8	43.7	25.4
	Male	22581	37.1	10.4	0.89	0.89	722.6	71.9	13.0	20.4	48.5	18.0
Ethnicity	African American	5133	31.8	10.5	0.89	0.89	685.0	79.1	27.5	27.6	36.2	8.6
	American Indian	8711	37.2	10.5	0.90	0.90	722.4	73.4	13.6	19.7	48.1	18.6
	Hispanic	4979	33.8	10.6	0.89	0.89	699.6	77.7	21.4	25.2	41.8	11.5
	Asian	819	42.5	10.6	0.91	0.91	759.5	74.7	7.0	11.8	41.6	39.6
	Pacific Islander	92	35.1	11.7	0.91	0.92	704.9	89.5	25.0	14.1	44.6	16.3
	White	25757	39.8	10.5	0.90	0.90	740.1	72.7	9.4	15.4	48.1	27.0
	Other	693	38.0	10.4	0.89	0.89	727.5	72.7	12.8	15.6	53.0	18.6
IEP	No	41503	38.8	10.6	0.90	0.90	733.7	73.2	10.8	17.3	48.0	23.9
	Yes	4681	28.5	9.4	0.86	0.86	661.5	77.7	38.2	29.9	29.0	2.9
ELL	No	45594	37.9	10.9	0.90	0.91	727.2	76.4	13.2	18.4	46.4	22.0
	Yes	590	28.5	10.1	0.88	0.88	660.5	83.3	39.0	31.0	24.9	5.1
FLS	No	19748	41.8	10.0	0.90	0.90	753.2	69.3	6.5	11.8	48.3	33.3
	Yes	26436	34.8	10.6	0.89	0.89	706.4	76.0	18.8	23.6	44.4	13.2

Table 3.18
Subgroup Results: Grade 7 Geography

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		43750	29.4	8.3	0.88	0.88	768.9	88.3	3.3	15.8	62.8	18.1
Gender	Female	22535	29.9	8.7	0.89	0.89	773.8	93.5	3.8	15.2	59.5	21.6
	Male	21215	28.9	8.0	0.87	0.87	763.6	82.2	2.9	16.5	66.2	14.4
Ethnicity	African American	4564	24.5	8.4	0.87	0.87	718.9	91.3	8.4	30.3	54.9	6.4
	American Indian	8136	28.5	8.2	0.87	0.87	760.1	85.6	3.5	17.3	65.2	14.0
	Hispanic	4402	26.5	8.3	0.87	0.87	740.3	87.7	5.0	23.4	61.8	9.8
	Asian	817	33.2	7.6	0.88	0.88	810.6	84.6	1.3	7.8	56.4	34.4
	Pacific Islander	67	27.8	8.3	0.87	0.88	752.9	82.4	3.0	20.9	61.2	14.9
	White	24928	30.9	7.9	0.87	0.87	784.8	83.8	2.1	11.6	63.6	22.7
	Other	836	28.9	7.8	0.86	0.86	764.3	80.1	2.9	15.9	68.1	13.2
IEP	No	41102	30.0	8.0	0.87	0.87	775.1	84.0	2.4	14.1	64.4	19.1
	Yes	2648	20.3	7.9	0.85	0.85	672.2	97.5	17.6	42.1	37.9	2.4
ELL	No	43522	29.4	8.3	0.88	0.88	769.3	88.1	3.3	15.7	62.9	18.2
	Yes	228	21.0	7.5	0.83	0.83	685.3	85.1	11.0	45.6	40.4	3.1
FLS	No	20973	32.2	7.4	0.86	0.87	798.0	79.6	1.4	8.6	63.0	27.1
	Yes	22777	26.8	8.3	0.87	0.87	742.1	87.5	5.1	22.5	62.5	9.8

Table 3.19
Subgroup Results: Grade 8 History

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		43806	28.7	8.7	0.89	0.89	726.8	83.1	8.6	23.4	57.5	10.4
Gender	Female	22296	29.1	9.0	0.90	0.90	730.9	89.0	9.4	21.5	56.2	12.9
	Male	21510	28.2	8.2	0.87	0.88	722.6	76.3	7.8	25.4	59.0	7.8
Ethnicity	African American	4658	24.4	8.5	0.87	0.88	686.9	85.5	17.4	34.7	44.2	3.7
	American Indian	8186	27.8	8.4	0.88	0.88	718.6	79.3	9.2	26.2	57.1	7.5
	Hispanic	4253	25.6	8.5	0.88	0.88	698.9	81.2	14.0	31.8	49.6	4.6
	Asian	763	31.9	8.5	0.90	0.90	757.1	84.1	5.9	14.2	59.6	20.3
	Pacific Islander	100	25.9	8.6	0.88	0.88	697.3	86.5	15.0	32.0	49.0	4.0
	White	25014	30.2	8.3	0.89	0.89	740.9	80.3	6.0	19.2	61.4	13.4
	Other	832	28.5	8.2	0.88	0.88	725.6	75.3	7.6	24.0	59.9	8.5
IEP	No	39706	29.6	8.3	0.88	0.88	735.3	78.4	6.1	21.6	60.9	11.4
	Yes	4100	19.9	7.4	0.83	0.83	644.8	83.2	32.7	41.6	25.0	0.8
ELL	No	43494	28.7	8.6	0.89	0.89	727.5	82.7	8.4	23.3	57.8	10.5
	Yes	312	19.1	7.8	0.85	0.85	632.5	91.0	38.5	40.1	20.8	0.6
FLS	No	21668	31.3	8.1	0.88	0.88	751.2	78.1	4.5	16.3	63.2	16.0
	Yes	22138	26.1	8.4	0.88	0.88	703.0	80.9	12.6	30.4	52.0	5.0

Table 3.20
Subgroup Results: Grade 5 Writing

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores ^a		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		44543	40.7	8.7	-	-	44.6	8.5	3.4	13.7	69.1	13.7
Gender	Female	22699	39.1	8.7	-	-	43.0	8.7	4.7	17.9	67.6	9.8
	Male	21844	42.4	8.2	-	-	46.2	8.0	2.0	9.4	70.7	17.8
Ethnicity	African American	4979	38.3	8.7	-	-	42.2	8.7	5.2	19.9	66.7	8.2
	American Indian	8332	40.1	8.5	-	-	43.9	8.4	3.9	14.5	70.3	11.2
	Hispanic	4715	39.0	8.3	-	-	42.9	8.3	4.3	17.1	69.8	8.7
	Asian	790	44.1	8.5	-	-	47.8	8.2	1.8	6.5	66.3	25.4
	Pacific Islander	85	40.5	7.2	-	-	44.4	7.1	2.3	10.5	80.2	7.0
	White	25065	41.6	8.6	-	-	45.4	8.4	2.8	11.9	69.0	16.3
	Other	577	40.7	8.4	-	-	44.6	8.2	3.4	12.3	72.2	12.0
IEP	No	41129	41.5	8.1	-	-	45.3	8.0	15.4	34.3	46.4	3.9
	Yes	3414	31.5	9.3	-	-	35.4	9.3	1.3	10.1	73.2	15.5
ELL	No	44205	40.7	8.6	-	-	44.6	8.5	3.2	13.2	69.4	14.2
	Yes	338	34.7	8.8	-	-	38.6	8.8	8.0	25.1	62.7	4.1
FLS	No	19277	43.0	8.3	-	-	46.9	8.0	1.7	8.6	69.5	20.2
	Yes	25266	38.9	8.5	-	-	42.8	8.5	4.7	17.6	68.8	8.9

