DRAFT

Oklahoma School Testing Program

2010 Technical Report

Achieving Classroom Excellence

End-of-Instruction

Assessments

Submitted to
The Oklahoma State Department of Education
October 2010

Executive Summary

Introduction

The Oklahoma School Testing Program (OSTP) is a state-wide assessment program that includes the End-of-Instruction (EOI) assessments, where students who complete an area of instruction must also take the corresponding state-wide, standardized assessment. The subjects included within this testing program are Algebra I, Algebra II, Geometry, Biology I, English II, English III, and U.S. History. Each test is a measure of a student's knowledge relative to the *Priority Academic Student Skills (PASS)*, Oklahoma's content standards. These tests are part of the Achieving Classroom Excellence (ACE) legislation passed in 2005 as amended in 2006, which outlines the curriculum, the competencies, and the testing requirements for students to receive a high school diploma from the state of Oklahoma. Algebra I, English II, Biology I, and U.S. History were existing tests in the program with Algebra II, Geometry, and English III added as operational tests for the 2007-2008 testing cycle. These End-of-Instruction tests are administered in Winter, Trimester, Spring, and Summer. The OSTP was established to improve academic achievement for all Oklahoma students and it also meets the requirements of the No Child Left Behind Act (NCLB), which was introduced by the Federal Government in 2001. In 2006, Pearson was contracted by the Oklahoma State Department of Education (SDE) to develop, administer, and maintain the OSTP-ACE EOI tests. This report provides technical details of work accomplished through the end of 2010 on these tests.

Purpose

The purpose of this Technical Report is to provide objective information regarding technical aspects of the OSTP-ACE EOI assessments. This volume is intended to be one source of information to Oklahoma K-12 educational stakeholders (including testing coordinators, educators, parents, and other interested citizens) about the development, implementation, scoring, and technical attributes of the OSTP-ACE EOI assessments. Other sources of information regarding the OSTP-ACE EOI tests—administered mostly online, with some paper formatted tests available—include the administration manuals, interpretation manuals, student-, teacher-, and parent guides, implementation materials, and training materials.

The information provided here fulfills legal, professional, and scientific guidelines (AERA, APA, & NCME, 1999) for technical reports of large-scale educational assessments and is intended for use by qualified users within schools who use the OSTP-ACE EOI assessments and interpret the results. Specifically, information was selected for inclusion in this report based on NCLB requirements and the following Standards for Educational and Psychological Testing:

- Standards 6.1 – 6.15 Supporting Documentation for Tests
- Standards 10.1—10.12 Testing Individuals with Disabilities
- Standards13.1—13.19 Educational Testing and Assessment

This technical report provides accurate, complete, current, and clear documentation of the OSTP-ACE EOI development methods, data analysis, and results as is appropriate for use by qualified users and technical experts. Section 1 provides an overview of the test design, test content, and content standards. Section 2 provides summary information about the test administration. Section 3 details the classical item analyses and reliability results, and Section 4 details the calibration, equating, scaling analyses, and results. Section 5 provides the results of the classification accuracy and classifications studies. Finally, Section 6

provides higher-level summaries of all the tests included in the OSTP-ACE EOI testing program.

Information provided in this report presents valuable information about the OSTP-ACE EOI assessments regarding:

1. Content standards,
2. Content of the tests,
3. Test form design,
4. Administration of the tests,
5. Identification of ineffective items,
6. Detection of item bias,
7. Reliability of the tests,
8. Calibration of the tests,
9. Equating of tests,
10. Scaling and scoring of the tests, and
11. Decision accuracy and classification.

Each of these facets in the OSTP-ACE EOI assessments development and use cycle is critical to validity of test scores and interpretation of results. This technical report covers all of these topics for the 2009-10 testing year.

Table of Contents

Section 1 1
1.1 Overview of the OSTP-ACE EOI Assessments 1
1.1.a Purpose 1
1.1.b *PASS* Content Standards 1
1.2 Summary of Test Development and Content Validity 3
1.2.a Aligning Test to *PASS* Content Standards 3
1.2.b Item Pool Development and Selection 4
1.2.c Configuration of the Seven Tests 5
1.2.d Operational and Field Test Items by Content Area 6
Section 2 14
2.1 Packaging and Shipping 14
2.2 Materials Return 14
2.3 Materials Discrepancies Process 15
2.4 Processing Assessment Materials Returned by Schools 15
Section 3 16
3.1 Sampling Plan and Field Test Design 16
3.1.a Sampling Plan 16
3.1.b Field Test Design 16
3.1.c Data Receipt Activities 16
3.2 Classical Item Analyses 18
3.2.a Test-Level Summaries of Classical Item Analyses 18
3.3 Procedures for Detecting Item Bias 19
3.3.a Differential Item Functioning Results 20
3.4 Data Review 21
3.4.a Results of Data Review 22
3.5 Test Reliability 23
3.6 Test Reliability by Subgroup 24
3.7 Inter-rater Reliability 25
Section 4 29
4.1 Item Response Theory (IRT) models 29
4.2 Assessment of IRT Fit to the model 29
4.2.a Calibration and IRT Fit Results 30
4.2.a.i Winter/Trimester 2009-10 30
4.2.a.ii Spring 2010 31
4.3 Calibration and Equating 32
4.3.a Common Linking Items for Spring 2010 32
4.4 Item Stability Evaluation Methods 33
4.4.a Results of the Item Parameter Item Stability Check 34
4.5 Scaling and Scoring Results 34

Section 5 55
5.1 Classification Consistency and Accuracy 55
Section 6 60
6.1 Descriptive Statistics 60
6.2 Performance Level Distribution 65
6.3 Conditional Standard Error of Measurement 65
6.4 Standard Error of Measurement 66
References 68
Appendix A 69
Appendix B 80
Appendix C 95

List of Tables

Table 1.1. Oklahoma Content Standards by Subject 2
Table 1.2. Criteria for Aligning the Test with *PASS* Standards and Objectives. 4
Table 1.3. Percentage of Items by Depth of Knowledge Levels 5
Table 1.4. Configuration of the OSTP-ACE EOI Tests for Winter/Trimester 2009-10 6
Table 1.6. Number of Common Linking Items per Subject for Spring 2010 6
Table 1.7. Number of Items and Points by Content Standard for Algebra I 7
Table 1.8. Number of Items and Points by Content Standard for Algebra II 8
Table 1.9. Number of Items and Points by Content Standard for Geometry 9
Table 1.10. Number of Items and Points by Content Standard for Biology I 10
Table 1.11. Number of Items and Points by Content Standard for English II 11
Table 1.12. Number of Items and Points by Content Standard for English III 12
Table 1.13. Number of Items and Points by Content Standard for U.S. History 13
Table 3.1. Demographic Characteristics of Calibration and Equating Sample for Winter/Trimester 2009-10 17
Table 3.2. Demographic Characteristics of Calibration and Equating Sample for Spring 2010. 17
Table 3.3. Test-Level Summaries of Classical Item Analyses for Winter/Trimester 2009-10 and Spring 2010 19
Table 3.4. DIF Flag Incidence Across All OSTP-ACE EOI Field Test Items for Winter/Trimester 2009-10 and Spring 2010 21
Table 3.5. Number of Items Per Subject Flagged and Rejected During Winter/Trimester 2009–2010 and Spring 2010 Field Test Data Review 23
Table 3.6. Cronbach's Alpha for Winter/Trimester 2009-10 and Spring 2010 Administrations by Subject 24
Table 3.7. Test Reliability by Subgroup for Spring 2010 25
Table 3.8. Inter-rater Reliability for English II Operational Writing Prompts for Winter/Trimester 2009-10 and Spring 2010 27
Table 3.9. Inter-rater Reliability for English III Operational Writing Prompts for Winter/Trimester 2009-10 and Spring 2010 28
Table 4.1. Number of Common Linking Items Per Subject for Spring 2010 33
Table 4.2. LOSS, HOSS, and Scaling Constants by Subject 35
Table 4.3. Performance-Level Cut Scores by Subject 35
Table 4.4. Raw Score to Scale Score Conversion Tables for Winter/Trimester 2009-10 36
Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 42
Table 5.1. Estimates of Accuracy and Consistency of Performance Classification for Winter/Trimester 2009-10 56

Table 5.2. Estimates of Accuracy and Consistency of Performance Classification for Spring 2010 56
Table 5.3. Accuracy and Consistency Estimates by Cut Score: False Positive- and False Negative Rates for Winter/Trimester 2009-10 58
Table 5.4. Accuracy and Consistency Estimates by Cut Score: False Positive- and False Negative Rates for Spring 2010 59
Table 6.1. Descriptive Statistics of the Scale Scores for Winter/Trimester 2009-10 - Overall. 60
Table 6.2. Descriptive Statistics of the Scale Scores for Winter/Trimester 2009-10 by Gender 60
Table 6.3. Descriptive Statistics of the Scale Scores for Winter/Trimester 2009-10 by Race/Ethnicity 60
Table 6.4. Descriptive Statistics of the Scale Scores for Winter/Trimester 2009-10 by Free/Reduced Lunch Status 61
Table 6.5. Descriptive Statistics of the Scale Scores for Spring 2010 - Overall 62
Table 6.6. Descriptive Statistics of the Scale Scores for Spring 2010 by Gender 62
Table 6.7. Descriptive Statistics of the Scale Scores for Spring 2010 by Race/Ethnicity 63
Table 6.8. Descriptive Statistics of the Scale Scores for Spring 2010 by Free/Reduced Lunch Status 64
Table 6.9. Percentage of Students by Performance Level for Winter/Trimester 2009-10 and Spring 2010 65
Table 6.10. Overall Estimates of *SEM* by Subject 67

Section 1

Overview of the Oklahoma School Testing Program (OSTP) Achieving Classroom Excellence (ACE) End-of-Instruction (EOI) Assessments

1.1 Overview of the OSTP-ACE EOI Assessments

The Achieving Classroom Excellence End-of-Instruction assessment is a state-mandated, secondary-level, criterion-referenced testing program used to assess student proficiency at the End-of-Instruction in Algebra I, Algebra II, Geometry, Biology I, English II, English III, and U.S. History. The Oklahoma ACE EOI tests are used to assess student proficiency relative to a specific set of academic skills established by committees of Oklahoma educators. This special set of skills is referred to as the *Priority Academic Student Skills* (*PASS*), which represents skills that students are expected to master by the End-of-Instruction for each subject. All secondary-level students, who have completed instruction in Algebra I, Algebra II, Geometry, Biology I, English II, English III, and U.S. History must take the corresponding Oklahoma ACE EOI tests. The Spring 2009 administration was the first administration with graduation requirements attached to them for the incoming freshman students. For these students and future students, to graduate with a high school diploma from the State of Oklahoma, students must score proficient or above in Algebra I and English II, and two of the following five: Algebra II, Biology I, English III, Geometry, or U.S. History. Students who fail to earn a proficient score are permitted to retake these tests. All *PASS* standards and objectives are measured exclusively by multiple-choice items, except for English II and English III, each of which include one writing prompt. The Winter/Trimester 2009-10 and Spring 2010 OSTP-ACE EOI Algebra I, Algebra II, Geometry, Biology I, English II, English III, and U.S. History assessments were developed by Pearson in collaboration with the Oklahoma State Department of Education (SDE) and were administered by SDE.

Pearson scored, equated, and scaled the assessments. There was one form administered in Winter/Trimester 2009-10 for each subject. In the Spring 2010 administration, there were two core operational forms with 11 field test forms for Algebra I, Algebra II, Geometry, Biology I, and U.S. History and 16 field test forms for English II and English III. Each test form was embedded with field test items to add to the item pool. In addition, an equivalent form from one of the previous administrations was designated as a breach form and a Braille test was built for each subject using the Winter/Trimester 2009-10 test forms and then used again in the Spring 2010 administration. A student could receive an equivalent form for various reasons, including becoming ill during test administration or experiencing some kind of security breach. The State Department of Education Office of Accountability and Assessments determines eligibility for an equivalent form on a case-by-case basis. These students' responses were scored and reported using the scoring tables from the form's previous administration.

1.1.a Purpose

Pearson developed the 2009-10 OSTP-ACE EOI assessments to measure the Oklahoma *PASS* content standards, as listed in the following section. The objectives associated with content and/or process standards tested are provided in Appendix A.

1.1.b *PASS* Content Standards

The Oklahoma Content Standards are shown in Table 1.1.

Table 1.1. Oklahoma Content Standards by Subject

Algebra I	
Standard 1.	Number Sense and Algebraic Operations
Standard 2.	Relations and Functions
Standard 3.	Data Analysis, Probability & Statistics
Algebra II	
Standard 1.	Number Sense and Algebraic Operations
Standard 2.	Relations and Functions
Standard 3.	Data Analysis, Probability, & Statistics
Geometry	
Standard 1.	Logical Reasoning
Standard 2.	Properties of 2-Dimensional Figures
Standard 3.	Triangles and Trigonometric Ratios
Standard 4.	Properties of 3-Dimensional Figures
Standard 5.	Coordinate Geometry
Biology I	
PASS Process/Inquiry Standards and Objectives	
Process 1.	Observe and Measure
Process 2.	Classify
Process 3.	Experiment
Process 4.	Interpret and Communicate
Process 5.	Model
PASS Content Standards and Objectives	
Standard 1.	The Cell
Standard 2.	The Molecular Basis of Heredity
Standard 3.	Biological Diversity
Standard 4.	The Interdependence of Organisms
Standard 5.	Matter/Energy/Organization in Living Systems
Standard 6.	The Behavior of Organisms
English II	
Reading/Literature:	
Standard 1.	Vocabulary
Standard 2.	Comprehension
Standard 3.	Literature
Standard 4.	Research and Information
Writing/Grammar/Usage and Mechanics:	
Standard 1/2.	Writing
Standard 3.	Grammar/Usage and Mechanics
English III	
Reading/Literature:	
Standard 1.	Vocabulary
Standard 2.	Comprehension
Standard 3.	Literature
Standard 4.	Research and Information
Writing/Grammar/Usage and Mechanics:	
Standard 1/2.	Writing
Standard 3.	Grammar/Usage and Mechanics

Table 1.1. Oklahoma Content Standards by Subject (cont.)

U.S. History	
Standard 1.	Civil War/Reconstruction Era
Standard 2.	Impact of Immigration and Industrialization
Standard 3.	Imperialism, World War I, and Isolationism
Standard 4.	United States During the 1920s and 1930s
Standard 5.	World War II
Standard 6.	United States Since World War II

1.2 Summary of Test Development and Content Validity

To ensure content validity of the Oklahoma ACE EOI tests, Pearson content experts closely study the Oklahoma *Priority Academic Student Skills (PASS)* and work with Oklahoma content area specialists, teachers, and assessment experts to develop a pool of items that measure Oklahoma's Assessment Frameworks (i.e., *PASS*) for each subject. Once the need for field test items was determined, based on the availability of items for future test construction, a pool of items that measured Oklahoma's *PASS* in each subject was developed. These items were developed under universal design guidelines set by the SDE and carefully reviewed and discussed by Content and Bias/Sensitivity Review Committees to evaluate not only content validity, but also plain language and the quality and appropriateness of the items. These committees were comprised of Oklahoma teachers and SDE staff. The committees' recommendations were used to select and/or revise items from the item pool used to construct the field test portions of the Winter/Trimester 2009-10 and the Spring 2010 assessments.

1.2.a Aligning Test to *PASS* Content Standards

In addition to the test Blueprints provided by SDE, Table 1.2 describes four criteria for test alignment with the *PASS* Standards and Objectives.

Table 1.2. Criteria for Aligning the Test with *PASS* Standards and Objectives.

1. Categorical Concurrence	The test is constructed so that there are at least six items measuring each *PASS* standard with the content category consistent with the related standard. The number of items, six, is based on estimating the number of items that could produce a reasonably reliable estimate of a student's mastery of the content measured.
2. Range-of-Knowledge	The test is constructed so that at least 50% of the objectives for a *PASS* standard have at least one corresponding assessment item.
3. Balance-of-Representation	The test is constructed according to the alignment blueprint, which reflects the degree of representation given on the test to each *PASS* standard and objective in terms of the percent of total test items measuring each standard and the number of test items measuring each objective.
4. Source-of-Challenge	Each test item is constructed in such a way that the major cognitive demand comes directly from the targeted *PASS* skill or concept being assessed, not from specialized knowledge or cultural background that the test-taker may bring to the testing situation.