^a Weighted composite scores

Table 3.21
Subgroup Results: Grade 8 Writing

Group	Subgroup	Valid N	Raw Scores		Alpha	Stratified Alpha	Scaled Scores ^a		Percent in Achievement Level			
			Mean	SD			Mean	SD	Unsat	Lim	Sat	Adv
Overall		42290	43.6	7.9	-	-	45.7	7.4	1.9	8.7	75.4	14.0
Gender	Female	21482	41.7	8.0	-	-	43.9	7.5	2.8	12.3	76.2	8.7
	Male	20808	45.6	7.3	-	-	47.6	6.8	1.0	4.9	74.6	19.5
Ethnicity	African American	4443	41.0	8.1	-	-	43.3	7.6	3.4	14.5	74.5	7.6
	American Indian	7942	43.0	7.8	-	-	45.1	7.3	2.1	9.6	76.6	11.7
	Hispanic	4074	41.5	8.0	-	-	43.7	7.5	2.9	12.2	76.7	8.3
	Asian	741	45.6	7.9	-	-	47.5	7.3	1.5	6.0	69.9	22.6
	Pacific Islander	80	42.4	9.2	-	-	44.5	8.6	3.7	14.8	65.4	16.0
	White	24265	44.6	7.7	-	-	46.6	7.2	1.5	6.8	74.9	16.7
	Other	745	43.5	6.9	-	-	45.6	6.5	1.1	7.2	82.5	9.2
IEP	No	39301	44.4	7.4	-	-	46.4	6.9	0.8	4.9	78.3	15.9
	Yes	2989	34.2	8.3	-	-	36.9	7.8	8.8	31.3	57.7	2.2
ELL	No	42011	43.7	7.9	-	-	45.8	7.4	1.8	8.1	75.6	14.5
	Yes	279	35.2	8.6	-	-	37.8	8.1	5.3	21.8	70.3	2.5
FLS	No	20914	45.6	7.4	-	-	47.5	6.9	1.0	5.0	74.5	19.5
	Yes	21376	41.7	8.0	-	-	43.9	7.5	2.8	12.3	76.3	8.6

^a Weighted composite scores

CHAPTER IV. PERFORMANCE STANDARDS

Performance standards represent the criteria which specify a minimum score a student must achieve on the statewide assessment to be placed into a given performance level. In Oklahoma, four performance levels (i.e., unsatisfactory, limited knowledge, satisfactory, and advanced) were previously established for grades 5 and 8 in 2001, for grades 3 and 4 in 2005, and for grades 6 and 7 in 2006. However, to increase rigor by raising standards for grades 3 through 8 student’s achievement on the OCCT as a means to be more competitive at the national and international levels, to vertically align proficiency expectations for students on the OCCT tests across grades 3 through 8, and to align student expectations on the OCCT more closely with student expectations for the National Assessment of Educational Progress (NAEP), revised performance standards (unsatisfactory, limited knowledge, proficient, and advanced) for reading and mathematics were established in 2009.

The workshop to set new academic achievement level cutpoints for grades 3 through 8 in reading and mathematics was held June 15-18, 2009 in Oklahoma City. Thirty seven educational stakeholders from Oklahoma participated in recommending cut scores for the OCCT. Committee members were primarily selected to span grades 3 through 8, although a small number of higher education teachers and those from the business community who are knowledgeable of education in Oklahoma were also selected. The standard setting method known as the Bookmark procedure (Lewis, Mitzel, & Green, 1996), which is the same procedure used in the previous setting of performance level cut scores in reading and mathematics for grades 3 through 8, was employed. The details of the standard setting materials, procedures, methods, and results were reported in the *OCCT Standard Setting: Technical Report* (SDE, 2009). Table 4.1 summarizes the final scaled score ranges for the achievement levels.

Table 4.1 Final Scaled Score Ranges for Reading and Mathematics

Subject	Grade	Unsatisfactory	Limited Knowledge	Proficient	Advanced
Reading	3	400-648	649-699	700-890	891-990
	4	400-657	658-699	700-844	845-990
	5	400-640	641-699	700-829	830-990
	6	400-646	647-699	700-827	828-990
	7	400-667	668-699	700-801	802-990
	8	400-654	655-699	700-832	833-990
Mathematics	3	400-635	636-699	700-797	798-990
	4	400-638	639-699	700-815	816-990
	5	400-641	642-699	700-766	767-990
	6	400-659	660-699	700-753	754-990
	7	400-666	667-699	700-765	766-990
	8	400-661	662-699	700-770	771-990

REFERENCES

- Brennan, R.L. (2004). BB-CLASS [Computer Software]. Iowa City, IA.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Harcourt Brace Jovanovich College Publishers.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297 – 334.
- Data Recognition Corporation (2008). iTEMs [Computer Software]. Maple Grove, MN.
- Dorans, N. J., & Holland, P.W. (1993). DIF Detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Ed.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum.
- Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement*, *27*, 345 – 359.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum.
- Kolen, M.J., Zeng, L., & Hanson, B.A. (1996). Conditional standard errors of measurement for scale scores using IRT. *Journal of Educational Measurement*, *2*, 129-140.
- Lewis, D. M., Mitzel, H. C. & Green, D. R. (1996). Standard Setting: A Bookmark Approach. In D. R. Green (Chair), IRT-based standard-setting procedures utilizing behavioral anchoring. Symposium conducted at the Council of Chief State School Officers National Conference on Large Scale Assessment, Phoenix, AZ.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, *32*(2), 179 – 197.
- Lord, F. M. (1965). A strong true-score theory with applications. *Psychometrika*, *30*, 239 – 297.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*, 719–748.
- Muraki, E., & Bock, R.D. (2003). PARSCALE 4 [Computer Software]. Chicago: Scientific Software International.
- Qualls, A. L. (1995). Estimating the Reliability of a Test Containing Multiple Item Formats. *Applied Measurement in Education*, *8*, 111-120.
- SDE (2009). *OCCT Standard Setting: Technical Report*. OK: Oklahoma City.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*(2), 201–210.

Thissen, D., Pommerich, M., Billeaud, K., & Williams, V. S. L. (1995). Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measurement, 19*, 39-49.

Zwick, R., & Ercikan, K. (1989). Analysis of differential item functioning in the NAPE history assessment. *Journal of Educational Measurement, 26*, 55-66.

APPENDIX A. DATA REVIEW RESULTS

DRC Test Development and Psychometric Services staff conducted a Data Review with SDE assessment and curriculum staff on July 20, 2009, to evaluate the statistical performance of items field-tested in 2009. For each content area and grade, Senior Content Specialists prepared 80 field-tested item cards. The items with poor statistics were flagged for SDE’s review. DRC opened the review with a presentation on the interpretation of the item statistics. DRC psychometricians then advised participants and answered questions as needed. The results of the item data review are shown below in Table A.1.

Table A.1
Data Review Results

Subject	Grade	Accept	Accept W/R*	Accept Total	Percent Accept	Reject	Total
Mathematics	3	73	3	76	95%	4	80
	4	70	10	80	100%	0	80
	5	67	12	79	99%	1	80
	6	75	5	80	100%	0	80
	7	72	4	76	95%	4	80
	8	71	6	77	96%	3	80
Reading	3	76	0	76	95%	4	80
	4	74	0	74	93%	6	80
	5	77	0	77	96%	3	80
	6	74	0	74	93%	6	80
	7	72	0	72	90%	8	80
	8	80	0	80	100%	0	80
Social Studies	5	70	1	71	89%	9	80
	7	71	3	74	93%	6	80
	8	68	1	69	86%	11	80
Science	5	71	6	77	96%	3	80
	8	59	11	70	88%	10	80

* Items may be edited and returned to the pool for future field testing.

APPENDIX B. RAW-TO-SCALED SCORE CONVERSION TABLES AND FREQUENCY DISTRIBUTIONS*

* The frequency analysis was based on the final data files with students who took the standard OCCT.

Table B.1
Raw-to-Scaled Score Table and Frequency Distribution: Grade 3 Reading

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	45	U	1	0.00	1	0.00
1	400	45	U	0	0.00	1	0.00
2	400	45	U	1	0.00	2	0.00
3	400	45	U	3	0.01	5	0.01
4	400	45	U	1	0.00	6	0.01
5	400	45	U	3	0.01	9	0.02
6	400	45	U	8	0.02	17	0.04
7	400	45	U	11	0.03	28	0.06
8	400	45	U	17	0.04	45	0.10
9	400	45	U	36	0.08	81	0.19
10	432	53	U	62	0.14	143	0.33
11	472	57	U	82	0.19	225	0.52
12	498	58	U	99	0.23	324	0.74
13	517	56	U	128	0.29	452	1.04
14	533	53	U	152	0.35	604	1.38
15	546	48	U	231	0.53	835	1.91
16	558	44	U	234	0.54	1069	2.45
17	569	39	U	277	0.63	1346	3.09
18	579	36	U	308	0.71	1654	3.79
19	588	33	U	336	0.77	1990	4.56
20	596	30	U	393	0.90	2383	5.46
21	604	29	U	412	0.94	2795	6.41
22	612	27	U	516	1.18	3311	7.59
23	620	26	U	565	1.30	3876	8.89
24	628	25	U	646	1.48	4522	10.37
25	635	25	U	649	1.49	5171	11.85
26	642	24	U	798	1.83	5969	13.68
27	649	24	L	847	1.94	6816	15.62
28	656	24	L	936	2.15	7752	17.77
29	664	24	L	1110	2.54	8862	20.31
30	671	23	L	1242	2.85	10104	23.16
31	678	23	L	1273	2.92	11377	26.08
32	685	23	L	1424	3.26	12801	29.34
33	693	23	L	1536	3.52	14337	32.87
34	700	23	S	1676	3.84	16013	36.71
35	708	24	S	1771	4.06	17784	40.77
36	716	24	S	1859	4.26	19643	45.03
37	724	24	S	2000	4.58	21643	49.61
38	732	25	S	2077	4.76	23720	54.37
39	741	25	S	2201	5.05	25921	59.42
40	750	27	S	2196	5.03	28117	64.45
41	760	28	S	2334	5.35	30451	69.80
42	771	30	S	2272	5.21	32723	75.01
43	783	32	S	2173	4.98	34896	79.99
44	796	36	S	2108	4.83	37004	84.83
45	812	41	S	1923	4.41	38927	89.24
46	831	46	S	1717	3.94	40644	93.17
47	855	52	S	1357	3.11	42001	96.28
48	891	54	A	959	2.20	42960	98.48
49	956	44	A	504	1.16	43464	99.64
50	990	44	A	159	0.36	43623	100.00

Table B.2
Raw-to-Scaled Score Table and Frequency Distribution: Grade 4 Reading

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	45	U	0	0.00	0	0.00
1	400	45	U	1	0.00	1	0.00
2	400	45	U	3	0.01	4	0.01
3	400	45	U	1	0.00	5	0.01
4	400	45	U	1	0.00	6	0.01
5	400	45	U	4	0.01	10	0.02
6	400	45	U	6	0.01	16	0.04
7	400	45	U	4	0.01	20	0.05
8	400	45	U	20	0.05	40	0.09
9	400	45	U	19	0.04	59	0.14
10	436	52	U	44	0.10	103	0.24
11	472	55	U	65	0.15	168	0.40
12	496	56	U	89	0.21	257	0.61
13	515	54	U	126	0.30	383	0.90
14	530	51	U	151	0.36	534	1.26
15	543	47	U	173	0.41	707	1.67
16	555	42	U	189	0.45	896	2.12
17	566	38	U	232	0.55	1128	2.66
18	575	35	U	260	0.61	1388	3.28
19	584	32	U	305	0.72	1693	4.00
20	592	29	U	323	0.76	2016	4.76
21	600	27	U	403	0.95	2419	5.71
22	608	26	U	405	0.96	2824	6.67
23	615	24	U	451	1.07	3275	7.74
24	622	23	U	534	1.26	3809	9.00
25	628	22	U	532	1.26	4341	10.25
26	634	22	U	637	1.50	4978	11.76
27	641	21	U	715	1.69	5693	13.45
28	647	20	U	762	1.80	6455	15.25
29	652	20	U	832	1.97	7287	17.21
30	658	20	L	866	2.05	8153	19.26
31	664	19	L	970	2.29	9123	21.55
32	670	19	L	1105	2.61	10228	24.16
33	675	19	L	1135	2.68	11363	26.84
34	681	19	L	1300	3.07	12663	29.91
35	687	20	L	1359	3.21	14022	33.12
36	694	20	L	1508	3.56	15530	36.68
37	700	20	S	1634	3.86	17164	40.54
38	707	21	S	1797	4.24	18961	44.79
39	714	22	S	1972	4.66	20933	49.45
40	722	23	S	2144	5.06	23077	54.51
41	730	24	S	2273	5.37	25350	59.88
42	739	26	S	2429	5.74	27779	65.62
43	750	29	S	2517	5.95	30296	71.56
44	761	32	S	2629	6.21	32925	77.77
45	775	37	S	2502	5.91	35427	83.68
46	792	44	S	2313	5.46	37740	89.15
47	814	52	S	1967	4.65	39707	93.79
48	845	60	A	1483	3.50	41190	97.30
49	898	62	A	857	2.02	42047	99.32
50	990	62	A	288	0.68	42335	100.00

Table B.3
Raw-to-Scaled Score Table and Frequency Distribution: Grade 5 Reading

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	42	U	3	0.01	3	0.01
1	400	42	U	5	0.01	8	0.02
2	400	42	U	5	0.01	13	0.03
3	400	42	U	5	0.01	18	0.04
4	400	42	U	9	0.02	27	0.06
5	400	42	U	4	0.01	31	0.07
6	400	42	U	5	0.01	36	0.09
7	400	42	U	8	0.02	44	0.10
8	400	42	U	18	0.04	62	0.15
9	400	42	U	23	0.05	85	0.20
10	433	48	U	45	0.11	130	0.31
11	466	52	U	67	0.16	197	0.47
12	490	53	U	83	0.20	280	0.66
13	508	52	U	111	0.26	391	0.93
14	522	49	U	128	0.30	519	1.23
15	535	45	U	163	0.39	682	1.62
16	546	41	U	197	0.47	879	2.09
17	556	37	U	222	0.53	1101	2.61
18	566	34	U	220	0.52	1321	3.14
19	575	31	U	245	0.58	1566	3.72
20	583	29	U	297	0.71	1863	4.42
21	591	28	U	362	0.86	2225	5.28
22	599	26	U	413	0.98	2638	6.26
23	606	25	U	416	0.99	3054	7.25
24	613	25	U	463	1.10	3517	8.35
25	620	24	U	584	1.39	4101	9.74
26	627	24	U	621	1.47	4722	11.21
27	634	24	U	682	1.62	5404	12.83
28	641	23	L	789	1.87	6193	14.71
29	648	23	L	859	2.04	7052	16.75
30	656	23	L	990	2.35	8042	19.10
31	663	23	L	1032	2.45	9074	21.55
32	670	23	L	1173	2.79	10247	24.34
33	677	23	L	1270	3.02	11517	27.35
34	685	23	L	1419	3.37	12936	30.72
35	692	23	L	1577	3.75	14513	34.47
36	700	23	S	1639	3.89	16152	38.36
37	708	23	S	1755	4.17	17907	42.53
38	716	24	S	1912	4.54	19819	47.07
39	725	24	S	2003	4.76	21822	51.83
40	733	25	S	2199	5.22	24021	57.05
41	743	26	S	2357	5.60	26378	62.65
42	753	27	S	2399	5.70	28777	68.34
43	763	30	S	2394	5.69	31171	74.03
44	776	33	S	2436	5.79	33607	79.81
45	790	38	S	2405	5.71	36012	85.52
46	807	44	S	2200	5.22	38212	90.75
47	830	51	A	1778	4.22	39990	94.97
48	862	57	A	1295	3.08	41285	98.05
49	920	54	A	635	1.51	41920	99.56
50	990	54	A	187	0.44	42107	100.00