1.2.b Item Pool Development and Selection

The source of the operational items included a pool of previously field-tested or operationally-administered items ranging from the Spring 2005 through the Spring 2009 administrations for Algebra I, Biology I, English II, and U.S. History and from the census Spring 2007 field test through the Spring 2009 embedded field test for Algebra II, Geometry, and English III. Note that the items were calibrated live using data from the operational administrations to estimate parameters for these items.

The ACE EOI tests for the Winter/Trimester 2009-10 and Spring 2010 cycle were built by including previously field-tested and operational items. To equate the forms across years, the entire set of operational items served as anchors or links to the base scale. Equating is necessary to account for slight year-to-year differences in form difficulty and to maintain comparability across years. Details of the equating procedures applied are provided in a subsequent section in this document. Content experts also targeted the percentage of items measuring various Depth of Knowledge (DOK) levels for assembling the tests. Table 1.3 provides the DOK level percentages for the Winter/Trimester 2009-10 and Spring 2010 operational assessments. Notice that the actual percentage is close but not exactly within the target percentages in the operational test for some content areas.

Table 1.3. Percentage of Items by Depth of Knowledge Levels

Test Session	DOK Level	Target DOK	Actual			
			Algebra I	Algebra II	Geometry	Biology I
Winter/ Trimester 2009-10	1	10-15	14.55	14.55	14.55	13.33
	2	60-70	60.00	65.45	63.64	68.33
	3/4	15-25	25.45	20.00	21.82	18.33
Spring 2010 Core A	1	10-15	12.73	14.55	12.73	15.00
	2	60-70	67.27	63.64	63.64	65.00
	3/4	15-25	20.00	21.82	23.64	20.00
Spring 2010 Core B	1	10-15	12.73	10.91	12.73	11.67
	2	60-70	65.45	67.27	61.82	68.33
	3/4	15-25	21.82	21.82	25.45	20.00

Note: All values are in percentages.

Table 1.3. Percentage of Items by Depth of Knowledge Levels (cont.)

Test Session	DOK Level	Target DOK	Actual		
			English II	English III	U.S. History
Winter/ Trimester 2009-10	1	10-15	11.48	9.52	11.67
	2	60-70	68.85	69.84	66.67
	3/4	15-25	19.67	20.63	21.67
Spring 2010 Core A	1	10-15	13.11	14.29	11.67
	2	60-70	60.66	63.49	65.00
	3/4	15-25	26.23	22.22	23.33
Spring 2010 Core B	1	10-15	8.20	7.94	10.00
	2	60-70	70.49	71.43	65.00
	3/4	15-25	21.31	20.63	25.00

Note: All values are in percentages.

1.2.c Configuration of the Seven Tests

Table 1.4 and Table 1.5 provide overviews of the number of operational and field test items for the Winter/Trimester 2009-10 and Spring 2010 OSTP-ACE EOI assessments. The Spring 2010 test was comprised of two dual core, operationally scored forms for each subject. While most items were unique to each form, there were at least 15 items in common across the core forms for use during calibration, scaling, and equating. The number of common linking items per subject is presented in Table 1.6. Field test items were embedded in the operational test forms for all content areas to build the item bank for future use. The forms in the Spring 2010 assessments were randomly assigned within classrooms to obtain randomly-equivalent samples of examinees for the field test items.

Table 1.4. Configuration of the OSTP-ACE EOI Tests for Winter/Trimester 2009-10

					Maximum Possible Points on Test Items (Per Form)			
		Item Counts (Per Form)			OP		FT	
Subject	Forms	OP	FT	Test	MC	CR	MC	CR
Algebra I	1	55	20	75	55	0	20	0
Algebra II	1	55	20	75	55	0	20	0
Biology I	1	60	20	80	60	0	20	0
English II	1	60/1*	20	80/1*	60	6	20	0
English III	1	62/1*	20	82/1*	62	10	20	0
Geometry	1	55	20	75	55	0	20	0
U.S. History	1	60	20	80	60	0	20	0

Note: OP = Operational; FT = Field Test; MC = Multiple Choice; CR = Constructed Response; * = multiple choice/constructed response.

Table 1.5. Configuration of the OSTP-ACE/EOI Tests for Spring 2010

					Maximum Possible Points on Test Items (Per Form)			
		Item Counts (Per Form)			OP		FT	
Subject	Forms	OP**	FT	Test	MC	CR	MC	CR
Algebra I	11	55	20	75	55	0	20	0
Algebra II	11	55	20	75	55	0	20	0
Biology I	11	60	20	80	60	0	20	0
English II	16	60/1*	20/1*	80/2*	60	6	20	6
English III	16	62/1*	20/1*	82/2*	62	10	20	10
Geometry	11	55	20	75	55	0	20	0
U.S. History	11	60	20	80	60	0	20	0

Note: OP = Operational; FT = Field Test; MC = Multiple Choice; CR = Constructed Response; * = multiple choice/constructed response; **=by Core Form (some items were common across forms).

Table 1.6. Number of Common Linking Items per Subject for Spring 2010

Subject	No. of CL Items	Total No. of Items*
Algebra I	19	91
Algebra II	21	89
Biology I	20	100
English II	20	100
English III	20	104
Geometry	20	90
U.S. History	19	101

Note: No. = Number; CL = common linking items; *= Number of unique operational items.

1.2.d Operational and Field Test Items by Content Area

Algebra I. The Winter/Trimester 2009-10 Algebra I administration was comprised of one form with 55 operational items and 20 field test items. All operational items were considered anchor items on this form, selected from available items in the item bank. There were two core forms and 11 field test sets in the Spring 2010 administration. Each of the forms contained 55 operational items and 20 field test items, totaling 75 items per form. The

number of items and maximum points possible by content standard is shown in Table 1.7. Algebra I scores were reported by content standard and at the objective level. There were nine or more operational items in each reported category. Each item was mapped to one content standard and one objective per content standard.

Table 1.7. Number of Items and Points by Content Standard for Algebra I

	Content Standard							
	1		2		3		Total	
Form	Items	Points	Items	Points	Items	Points	Items	Points
Winter 2009-10								
Operational	15	15	31	31	9	9	55	55
FT Form 1	8	8	7	7	5	5	20	20
Spring 2010								
Core A	15	15	31	31	9	9	55	55
Core B	15	15	31	31	9	9	55	55
FT Form 1	6	6	10	10	4	4	20	20
FT Form 2	4	4	14	14	2	2	20	20
FT Form 3	8	8	10	10	2	2	20	20
FT Form 4	5	5	13	13	2	2	20	20
FT Form 5	6	6	12	12	2	2	20	20
FT Form 6	5	5	13	13	2	2	20	20
FT Form 7	5	5	12	12	3	3	20	20
FT Form 8	5	5	13	13	2	2	20	20
FT Form 9	4	4	13	13	3	3	20	20
FT Form 10	5	5	13	13	2	2	20	20
FT Form 11	4	4	14	14	2	2	20	20

Note: FT = Field Test.

Algebra II. The Winter/Trimester 2009-10 Algebra II administration was comprised of one form with 55 operational items and 20 field test items. All operational items were considered anchor items on this form, selected from available items in the item bank. There were two core forms and 11 field test sets in the Spring 2010 administration. Each of the forms contained 55 operational items and 20 field test items, totaling 75 items per form. The number of items and maximum points possible by content standard is shown in Table 1.8. Algebra II scores were reported by content standard and at the objective level. There were nine or more operational items in each reported category. Each item was mapped to one content standard and one objective per content standard.

Table 1.8. Number of Items and Points by Content Standard for Algebra II

	Content Standard							
	1		2		3		Total	
Form	Items	Points	Items	Points	Items	Points	Items	Points
Winter 2009-10								
Operational	15	15	31	31	9	9	55	55
FT Form 1	2	2	15	15	3	3	20	20
Spring 2010								
Core A	15	15	31	31	9	9	55	55
Core B	15	15	31	31	9	9	55	55
FT Form 1	8	8	10	10	2	2	20	20
FT Form 2	7	7	10	10	3	3	20	20
FT Form 3	6	6	12	12	2	2	20	20
FT Form 4	6	6	11	11	3	3	20	20
FT Form 5	6	6	12	12	2	2	20	20
FT Form 6	5	5	9	9	6	6	20	20
FT Form 7	5	5	14	14	1	1	20	20
FT Form 8	7	7	11	11	2	2	20	20
FT Form 9	7	7	11	11	2	2	20	20
FT Form 10	7	7	10	10	3	3	20	20
FT Form 11	5	5	11	11	4	4	20	20

Note: FT = Field Test.

Geometry. The Winter/Trimester 2009-10 Geometry administration was comprised of one form with 55 operational items and 20 field test items. All operational items were considered anchor items on this form, selected from available items in the item bank. There were two core forms and 11 field test sets in the Spring 2010 administration. Each of the forms contained 55 operational items and 20 field test items, totaling 75 items per form. The number of items and maximum points possible by content standard is shown in Table 1.9. Geometry scores were reported by content standard and at the objective level. There were six or more items in each reported category. Each item was mapped to one content standard and one objective per content standard.

Table 1.9. Number of Items and Points by Content Standard for Geometry

	Content Standard											
	1		2		3		4		5		Total	
Form	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts
Winter 2009-10												
Operational	6	6	20	20	12	12	10	10	7	7	55	55
FT Form 1	2	2	9	9	4	4	3	3	2	2	20	20
Spring 2010												
Core A	6	6	20	20	12	12	10	10	7	7	55	55
Core B	6	6	20	20	12	12	10	10	7	7	55	55
FT Form 1	3	3	7	7	4	4	4	4	2	2	20	20
FT Form 2	3	3	6	6	3	3	3	3	5	5	20	20
FT Form 3	2	2	6	6	5	5	3	3	4	4	20	20
FT Form 4	2	2	9	9	4	4	4	4	1	1	20	20
FT Form 5	1	1	6	6	7	7	4	4	2	2	20	20
FT Form 6	2	2	7	7	4	4	4	4	3	3	20	20
FT Form 7	2	2	6	6	6	6	4	4	2	2	20	20
FT Form 8	2	2	7	7	3	3	6	6	2	2	20	20
FT Form 9	2	2	8	8	4	4	5	5	1	1	20	20
FT Form 10	3	3	5	5	5	5	4	4	3	3	20	20
FT Form 11	2	2	8	8	4	4	4	4	2	2	20	20

Note: Its = Number of Items; Pts = Number of Points; FT = Field Test.

Biology I. The Winter/Trimester 2009-10 Biology I administration was comprised of one form with 60 operational items and 20 field test items. All operational items were considered anchor items on this form, selected from available items in the item bank. There were two core forms and 11 field test sets in the Spring 2010 administration. Each of the forms contained 60 operational items and 20 field test items, totaling 80 items per form. The number of items and the maximum number points possible by content standard is shown in Table 1.10. Biology I scores were reported for content and process standards at the standard level. Each reported process standard has eight or more items and each content standard has eight or more items. Unlike other subjects, all items in Biology I were primarily mapped to process standards. All items (except safety items) were also mapped to content standards.

Table 1.10. Number of Items and Points by Content Standard for Biology I

	Content Standard													
	1		2		3		4		5		6		Total*	
Form	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts
Winter 2009-10														
Operational	9	9	9	9	9	9	12	12	9	9	8	8	56	56
FT Form 1	1	1	2	2	8	8	3	3	4	4	1	1	19	19
Spring 2010														
Core A	9	9	9	9	9	9	12	12	9	9	8	8	56	56
Core B	9	9	9	9	9	9	12	12	9	9	8	8	56	56
FT Form 1	3	3	2	2	3	3	3	3	5	5	3	3	19	19
FT Form 2	1	1	2	2	3	3	5	5	3	3	5	5	19	19
FT Form 3	3	3	2	2	2	2	4	4	3	3	5	5	19	19
FT Form 4	1	1	5	5	2	2	5	5	1	1	5	5	19	19
FT Form 5	2	2	2	2	3	3	3	3	2	2	7	7	19	19
FT Form 6	2	2	4	4	3	3	4	4	2	2	5	5	20	20
FT Form 7	4	4	2	2	2	2	3	3	3	3	5	5	19	19
FT Form 8	-	-	4	4	1	1	6	6	2	2	6	6	19	19
FT Form 9	3	3	2	2	4	4	4	4	1	1	6	6	20	20
FT Form 10	3	3	3	3	2	2	4	4	1	1	7	7	20	20
FT Form 11	2	2	5	5	1	1	4	4	4	4	4	4	20	20

Note: Its = Number of Items; Pts = Number of Points; FT = Field Test; Some totals for OP forms and FT forms are less than 60 (for OP) and 20 (for FT) due to dual item alignment – an item does not map to a content standard, but maps to a process.

English II. The Winter/Trimester 2009-10 English II administration was comprised of one form with 60 operational MC items, 1 open-ended writing prompt, 20 field test MC items, and 1 field test open-ended writing prompt. All operational items were considered anchor items on this form, selected from available items in the item bank. There were two core forms and 16 field test sets in the Spring 2010 administration. Each of the forms contained 60 operational MC items, 1 operational open-ended writing prompt, 20 field test MC items, and one field test open-ended writing prompt, totaling 82 items per form. Table 1.11 lists the number of items and the maximum possible number of points by content standard in the Winter/Trimester 2009-10 and Spring 2010 forms. English II scores were reported at the content standard level. Each item was mapped to one content standard and one objective. The writing prompts in English II were scored analytically on five traits with a maximum of four score points per trait. The scores in the analytic traits were reported in the Writing report. The trait scores were weighted differentially to derive a composite score that ranged from 1 to 6. The composite scores contributed to the English II total score.

Table 1.11. Number of Items and Points by Content Standard for English II

	Content Standard													
	R1		R2		R3		R4		W1/W2		W3		Total	
Form	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts
Winter 2009-10														
Operational	7	7	17	17	18	18	6	6	1	6	12	12	61	66
FT Form 1	2	2	6	6	1	1	3	3	1	6	8	8	21	26
Spring 2010														
Core A	8	8	17	17	17	17	6	6	1	6	12	12	61	66
Core B	6	6	18	18	18	18	6	6	1	6	12	12	61	66
FT Form 1	1	1	7	7	8	8	4	4	1	6	-	-	21	26
FT Form 2	1	1	5	5	5	5	1	1	1	6	8	8	21	26
FT Form 3	4	4	7	7	7	7	2	2	1	6	-	-	21	26
FT Form 4	2	2	5	5	3	3	2	2	1	6	8	8	21	26
FT Form 5	3	3	6	6	8	8	3	3	1	6	-	-	21	26
FT Form 6	3	3	8	8	8	8	1	1	1	6	-	-	21	26
FT Form 7	2	2	7	7	9	9	2	2	1	6	-	-	21	26
FT Form 8	1	1	8	8	9	9	2	2	1	6	-	-	21	26
FT Form 9	2	2	5	5	4	4	1	1	1	6	8	8	21	26
FT Form 10	2	2	7	7	7	7	4	4	1	6	-	-	21	26
FT Form 11	3	3	11	11	3	3	3	3	1	6	-	-	21	26
FT Form 12	2	2	9	9	6	6	3	3	1	6	-	-	21	26
FT Form 13	2	2	4	4	4	4	2	2	1	6	8	8	21	26
FT Form 14	3	3	6	6	7	7	4	4	1	6	-	-	21	26
FT Form 15	2	2	9	9	3	3	6	6	1	6	-	-	21	26
FT Form 16	2	2	8	8	8	8	2	2	1	6	-	-	21	26

Note: Its = Number of Items; Pts = Number of Points; FT = Field Test.

English III. The Winter/Trimester 2009-10 English III administration was comprised of one form with 62 operational MC items, 1 open-ended writing prompt, 20 field test MC items, and 1 field test open-ended writing prompt. All operational items were considered anchor items on this form, selected from available items in the item bank. There were two core forms and 16 field test sets in the Spring 2010 administration. Each of the forms contained a set of 62 operational MC items, 1 operational open-ended writing prompt, 20 field test MC items, and 1 field test open-ended writing prompt, totaling 84 items per form. Table 1.12 lists the number of items and the maximum possible number of points by content standard in the Winter/Trimester 2009-10 and Spring 2010 tests. English III scores were reported at the content standard level. Each item was mapped to one content standard and one objective. The writing prompts in English III were scored analytically on five traits with a maximum of four score points for each trait. The scores in the analytic traits were reported in the Writing report. The trait scores were weighted differentially to derive a composite score that ranged from 1 to 10. The composite scores contributed to the English III total score.

Table 1.12. Number of Items and Points by Content Standard for English III

	Content Standard												Total	
	R1		R2		R3		R4		W1/W2		W3			
Form	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts
Winter 2009-10														
Operational	6	6	19	19	17	17	6	6	1	10	14	14	63	72
FT Form 1	3	3	6	6	8	8	3	3	1	10	-	-	21	30
Spring 2010														
Core A	6	6	19	19	17	17	6	6	1	10	14	14	63	72
Core B	6	6	18	18	18	18	6	6	1	10	14	14	63	72
FT Form 1	3	3	8	8	6	6	3	3	1	10	-	-	21	30
FT Form 2	2	2	5	5	5	5	2	2	1	10	6	6	21	30
FT Form 3	4	4	5	5	8	8	3	3	1	10	-	-	21	30
FT Form 4	2	2	6	6	4	4	2	2	1	10	6	6	21	30
FT Form 5	2	2	9	9	6	6	3	3	1	10	-	-	21	30
FT Form 6	1	1	5	5	5	5	1	1	1	10	8	8	21	30
FT Form 7	2	2	8	8	7	7	3	3	1	10	-	-	21	30
FT Form 8	2	2	3	3	5	5	2	2	1	10	8	8	21	30
FT Form 9	2	2	9	9	4	4	1	1	1	10	4	4	21	30
FT Form 10	2	2	8	8	6	6	4	4	1	10	-	-	21	30
FT Form 11	2	2	6	6	6	6	2	2	1	10	4	4	21	30
FT Form 12	2	2	9	9	7	7	2	2	1	10	-	-	21	30
FT Form 13	2	2	5	5	5	5	4	4	1	10	4	4	21	30
FT Form 14	3	3	8	8	5	5	4	4	1	10	-	-	21	30
FT Form 15	2	2	6	6	8	8	.	.	1	10	4	4	21	30
FT Form 16	4	4	8	8	5	5	3	3	1	10	-	-	21	30

Note: Its = Number of Items; Pts = Number of Points; FT = Field Test.

U.S. History. The Winter/Trimester 2009-10 U.S. History administration was comprised of one form with 60 operational items and 20 field test items. All operational items were considered anchor items on this form, selected from available items in the item bank. There were two core forms and 11 field test sets in the Spring 2010 administration. Each of the forms contained a set of 60 operational items and 20 field test items, totaling 80 items per form. The number of items and maximum points possible by content standard in Winter/Trimester 2009-10 and Spring 2010 are shown in Table 1.13. U.S. History scores were reported only at the content standard level and each reported standard had six or more items.

Table 1.13. Number of Items and Points by Content Standard for U.S. History

	Content Standard													
	1		2		3		4		5		6		Total	
Form	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts	Its	Pts
Winter 2009-10														
Operational	6	6	9	9	9	9	12	12	9	9	15	15	60	60
FT Form 1	1	1	4	4	2	2	4	4	2	2	7	7	20	20
Spring 2010														
Core A	6	6	9	9	9	9	12	12	9	9	15	15	60	60
Core B	6	6	9	9	9	9	12	12	9	9	15	15	60	60
FT Form 1	3	3	2	2	3	3	5	5	3	3	4	4	20	20
FT Form 2	3	3	2	2	2	2	5	5	3	3	5	5	20	20
FT Form 3	2	2	4	4	4	4	5	5	3	3	2	2	20	20
FT Form 4	2	2	3	3	2	2	5	5	3	3	5	5	20	20
FT Form 5	-	-	3	3	4	4	5	5	3	3	5	5	20	20
FT Form 6	3	3	2	2	3	3	5	5	2	2	5	5	20	20
FT Form 7	2	2	3	3	3	3	4	4	4	4	4	4	20	20
FT Form 8	2	2	3	3	3	3	4	4	2	2	6	6	20	20
FT Form 9	2	2	3	3	2	2	4	4	4	4	5	5	20	20
FT Form 10	2	2	4	4	3	3	4	4	2	2	5	5	20	20
FT Form 11	2	2	2	2	3	3	5	5	4	4	4	4	20	20

Note: Its = Number of Items; Pts = Number of Points; FT = Field Test.

Section 2

Administration of the ACE EOI assessments

Valid and reliable assessment requires that assessments are first constructed in alignment with the Oklahoma content standards and then administered and scored according to sound measurement principles. Sound assessment practices require that schools administer all assessments in a consistent manner across the state so that all students have a fair and equitable opportunity for a score that accurately reflects their achievement in each subject.

The schools play a key role in administering the OSTP-ACE EOI assessments in a manner consistent with established procedures, monitoring the fair administration of the assessment, and working with the SDE office to address deviations from established assessment administration procedures. The role that district and school faculty members play is essential in the fair and equitable administration of successful ACE EOI assessments.

2.1 Packaging and Shipping

To provide OSTP-ACE EOI with secure and dependable services for the shipping of the Oklahoma assessment materials, Pearson's Warehousing and Transportation Department maintains the quality and security of material distribution and return by using such methods as sealed trailers and hiring reputable carriers with the ability to immediately trace shipments. Pearson uses all available tracking capabilities to provide status information and early opportunities for corrective action.

Materials are packaged by school and delivered to the district coordinators. Each shipment to a district contains a shipping document set that includes a packing list for each school's materials and a pallet map that shows the identity and pallet assignment of each carton.

Materials are packaged using information provided by the Assessment Coordinators through Pearson's SchoolHouse™ website, and optionally with data received directly from Oklahoma. Oklahoma educators also use the SchoolHouse™ site to provide Pearson with the Pre-Identification information needed to print the student identification section on answer documents. Bar-coding of all secure materials during the pre-packaging effort allows for accurate tracking of these materials through the entire packing, delivery, and return process. It also permits Pearson to inventory all materials throughout the packaging and delivery process along with the ability to provide the customer with status updates at any time. Use of handheld radio-frequency scanners in the packaging process help to eliminate the possibility of packing the wrong materials. The proprietary "pick-and-pack" process prompts packaging personnel as to what materials are to go in which shipping box. If the packer tries to pack the wrong item (or number of items into a shipping carton), the system signals an alert.

2.2 Materials Return

Test administration handbooks provide clear instructions on how to assemble, box, label, and return testing materials after test administration. Because of the criticality of used test materials and quantities often involved, safety is also a major concern, not only for the materials but for the people moving them. Only single-column boxes are used to distribute and collect test materials, so the weight of each carton is kept to a reasonable and manageable limit.

Paper bands are provided to group and secure used student response booklets for scoring. Color-coded return mailing labels with detailed return information (district address and code number, receipt address, box *x* of *y*, shipper's tracking number, etc.) are also provided. These labels facilitate accurate and efficient sorting of each carton and its contents upon receipt at Pearson.

2.3 Materials Discrepancies Process

The image scanning process enables Pearson to concurrently capture Optical Mark Read (OMR) responses, images, and security information electronically. All scorable material discrepancies are captured, investigated by our Oklahoma Call Center team, reported, and resolved prior to a batch passing through a clean post edit and images being released for scoring.

As scanning of materials progresses, any discrepancies in materials received versus shipped are reported immediately to the SDE and scoring will begin. This system allows Pearson to proceed in scoring clean batches while any discrepant material issues are being resolved. As discrepant materials are received, they will be processed. Data from discrepant material receipts are captured in the same database as all other material receipts resulting in a complete record of materials for each school. As batches clear the clean post edit, clipped images are prepared and distributed for scoring. The Oklahoma Call Center Team notifies the SDE regarding unresolved material discrepancies within 24 hours after Pearson's initial attempt to contact the school principal. Within one week after materials are returned, Pearson's Service Center Team also notifies the SDE of any missing or incomplete shipments from schools that received testing materials.

Resolution of missing secure test materials and used answer booklets. Pearson provides updates on a daily basis to the initial discrepancy reports, in response to SDE specifications and requests. The Oklahoma Call Center team makes every attempt to resolve all discrepancies involving secure test books and used answer booklets in a timely manner. Using daily, updated discrepancy reports, Pearson is in constant contact with the respective districts/schools. Pearson and the SDE work out details on specific approaches to resolution of material return discrepancies, and what steps will be taken if "lost" secure test books and/or used answer documents are not found and remain unreturned to Pearson.

2.4 Processing Assessment Materials Returned by Schools

Pearson's receipt system provides for the logging of materials within 24 hours of receipt and the readiness of materials for scanning within 72 hours of receipt. District status is available from a web-based system accessible by SDE. In addition, the Oklahoma Call Center is able to provide receipt status information if required. The receipt notification website's database is updated daily to allow for accurate information being presented to inquiring district/school personnel. As with initial shipping, the secure and accurate receipt of test materials is a priority with Pearson. Quality assurance procedures provide that all materials are checked in using pre-defined procedures. Materials are handled in a highly secure manner from the time of receipt until final storage and shredding. The receipt of all secure materials is verified through the scanning of barcodes and the comparison of this data to that in security files established during the initial shipment of Oklahoma test materials to the district assessment coordinators.

Section 3

Classical Item Analysis and Results

3.1 Sampling Plan and Field Test Design

3.1.a Sampling Plan

Population data were used for classical and item response theory (IRT) analyses for all Winter/Trimester 2009-10 and Spring 2010 tests. All students who complete a course with an End-of-Instruction test associated with it must also take the test.

3.1.b Field Test Design

New items are field-tested to build up the item bank for future high stakes administrations. The overall field test design used by Pearson was an embedded field test design where newly-developed field test items were embedded throughout the test. The advantage of an embedded field test design is that test-takers do not know where the field test items are located and therefore will treat each item as a scored item. Twenty multiple choice field test items per form (Winter/Trimester 2009 and Spring 2010) and one open-ended field test item per form (English II and English III; Spring 2010 only) were placed in common positions across forms and administrations. Field test items were prioritized for inclusion on forms based on current item bank analyses.

3.1.c Data Receipt Activities

After all tests were scored, a data file was provided for item analyses and calibration. A data clean-up process that removed invalid cases, ineligible responses, absent students, and second time test takers was completed. A statistical key check was also performed at this time. This 'cleaned' sample was used for classical item analyses, calibration, and equating. Upon receipt of data, a research scientist inspected several data fields to determine if the data met expectations, including:

- Student ID
- Demographic fields
- Form identification fields
- Raw response fields
- Scored response fields
- Total score and subscore fields
- Fields used to implement exclusion from analysis rules

Exclusion Rules. Following data inspection and clean-up, exclusionary rules were applied to form the final sample that was used for classical item analyses, calibration, and equating. Any student who had attempted at least five responses was included in the data analyses. The demographic breakdown of the students in the Winter/Trimester 2009-10 and Spring 2010 item analysis and calibration sample appear in Table 3.1 and Table 3.2, respectively.

Table 3.1. Demographic Characteristics of Calibration and Equating Sample for Winter/Trimester 2009-10

Subject	Total	Male	Female	African American	Native American	Hispanic	Asian	Pacific Islander	White	Other
Algebra I	1,692	835	857	335	262	173	42	4	869	7
Algebra II	2,038	1,020	1,018	427	224	148	61	1	1,169	8
Biology I	2,379	1,181	1,198	509	351	170	60	3	1,273	13
English II	2,590	1,301	1,289	399	370	181	62	4	1,560	14
English III	2,766	1,367	1,399	414	415	227	76	4	1,617	13
Geometry	2,127	1,014	1,113	369	278	176	47	7	1,235	15
U.S. History	2,032	1,001	1,031	340	331	169	35	4	1,143	10

Note: Gender and Ethnicity values may not add to the total due to missing responses.

Table 3.2. Demographic Characteristics of Calibration and Equating Sample for Spring 2010

Subject	Total	Male	Female	African American	Native American	Hispanic	Asian	Pacific Islander	White	Other
Algebra I	23,791	11,621	12,170	2,132	5,015	2,434	498	91	13,521	100
Algebra II	24,042	11,697	12,345	2,134	4,588	1,803	587	80	14,788	62
Biology I	29,828	14,747	15,081	2,529	5,816	2,881	582	112	17,815	93
English II	35,463	17,532	17,928	3,457	6,748	3,324	834	96	20,885	119
English III	35,556	17,757	17,799	3,456	6,681	3,102	761	102	21,329	125
Geometry	28,425	14,109	14,316	2,743	5,385	2,446	559	109	17,095	88
U.S. History	29,328	14,379	14,949	2,595	5,422	2,656	721	95	17,754	85

Note: Gender and Ethnicity values may not add to the total due to missing responses.

Statistical Key Check. Administering items that have only one correct key and are correctly scored is critical for accurate assessment of student performance. To screen for potentially problematic items, a statistical key check was conducted and items were flagged that met any of the following criteria:

- Less than 200 students responded to the item
- Correct response *p*-value less than 0.20
- Correct response uncorrected point-biserial correlation less than 0.20
- Distractor *p*-value greater than or equal to 0.40
- Distractor point-biserial correlation greater than or equal to 0.05

Any flagged operational items are submitted for key review by the appropriate Pearson content specialist. Any flagged items that are identified by content experts as having key issues are submitted to SDE for review before dropping the item from the operational scoring. There were no items identified in the Winter/Trimester 2009-10 or Spring 2009 administrations as having a key issue. Once the keys were verified, classical item analyses were conducted.

3.2 Classical Item Analyses

Following completion of the data receipt activities and statistical key check, the following classical item analyses were conducted for operational and field test items:

- Total case count
- Summary demographic statistics (e.g., males, females, African American, White, Hispanic, Asian, Pacific Islander, Native American, and Other)
- Frequency distributions for all multiple choice items and frequency distributions of score ratings and condition codes for writing prompts
    - Percentage of students in different multiple choice categories and, for the writing prompt, in different score categories (overall and broken down by gender and ethnicity)
- Item *p*-value
    - Mean item *p*-value
- Item-test point-biserial correlation
    - Mean item-test point-biserial correlation
    - Point-biserial correlation by response option (overall and broken down by gender and ethnicity)
- Omit percentage per item
    - Not reached analysis results per item
- Mean score by response option (overall and broken down by gender and ethnicity)

Once the keys were verified and the item analysis results reviewed, the data were used for calibration and equating.

3.2.a Test-Level Summaries of Classical Item Analyses

The test-level raw score descriptive statistics for the calibration samples are shown in Table 3.3. Note that students whose tests were invalidated and those students taking the test for a second time were excluded. The operational test results indicate that the omit rates were smaller than 1% for all subjects. The mean raw score and the mean percent of the maximum raw scores were relatively similar for both administrations. As indicated in the test

configuration section, there were multiple forms with a duplicate set of operational items and a unique set of field test items in the Winter/Trimester 2009-10 and Spring 2010 tests. A separate item analysis by test form indicated that, in both administrations, the omit rates were below 1% for all content areas. The mean percent of the maximum possible raw score across forms indicates that the forms were relatively similar in difficulty for all content areas.

Table 3.3. Test-Level Summaries of Classical Item Analyses for Winter/Trimester 2009-10 and Spring 2010

Subject and Administration	Sample Size	Mean	Mean % of Max	Items / Points	Mean *p*	Mean r_{pb}	Omit Min	Omit Max
Algebra I-W09	1,692	30.04	0.55	55	0.55	0.37	0.00	0.35
Algebra I-S10 CA	16,239	33.34	0.61	55	0.61	0.40	0.01	0.11
Algebra I-S10 CB	13,510	33.19	0.60	55	0.60	0.41	0.01	0.06
Algebra II-W09	2,038	32.57	0.59	55	0.59	0.40	0.00	0.20
Algebra II-S10 CA	15,456	31.66	0.58	55	0.58	0.45	0.01	0.08
Algebra II-S10 CB	12,519	31.19	0.57	55	0.57	0.43	0.01	0.09
Biology I-W09	2,379	40.16	0.67	60	0.67	0.39	0.00	0.34
Biology I-S10 CA	19,155	39.35	0.66	60	0.66	0.37	0.01	0.07
Biology I-S10 CB	15,933	40.57	0.68	60	0.68	0.37	0.01	0.06
English II-W09	2,590	46.90	0.71	61/66	0.72	0.36	0.00	0.62
English II-S10 CA	17,614	49.84	0.76	61/66	0.77	0.37	0.01	0.12
English II-S10 CB	17,849	47.73	0.72	61/66	0.73	0.36	0.01	0.13
English III-W09	2,766	44.64	0.62	63/72	0.63	0.40	0.00	0.29
English III-S10 CA	17,911	46.45	0.65	63/72	0.65	0.40	0.02	0.20
English III-S10 CB	17,645	45.94	0.64	63/72	0.65	0.39	0.01	0.18
Geometry-W09	2,127	35.74	0.65	55	0.65	0.41	0.00	0.19
Geometry-S10 CA	17,847	35.76	0.65	55	0.65	0.44	0.02	0.22
Geometry-S10 CA	14,846	36.07	0.66	55	0.66	0.44	0.02	0.16
U.S. History-W09	2,032	38.15	0.64	60	0.64	0.38	0.00	0.15
U.S. History-S10 CA	17,606	40.25	0.67	60	0.67	0.41	0.01	0.05
U.S. History-S10 CB	14,644	39.84	0.66	60	0.66	0.40	0.01	0.06

Note: W09 = Winter/Trimester 2009-10; S10 CA = Spring 2010 Core A; S10 CB = Spring 2010 Core B; r_{pb} = point biserial correlation.