Table B.4
Raw-to-Scaled Score Table and Frequency Distribution: Grade 6 Reading

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	44	U	0	0.00	0	0.00
1	400	44	U	0	0.00	0	0.00
2	400	44	U	1	0.00	1	0.00
3	400	44	U	2	0.00	3	0.01
4	400	44	U	1	0.00	4	0.01
5	400	44	U	2	0.00	6	0.01
6	400	44	U	11	0.03	17	0.04
7	400	44	U	9	0.02	26	0.06
8	400	44	U	19	0.05	45	0.11
9	400	44	U	40	0.10	85	0.21
10	414	52	U	72	0.18	157	0.39
11	464	57	U	89	0.22	246	0.60
12	494	59	U	117	0.29	363	0.89
13	515	58	U	144	0.35	507	1.24
14	531	54	U	176	0.43	683	1.68
15	545	50	U	199	0.49	882	2.16
16	556	44	U	241	0.59	1123	2.76
17	566	40	U	287	0.70	1410	3.46
18	575	35	U	308	0.76	1718	4.22
19	584	32	U	334	0.82	2052	5.03
20	592	29	U	371	0.91	2423	5.95
21	600	27	U	400	0.98	2823	6.93
22	607	26	U	423	1.04	3246	7.96
23	614	25	U	446	1.09	3692	9.06
24	621	24	U	533	1.31	4225	10.37
25	628	23	U	554	1.36	4779	11.73
26	634	22	U	614	1.51	5393	13.23
27	641	22	U	666	1.63	6059	14.87
28	647	22	L	752	1.85	6811	16.71
29	654	21	L	802	1.97	7613	18.68
30	660	21	L	880	2.16	8493	20.84
31	666	21	L	933	2.29	9426	23.13
32	673	21	L	1010	2.48	10436	25.61
33	679	21	L	1179	2.89	11615	28.50
34	686	21	L	1282	3.15	12897	31.64
35	693	22	L	1434	3.52	14331	35.16
36	700	22	S	1637	4.02	15968	39.18
37	707	23	S	1684	4.13	17652	43.31
38	715	23	S	1860	4.56	19512	47.88
39	723	24	S	1932	4.74	21444	52.62
40	731	25	S	2016	4.95	23460	57.56
41	741	26	S	2265	5.56	25725	63.12
42	751	28	S	2345	5.75	28070	68.87
43	762	31	S	2334	5.73	30404	74.60
44	774	34	S	2428	5.96	32832	80.56
45	789	38	S	2350	5.77	35182	86.32
46	806	44	S	2104	5.16	37286	91.49
47	828	52	A	1677	4.11	38963	95.60
48	860	58	A	1139	2.79	40102	98.40
49	916	56	A	535	1.31	40637	99.71
50	990	56	A	119	0.29	40756	100.00

Table B.5
Raw-to-Scaled Score Table and Frequency Distribution: Grade 7 Reading

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	57	U	1	0.00	1	0.00
1	400	57	U	2	0.01	3	0.01
2	400	57	U	1	0.00	4	0.01
3	400	57	U	3	0.01	7	0.02
4	400	57	U	3	0.01	10	0.03
5	400	57	U	1	0.00	11	0.03
6	400	57	U	3	0.01	14	0.04
7	400	57	U	13	0.03	27	0.07
8	400	57	U	10	0.03	37	0.09
9	400	57	U	24	0.06	61	0.15
10	469	63	U	41	0.10	102	0.26
11	502	65	U	63	0.16	165	0.41
12	524	63	U	83	0.21	248	0.62
13	540	58	U	103	0.26	351	0.88
14	554	53	U	130	0.33	481	1.21
15	566	47	U	184	0.46	665	1.67
16	576	41	U	207	0.52	872	2.19
17	585	36	U	229	0.58	1101	2.77
18	594	32	U	293	0.74	1394	3.50
19	602	29	U	320	0.80	1714	4.30
20	610	27	U	337	0.85	2051	5.15
21	617	25	U	403	1.01	2454	6.16
22	624	24	U	446	1.12	2900	7.28
23	630	23	U	477	1.20	3377	8.48
24	637	22	U	514	1.29	3891	9.77
25	643	22	U	584	1.47	4475	11.24
26	649	21	U	724	1.82	5199	13.06
27	655	21	U	680	1.71	5879	14.77
28	662	21	U	792	1.99	6671	16.75
29	668	21	L	804	2.02	7475	18.77
30	674	21	L	984	2.47	8459	21.25
31	680	21	L	1075	2.70	9534	23.95
32	687	21	L	1194	3.00	10728	26.94
33	693	21	L	1224	3.07	11952	30.02
34	700	21	S	1377	3.46	13329	33.48
35	707	21	S	1536	3.86	14865	37.34
36	714	22	S	1626	4.08	16491	41.42
37	721	22	S	1712	4.30	18203	45.72
38	729	23	S	1924	4.83	20127	50.55
39	737	23	S	2049	5.15	22176	55.70
40	745	24	S	2165	5.44	24341	61.14
41	754	26	S	2299	5.77	26640	66.91
42	764	28	S	2233	5.61	28873	72.52
43	775	30	S	2294	5.76	31167	78.28
44	787	33	S	2171	5.45	33338	83.73
45	802	38	A	2017	5.07	35355	88.80
46	819	44	A	1726	4.34	37081	93.13
47	842	50	A	1352	3.40	38433	96.53
48	875	55	A	825	2.07	39258	98.60
49	934	49	A	431	1.08	39689	99.68
50	990	49	A	126	0.32	39815	100.00

Table B.6
Raw-to-Scaled Score Table and Frequency Distribution: Grade 8 Reading

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	43	U	0	0.00	0	0.00
1	400	43	U	2	0.00	2	0.00
2	400	43	U	2	0.00	4	0.01
3	400	43	U	0	0.00	4	0.01
4	400	43	U	2	0.00	6	0.01
5	400	43	U	1	0.00	7	0.02
6	400	43	U	2	0.00	9	0.02
7	400	43	U	5	0.01	14	0.03
8	400	43	U	10	0.02	24	0.06
9	400	43	U	16	0.04	40	0.10
10	400	43	U	19	0.05	59	0.15
11	414	50	U	42	0.10	101	0.25
12	458	55	U	62	0.15	163	0.40
13	486	58	U	89	0.22	252	0.63
14	508	57	U	129	0.32	381	0.95
15	525	55	U	143	0.35	524	1.30
16	539	51	U	192	0.48	716	1.78
17	552	46	U	204	0.51	920	2.28
18	563	42	U	217	0.54	1137	2.82
19	574	38	U	279	0.69	1416	3.51
20	583	35	U	297	0.74	1713	4.25
21	593	32	U	341	0.85	2054	5.10
22	601	30	U	393	0.98	2447	6.07
23	609	29	U	430	1.07	2877	7.14
24	617	27	U	470	1.17	3347	8.31
25	625	26	U	552	1.37	3899	9.68
26	632	26	U	638	1.58	4537	11.26
27	640	25	U	676	1.68	5213	12.94
28	647	25	U	774	1.92	5987	14.86
29	655	25	L	898	2.23	6885	17.09
30	662	24	L	1037	2.57	7922	19.66
31	669	24	L	1113	2.76	9035	22.43
32	677	24	L	1284	3.19	10319	25.61
33	684	24	L	1363	3.38	11682	29.00
34	692	25	L	1516	3.76	13198	32.76
35	700	25	S	1709	4.24	14907	37.00
36	708	25	S	1744	4.33	16651	41.33
37	717	26	S	1849	4.59	18500	45.92
38	726	27	S	2019	5.01	20519	50.93
39	735	28	S	2176	5.40	22695	56.33
40	746	29	S	2217	5.50	24912	61.84
41	756	30	S	2390	5.93	27302	67.77
42	768	32	S	2362	5.86	29664	73.63
43	781	35	S	2406	5.97	32070	79.60
44	796	39	S	2193	5.44	34263	85.05
45	813	43	S	1948	4.84	36211	89.88
46	833	49	A	1639	4.07	37850	93.95
47	859	55	A	1232	3.06	39082	97.01
48	897	55	A	774	1.92	39856	98.93
49	966	42	A	350	0.87	40206	99.80
50	990	42	A	81	0.20	40287	100.00

Table B.7
Raw-to-Scaled Score Table and Frequency Distribution: Grade 3 Mathematics