3.3 Procedures for Detecting Item Bias

One of the goals of the OSTP-ACE EOI assessments is to assemble a set of items that provides a measure of a student's ability that is as fair and accurate as possible for all subgroups within the population. Differential item functioning (DIF) analysis refers to statistical procedures that assess whether items are differentially difficult for different groups of examinees. DIF procedures typically control for overall between-group differences on a criterion, usually total test scores. Between-group performance on each item is then compared within sets of examinees having the same total test scores. 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 or statistical Type I error. As a result, DIF statistics are used only to identify potential sources of item bias. Subsequent review by content experts and bias committees are required to determine the source and meaning of performance differences.

For the OSTP-ACE EOI test DIF analyses, DIF statistics were estimated for all major subgroups of students with sufficient sample size: African American, Hispanic, Asian, Native American, and Female. Field test items with statistically-significant differences in performance were flagged so that items could be carefully examined for possible biased or unfair content that was undetected in earlier fairness and bias content review meetings held prior to form construction.

Pearson used the Mantel-Haenszel (MH) chi-square approach for detecting DIF in multiple choice and open-ended items. Pearson calculated the Mantel-Haenszel statistic (MH D-DIF; Holland & Thayer 1988) to measure the degree and magnitude of DIF. The student group of interest is the *focal* group, and the group to which performance on the item is being compared is the *reference* group. The reference groups for these DIF analyses were White for race and male for gender. The focal groups were females and minority race groups.

Items were separated into one of three categories on the basis of DIF statistics (Holland and Thayer 1988; Dorans and Holland 1993): negligible DIF (category A), intermediate DIF (category B), and large DIF (category C). The items in category C, which exhibit significant DIF, are of primary concern. The item classifications are based on the Mantel-Haenszel chi-square and the MH delta (Δ) value. Positive values of delta indicate that the item is easier for the focal group, and a negative value of delta indicates that the item is more difficult for the focal group. The item classifications are made as follows (Michaelides, 2008):

- The item is classified as C category if the MH D-DIF is significantly greater than 1.0 in absolute value, and its absolute value is at least 1.5.
- The item is classified as B category if the MH D-DIF is significantly different from zero, its absolute value is at least 1.0, and its absolute value is either less than 1.5 or not significantly greater than 1.0.
- The item is classified as A category if the MH D-DIF is not significantly different from zero ($p \geq 0.05$), or if its absolute value is less than 1.0.

3.3.a Differential Item Functioning Results

The data in Table 3.4 summarizes the number of items in DIF categories for the seven subjects for the Winter/Trimester 2009-10 and Spring 2010 administrations. The results presented in this table are for field test items only. Items flagged for DIF were placed before expert content specialist committees during the Spring 2010 field test data review as described in the Section 3.4. Field test items that exhibit bias as a result of the content of the item were removed from the item bank excluding them from future use.

Table 3.4. DIF Flag Incidence Across All OSTP-ACE EOI Field Test Items for Winter/Trimester 2009-10 and Spring 2010

Subject	Total FT Items	Native American	Asian	African American	Hispanic	Female
Winter 2009-10						
Algebra I	20	0	0	6	1	4
Algebra II	20	0	1	0	3	0
Geometry	20	0	0	0	0	0
Biology I	20	0	0	1	4	1
English II	20	1	1	0	2	3
English III	20	0	3	3	1	2
U.S. History	20	0	0	3	4	2
Spring 2010						
Algebra I	220	0	3	17	6	10
Algebra II	220	1	4	12	10	7
Geometry	220	2	1	13	13	16
Biology I	220	0	1	9	8	14
English II	320	6	9	31	47	30
English III	320	1	3	28	24	24
U.S. History	219	0	6	18	8	18

3.4 Data Review

Data review represents a critical step in the test development cycle. At the data review meeting, SDE and Pearson staff had the opportunity to review actual student performance on the newly-developed and field-tested multiple choice items across the seven subjects based on the Winter/Trimester 2009-10 and Spring 2010 field test administrations. The data review focused on the content validity, curricular alignment, and statistical functioning of field-tested items prior to selection for operational test forms. The field test results used in the data review provided evidence that the items were designed to yield valid results and were accessible for use by the widest possible range of students. The review of student performance should provide evidence regarding the fulfillment of requirement 200.2(b)(2)of NCLB. The purpose of the review meeting was to ensure that psychometrically-sound, fair, and aligned items are used in the construction of the ACE EOI assessments and entered into the respective item banks. Pearson provided technical and psychometric expertise to provide a clear explanation about the content of the items, the field test process, the scoring process, and the resulting field test data to ensure the success of these meetings and the defensibility of the program.

Data review meetings were a collaborative effort between SDE and Pearson. SDE administrators and content specialists attended the meeting facilitated by Pearson content specialists and research scientists who trained the SDE staff on how to interpret and review the field test data. Meeting materials included a document explaining the flagging criteria, a document containing flagged items, and the item images. Pearson discussed with SDE the analyses performed and the criteria for flagging the items. Flagged items were then reviewed and decisions were made as to whether to accept the item, accept the item with revisions, or reject the item. Review of the data included presentation of *p*-value, point-biserial correlation, point-biserial correlation by response option, response distributions, mean overall score by response option, and indications of item DIF and IRT misfit. Items failing to

meet the requirements of sound technical data were carefully considered for rejection by the review panel, thereby enhancing the reliability and improving the validity of the items left in the bank for future use. While the panel used the data as a tool to inform their judgments, the panel (and not the data alone) made the final determination as to the appropriateness or fairness of the assessment items. The flagging criteria for the ACE EOI assessments are as follows:

- *p*-value < .25 or > .90
- point-biserial correlation < .15
- distractor point-biserial correlation > .05
- differential item functioning (DIF): test item biases for subgroups
- IRT misfit as flagged by the Q1 index (see section 4.2)

Bias Review. One aspect of the data review meetings was to assess potential bias based on DIF results and item content. Although bias in the items had been avoided through writer training and review processes, there is always the potential for bias to be detected through statistical analysis. It is important to include this step in the development cycle because SDE and Pearson wish to avoid inclusion of an item that is biased in some way against a group, because the item may lead to inequitable test results. As described earlier, all field test items were analyzed statistically for DIF using the field test data. A Pearson research scientist explained the meaning, in terms of level, and the direction of the DIF flags. The data review panel reviewed the item content, the percentage of students selecting each response option, and the point-biserial correlation for each response option by gender and ethnicity for all items flagged for DIF. The data review panel was then asked if there was context (for example, cultural barriers) or language in an item that might result in bias (i.e., an explanation for the existence of the statistical DIF flag).

3.4.a Results of Data Review

The number of items inspected during data review as a result of the item meeting the statistical flagging criteria for the classical item analyses, DIF, and IRT procedures is presented in Table 3.5.

Table 3.5. Number of Items Per Subject Flagged and Rejected During Winter/Trimester 2009–2010 and Spring 2010 Field Test Data Review

Subject	No. of FT Items	No. Flagged	Rejected	Accepted	Accepted with Edits
Winter 2009-10					
Algebra I	20	14	1	6	13
Algebra II	20	8	0	12	8
Geometry	20	6	0	14	6
Biology I	20	11	2	9	9
English II	20	11	1	9	10
English III	20	8	1	12	7
U.S. History	20	10	1	10	9
Spring 2010					
Algebra I	220	74	8	146	66
Algebra II	220	84	5	136	79
Geometry	220	90	6	130	84
Biology I	220	60	10	160	50
English II	320	162	45	158	117
English III	320	111	41	209	70
U.S. History	219	94	10	125	84

3.5 Test Reliability

The reliability of a test provides an estimate of the extent to which an assessment will yield the same results when administered in different times, locations, or samples, when the two administrations do not differ in relevant variables. The reliability coefficient is an index of consistency of test results. Reliability coefficients are usually forms of correlation coefficients and must be interpreted within the context and design of the assessment and of the reliability study. Cronbach's alpha is a commonly-used internal consistency measure, which is derived from analysis of the consistency of the performance of individuals on items in a test administration. Cronbach's alpha is calculated as shown in equation (1). In this formula, s_i^2 denotes the estimated variance for each item, with items indexed i = 1, 2, … k, and s^2_{sum} denotes the variance for the sum of all k items:

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{k} s_i^2}{s_{sum}^2}\right). \tag{1}$$

Cronbach's alpha was estimated for each of the content areas for the operational portion of the test.

Table 3.6 presents Cronbach's alpha for the operational tests by subject area for the Winter/Trimester 2009-10 and Spring 2010 ACE EOI administrations. These reliability coefficients indicate that the OSTP-ACE EOI assessments had strong internal consistency and that the tests produce relatively stable scores.

Table 3.6. Cronbach's Alpha for Winter/Trimester 2009-10 and Spring 2010 Administrations by Subject

Subject	Administration and Form	Alpha
Algebra I	Winter 2009-10	0.89
	Spring 2010 – Core A	0.91
	Spring 2010 – Core B	0.91
Algebra II	Winter 2009-10	0.90
	Spring 2010 – Core A	0.92
	Spring 2010 – Core B	0.92
Biology I	Winter 2009-10	0.90
	Spring 2010 – Core A	0.89
	Spring 2010 – Core B	0.89
English II	Winter 2009-10	0.88
	Spring 2010 – Core A	0.89
	Spring 2010 – Core B	0.88
English III	Winter 2009-10	0.91
	Spring 2010 – Core A	0.92
	Spring 2010 – Core B	0.90
Geometry	Winter 2009-10	0.91
	Spring 2010 – Core A	0.92
	Spring 2010 – Core B	0.92
U.S. History	Winter 2009-10	0.90
	Spring 2010 – Core A	0.92
	Spring 2010 – Core B	0.91

3.6 Test Reliability by Subgroup

Table 3.7 addresses the reliability analysis results by the different reporting subgroups for the OSTP-ACE EOI assessments for Spring 2010 for each core form. Table 3.7 illustrates the subject, the subgroups, the number of students used in the analyses and the associated Cronbach's Alpha for each subject and subgroup. In all instances, the reliability coefficients are well above the accepted lower limit of .70.

Table 3.7. Test Reliability by Subgroup for Spring 2010

Subject	Core	Male	Female	African-American	Native American	Hispanic	Asian	White
Algebra I	A	0.91	0.90	0.89	0.90	0.89	0.93	0.90
	B	0.91	0.90	0.88	0.89	0.89	0.92	0.91
Algebra II	A	0.93	0.92	0.89	0.91	0.91	0.94	0.93
	B	0.92	0.91	0.89	0.90	0.91	0.93	0.91
Biology I	A	0.90	0.88	0.87	0.88	0.89	0.91	0.88
	B	0.90	0.89	0.88	0.89	0.89	0.91	0.88
English II	A	0.89	0.88	0.87	0.88	0.89	0.91	0.88
	B	0.90	0.89	0.88	0.89	0.89	0.91	0.88
English III	A	0.90	0.88	0.87	0.88	0.89	0.91	0.88
	B	0.90	0.89	0.88	0.89	0.89	0.91	0.88
Geometry	A	0.93	0.92	0.91	0.91	0.91	0.94	0.92
	B	0.92	0.92	0.91	0.91	0.91	0.94	0.92
U.S. History	A	0.92	0.91	0.91	0.91	0.91	0.92	0.91
	B	0.92	0.90	0.90	0.90	0.91	0.91	0.90

Table 3.7. Test Reliability by Subgroup for Spring 2010 (cont.)

Subject	Core	English Language Learner	Individual Education Plan	Economically Disadvantaged
Algebra I	A	0.88	0.89	0.89
	B	0.88	0.89	0.89
Algebra II	A	0.92	0.88	0.91
	B	0.89	0.88	0.90
Biology I	A	0.82	0.88	0.88
	B	0.86	0.90	0.89
English II	A	0.82	0.88	0.88
	B	0.86	0.90	0.89
English III	A	0.82	0.88	0.88
	B	0.86	0.90	0.89
Geometry	A	0.91	0.89	0.91
	B	0.89	0.90	0.91
U.S. History	A	0.87	0.92	0.91
	B	0.86	0.91	0.90

3.7 Inter-rater Reliability

Inter-rater reliability is referred to as the degree of agreement among scorers that allows for the scores to be interpreted as reasonably intended by the test developer (AERA, APA and NCME, 1999). The Winter/Trimester 2009-10 English II and English III tests contained one operational writing prompt each and the Spring 2010 tests contained one writing prompt per core form. Raters were trained to implement the scoring rubrics, anchor papers, check sets, and resolution reading. The items were analytically scored by two raters on five strands in both administrations. The final writing score for a student in a given strand is the average of the two scores. The inter-rater reliability coefficients for the operational prompt are presented in Table 3.8 for English II and Table 3.9 for English III. The results show that exact and adjacent rater agreement on trait scores for both the Winter/Trimester 2009-10 and

Spring 2010 operational writing prompts were reasonably high. The weighted Kappa statistic (Kraemer, 1982) is an indication of inter-rater reliability after correcting for chance. The Kappa values for the OSTP-ACE EOI Winter/Trimester 2009-10 and Spring 2010 operational writing prompts are within the moderate range.