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	47	U	0	0.00	0	0.00
1	400	47	U	0	0.00	0	0.00
2	400	47	U	0	0.00	0	0.00
3	400	47	U	0	0.00	0	0.00
4	400	47	U	5	0.01	5	0.01
5	400	47	U	5	0.01	10	0.02
6	400	47	U	26	0.06	36	0.08
7	400	47	U	45	0.10	81	0.18
8	400	47	U	69	0.16	150	0.34
9	431	56	U	102	0.23	252	0.57
10	475	61	U	139	0.31	391	0.88
11	504	62	U	193	0.44	584	1.32
12	527	60	U	304	0.69	888	2.01
13	546	56	U	353	0.80	1241	2.81
14	561	51	U	399	0.90	1640	3.71
15	575	46	U	476	1.08	2116	4.79
16	587	41	U	511	1.16	2627	5.94
17	599	37	U	580	1.31	3207	7.25
18	609	34	U	607	1.37	3814	8.62
19	618	31	U	700	1.58	4514	10.21
20	627	29	U	809	1.83	5323	12.04
21	636	27	L	834	1.89	6157	13.92
22	644	26	L	908	2.05	7065	15.98
23	651	25	L	947	2.14	8012	18.12
24	659	24	L	1015	2.30	9027	20.41
25	666	23	L	992	2.24	10019	22.66
26	673	22	L	1135	2.57	11154	25.22
27	680	22	L	1175	2.66	12329	27.88
28	686	22	L	1210	2.74	13539	30.62
29	693	21	L	1347	3.05	14886	33.66
30	700	21	S	1417	3.20	16303	36.87
31	707	21	S	1476	3.34	17779	40.20
32	714	21	S	1603	3.62	19382	43.83
33	721	21	S	1630	3.69	21012	47.52
34	728	22	S	1729	3.91	22741	51.43
35	736	22	S	1834	4.15	24575	55.57
36	744	23	S	1910	4.32	26485	59.89
37	752	24	S	2046	4.63	28531	64.52
38	762	26	S	2074	4.69	30605	69.21
39	772	29	S	2172	4.91	32777	74.12
40	784	33	S	2214	5.01	34991	79.13
41	798	39	A	2222	5.02	37213	84.15
42	816	48	A	2271	5.14	39484	89.29
43	841	58	A	2083	4.71	41567	94.00
44	884	65	A	1751	3.96	43318	97.96
45	990	65	A	903	2.04	44221	100.00

Table B.8
Raw-to-Scaled Score Table and Frequency Distribution: Grade 4 Mathematics

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	48	U	0	0.00	0	0.00
1	400	48	U	2	0.00	2	0.00
2	400	48	U	1	0.00	3	0.01
3	400	48	U	0	0.00	3	0.01
4	400	48	U	1	0.00	4	0.01
5	400	48	U	4	0.01	8	0.02
6	400	48	U	7	0.02	15	0.03
7	400	48	U	18	0.04	33	0.08
8	400	48	U	44	0.10	77	0.18
9	400	48	U	66	0.15	143	0.33
10	435	56	U	101	0.24	244	0.57
11	475	61	U	143	0.33	387	0.90
12	503	62	U	199	0.46	586	1.37
13	525	61	U	296	0.69	882	2.06
14	544	58	U	309	0.72	1191	2.78
15	559	53	U	384	0.90	1575	3.67
16	574	48	U	435	1.01	2010	4.69
17	586	44	U	503	1.17	2513	5.86
18	598	40	U	562	1.31	3075	7.17
19	609	37	U	718	1.68	3793	8.85
20	620	35	U	773	1.80	4566	10.65
21	629	33	U	833	1.94	5399	12.60
22	639	31	L	934	2.18	6333	14.77
23	648	30	L	972	2.27	7305	17.04
24	657	29	L	1148	2.68	8453	19.72
25	666	28	L	1261	2.94	9714	22.66
26	675	27	L	1347	3.14	11061	25.80
27	683	27	L	1431	3.34	12492	29.14
28	692	26	L	1489	3.47	13981	32.62
29	700	26	S	1695	3.95	15676	36.57
30	708	26	S	1756	4.10	17432	40.67
31	717	26	S	1817	4.24	19249	44.91
32	726	26	S	1864	4.35	21113	49.25
33	735	27	S	1884	4.40	22997	53.65
34	744	27	S	1983	4.63	24980	58.28
35	754	28	S	2108	4.92	27088	63.19
36	764	29	S	2183	5.09	29271	68.29
37	775	30	S	2127	4.96	31398	73.25
38	787	32	S	2172	5.07	33570	78.32
39	800	35	S	1989	4.64	35559	82.96
40	816	39	A	1878	4.38	37437	87.34
41	834	44	A	1766	4.12	39203	91.46
42	857	48	A	1494	3.49	40697	94.94
43	888	51	A	1156	2.70	41853	97.64
44	941	45	A	689	1.61	42542	99.25
45	990	45	A	323	0.75	42865	100.00

Table B.9
Raw-to-Scaled Score Table and Frequency Distribution: Grade 5 Mathematics

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	57	U	1	0.00	1	0.00
1	400	57	U	4	0.01	5	0.01
2	400	57	U	4	0.01	9	0.02
3	400	57	U	5	0.01	14	0.03
4	400	57	U	4	0.01	18	0.04
5	400	57	U	10	0.02	28	0.07
6	400	57	U	17	0.04	45	0.11
7	400	57	U	24	0.06	69	0.16
8	400	57	U	67	0.16	136	0.32
9	400	57	U	88	0.21	224	0.53
10	452	66	U	150	0.35	374	0.88
11	495	70	U	199	0.47	573	1.35
12	524	70	U	291	0.69	864	2.04
13	547	67	U	346	0.82	1210	2.85
14	566	62	U	468	1.10	1678	3.95
15	583	55	U	567	1.34	2245	5.29
16	597	49	U	643	1.52	2888	6.81
17	610	44	U	788	1.86	3676	8.66
18	621	39	U	851	2.01	4527	10.67
19	632	36	U	996	2.35	5523	13.02
20	642	33	L	1070	2.52	6593	15.54
21	652	31	L	1136	2.68	7729	18.22
22	661	29	L	1358	3.20	9087	21.42
23	669	27	L	1338	3.15	10425	24.57
24	677	26	L	1421	3.35	11846	27.92
25	685	25	L	1477	3.48	13323	31.40
26	693	24	L	1572	3.70	14895	35.10
27	700	23	S	1509	3.56	16404	38.66
28	707	22	S	1635	3.85	18039	42.51
29	714	22	S	1680	3.96	19719	46.47
30	722	22	S	1662	3.92	21381	50.39
31	729	21	S	1698	4.00	23079	54.39
32	736	21	S	1756	4.14	24835	58.53
33	743	21	S	1638	3.86	26473	62.39
34	751	22	S	1659	3.91	28132	66.30
35	758	22	S	1687	3.98	29819	70.28
36	767	23	A	1737	4.09	31556	74.37
37	775	24	A	1651	3.89	33207	78.26
38	785	26	A	1555	3.66	34762	81.93
39	795	29	A	1517	3.58	36279	85.50
40	807	33	A	1506	3.55	37785	89.05
41	822	39	A	1314	3.10	39099	92.15
42	841	46	A	1269	2.99	40368	95.14
43	868	52	A	1025	2.42	41393	97.56
44	920	51	A	731	1.72	42124	99.28
45	990	51	A	306	0.72	42430	100.00

Table B.10
Raw-to-Scaled Score Table and Frequency Distribution: Grade 6 Mathematics