Table 3.8. Inter-rater Reliability for English II Operational Writing Prompts for Winter/Trimester 2009-10 and Spring 2010

			Point Discrepancy Percentages							Agreement Percentages			
Trait	Max Points	Valid N	-3	-2	-1	0	1	2	3	Exact	Adjacent	+/- 2 or more	Kappa
Winter/Trimester 2009-10													
1	4	2,345	0.00	0.26	15.82	67.29	16.12	0.51	0.00	67.29	31.94	0.77	0.38
2	4	2,345	0.00	0.34	17.53	64.90	16.55	0.68	0.00	64.90	34.08	1.02	0.37
3	4	2,345	0.00	0.26	15.65	68.14	15.82	0.13	0.00	68.14	31.47	0.39	0.38
4	4	2,345	0.00	0.26	15.57	65.84	17.87	0.47	0.00	65.84	33.44	0.73	0.38
5	4	2,345	0.00	0.43	16.33	65.12	17.87	0.26	0.00	65.12	34.20	0.69	0.37
Spring 2010 Core Form A													
1	4	17,356	0.02	0.47	16.69	65.64	16.70	0.46	0.01	65.64	33.39	0.96	0.28
2	4	17,356	0.01	0.53	16.90	65.30	16.79	0.46	0.01	65.30	33.69	1.01	0.28
3	4	17,356	0.00	0.28	16.44	66.71	16.28	0.29	0.00	66.71	32.72	0.57	0.29
4	4	17,356	0.01	0.50	16.51	66.16	16.50	0.32	0.00	66.16	33.01	0.83	0.31
5	4	17,356	0.01	0.58	17.67	63.44	17.82	0.49	0.00	63.44	35.49	1.08	0.28
Spring 2010 Core Form B													
1	4	17,571	0.01	0.35	16.41	66.39	16.48	0.37	0.01	66.39	32.89	0.74	0.29
2	4	17,571	0.01	0.36	17.09	65.19	16.86	0.49	0.00	65.19	33.95	0.86	0.28
3	4	17,571	0.00	0.33	16.89	66.07	16.35	0.36	0.00	66.07	33.24	0.69	0.29
4	4	17,571	0.01	0.40	16.64	66.06	16.45	0.44	0.01	66.06	33.09	0.86	0.32
5	4	17,571	0.00	0.59	18.31	62.73	17.82	0.53	0.02	62.73	36.13	1.14	0.27

Table 3.9. Inter-rater Reliability for English III Operational Writing Prompts for Winter/Trimester 2009-10 and Spring 2010

			Point Discrepancy Percentages							Agreement Percentages			
Trait	Max Points	Valid N	-3	-2	-1	0	1	2	3	Exact	Adjacent	+/- 2 or more	Kappa
Winter/Trimester 2009-10													
1	4	2,433	0.04	1.19	18.82	60.09	19.36	0.49	0.00	60.09	38.18	1.72	0.36
2	4	2,433	0.04	1.07	20.96	58.28	18.70	0.95	0.00	58.28	39.66	2.06	0.32
3	4	2,433		0.70	18.91	61.98	17.80	0.62	0.00	61.98	36.71	1.32	0.32
4	4	2,433	0.08	1.07	20.10	57.30	20.39	1.07	0.00	57.30	40.49	2.22	0.30
5	4	2,433	0.04	1.44	20.96	53.93	21.99	1.64	0.00	53.93	42.95	3.12	0.29
Spring 2010 Core Form A													
1	4	17,533	0.01	0.54	17.16	64.44	17.33	0.52	0.00	64.44	34.49	1.07	0.33
2	4	17,533	0.02	0.70	17.70	62.80	18.11	0.67	0.01	62.80	35.81	1.40	0.32
3	4	17,533	0.01	0.48	17.50	64.35	17.21	0.44	0.01	64.35	34.71	0.94	0.32
4	4	17,533	0.02	0.83	18.14	62.16	18.09	0.76	0.00	62.16	36.23	1.61	0.33
5	4	17,533	0.01	0.95	19.83	58.86	19.50	0.85	0.00	58.86	39.33	1.81	0.30
Spring 2010 Core Form B													
1	4	17,270	0.00	0.6	16.71	64.77	17.31	0.61	0.01	64.77	34.02	1.22	0.32
2	4	17,270	0.00	0.64	17.63	63.10	17.94	0.68	0.01	63.10	35.57	1.33	0.32
3	4	17,270	0.00	0.50	16.75	64.80	17.42	0.53	0.00	64.80	34.17	1.03	0.32
4	4	17,270	0.00	0.73	17.73	62.54	18.22	0.78	0.00	62.54	35.95	1.51	0.33
5	4	17,270	0.01	0.89	19.11	59.64	19.36	0.98	0.01	59.64	38.47	1.89	0.30

Section 4

Calibration, Equating, and Scaling

4.1 Item Response Theory (IRT) models

Dichotomous Item Response Theory Model. The three-parameter logistic (3-PL) item response theory (IRT) model (Lord & Novick, 1968) was used for calibrating the dichotomously-scored multiple choice items. In the 3-PL model (Lord, 1980), the probability that a student with an ability level of θ responds correctly to item *i* is

$$P_i(\theta) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i(\theta - b_i)}}, \tag{2}$$

where a_i is the item discrimination parameter, b_i is the item difficulty parameter, c_i is the lower asymptote parameter, and *D* is a scaling constant, which is traditional equal to 1.7. With multiple-choice items it is assumed that, due to guessing, examinees with very low ability levels have a probability greater than zero of responding correctly to an item. This probability is represented in the 3-PL model by the c_i parameter.

Polytomous Item Response Theory Model. For calibrating the polytomously-scored constructed response or open-ended (OE) writing prompt items, the Generalized Partial Credit (GPC; Muraki, 1997) model was used. In the GPC model, the probability that a student with ability level θ will have a score in the kth category of the ith item is

$$P_{ik}(\theta) = \frac{\exp\left[\sum_{v=1}^{k} Da_i(\theta - b_{iv})\right]}{\sum_{c=1}^{m_i} \exp\left[\sum_{v=1}^{c} Da_i(\theta - b_{iv})\right]}, \tag{3}$$

where m_i is the total score levels for item *i* for *k* = *v* category responses, a_i is the slope parameter (or Da_i), and b_{iv} is the category intersection parameters (or (b_i – d_{iv}) where b_i is location/difficulty and d_{iv} is the threshold parameters representing category boundaries relative to the item location parameter).

The IRT models were implemented using MULTILOG 7.0 (Thissen, Chen, & Bock, 2003). MULTILOG estimates parameters simultaneously for dichotomous and polytomous items via marginal maximum likelihood procedures and implements the GPC model with the appropriate parameter coding. All item and student ability calibrations were independently conducted and verified by at least two Pearson research scientists.

4.2 Assessment of IRT Fit to the model

Item fit was assessed using the Yen's (1981, 1984) Q_1 item fit index, which approximately follows a χ^2 distribution:

$$Q_{1i} = \sum_{r=1}^{10} \frac{N_r (O_{ir} - E_{ir})^2}{E_{ir}(1 - E_{ir})}, \tag{4}$$

where Q_{1i} is the fit of the ith item, N_r is the number of examinees per cell, O_{ir} is the observed proportion of examinees in cell r that correctly answered item i, and E_{ir} is the expected portion of examinees in cell r that correctly answered item i. The expected proportions are computed using ability- and item parameter estimates in Equations (2) and (3) and summing over examinees in cell r:

$$E_{ir} = \frac{1}{N_{ir}} \sum_{k \varepsilon r}^{N_{ir}} P_i(\hat{\theta}_k) \ . \tag{5}$$

Because chi-square statistics are affected by sample size and associated degrees of freedom, the following standardization of the Q_1 statistic was used:

$$Z_j = \frac{Q_{1i} - df}{\sqrt{(2df)}} \ . \tag{6}$$

The Z-statistic is an index of the degree to which observed proportions of item scores are similar to the proportions that would be expected, given the estimated ability- and item parameters. Large differences between expected and observed item performance may indicate poor item fit. To assess item fit, a critical Z-value is determined. Items with Z-values that are larger than this critical Z-value have poor item fit. The item characteristic curves, classical item statistics, and item content were reviewed for items flagged by Q_1. An internally-developed software program, Q1Static, was used to compute the Q_1 item fit index.

Operational items flagged by Q_1 that were not flagged by the classical item statistics and had reasonable IRT parameter estimates were not reviewed further. If any operational items were also flagged by classical item statistics and/or had poor IRT parameter estimates (e.g., low *a* parameter), the items were reviewed by Pearson content specialists. Any item that was potentially mis-keyed was presented to SDE to make a decision regarding whether to keep or remove the item. No such incidences occurred for operational items administered in Winter/Trimester 2009-10 or Spring 2010.

4.2.a Calibration and IRT Fit Results

4.2.a.i Winter/Trimester 2009-10

Algebra I. For the Winter/Trimester 2009-10 Algebra I assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. Six Algebra I items were flagged for further review based on their fit statistics.

Algebra II. For the Winter/Trimester 2009-10 Algebra II assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. Five Algebra II items were flagged for further review based on their fit statistics.

Biology I. For the Winter/Trimester 2009-10 Biology I assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. Six Biology I items were flagged for further review based on their fit statistics.

English II. For the Winter/Trimester 2009-10 English II assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. Nine English II items were flagged for further review based on their fit statistics.

English III. For the Winter/Trimester 2009-10 English III assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. No English III items were flagged for further review based on their fit statistics.

Geometry. For the Winter/Trimester 2009-10 Geometry assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. Eleven Geometry items were flagged for further review based on their fit statistics.

U.S. History. For the Winter/Trimester 2009-10 U.S. History assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. Seven U.S. History items were flagged for further review based on their fit statistics.

For each item that was flagged based on its model fit indices, a careful review of both CTT and IRT item statistics was conducted to determine whether the item should be dropped from calibration, scaling, equating, or scoring. No items were dropped from any of the Winter/Trimester 2009-10 ACE EOI assessments for calibration, equating, or scoring as a result of the Q_1 results.

4.2.a.ii Spring 2010

Algebra I. For the Spring 2010 Algebra I assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. One Algebra I item was flagged for further review based on its fit statistics.

Algebra II. For the Spring 2010 Algebra II assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. No Algebra II items were flagged for further review based on their fit statistics.

Biology I. For the Spring 2010 Biology I assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. No Biology I items were flagged for further review based on their fit statistics.

English II. For the Spring 2010 English II assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. No English II items were flagged for further review based on their fit statistics.

English III. For the Spring 2010 English III assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. Two English III items were flagged for further review based on their fit statistics.

Geometry. For the Spring 2010 Geometry assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. One Geometry item was flagged for further review based on its fit statistics.

U.S. History. For the Spring 2010 U.S. History assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. No U.S. History items were flagged for further review based on their fit statistics.

For each item that was flagged based on its model fit indices, a careful review of both CTT and IRT item statistics was conducted to determine whether the item should be dropped from calibration, scaling, equating, or scoring. No items were dropped from any of the Spring 2010 ACE EOI assessments for calibration, equating, or scoring as a result of the Q_1 results.

Field Test Items. The field test items across all subjects were evaluated using the Q_1 statistic to evaluate the extent to which the obtained proportions of item scores are close to the proportions that would be expected based on the estimated thetas and item parameters. Any field test items flagged by Q_1 were included in the data review for review by contest specialists from Pearson and SDE (for more on data review, please see Section 3.4).

4.3 Calibration and Equating

The 3-PL model was used for calibration of Algebra I, Algebra II, Geometry, Biology I, and U.S. History because all of these tests consist of only multiple choice items. Because English II and English III have multiple choice and constructed response items, a simultaneous calibration with the 3-PL and GPC models was implemented.

A common item, non-equivalent groups (CINEG) design was used for all content areas to link the current test forms (i.e., Winter/Trimester 2009-10 and Spring 2010) to the base scale. Typically, for the CINEG design common, or anchor, items are selected to be representative of the test content in terms of difficulty and the test blueprint. For the Winter 2009 and Spring 2010 tests, all operational items were used as common or anchor items to link to he base year scale. The Stocking and Lord (1983) procedure, which estimates the equating transformation constants by minimizing the distance between the test characteristic curves of the common items, was used to equate the tests to the base year.

Equating was conducted employing using freely-available software, STUIRT (Kim & Kolen, 2004). Prior to conducting the equating, anchor item stability checks were performed to eliminate the impact of item drift on equating.

4.3.a Common Linking Items for Spring 2010

Table 4.1 presents the number and percentage of common linking items by subject for the Spring 2010 administration. The common linking items were necessary as a result of two core operational forms for the Spring 2010 administration. The common linking items were used for simultaneous calibration during the IRT item parameter estimation to keep the items on the same scale. The common linking set was comprised of approximately 20 items or greater than 25% of all operational items, and counts vary by subject. In addition, the common linking set was proportionally representative of the total test in terms of content assessed and mimicked the difficulty of the overall test as well.

Table 4.1. Number of Common Linking Items Per Subject for Spring 2010

Subject	Number of Items on Test	Number of Linking Items	Percent of Test
Algebra I	55	19	35%
Algebra II	55	21	38%
Biology I	60	20	33%
English II	61	20	33%
English III	63	20	32%
Geometry	55	20	36%
U.S. History	60	19	32%

4.4 Item Stability Evaluation Methods

Despite the careful selection and placement of the operational items, it is possible for these items to perform differentially across administrations. Dramatic changes in item parameter values can result in systematic errors in equating results (Kolen & Brennan, 2004). As a result, prior to finalizing the equating constants, Pearson evaluated changes in the item parameters from the item bank to the Winter 2009 and Spring 2010 administration. The process used in this evaluation is called an item parameter stability check.

The item parameter stability check that Pearson performed is an iterative approach, which uses a method that is similar to the one used to check for differential item functioning. This method is called the d^2 procedure. The steps taken were as follows:

1) Use a theoretically-weighted posterior θ distribution, $g(\theta_k)$, with 40 quadrature points.
2) Place the current linking item parameters on the baseline scale by computing Stocking & Lord (SL) constants using STUIRT and all (k) linking items.
3) Apply the SL linking constants to the current item parameters and compute the current raw score to scale score table. The results based on all k linking items will comprise the original table.
4) For each linking item, calculate the weighted sum of the squared deviation (d^2) between the item characteristic curves.
   a) Apply the SL constants to the estimated ability levels ($\hat{\theta}$) associated with the standard normal θ distribution used to generate the SL constants.
   b) For each anchor item, calculate a weighted sum of the squared deviations between the ICCs (d^2) based on the old (x) and new (y) parameter estimates at each point in the θ distribution multiplied by the theoretically-weighted distribution.

$$d_i^{\,2} = \sum^{k} \left[P_{ix}(\theta_k) - P_{iy}(\theta_k) \right]^2 \bullet g(\theta_k) \tag{7}$$

   c) Review and sort the items in descending (largest to smallest) order according to the d^2 estimate.
   d) Step 4c) results in the item with the largest area at the top.
      i) Drop the item with the largest d^2 from the linking set.
      ii) Repeat steps 2) through 4c) until 10 items are dropped computing 11 raw score to scale score tables for comparative purposes.

e) Review the raw score to scale score tables and keep the raw score to scale score table where the raw to scale tables across iterations do not differ at all of the cut score points. The raw score to scale score table before the last iteration becomes the final table.

Before removing any item from the item parameter stability check, the following additional characteristics were examined: 1) prior and current year *p*-values and point-biserial correlations, 2) prior and current year IRT parameter estimates, 3) prior and current year item sequence, 4) standard and objective/skill of the item, 5) impact on blueprint representation, 6) passage ID/title for items linked to a stimulus, and 7) content review of the actual item. Decisions about whether to keep or remove an item were evaluated on a per item basis. If an item (note, only one item can be removed at a time) was removed from the, the process (beginning at the equating step) was be repeated until there were no further items to be removed (the raw score to scale score table has stabilized or the item is judged that it should be included in the equating set; for example, a portion of the blueprint is not represented if the item is removed).

4.4.a Results of the Item Parameter Item Stability Check

Once the anchor set was finalized, the equating constants obtained from the final Stocking and Lord (1983) run were applied to the non-anchor operational items for computation of raw score to scale score tables. For Winter/Trimester 2009-10, three items were removed from Algebra II and Biology I, one item from Algebra I, English II, Geometry, and U.S. History, and zero items from English III. For Spring 2010, there were two anchor items removed from Geometry, one item from English II, and zero items from Algebra I, Algebra II, Biology I, and U.S. History. Any item removed from the item parameter stability check set still contributed to student scores.

4.5 Scaling and Scoring Results

The Lowest Obtainable Scale Score (LOSS), Highest Obtainable Scale Score (HOSS), and final scaling constants for each of the subjects are shown in Table 4.2. The scaling constants, *M*1 (multiplicative) and *M*2 (additive), place the true scores associated with each raw score point onto the reporting or operational scale using a straightforward linear transformation:

$$\text{Scale Score} = \left(\hat{\tau} \times M1\right) + M2 \qquad (8)$$

where, $\hat{\tau}$ = true score.

The raw score to number-correct scales scores were generated from equated parameter estimates using a freely-available software program, POLYEQUATE (Kolen, 2004). Each scale score on the assessment is associated with a performance level that describes the types of behavior, knowledge, and skill a student in this score level is likely to be able to do. For the ACE EOI assessments, there are three cut scores that divide scores into four performance levels: Unsatisfactory, Limited Knowledge, Proficient, and Advanced. The cut scores for each of the tests appear in Table 4.3. In addition, a conditional standard error of measurement (CSEM; please see Section 6.3 for computation of CSEM) was computed for each of the raw score points. The resulting raw score to scale score conversions, CSEMs, as well as the performance levels for Algebra I, Algebra II, Biology I, English II, English III, Geometry, and

U.S. History are shown in Table 4.4 and Table 4.5 for Winter/Trimester 2009-10 and Spring 2010, respectively.

Table 4.2. LOSS, HOSS, and Scaling Constants by Subject

Subject	LOSS	HOSS	*M*1	*M*2
Algebra I	490	999	58.0000	723.8000
Algebra II	440	999	77.1164	692.2381
Biology I	440	999	76.49429	716.76173
English II	440	999	84.80517	734.90335
English III	440	999	74.32896	736.1256
Geometry	440	999	75.51595	721.9844
US History	440	999	77.92698	722.20515

Table 4.3. Performance-Level Cut Scores by Subject

	Cut Scores		
Subject	Limited Knowledge	Proficient	Advanced
Algebra I	662	700	762
Algebra II	651	696	774
Biology I	627	691	775
English II	588	693	797
English III	649	695	795
Geometry	635	695	774
U.S. History	603	689	747

Table 4.4. Raw Score to Scale Score Conversion Tables for Winter/Trimester 2009-10

Raw Score	Algebra I			Biology I			U.S. History			English II		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
0	490	58	1	440	44	1	440	50	1	440	37	1
1	490	58	1	440	44	1	440	50	1	440	37	1
2	490	58	1	440	44	1	440	50	1	440	37	1
3	490	58	1	440	44	1	440	50	1	440	37	1
4	490	58	1	440	44	1	440	50	1	440	37	1
5	490	58	1	440	44	1	440	50	1	440	37	1
6	490	58	1	440	44	1	440	50	1	440	37	1
7	490	58	1	440	44	1	440	50	1	440	37	1
8	490	58	1	440	44	1	440	50	1	440	37	1
9	490	58	1	440	44	1	440	50	1	440	37	1
10	490	58	1	440	44	1	440	50	1	440	37	1
11	490	58	1	440	44	1	440	50	1	440	37	1
12	490	58	1	440	44	1	440	50	1	440	37	1
13	547	61	1	440	44	1	440	50	1	440	37	1
14	584	63	1	456	47	1	440	50	1	440	37	1
15	607	62	1	488	51	1	452	51	1	440	37	1
16	624	58	1	513	54	1	492	56	1	461	40	1
17	637	53	1	533	55	1	520	59	1	484	44	1
18	648	47	1	550	54	1	542	60	1	503	46	1
19	662	41	2	565	51	1	559	58	1	519	47	1
20	666	35	2	578	48	1	575	55	1	534	46	1
21	673	30	2	589	45	1	588	51	1	546	45	1
22	679	26	2	600	41	1	603	47	2	558	43	1
23	685	23	2	610	38	1	611	43	2	569	40	1
24	691	21	2	627	36	2	620	39	2	578	38	1
25	700	19	3	628	33	2	629	36	2	588	35	2
26	701	18	3	637	31	2	638	33	2	596	33	2
27	705	17	3	645	30	2	646	30	2	604	31	2
28	710	16	3	652	28	2	653	28	2	612	30	2

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.4. Raw Score to Scale Score Conversion Tables for Winter/Trimester 2009-10 (cont.)

Raw Score	Algebra I			Biology I			U.S. History			English II		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
29	714	15	3	660	27	2	660	27	2	619	28	2
30	718	15	3	667	26	2	667	25	2	626	27	2
31	722	14	3	674	25	2	673	24	2	633	26	2
32	726	14	3	681	25	2	679	23	2	640	26	2
33	729	13	3	691	24	3	689	22	3	646	25	2
34	733	13	3	694	23	3	691	22	3	653	25	2
35	737	13	3	700	23	3	697	21	3	659	24	2
36	740	13	3	706	22	3	703	21	3	666	24	2
37	744	13	3	713	22	3	708	20	3	672	24	2
38	748	13	3	719	22	3	714	20	3	678	23	2
39	752	13	3	725	21	3	719	20	3	684	23	2
40	756	13	3	731	21	3	725	20	3	693	23	3
41	762	13	4	737	21	3	730	20	3	697	23	3
42	764	13	4	743	21	3	736	20	3	704	23	3
43	768	13	4	750	21	3	747	20	4	710	23	3
44	773	14	4	756	21	3	748	20	4	717	24	3
45	777	14	4	762	21	3	754	21	4	724	24	3
46	783	15	4	775	22	4	760	21	4	731	24	3
47	788	16	4	776	22	4	767	22	4	738	24	3
48	794	18	4	783	23	4	774	23	4	745	25	3
49	801	21	4	791	24	4	782	24	4	753	25	3
50	809	25	4	799	25	4	790	26	4	761	26	3
51	819	32	4	808	27	4	799	28	4	769	27	3
52	833	42	4	818	29	4	810	31	4	778	27	3
53	853	54	4	829	32	4	821	34	4	787	28	3
54	892	63	4	841	36	4	835	39	4	797	30	4
55	999	47	4	856	41	4	851	44	4	808	31	4
56	-	-	-	875	46	4	872	49	4	819	33	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.4. Raw Score to Scale Score Conversion Tables for Winter/Trimester 2009-10 (cont.)

Raw Score	Algebra I			Biology I			U.S. History			English II		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
57	-	-	-	900	50	4	900	52	4	832	34	4
58	-	-	-	937	49	4	944	49	4	846	37	4
59	-	-	-	999	35	4	999	39	4	861	40	4
60	-	-	-	999	35	4	999	39	4	878	43	4
61	-	-	-	-	-	-	-	-	-	899	45	4
62	-	-	-	-	-	-	-	-	-	923	46	4
63	-	-	-	-	-	-	-	-	-	955	42	4
64	-	-	-	-	-	-	-	-	-	999	32	4
65	-	-	-	-	-	-	-	-	-	999	32	4
66	-	-	-	-	-	-	-	-	-	999	32	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.4. Raw Score to Scale Score Conversion Tables for Winter/Trimester 2009-10 (cont.)

Raw Score	Algebra II			Geometry			English III		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
0	440	67	1	440	61	1	440	52	1
1	440	67	1	440	61	1	440	52	1
2	440	67	1	440	61	1	440	52	1
3	440	67	1	440	61	1	440	52	1
4	440	67	1	440	61	1	440	52	1
5	440	67	1	440	61	1	440	52	1
6	440	67	1	440	61	1	440	52	1
7	440	67	1	440	61	1	440	52	1
8	440	67	1	440	61	1	440	52	1
9	440	67	1	440	61	1	440	52	1
10	440	67	1	440	61	1	440	52	1
11	440	67	1	440	61	1	440	52	1
12	440	67	1	440	61	1	440	52	1
13	512	71	1	486	65	1	440	52	1
14	552	74	1	530	69	1	461	54	1
15	577	72	1	559	69	1	504	59	1
16	595	68	1	580	67	1	531	61	1
17	610	62	1	597	62	1	550	60	1
18	623	55	1	612	56	1	566	58	1
19	634	49	1	624	50	1	580	55	1
20	651	43	2	635	44	2	591	51	1
21	652	38	2	645	39	2	602	46	1
22	661	34	2	654	35	2	611	42	1
23	669	31	2	662	32	2	619	38	1
24	676	28	2	670	29	2	627	34	1
25	683	27	2	677	27	2	634	31	1
26	696	25	3	684	25	2	641	29	1
27	697	24	3	695	24	3	649	27	2
28	704	23	3	697	22	3	653	25	2

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.4. Raw Score to Scale Score Conversion Tables for Winter/Trimester 2009-10 (cont.)

Raw Score	Algebra II			Geometry			English III		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
29	710	22	3	703	21	3	659	24	2
30	716	22	3	709	21	3	664	23	2
31	722	21	3	715	20	3	669	22	2
32	728	21	3	720	19	3	674	21	2
33	734	20	3	725	18	3	679	21	2
34	740	20	3	731	18	3	684	20	2
35	746	19	3	736	18	3	689	20	2
36	751	19	3	741	17	3	695	19	3
37	757	19	3	746	17	3	699	19	3
38	763	19	3	751	17	3	703	19	3
39	768	19	3	756	16	3	708	19	3
40	774	19	4	761	16	3	713	19	3
41	780	19	4	766	16	3	717	19	3
42	786	19	4	774	16	4	722	18	3
43	793	20	4	777	17	4	727	18	3
44	799	20	4	782	17	4	731	18	3
45	806	21	4	788	18	4	736	19	3
46	814	22	4	794	19	4	741	19	3
47	822	23	4	801	20	4	746	19	3
48	830	25	4	808	22	4	751	19	3
49	840	28	4	817	25	4	756	19	3
50	852	31	4	827	30	4	761	20	3
51	866	36	4	839	35	4	767	20	3
52	883	41	4	856	43	4	772	20	3
53	909	44	4	881	50	4	778	21	3
54	956	39	4	929	50	4	784	21	3
55	999	31	4	999	40	4	795	22	4
56	-	-	-	-	-	-	797	22	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.4. Raw Score to Scale Score Conversion Tables for Winter/Trimester 2009-10 (cont.)

Raw Score	Algebra II			Geometry			English III		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
57	-	-	-	-	-	-	804	23	4
58	-	-	-	-	-	-	811	24	4
59	-	-	-	-	-	-	819	25	4
60	-	-	-	-	-	-	827	26	4
61	-	-	-	-	-	-	836	28	4
62	-	-	-	-	-	-	846	29	4
63	-	-	-	-	-	-	856	31	4
64	-	-	-	-	-	-	868	34	4
65	-	-	-	-	-	-	881	36	4
66	-	-	-	-	-	-	896	39	4
67	-	-	-	-	-	-	914	41	4
68	-	-	-	-	-	-	936	41	4
69	-	-	-	-	-	-	964	36	4
70	-	-	-	-	-	-	999	28	4
71	-	-	-	-	-	-	999	28	4
72	-	-	-	-	-	-	999	28	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010

Raw Score	Biology I Core A			Biology I Core B			U.S. History Core A			U.S. History Core B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
0	440	44	1	440	42	1	440	51	1	440	47	1
1	440	44	1	440	42	1	440	51	1	440	47	1
2	440	44	1	440	42	1	440	51	1	440	47	1
3	440	44	1	440	42	1	440	51	1	440	47	1
4	440	44	1	440	42	1	440	51	1	440	47	1
5	440	44	1	440	42	1	440	51	1	440	47	1
6	440	44	1	440	42	1	440	51	1	440	47	1
7	440	44	1	440	42	1	440	51	1	440	47	1
8	440	44	1	440	42	1	440	51	1	440	47	1
9	440	44	1	440	42	1	440	51	1	440	47	1
10	440	44	1	440	42	1	440	51	1	440	47	1
11	440	44	1	440	42	1	440	51	1	440	47	1
12	440	44	1	440	42	1	440	51	1	440	47	1
13	440	44	1	440	42	1	440	51	1	440	47	1
14	440	44	1	469	46	1	452	52	1	465	50	1
15	472	49	1	496	50	1	496	57	1	498	55	1
16	499	53	1	517	52	1	525	60	1	523	57	1
17	521	55	1	534	52	1	547	60	1	543	57	1
18	540	55	1	550	50	1	564	58	1	559	55	1
19	556	54	1	563	48	1	578	54	1	573	52	1
20	570	51	1	575	45	1	590	50	1	586	48	1
21	583	48	1	586	42	1	603	45	2	603	45	2
22	595	45	1	596	39	1	611	40	2	607	41	2
23	607	42	1	606	36	1	620	36	2	616	37	2
24	617	39	1	614	34	1	628	33	2	625	34	2
25	627	37	2	627	32	2	635	30	2	633	32	2
26	636	34	2	631	30	2	642	28	2	641	30	2
27	645	33	2	638	29	2	649	26	2	648	28	2
28	654	31	2	646	28	2	655	25	2	655	26	2

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	Biology I Core A			Biology I Core B			U.S. History Core A			U.S. History Core B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
29	662	29	2	653	27	2	661	23	2	661	25	2
30	670	28	2	660	26	2	667	22	2	668	24	2
31	677	27	2	667	25	2	673	21	2	674	23	2
32	684	26	2	673	24	2	678	21	2	679	22	2
33	691	25	3	680	24	2	684	20	2	689	21	3
34	698	24	3	691	24	3	689	20	3	691	21	3
35	705	24	3	693	23	3	694	19	3	696	20	3
36	711	23	3	699	23	3	699	19	3	702	20	3
37	718	23	3	706	23	3	704	19	3	707	20	3
38	724	22	3	712	23	3	709	19	3	712	19	3
39	731	22	3	719	23	3	714	18	3	718	19	3
40	737	22	3	726	23	3	720	18	3	723	19	3
41	743	22	3	732	22	3	725	18	3	728	19	3
42	750	22	3	739	22	3	730	18	3	734	19	3
43	756	22	3	746	23	3	735	19	3	740	19	3
44	763	22	3	753	23	3	741	19	3	747	20	4
45	775	22	4	760	23	3	747	19	4	751	20	4
46	777	22	4	767	23	3	752	19	4	757	20	4
47	784	22	4	775	24	4	759	20	4	764	21	4
48	791	23	4	783	25	4	765	21	4	771	22	4
49	799	24	4	791	26	4	772	21	4	778	23	4
50	807	25	4	800	27	4	779	22	4	786	24	4
51	816	27	4	810	29	4	787	24	4	794	26	4
52	826	29	4	821	31	4	795	26	4	804	28	4
53	837	32	4	833	34	4	805	29	4	814	32	4
54	850	36	4	847	37	4	816	33	4	827	36	4
55	865	41	4	863	42	4	829	38	4	842	42	4
56	884	45	4	883	46	4	846	44	4	860	48	4
57	910	48	4	910	48	4	867	49	4	885	54	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	Biology I Core A			Biology I Core B			U.S. History Core A			U.S. History Core B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
58	953	44	4	952	44	4	900	52	4	925	54	4
59	999	36	4	999	35	4	961	44	4	999	41	4
60	999	36	4	999	35	4	999	36	4	999	41	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	Geometry Core A			Geometry Core B			Algebra II Core A			Algebra II Core B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
0	440	59	1	440	58	1	440	71	1	440	71	1
1	440	59	1	440	58	1	440	71	1	440	71	1
2	440	59	1	440	58	1	440	71	1	440	71	1
3	440	59	1	440	58	1	440	71	1	440	71	1
4	440	59	1	440	58	1	440	71	1	440	71	1
5	440	59	1	440	58	1	440	71	1	440	71	1
6	440	59	1	440	58	1	440	71	1	440	71	1
7	440	59	1	440	58	1	440	71	1	440	71	1
8	440	59	1	440	58	1	440	71	1	440	71	1
9	440	59	1	440	58	1	440	71	1	440	71	1
10	440	59	1	440	58	1	440	71	1	440	71	1
11	483	62	1	440	58	1	440	71	1	440	71	1
12	531	66	1	493	63	1	480	72	1	440	71	1
13	559	66	1	532	66	1	546	76	1	526	76	1
14	578	63	1	558	66	1	578	76	1	564	77	1
15	594	58	1	577	63	1	599	72	1	588	75	1
16	606	52	1	593	58	1	614	65	1	606	70	1
17	617	46	1	607	52	1	627	58	1	620	63	1
18	635	40	2	619	47	1	638	50	1	632	55	1
19	636	36	2	635	42	2	651	43	2	651	48	2
20	644	32	2	639	37	2	655	37	2	652	41	2
21	652	29	2	648	33	2	663	32	2	660	36	2
22	659	27	2	656	30	2	670	29	2	668	32	2
23	666	25	2	663	28	2	676	26	2	676	29	2
24	673	24	2	671	26	2	683	24	2	683	26	2
25	679	23	2	678	25	2	689	22	2	689	25	2
26	685	22	2	684	23	2	696	21	3	696	23	3
27	695	21	3	695	22	3	700	20	3	702	22	3
28	697	20	3	697	22	3	705	20	3	708	21	3