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	68	U	0	0.00	0	0.00
1	400	68	U	0	0.00	0	0.00
2	400	68	U	1	0.00	1	0.00
3	400	68	U	2	0.00	3	0.01
4	400	68	U	3	0.01	6	0.01
5	400	68	U	3	0.01	9	0.02
6	400	68	U	10	0.02	19	0.05
7	400	68	U	21	0.05	40	0.10
8	400	68	U	76	0.19	116	0.28
9	440	68	U	72	0.18	188	0.46
10	501	73	U	156	0.38	344	0.84
11	534	72	U	225	0.55	569	1.39
12	556	68	U	315	0.77	884	2.16
13	574	61	U	415	1.01	1299	3.17
14	588	54	U	466	1.14	1765	4.30
15	600	46	U	511	1.25	2276	5.55
16	611	40	U	631	1.54	2907	7.09
17	621	35	U	698	1.70	3605	8.79
18	630	31	U	825	2.01	4430	10.80
19	638	28	U	848	2.07	5278	12.87
20	646	26	U	909	2.22	6187	15.09
21	653	25	U	998	2.43	7185	17.52
22	660	23	L	1086	2.65	8271	20.17
23	667	22	L	1199	2.92	9470	23.09
24	674	22	L	1224	2.98	10694	26.07
25	681	21	L	1313	3.20	12007	29.28
26	687	21	L	1361	3.32	13368	32.59
27	694	20	L	1424	3.47	14792	36.07
28	700	20	S	1391	3.39	16183	39.46
29	706	20	S	1602	3.91	17785	43.36
30	713	20	S	1643	4.01	19428	47.37
31	719	20	S	1581	3.85	21009	51.22
32	726	20	S	1672	4.08	22681	55.30
33	733	20	S	1695	4.13	24376	59.43
34	740	20	S	1656	4.04	26032	63.47
35	747	21	S	1678	4.09	27710	67.56
36	754	21	A	1681	4.10	29391	71.66
37	763	23	A	1728	4.21	31119	75.87
38	771	24	A	1727	4.21	32846	80.08
39	781	27	A	1625	3.96	34471	84.05
40	792	31	A	1583	3.86	36054	87.91
41	806	37	A	1522	3.71	37576	91.62
42	822	45	A	1362	3.32	38938	94.94
43	846	56	A	1065	2.60	40003	97.53
44	886	63	A	676	1.65	40679	99.18
45	990	63	A	335	0.82	41014	100.00

Table B.11
Raw-to-Scaled Score Table and Frequency Distribution: Grade 7 Mathematics

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	57	U	0	0.00	0	0.00
1	400	57	U	4	0.01	4	0.01
2	400	57	U	3	0.01	7	0.02
3	400	57	U	2	0.01	9	0.02
4	400	57	U	6	0.02	15	0.04
5	400	57	U	6	0.02	21	0.05
6	400	57	U	24	0.06	45	0.11
7	400	57	U	27	0.07	72	0.18
8	400	57	U	61	0.15	133	0.33
9	400	57	U	96	0.24	229	0.57
10	457	65	U	132	0.33	361	0.90
11	497	69	U	235	0.59	596	1.49
12	525	69	U	265	0.66	861	2.15
13	546	66	U	382	0.96	1243	3.11
14	564	61	U	439	1.10	1682	4.21
15	579	55	U	549	1.37	2231	5.58
16	593	49	U	672	1.68	2903	7.26
17	606	44	U	736	1.84	3639	9.10
18	617	40	U	855	2.14	4494	11.24
19	628	37	U	982	2.46	5476	13.70
20	638	34	U	1002	2.51	6478	16.20
21	648	32	U	1119	2.80	7597	19.00
22	658	31	U	1253	3.13	8850	22.13
23	667	29	L	1370	3.43	10220	25.56
24	675	28	L	1418	3.55	11638	29.11
25	684	27	L	1547	3.87	13185	32.98
26	692	26	L	1645	4.11	14830	37.09
27	700	26	S	1665	4.16	16495	41.26
28	708	25	S	1733	4.33	18228	45.59
29	716	25	S	1779	4.45	20007	50.04
30	724	24	S	1879	4.70	21886	54.74
31	732	24	S	1733	4.33	23619	59.07
32	740	24	S	1799	4.50	25418	63.57
33	749	25	S	1708	4.27	27126	67.85
34	757	25	S	1669	4.17	28795	72.02
35	766	25	A	1653	4.13	30448	76.15
36	776	26	A	1578	3.95	32026	80.10
37	786	28	A	1546	3.87	33572	83.97
38	797	29	A	1395	3.49	34967	87.46
39	809	32	A	1273	3.18	36240	90.64
40	823	36	A	1135	2.84	37375	93.48
41	839	41	A	916	2.29	38291	95.77
42	860	47	A	704	1.76	38995	97.53
43	890	50	A	542	1.36	39537	98.89
44	945	43	A	302	0.76	39839	99.64
45	990	43	A	143	0.36	39982	100.00

Table B.12
Raw-to-Scaled Score Table and Frequency Distribution: Grade 8 Mathematics

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	60	U	3	0.01	3	0.01
1	400	60	U	0	0.00	3	0.01
2	400	60	U	0	0.00	3	0.01
3	400	60	U	0	0.00	3	0.01
4	400	60	U	2	0.00	5	0.01
5	400	60	U	3	0.01	8	0.02
6	400	60	U	13	0.03	21	0.05
7	400	60	U	18	0.04	39	0.10
8	400	60	U	51	0.13	90	0.22
9	400	60	U	68	0.17	158	0.39
10	400	60	U	121	0.30	279	0.69
11	435	69	U	185	0.46	464	1.15
12	493	74	U	276	0.69	740	1.84
13	528	74	U	347	0.86	1087	2.70
14	553	71	U	414	1.03	1501	3.73
15	572	66	U	481	1.19	1982	4.92
16	589	59	U	664	1.65	2646	6.57
17	602	52	U	733	1.82	3379	8.39
18	615	46	U	793	1.97	4172	10.35
19	626	40	U	994	2.47	5166	12.82
20	636	36	U	1037	2.57	6203	15.40
21	645	32	U	1155	2.87	7358	18.26
22	654	30	U	1153	2.86	8511	21.12
23	662	28	L	1280	3.18	9791	24.30
24	670	26	L	1417	3.52	11208	27.82
25	678	25	L	1365	3.39	12573	31.21
26	685	24	L	1487	3.69	14060	34.90
27	693	24	L	1512	3.75	15572	38.65
28	700	23	S	1559	3.87	17131	42.52
29	707	23	S	1529	3.79	18660	46.31
30	714	23	S	1623	4.03	20283	50.34
31	722	23	S	1614	4.01	21897	54.35
32	729	23	S	1724	4.28	23621	58.63
33	737	23	S	1630	4.05	25251	62.67
34	744	23	S	1639	4.07	26890	66.74
35	753	24	S	1528	3.79	28418	70.53
36	761	25	S	1622	4.03	30040	74.56
37	771	26	A	1673	4.15	31713	78.71
38	781	28	A	1496	3.71	33209	82.42
39	792	31	A	1524	3.78	34733	86.21
40	806	35	A	1317	3.27	36050	89.48
41	821	40	A	1278	3.17	37328	92.65
42	841	47	A	1064	2.64	38392	95.29
43	869	53	A	953	2.37	39345	97.65
44	918	52	A	589	1.46	39934	99.12
45	990	52	A	356	0.88	40290	100.00

Table B.13
Raw-to-Scaled Score Table and Frequency Distribution: Grade 5 Science

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	82	U	0	0.00	0	0.00
1	400	82	U	3	0.01	3	0.01
2	400	82	U	4	0.01	7	0.02
3	400	82	U	0	0.00	7	0.02
4	400	82	U	2	0.00	9	0.02
5	400	82	U	9	0.02	18	0.04
6	400	82	U	18	0.04	36	0.08
7	400	82	U	36	0.08	72	0.17
8	400	82	U	62	0.14	134	0.31
9	505	90	U	92	0.21	226	0.53
10	555	91	U	147	0.34	373	0.87
11	584	86	U	220	0.51	593	1.38
12	604	77	U	287	0.67	880	2.05
13	620	67	U	360	0.84	1240	2.89
14	634	56	U	431	1.00	1671	3.89
15	646	47	L	486	1.13	2157	5.02
16	656	40	L	598	1.39	2755	6.42
17	666	35	L	648	1.51	3403	7.93
18	675	31	L	721	1.68	4124	9.61
19	683	28	L	802	1.87	4926	11.48
20	691	26	L	972	2.26	5898	13.74
21	699	25	L	1035	2.41	6933	16.15
22	706	24	S	1064	2.48	7997	18.63
23	713	23	S	1199	2.79	9196	21.42
24	720	22	S	1248	2.91	10444	24.33
25	727	22	S	1388	3.23	11832	27.56
26	734	22	S	1463	3.41	13295	30.97
27	741	21	S	1547	3.60	14842	34.58
28	748	21	S	1665	3.88	16507	38.45
29	755	21	S	1771	4.13	18278	42.58
30	762	21	S	1787	4.16	20065	46.74
31	769	21	S	1952	4.55	22017	51.29
32	776	21	S	1973	4.60	23990	55.89
33	784	22	S	1966	4.58	25956	60.47
34	792	22	S	1914	4.46	27870	64.93
35	800	22	S	1933	4.50	29803	69.43
36	808	23	S	2057	4.79	31860	74.22
37	818	24	A	1924	4.48	33784	78.70
38	827	26	A	1893	4.41	35677	83.11
39	838	28	A	1717	4.00	37394	87.11
40	851	31	A	1600	3.73	38994	90.84
41	865	34	A	1382	3.22	40376	94.06
42	883	38	A	1114	2.60	41490	96.65
43	908	40	A	794	1.85	42284	98.50
44	948	36	A	474	1.10	42758	99.61
45	990	36	A	168	0.39	42926	100.00