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	Geometry Core A			Geometry Core B			Algebra II Core A			Algebra II Core B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
29	703	20	3	703	21	3	710	19	3	713	21	3
30	709	19	3	708	20	3	716	19	3	719	20	3
31	714	19	3	714	20	3	721	18	3	724	19	3
32	720	19	3	720	19	3	726	18	3	730	19	3
33	725	18	3	725	19	3	731	18	3	735	19	3
34	731	18	3	731	18	3	736	18	3	741	18	3
35	736	18	3	736	18	3	741	17	3	746	18	3
36	742	18	3	742	18	3	746	17	3	751	18	3
37	747	18	3	747	18	3	752	17	3	756	18	3
38	753	17	3	752	18	3	757	17	3	762	18	3
39	758	17	3	758	17	3	762	17	3	767	18	3
40	764	17	3	763	17	3	768	17	3	774	18	4
41	774	17	4	774	18	4	774	18	4	778	18	4
42	775	18	4	775	18	4	779	18	4	784	18	4
43	781	18	4	780	18	4	785	18	4	790	19	4
44	787	18	4	787	18	4	791	19	4	796	19	4
45	794	19	4	793	19	4	798	19	4	803	20	4
46	800	20	4	800	20	4	804	20	4	810	21	4
47	808	21	4	807	21	4	812	21	4	818	22	4
48	816	23	4	815	23	4	820	23	4	826	24	4
49	825	25	4	825	26	4	829	25	4	836	26	4
50	835	29	4	835	29	4	839	28	4	847	29	4
51	848	33	4	848	34	4	852	32	4	860	33	4
52	865	40	4	865	40	4	867	38	4	876	38	4
53	888	46	4	889	47	4	888	45	4	898	43	4
54	928	48	4	931	47	4	924	48	4	936	44	4
55	999	35	4	999	35	4	999	33	4	999	32	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	English III Core A Prompt A			English III Core A Prompt B			English III Core B Prompt A			English III Core B Prompt B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
0	440	50	1	440	49	1	440	47	1	440	46	1
1	440	50	1	440	49	1	440	47	1	440	46	1
2	440	50	1	440	49	1	440	47	1	440	46	1
3	440	50	1	440	49	1	440	47	1	440	46	1
4	440	50	1	440	49	1	440	47	1	440	46	1
5	440	50	1	440	49	1	440	47	1	440	46	1
6	440	50	1	440	49	1	440	47	1	440	46	1
7	440	50	1	440	49	1	440	47	1	440	46	1
8	440	50	1	440	49	1	440	47	1	440	46	1
9	440	50	1	440	49	1	440	47	1	440	46	1
10	440	50	1	440	49	1	440	47	1	440	46	1
11	440	50	1	440	49	1	440	47	1	440	46	1
12	440	50	1	440	49	1	440	47	1	440	46	1
13	454	51	1	458	51	1	487	52	1	489	52	1
14	502	56	1	504	55	1	515	55	1	517	55	1
15	529	58	1	529	57	1	536	55	1	536	55	1
16	548	58	1	547	57	1	551	54	1	551	54	1
17	562	56	1	561	55	1	564	51	1	563	51	1
18	575	53	1	573	52	1	575	48	1	574	47	1
19	586	49	1	584	48	1	585	44	1	584	43	1
20	596	44	1	594	44	1	594	40	1	593	39	1
21	605	40	1	603	40	1	603	37	1	601	36	1
22	613	37	1	611	36	1	611	34	1	609	33	1
23	621	33	1	619	33	1	618	31	1	616	31	1
24	628	31	1	626	30	1	625	29	1	623	29	1
25	635	29	1	633	28	1	632	27	1	630	27	1

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	English III Core A Prompt A			English III Core A Prompt B			English III Core B Prompt A			English III Core B Prompt B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
26	642	27	1	640	27	1	638	26	1	636	26	1
27	649	26	2	649	25	2	649	25	2	642	25	1
28	654	24	2	652	24	2	650	24	2	649	24	2
29	660	23	2	658	23	2	656	23	2	654	23	2
30	665	22	2	664	22	2	661	22	2	660	22	2
31	671	22	2	669	22	2	667	22	2	665	22	2
32	676	21	2	674	21	2	672	21	2	670	21	2
33	681	20	2	679	20	2	677	21	2	676	21	2
34	686	20	2	684	20	2	682	20	2	681	21	2
35	691	19	2	689	19	2	687	20	2	686	20	2
36	695	19	3	695	19	3	695	20	3	695	20	3
37	700	18	3	699	19	3	697	20	3	696	20	3
38	705	18	3	703	18	3	702	20	3	701	19	3
39	709	18	3	708	18	3	707	19	3	706	19	3
40	714	18	3	713	18	3	712	19	3	711	19	3
41	718	17	3	717	18	3	717	19	3	716	19	3
42	723	17	3	722	17	3	722	19	3	721	19	3
43	727	17	3	726	17	3	727	19	3	726	19	3
44	732	17	3	731	17	3	732	19	3	731	19	3
45	736	17	3	735	17	3	737	19	3	736	19	3
46	741	17	3	740	17	3	742	19	3	741	19	3
47	745	17	3	745	17	3	747	19	3	746	19	3
48	750	17	3	749	17	3	753	19	3	752	19	3
49	755	17	3	754	17	3	758	20	3	757	20	3
50	760	18	3	759	18	3	764	20	3	763	20	3
51	765	18	3	765	18	3	769	20	3	769	20	3
52	770	18	3	770	18	3	775	20	3	775	20	3
53	776	19	3	775	19	3	781	21	3	781	21	3

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	English III Core A Prompt A			English III Core A Prompt B			English III Core B Prompt A			English III Core B Prompt B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
54	781	19	3	781	19	3	787	21	3	787	21	3
55	787	20	3	787	20	3	795	21	4	795	22	4
56	795	20	4	795	20	4	801	22	4	801	22	4
57	800	21	4	801	21	4	808	23	4	808	22	4
58	808	22	4	808	22	4	815	23	4	816	23	4
59	815	23	4	816	22	4	823	24	4	824	24	4
60	823	23	4	824	23	4	831	25	4	832	24	4
61	832	24	4	833	24	4	840	26	4	841	25	4
62	841	26	4	842	25	4	849	27	4	850	27	4
63	851	27	4	852	27	4	859	28	4	860	28	4
64	862	29	4	863	28	4	870	30	4	871	30	4
65	874	31	4	875	31	4	882	32	4	883	32	4
66	887	33	4	888	33	4	895	35	4	896	34	4
67	902	36	4	903	35	4	910	37	4	911	36	4
68	919	37	4	920	37	4	928	38	4	929	37	4
69	941	37	4	941	36	4	952	36	4	952	35	4
70	972	32	4	970	31	4	986	28	4	984	28	4
71	999	25	4	999	25	4	999	25	4	999	24	4
72	999	25	4	999	25	4	999	25	4	999	24	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	English II Core A Prompt A			English II Core A Prompt B			English II Core B Prompt A			English II Core B Prompt B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
0	440	34	1	440	34	1	440	33	1	440	33	1
1	440	34	1	440	34	1	440	33	1	440	33	1
2	440	34	1	440	34	1	440	33	1	440	33	1
3	440	34	1	440	34	1	440	33	1	440	33	1
4	440	34	1	440	34	1	440	33	1	440	33	1
5	440	34	1	440	34	1	440	33	1	440	33	1
6	440	34	1	440	34	1	440	33	1	440	33	1
7	440	34	1	440	34	1	440	33	1	440	33	1
8	440	34	1	440	34	1	440	33	1	440	33	1
9	440	34	1	440	34	1	440	33	1	440	33	1
10	440	34	1	440	34	1	440	33	1	440	33	1
11	440	34	1	440	34	1	440	33	1	440	33	1
12	440	34	1	440	34	1	440	33	1	440	33	1
13	440	34	1	440	34	1	451	34	1	451	35	1
14	450	36	1	450	36	1	475	39	1	475	39	1
15	476	40	1	476	40	1	494	41	1	495	41	1
16	495	42	1	496	43	1	510	42	1	510	42	1
17	511	43	1	512	43	1	523	42	1	524	42	1
18	525	43	1	525	43	1	535	40	1	536	40	1
19	537	41	1	537	41	1	546	38	1	546	38	1
20	547	39	1	548	39	1	555	36	1	556	36	1
21	557	37	1	557	37	1	564	34	1	565	34	1
22	566	35	1	566	35	1	572	32	1	573	32	1
23	574	32	1	574	32	1	580	30	1	581	30	1
24	588	30	2	588	30	2	588	29	2	588	29	2
25	589	29	2	590	29	2	595	28	2	595	28	2

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	English II Core A Prompt A			English II Core A Prompt B			English II Core B Prompt A			English II Core B Prompt B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
26	596	27	2	597	27	2	602	27	2	602	27	2
27	603	26	2	604	26	2	608	26	2	609	26	2
28	609	25	2	610	25	2	615	25	2	616	25	2
29	616	24	2	617	25	2	621	25	2	622	25	2
30	622	24	2	623	24	2	628	24	2	629	24	2
31	628	23	2	629	23	2	634	24	2	635	24	2
32	634	23	2	635	23	2	640	24	2	641	24	2
33	640	23	2	641	23	2	646	24	2	647	24	2
34	646	22	2	647	22	2	652	23	2	653	23	2
35	652	22	2	653	22	2	659	23	2	660	23	2
36	658	22	2	659	22	2	665	23	2	666	23	2
37	664	22	2	665	22	2	671	23	2	672	23	2
38	670	22	2	671	22	2	677	23	2	678	23	2
39	676	22	2	677	22	2	684	23	2	685	23	2
40	682	22	2	683	22	2	693	23	3	693	23	3
41	693	22	3	693	22	3	697	24	3	698	24	3
42	694	22	3	695	22	3	703	24	3	704	24	3
43	700	22	3	701	22	3	710	24	3	711	24	3
44	707	22	3	708	22	3	717	24	3	718	24	3
45	713	22	3	714	22	3	724	24	3	725	24	3
46	720	22	3	721	22	3	731	25	3	732	25	3
47	726	23	3	728	23	3	738	25	3	739	25	3
48	733	23	3	735	23	3	746	25	3	747	26	3
49	741	23	3	742	23	3	754	26	3	755	26	3
50	748	24	3	749	24	3	762	27	3	763	27	3
51	756	25	3	757	24	3	771	27	3	772	28	3
52	764	25	3	766	25	3	780	28	3	781	29	3
53	773	26	3	774	26	3	797	29	4	797	30	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	English II Core A Prompt A			English II Core A Prompt B			English II Core B Prompt A			English II Core B Prompt B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
54	782	27	3	784	27	3	799	31	4	801	31	4
55	797	28	4	797	28	4	810	32	4	812	33	4
56	803	30	4	804	30	4	822	34	4	824	35	4
57	814	31	4	816	32	4	835	37	4	838	37	4
58	827	33	4	829	34	4	850	39	4	853	39	4
59	842	36	4	844	36	4	866	42	4	870	42	4
60	858	39	4	861	39	4	886	45	4	889	45	4
61	878	42	4	881	42	4	908	46	4	913	46	4
62	901	45	4	906	44	4	936	45	4	941	44	4
63	931	44	4	937	43	4	973	38	4	978	37	4
64	971	37	4	977	35	4	999	32	4	999	32	4
65	999	30	4	999	29	4	999	32	4	999	32	4
66	999	30	4	999	29	4	999	32	4	999	32	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	Algebra I Core A			Algebra I Core B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
0	490	50	1	490	55	1
1	490	50	1	490	55	1
2	490	50	1	490	55	1
3	490	50	1	490	55	1
4	490	50	1	490	55	1
5	490	50	1	490	55	1
6	490	50	1	490	55	1
7	490	50	1	490	55	1
8	490	50	1	490	55	1
9	490	50	1	490	55	1
10	490	50	1	490	55	1
11	490	50	1	490	55	1
12	538	55	1	513	57	1
13	572	57	1	565	60	1
14	594	57	1	594	61	1
15	611	54	1	613	59	1
16	625	49	1	628	54	1
17	636	44	1	639	48	1
18	646	39	1	649	42	1
19	655	34	1	662	36	2
20	662	30	2	666	31	2
21	669	26	2	673	27	2
22	676	24	2	679	24	2
23	682	21	2	685	21	2
24	687	20	2	690	20	2
25	692	18	2	696	18	2
26	700	17	3	700	17	3
27	702	17	3	705	16	3
28	707	16	3	709	16	3

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Table 4.5. Raw Score to Scale Score Conversion Tables for Spring 2010 (cont.)

Raw Score	Algebra I Core A			Algebra I Core B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
29	711	15	3	714	15	3
30	715	15	3	718	14	3
31	720	14	3	722	14	3
32	724	14	3	726	14	3
33	728	14	3	730	13	3
34	732	14	3	734	13	3
35	736	13	3	738	13	3
36	740	13	3	742	13	3
37	744	13	3	746	13	3
38	749	13	3	750	13	3
39	753	14	3	754	13	3
40	757	14	3	758	13	3
41	762	14	4	762	13	4
42	767	14	4	767	13	4
43	772	14	4	771	13	4
44	777	14	4	776	14	4
45	782	15	4	781	14	4
46	788	15	4	786	15	4
47	794	16	4	792	15	4
48	800	17	4	798	17	4
49	807	20	4	805	19	4
50	815	23	4	812	22	4
51	825	29	4	821	28	4
52	838	39	4	833	39	4
53	855	52	4	849	53	4
54	887	63	4	877	67	4
55	999	38	4	999	36	4

Note: *CSEM* = Conditional Standard Error of Measure; Perf. Level = Performance Level; 1 = Unsatisfactory, 2 = Limited Knowledge, 3 = Proficient, 4 = Advanced

Section 5

Classification Consistency and Accuracy Studies

5.1 Classification Consistency and Accuracy

Every test administration will result in some error in classifying examinees. The concept of the standard error of measurement (SEM) has implications for the intepretation of cut scores used to classify students into different performance levels. For example, a given student may have a true performance level greater than a cut score; however, due to random variations (measurement error), the student's observed test score may be below the cut score. As a result, the student may be classified as having a lower performance level. As discussed in Section 6.4, a student's observed score is most likely to fall within a standard error band around his or her true score. Thus, the classification of students into different performance levels can be imperfect; especially for the borderline students whose true scores lie close to the performance level cut scores.

According to Livingston and Lewis (1995, p. 180), the accuracy of a classification is "the extent to which the actual classifications of the test takers… agree with those that would be made on the basis of their true score" and are calculated from cross-tabulations between "classifications based on an observable variable and classifications based on an unobservable variable." Since the unobservable variable—the true score—is not available, Livingston and Lewis provide a method to estimate the true score distribution of a test and create the cross-tabulation of the true score and observed variable (raw score) classifications. Consistency is "the agreement between classifications based on two non-overlapping, equally-difficult forms of the test" (p. 180). Consistency is estimated using actual response data from a test and the test's reliability to statistically model two parallel forms of the test and compare the classifications on those alternate forms. There are three types of accuracy and consistency indices that can be generated using Livingston and Lewis' approach: overall, conditional on level, and by cut score.

The overall accuracy of performance level classifications is computed as a sum of the proportions on the diagonal of the joint distribution of true score- and observed score levels. Essentially, overall accuracy is a proportion (or percentage) of correct classifications across all levels. The overall consistency index is computed as the sum of the diagonal cells in a consistency table. Another way to express overall consistency is to use the kappa coefficient, as used in the inter-rater reliability studies in Section 3.7. Like the inter-rater reliability studies, kappa provides an estimate of agreement or the proportion of consistent classifications between two different tests after taking into account chance.

Consistency conditional on performance level is computed as the ratio between the proportion of correct classifications at the selected performance level (for example, proficient students who were classified as proficient) and the proportion of all the students classified into that level (total proportion of students who were considered proficient). Accuracy conditional on performance level is computed in a similar manner except that in the consistency table where both row and column marginal sums are the same, the accuracy table uses the sum based on estimated status as the total for computing accuracy conditional on performance level.

To evaluate decisions at specific cut scores, the joint distribution of all the performance levels are collapsed into dichotomized distributions around that specific cut score (for example collapsing Unsatisfactory and Limited Knowledge and then Proficient and Advanced to assess decisions at the Proficient cut score). The accuracy index at cut score is computed as the sum of the proportions of correct classifications around this selected cut score. The consistency at a specific cut score is obtained in a similar way, but by dichotomizing the distributions at the cut score performance level and between all other performance levels combined. Table 5.1 for Winter/Trimester 2009-10 and Table 5.2 for Spring 2010 present the overall accuracy and consistency indices for all of the ACE EOI tests.