Table B.14
Raw-to-Scaled Score Table and Frequency Distribution: Grade 8 Science

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	83	U	2	0.00	2	0.00
1	400	83	U	6	0.01	8	0.02
2	400	83	U	5	0.01	13	0.03
3	400	83	U	5	0.01	18	0.04
4	400	83	U	6	0.01	24	0.06
5	400	83	U	9	0.02	33	0.08
6	400	83	U	23	0.06	56	0.14
7	400	83	U	31	0.08	87	0.21
8	468	83	U	33	0.08	120	0.29
9	540	86	U	75	0.18	195	0.48
10	572	83	U	96	0.24	291	0.71
11	594	76	U	124	0.30	415	1.02
12	611	66	U	184	0.45	599	1.47
13	625	56	U	260	0.64	859	2.11
14	637	47	U	312	0.77	1171	2.87
15	647	40	L	415	1.02	1586	3.89
16	657	34	L	444	1.09	2030	4.98
17	666	31	L	568	1.39	2598	6.37
18	675	28	L	636	1.56	3234	7.94
19	682	26	L	711	1.74	3945	9.68
20	690	24	L	790	1.94	4735	11.62
21	697	23	L	896	2.20	5631	13.82
22	704	22	S	994	2.44	6625	16.26
23	711	22	S	1074	2.64	7699	18.89
24	717	21	S	1195	2.93	8894	21.82
25	724	21	S	1276	3.13	10170	24.96
26	730	20	S	1392	3.42	11562	28.37
27	737	20	S	1440	3.53	13002	31.90
28	743	20	S	1590	3.90	14592	35.81
29	750	20	S	1678	4.12	16270	39.92
30	757	20	S	1759	4.32	18029	44.24
31	763	20	S	1808	4.44	19837	48.68
32	770	20	S	1988	4.88	21825	53.55
33	778	20	S	1950	4.78	23775	58.34
34	785	21	S	1980	4.86	25755	63.20
35	793	21	S	2032	4.99	27787	68.18
36	802	22	S	2034	4.99	29821	73.17
37	811	23	S	1890	4.64	31711	77.81
38	820	24	S	1949	4.78	33660	82.60
39	831	26	A	1692	4.15	35352	86.75
40	843	29	A	1553	3.81	36905	90.56
41	856	33	A	1364	3.35	38269	93.90
42	873	37	A	1057	2.59	39326	96.50
43	896	42	A	800	1.96	40126	98.46
44	935	41	A	434	1.06	40560	99.53
45	990	41	A	193	0.47	40753	100.00

Table B.15
Raw-to-Scaled Score Table and Frequency Distribution: Grade 5 Social Studies

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	65	U	1	0.00	1	0.00
1	400	65	U	0	0.00	1	0.00
2	400	65	U	0	0.00	1	0.00
3	400	65	U	1	0.00	2	0.00
4	400	65	U	1	0.00	3	0.01
5	400	65	U	1	0.00	4	0.01
6	400	65	U	3	0.01	7	0.02
7	400	65	U	4	0.01	11	0.02
8	400	65	U	8	0.02	19	0.04
9	400	65	U	15	0.03	34	0.07
10	400	65	U	40	0.09	74	0.16
11	400	65	U	65	0.14	139	0.30
12	400	65	U	112	0.25	251	0.55
13	463	71	U	144	0.32	395	0.87
14	504	75	U	215	0.47	610	1.34
15	532	74	U	274	0.60	884	1.94
16	553	71	U	312	0.68	1196	2.62
17	570	66	U	422	0.92	1618	3.54
18	584	60	U	482	1.06	2100	4.60
19	597	54	U	526	1.15	2626	5.75
20	608	48	U	621	1.36	3247	7.11
21	618	43	U	613	1.34	3860	8.46
22	627	38	U	722	1.58	4582	10.04
23	636	35	U	786	1.72	5368	11.76
24	644	32	U	745	1.63	6113	13.39
25	652	30	L	865	1.89	6978	15.29
26	659	28	L	908	1.99	7886	17.28
27	666	26	L	960	2.10	8846	19.38
28	673	25	L	1022	2.24	9868	21.62
29	679	24	L	1077	2.36	10945	23.98
30	685	23	L	1158	2.54	12103	26.51
31	691	23	L	1214	2.66	13317	29.17
32	697	22	L	1253	2.74	14570	31.92
33	703	21	S	1242	2.72	15812	34.64
34	709	21	S	1326	2.90	17138	37.54
35	714	20	S	1363	2.99	18501	40.53
36	720	20	S	1370	3.00	19871	43.53
37	725	19	S	1408	3.08	21279	46.61
38	730	19	S	1423	3.12	22702	49.73
39	736	19	S	1447	3.17	24149	52.90
40	741	19	S	1497	3.28	25646	56.18
41	746	19	S	1462	3.20	27108	59.38
42	752	18	S	1506	3.30	28614	62.68
43	757	18	S	1479	3.24	30093	65.92
44	762	19	S	1406	3.08	31499	69.00
45	768	19	S	1419	3.11	32918	72.11
46	774	19	S	1406	3.08	34324	75.19
47	780	19	S	1323	2.90	35647	78.09
48	786	20	A	1321	2.89	36968	80.98
49	792	20	A	1214	2.66	38182	83.64

Table Continues

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
50	799	21	A	1202	2.63	39384	86.28
51	807	22	A	1186	2.60	40570	88.87
52	815	24	A	1091	2.39	41661	91.26
53	824	26	A	993	2.18	42654	93.44
54	834	29	A	876	1.92	43530	95.36
55	846	33	A	709	1.55	44239	96.91
56	860	37	A	586	1.28	44825	98.19
57	879	41	A	426	0.93	45251	99.13
58	905	44	A	249	0.55	45500	99.67
59	954	37	A	118	0.26	45618	99.93
60	990	37	A	31	0.07	45649	100.00

Table B.16
Raw-to-Scaled Score Table and Frequency Distribution: Grade 7 Geography

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	71	U	0	0.00	0	0.00
1	400	71	U	2	0.00	2	0.00
2	400	71	U	2	0.00	4	0.01
3	400	71	U	2	0.00	6	0.01
4	400	71	U	6	0.01	12	0.03
5	400	71	U	14	0.03	26	0.06
6	400	71	U	42	0.10	68	0.16
7	400	71	U	69	0.16	137	0.31
8	400	71	U	121	0.28	258	0.59
9	482	80	U	158	0.36	416	0.95
10	527	83	U	261	0.60	677	1.55
11	558	81	U	350	0.80	1027	2.35
12	582	75	U	422	0.97	1449	3.32
13	602	68	L	450	1.03	1899	4.35
14	618	60	L	569	1.30	2468	5.65
15	633	52	L	645	1.48	3113	7.13
16	646	46	L	662	1.52	3775	8.65
17	658	41	L	786	1.80	4561	10.45
18	668	37	L	825	1.89	5386	12.34
19	679	34	L	902	2.07	6288	14.41
20	688	32	L	966	2.21	7254	16.62
21	697	30	L	1099	2.52	8353	19.14
22	706	29	S	1158	2.65	9511	21.79
23	715	28	S	1292	2.96	10803	24.75
24	723	27	S	1311	3.00	12114	27.75
25	731	26	S	1416	3.24	13530	31.00
26	739	26	S	1466	3.36	14996	34.36
27	747	26	S	1565	3.59	16561	37.94
28	755	25	S	1714	3.93	18275	41.87
29	763	25	S	1695	3.88	19970	45.75
30	771	25	S	1914	4.38	21884	50.14
31	779	26	S	1837	4.21	23721	54.34
32	788	26	S	1987	4.55	25708	58.90
33	797	26	S	2031	4.65	27739	63.55
34	806	27	S	2031	4.65	29770	68.20
35	815	28	S	2021	4.63	31791	72.83
36	826	29	S	2000	4.58	33791	77.42
37	837	31	S	1957	4.48	35748	81.90
38	849	33	A	1805	4.14	37553	86.03
39	863	36	A	1678	3.84	39231	89.88
40	878	39	A	1419	3.25	40650	93.13
41	897	41	A	1207	2.77	41857	95.89
42	922	41	A	867	1.99	42724	97.88
43	956	36	A	543	1.24	43267	99.12
44	990	22	A	291	0.67	43558	99.79
45	990	22	A	91	0.21	43649	100.00

Table B.17
Raw-to-Scaled Score Table and Frequency Distribution: Grade 8 U.S. History

Raw Score	Scaled Score	CSEM	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	400	73	U	0	0.00	0	0.00
1	400	73	U	1	0.00	1	0.00
2	400	73	U	2	0.00	3	0.01
3	400	73	U	1	0.00	4	0.01
4	400	73	U	4	0.01	8	0.02
5	400	73	U	11	0.03	19	0.04
6	400	73	U	28	0.06	47	0.11
7	400	73	U	53	0.12	100	0.23
8	400	73	U	129	0.30	229	0.53
9	414	73	U	186	0.43	415	0.96
10	497	79	U	288	0.67	703	1.63
11	537	80	U	339	0.79	1042	2.41
12	563	77	U	494	1.14	1536	3.56
13	582	70	U	571	1.32	2107	4.88
14	598	62	U	736	1.71	2843	6.59
15	612	54	U	825	1.91	3668	8.50
16	624	46	L	822	1.90	4490	10.40
17	635	40	L	896	2.08	5386	12.48
18	645	35	L	993	2.30	6379	14.78
19	654	32	L	1056	2.45	7435	17.23
20	663	29	L	1163	2.69	8598	19.92
21	671	27	L	1193	2.76	9791	22.68
22	679	26	L	1288	2.98	11079	25.67
23	686	25	L	1318	3.05	12397	28.72
24	694	24	L	1358	3.15	13755	31.87
25	701	23	S	1496	3.47	15251	35.33
26	708	22	S	1509	3.50	16760	38.83
27	715	22	S	1540	3.57	18300	42.40
28	721	21	S	1584	3.67	19884	46.07
29	728	21	S	1585	3.67	21469	49.74
30	735	21	S	1701	3.94	23170	53.68
31	742	21	S	1730	4.01	24900	57.69
32	749	22	S	1783	4.13	26683	61.82
33	757	22	S	1777	4.12	28460	65.94
34	765	23	S	1760	4.08	30220	70.02
35	773	24	S	1811	4.20	32031	74.21
36	781	25	S	1799	4.17	33830	78.38
37	791	27	S	1669	3.87	35499	82.25
38	801	29	S	1666	3.86	37165	86.11
39	813	33	S	1470	3.41	38635	89.51
40	827	37	A	1315	3.05	39950	92.56
41	844	43	A	1215	2.81	41165	95.37
42	866	48	A	901	2.09	42066	97.46
43	898	50	A	670	1.55	42736	99.01
44	957	41	A	324	0.75	43060	99.76
45	990	41	A	102	0.24	43162	100.00

Table B.18
Composite Score Frequency Distribution: Grade 5 Writing

Composite Score	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
15	U	0	0.00	0	0.00
16	U	0	0.00	0	0.00
17	U	0	0.00	0	0.00
18	U	0	0.00	0	0.00
19	U	571	1.28	571	1.28
20	U	30	0.07	601	1.35
21	U	76	0.17	677	1.52
22	U	52	0.12	729	1.64
23	U	144	0.32	873	1.96
24	U	99	0.22	972	2.18
25	U	131	0.29	1103	2.48
26	L	247	0.55	1350	3.03
27	L	140	0.31	1490	3.35
28	L	209	0.47	1699	3.81
29	L	218	0.49	1917	4.30
30	L	289	0.65	2206	4.95
31	L	67	0.15	2273	5.10
32	L	617	1.39	2890	6.49
33	L	392	0.88	3282	7.37
34	L	3249	7.29	6531	14.66
35	L	753	1.69	7284	16.35
36	S	1468	3.30	8752	19.65
37	S	819	1.84	9571	21.49
38	S	1735	3.90	11306	25.38
39	S	806	1.81	12112	27.19
40	S	1288	2.89	13400	30.08
41	S	2031	4.56	15431	34.64
42	S	799	1.79	16230	36.44
43	S	1543	3.46	17773	39.90
44	S	2596	5.83	20369	45.73
45	S	892	2.00	21261	47.73
46	S	1595	3.58	22856	51.31
47	S	3002	6.74	25858	58.05
48	S	1825	4.10	27683	62.15
49	S	6292	14.13	33975	76.27
50	S	1274	2.86	35249	79.13
51	S	1898	4.26	37147	83.40
52	S	779	1.75	37926	85.14
53	S	445	1.00	38371	86.14
54	A	1319	2.96	39690	89.10
55	A	225	0.51	39915	89.61
56	A	739	1.66	40654	91.27
57	A	941	2.11	41595	93.38
58	A	330	0.74	41925	94.12
59	A	389	0.87	42314	95.00
60	A	2229	5.00	44543	100.00

Table B.19
Composite Score Frequency Distribution: Grade 8 Writing

Composite Score	Level	Frequency	Percent	Cumulative Frequency	Cumulative Percent
15	U	0	0.00	0	0.00
16	U	0	0.00	0	0.00
17	U	0	0.00	0	0.00
18	U	0	0.00	0	0.00
19	U	289	0.68	289	0.68
20	U	30	0.07	319	0.75
21	U	43	0.10	362	0.86
22	U	53	0.13	415	0.98
23	U	25	0.06	440	1.04
24	U	63	0.15	503	1.19
25	L	46	0.11	549	1.30
26	L	161	0.38	710	1.68
27	L	87	0.21	797	1.88
28	L	143	0.34	940	2.22
29	L	72	0.17	1012	2.39
30	L	97	0.23	1109	2.62
31	L	211	0.50	1320	3.12
32	L	216	0.51	1536	3.63
33	L	1226	2.90	2762	6.53
34	L	713	1.69	3475	8.22
35	L	736	1.74	4211	9.96
36	S	686	1.62	4897	11.58
37	S	735	1.74	5632	13.32
38	S	931	2.20	6563	15.52
39	S	1075	2.54	7638	18.06
40	S	1074	2.54	8712	20.60
41	S	1277	3.02	9989	23.62
42	S	1143	2.70	11132	26.32
43	S	1995	4.72	13127	31.04
44	S	1550	3.67	14677	34.71
45	S	2872	6.79	17549	41.50
46	S	2089	4.94	19638	46.44
47	S	8725	20.63	28363	67.07
48	S	1490	3.52	29853	70.59
49	S	1801	4.26	31654	74.85
50	S	964	2.28	32618	77.13
51	S	1443	3.41	34061	80.54
52	S	1126	2.66	35187	83.20
53	S	1152	2.72	36339	85.93
54	A	1337	3.16	37676	89.09
55	A	758	1.79	38434	90.88
56	A	591	1.40	39025	92.28
57	A	650	1.54	39675	93.82
58	A	548	1.30	40223	95.11
59	A	636	1.50	40859	96.62
60	A	1431	3.38	42290	100.00