Table 5.1. Estimates of Accuracy and Consistency of Performance Classification for Winter/Trimester 2009-10

Subject	Accuracy	Consistency	Kappa	False Positives	False Negatives
Algebra I	0.74	0.69	0.53	0.13	0.12
Algebra II	0.78	0.70	0.57	0.13	0.09
Biology I	0.76	0.71	0.58	0.15	0.10
English II	0.77	0.70	0.52	0.09	0.14
English III	0.77	0.74	0.59	0.16	0.07
Geometry	0.77	0.74	0.60	0.10	0.13
U.S. History	0.78	0.70	0.58	0.09	0.12

Table 5.2. Estimates of Accuracy and Consistency of Performance Classification for Spring 2010

Subject	Core	Accuracy	Consistency	Kappa	False Positives	False Negatives
Algebra I	A	0.77	0.72	0.59	0.04	0.19
	B	0.78	0.71	0.58	0.09	0.13
Algebra II	A	0.78	0.71	0.60	0.10	0.12
	B	0.78	0.71	0.59	0.09	0.13
Biology I	A	0.77	0.69	0.56	0.09	0.14
	B	0.76	0.71	0.56	0.14	0.11
English II	A	0.81	0.74	0.58	0.06	0.13
	B	0.77	0.71	0.55	0.16	0.07
English III	A	0.80	0.75	0.62	0.14	0.07
	B	0.78	0.74	0.58	0.17	0.05
Geometry	A	0.81	0.75	0.63	0.05	0.14
	B	0.79	0.75	0.62	0.11	0.10
U.S. History	A	0.81	0.75	0.63	0.08	0.11
	B	0.78	0.74	0.61	0.04	0.18

As shown in Table 5.1 and Table 5.2, the overall accuracy indices range between 74 and 78 percent for Winter/Trimester 2009-10 and 76 and 81 percent for Spring 2010 and overall consistency ranging between 69 and 74 percent for Winter/Trimester 2009-10 and 69 and 75 percent for Spring 2010. Kappa coefficients range from 0.52 and 0.60 for Winter/Trimester 2009-10 and 0.55 and 0.63 for Spring 2010. The rate of false positives range from 9 to 15 percent for Winter/Trimester 2009-10 and 4 to 17 percent for Spring 2010. The false negative rates range from 7 to 14 percent for Winter/Trimester 2009-10 and 5 to 19 percent for Spring 2010.

Table 5.3 and Table 5.4 provide the accuracy-, consistency-, false positive-, and false negative rates by cut score for Winter/Trimester 2009-10 and Spring 2010, respectively. The data in these tables reveal that the level of agreement for both accuracy and consistency is above 80 percent in all cases, with most above 90 percent. In general, the high rates of accuracy and consistency support the cut decisions made using these assessments. Similar to Table 5.1 and Table 5.2, the false positive and false negative rates were comparable for the Winter/Trimester 2009-10 and Spring 2010 administrations and are quite low.

The importance of the dichotomous categorization is particularly notable when they map onto pass/fail decisions for the assessments. For the EOI tests, the U+L/P+A is the important dichotomization, because it directly translates to the pass/fail decision point. Similar to other dichotomization distinctions, there are three main scenarios at this cut point: 1) observed performance is accurately reflective of the true ability level (i.e., the examinee passed and should have passed); 2) the true ability level is below the standard, but the observed test score is above the standard (i.e., a false positive); and 3) the true ability level is above the standard, but the observed test score is below the standard (i.e., a false negative). In examining Table 5.3 and Table 5.4, in Winter/Trimester 2009-10 Algebra I, for example, 89 percent of students were correctly classified as pass or fail based on their performance (scenario 1), 9 percent passed but their true performance is below the standard (scenario 2), and 3 percent failed although their true performance is above the standard (scenario 3). Overall, the accuracy rates for accurate classification are above 85% for the Winter/Trimester and Spring administrations for all subjects – students are appropriately (more than 85% of the time) categorized into pass/fail classifications based on their true ability using their observed score (raw score) as their classification score.

Table 5.3. Accuracy and Consistency Estimates by Cut Score: False Positive- and False Negative Rates for Winter/Trimester 2009-10

	Accuracy			Consistency			False Positives			False Negatives		
Subject	U / L+P+A	U+L / P+A	U+L+P / A	U / L+P+A	U+L / P+A	U+L+P / A	U / L+P+A	U+L / P+A	U+L+P / A	U / L+P+A	U+L / P+A	U+L+P / A
Algebra I	0.92	0.89	0.94	0.90	0.86	0.92	0.01	0.09	0.04	0.07	0.03	0.02
Algebra II	0.94	0.92	0.92	0.92	0.88	0.89	0.04	0.04	0.06	0.02	0.04	0.03
Biology I	0.96	0.91	0.88	0.95	0.89	0.87	0.02	0.01	0.11	0.02	0.07	0.01
English II	0.98	0.93	0.86	0.98	0.90	0.81	0.00	0.04	0.04	0.02	0.03	0.10
English III	0.95	0.92	0.90	0.94	0.90	0.89	0.01	0.06	0.10	0.04	0.02	0.01
Geometry	0.96	0.91	0.90	0.94	0.90	0.89	0.10	0.08	0.02	0.03	0.01	0.09
U.S. History	0.96	0.92	0.91	0.94	0.88	0.87	0.01	0.03	0.05	0.03	0.05	0.04

Note: U =Unsatisfactory; L = Limited Knowledge; P = Proficient; and A = Advanced.
Note: U / L+P+A = Unsatisfactory divided by Limited Knowledge plus Proficient plus Advanced; U+L / P+A = Unsatisfactory plus Limited Knowledge divided by Proficient plus Advanced; U+L+P / A = Unsatisfactory plus Limited Knowledge plus Proficient divided by Advanced.

Table 5.4. Accuracy and Consistency Estimates by Cut Score: False Positive- and False Negative Rates for Spring 2010

		Accuracy			Consistency			False Positives			False Negatives		
Subject	Core	U / L+P+A	U+L / P+A	U+L+P / A	U / L+P+A	U+L / P+A	U+L+P / A	U / L+P+A	U+L / P+A	U+L+P / A	U / L+P+A	U+L / P+A	U+L+P / A
Algebra I	A	0.94	0.91	0.92	0.92	0.89	0.90	0.01	0.01	0.02	0.05	0.08	0.06
	B	0.94	0.91	0.93	0.92	0.88	0.90	0.03	0.02	0.03	0.02	0.06	0.04
Algebra II	A	0.93	0.92	0.93	0.90	0.89	0.91	0.05	0.03	0.02	0.03	0. 06	0.04
	B	0.93	0.91	0.94	0.90	0.88	0.91	0.03	0.02	0.03	0.04	0.07	0.03
Biology I	A	0.95	0.91	0.90	0.94	0.88	0.87	0.00	0.05	0.04	0.05	0.04	0.06
	B	0.96	0.91	0.88	0.95	0.89	0.86	0.02	0.01	0.10	0.02	0.08	0.02
English II	A	0.99	0.95	0.87	0.99	0.92	0.83	0.00	0.03	0.03	0.01	0.03	0.10
	B	0.99	0.93	0.86	0.99	0.92	0.81	0.01	0.06	0.09	0.00	0.01	0.05
English III	A	0.96	0.93	0.90	0.95	0.92	0.88	0.00	0.05	0.08	0.03	0.02	0.02
	B	0.96	0.94	0.88	0.95	0.92	0.86	0.03	0.04	0.11	0.01	0.03	0.01
Geometry	A	0.96	0.93	0.92	0.95	0.91	0.88	0.00	0.02	0.04	0.04	0.06	0.05
	B	0.96	0.92	0.91	0.95	0.91	0.89	0.02	0.07	0.03	0.02	0.01	0.06
U.S. History	A	0.96	0.93	0.92	0.96	0.90	0.88	0.03	0.03	0.03	0.01	0.04	0.05
	B	0.96	0.91	0.90	0.96	0.90	0.88	0.00	0.01	0.02	0.03	0.08	0.07

Note: U =Unsatisfactory; L = Limited Knowledge; P = Proficient; and A = Advanced.
Note: U / L+P+A = Unsatisfactory divided by Limited Knowledge plus Proficient plus Advanced; U+L / P+A = Unsatisfactory plus Limited Knowledge divided by Proficient plus Advanced; U+L+P / A = Unsatisfactory plus Limited Knowledge plus Proficient divided by Advanced.

Section 6

Summary Statistics

6.1 Descriptive Statistics

The summary descriptive statistics of the scale scores for Winter/Trimester 2009-10 and Spring 2010 appears in Table 6.1 through Table 6.8. The scales scores presented exclude invalid student cases and second-time testers.

Table 6.1. Descriptive Statistics of the Scale Scores for Winter/Trimester 2009-10 - Overall

	Total			
Subject	***N***	**Mean**	***SD***	**Med.**
Algebra I	1,692	709.7	59.6	714
Algebra II	2,038	725.6	82.2	734
Biology I	2,379	734.3	83.8	737
English II	2,590	746.4	79.7	753
English III	2,766	738.0	77.5	741
Geometry	2,127	737.3	73.9	741
U.S. History	2,032	713.5	81.9	719

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.2. Descriptive Statistics of the Scale Scores for Winter/Trimester 2009-10 by Gender

	Female				Male			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Algebra I	857	710.5	55.5	718	835	708.9	63.5	714
Algebra II	1,018	724.5	78.9	728	1,020	726.7	85.4	734
Biology I	1,198	730.0	74.3	731	1,181	738.7	92.2	750
English II	1,289	751.9	75.0	753	1,301	740.9	83.7	745
English III	1,399	746.0	73.9	746	1,367	729.8	80.2	731
Geometry	1,113	735.6	72.2	741	1,014	739.2	75.7	746
U.S. History	1,031	702.7	75.8	708	1,001	724.5	86.4	730

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.3. Descriptive Statistics of the Scale Scores for Winter/Trimester 2009-10 by Race/Ethnicity

	African-American				Native American			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Algebra I	335	692.3	62.9	705	262	699.6	56.8	701
Algebra II	427	687.5	85.0	697	224	711.9	71.8	713
Biology I	509	701.3	84.2	706	351	728.7	74.7	737
English II	399	708.3	76.7	710	370	739.5	78.6	745
English III	414	696.3	77.4	703	415	728.9	72.3	731
Geometry	369	704.7	76.6	709	278	728.6	80.0	736
U.S. History	340	675.9	83.6	689	331	716.3	80.3	725

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.3. Descriptive Statistics of the Scale Scores for Winter/Trimester 2009-10 by Race/Ethnicity (cont.)

	Hispanic				Asian			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Algebra I	173	696.9	62.8	705	42	737.8	66.4	746
Algebra II	148	709.5	87.4	722	61	790.6	71.0	786
Biology I	170	699.1	99.5	713	60	758.7	100.0	756
English II	181	718.7	83.1	731	62	760.1	94.1	778
English III	227	710.4	79.9	717	76	782.8	80.4	778
Geometry	176	723.8	77.9	731	47	781.1	67.6	782
U.S. History	169	666.5	83.3	667	35	713.7	90.3	725

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.3. Descriptive Statistics of the Scale Scores for Winter/Trimester 2009-10 by Race/Ethnicity (cont.)

	White			
Subject	***N***	**Mean**	***SD***	**Med.**
Algebra I	869	720.6	55.5	722
Algebra II	1,169	740.7	76.1	746
Biology I	1,273	752.7	76.5	756
English II	1,560	760.6	75.2	761
English III	1,617	752.3	71.9	751
Geometry	1,235	749.4	66.6	751
U.S. History	1,143	731.0	74.8	736

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.4. Descriptive Statistics of the Scale Scores for Winter/Trimester 2009-10 by Free/Reduced Lunch Status

	Free/Reduced Lunch = Yes				Free/Reduced Lunch = No			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Algebra I	650	693.7	61.1	701	1,042	719.7	56.4	722
Algebra II	566	697.9	88.6	710	1,472	736.2	77.0	740
Biology I	596	704.1	87.5	713	1,783	744.4	80.1	750
English II	804	723.5	75.0	731	1,786	756.7	79.5	761
English III	891	712.9	77.8	717	1,875	749.9	74.5	751
Geometry	755	711.5	75.0	720	1,372	751.5	69.3	756
U.S. History	668	678.4	83.8	689	1,364	730.6	75.3	736

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.5. Descriptive Statistics of the Scale Scores for Spring 2010 - Overall

	Total			
Subject	***N***	**Mean**	***SD***	**Med.**
Core A				
Algebra I	16,239	726.9	57.6	732
Algebra II	15,456	718.0	89.2	726
Biology I	19,155	733.1	80.1	737
English II – OE A	8,881	760.7	78.0	764
English II – OE B	8,733	762.3	76.9	766
English III – OE A	8,964	746.8	73.1	750
English III – OE B	8,947	747.9	71.5	754
Geometry	17,847	741.9	79.8	747
U.S. History	17,606	724.7	78.7	730
Core B				
Algebra I	13,510	727.6	57.9	730
Algebra II	12,519	716.7	89.4	724
Biology I	15,933	734.1	78.7	739
English II – OE A	9,036	755.8	78.3	762
English II – OE B	8,811	756.6	76.6	755
English III – OE A	8,969	745.5	71.7	753
English III – OE B	8,676	745.9	71.2	752
Geometry	14,846	743.0	81.1	747
U.S. History	14,644	725.2	79.0	728

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.6. Descriptive Statistics of the Scale Scores for Spring 2010 by Gender

	Female				Male			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Core A								
Algebra I	8,315	728.5	55.1	732	7,924	725.2	60.1	732
Algebra II	7,896	720.7	84.9	726	7,560	715.1	93.4	721
Biology I	9,820	727.9	76.9	731	9,335	738.5	83.0	743
English II – OE A	4,498	769.0	77.8	769	4,382	752.2	77.3	756
English II – OE B	4,411	770.4	75.8	766	4,320	754.0	77.1	757
English III – OE A	4,506	756.7	70.1	760	4,458	736.7	74.7	741
English III – OE B	4,462	757.8	67.5	759	4,485	738.0	73.9	745
Geometry	8,969	741.3	77.5	742	8,878	742.5	82.1	747
U.S. History	8,902	712.7	74.6	714	8,704	736.9	81.0	741
Core B								
Algebra I	6,853	730.4	55.0	734	6,657	724.7	60.6	730
Algebra II	6,444	718.0	83.5	724	6,075	715.4	95.3	724
Biology I	8,022	729.1	74.9	732	7,911	739.2	82.1	746
English II – OE A	4,563	759.7	76.8	762	4,473	751.8	79.5	754
English II – OE B	4,455	762.6	75.1	763	4,356	750.4	77.6	755
English III – OE A	4,457	754.3	66.8	758	4,512	736.8	75.2	742
English III – OE B	4,374	754.4	66.2	757	4,302	737.3	74.9	746
Geometry	7,526	742.3	79.1	747	7,320	743.7	83.2	747
U.S. History	7,551	714.3	75.0	718	7,093	736.8	81.4	747

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.7. Descriptive Statistics of the Scale Scores for Spring 2010 by Race/Ethnicity

	African-American				Native American			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Core A								
Algebra I	1,439	699.2	58.8	707	3,387	719.4	56.4	724
Algebra II	1,281	678.4	85.3	689	2,962	700.0	87.1	710
Biology I	1,817	687.8	79.7	691	3,594	725.9	73.7	731
English II – OE A	847	724.8	77.4	726	1,676	753.6	72.3	756
English II – OE B	860	725.1	74.5	728	1,640	755.2	71.5	757
English III – OE A	883	707.7	74.7	709	1,644	738.6	70.7	741
English III – OE B	872	712.8	68.1	717	1,716	740.6	67.4	745
Geometry	1,687	698.4	83.1	703	3,370	729.6	73.9	731
U.S. History	1,540	681.3	81.6	689	3,275	716.3	75.1	720
Core B								
Algebra I	1,227	702.7	58.6	709	2,758	720.4	54.5	722
Algebra II	1,131	672.4	90.4	683	2,353	701.4	88.5	713
Biology I	1,523	687.9	77.6	693	2,991	726.8	74.5	732
English II – OE A	900	719.1	75.7	724	1,771	749.5	73.8	754
English II – OE B	850	720.9	81.0	725	1,661	748.2	71.3	747
English III – OE A	878	707.9	73.1	717	1,682	735.4	69.6	742
English III – OE B	823	706.1	72.9	706	1,639	739.3	68.5	746
Geometry	1,454	695.1	86.8	703	2,785	731.2	73.5	736
U.S. History	1,355	681.6	81.5	689	2,720	716.2	75.2	718

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.7. Descriptive Statistics of the Scale Scores for Spring 2010 by Race/Ethnicity (cont.)

	Hispanic				Asian			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Core A								
Algebra I	1,604	703.8	60.3	711	343	767.2	63.0	772
Algebra II	1,235	699.9	87.0	705	384	783.2	90.9	785
Biology I	1,860	689.4	87.7	698	414	756.3	88.1	756
English II – OE A	842	723.1	84.2	726	212	784.3	81.9	797
English II – OE B	809	724.3	81.9	721	202	780.3	85.4	784
English III – OE A	740	718.3	73.5	723	195	754.4	72.2	750
English III – OE B	807	723.2	70.8	731	174	767.7	74.0	765
Geometry	1,587	718.8	75.3	720	374	788.1	85.1	794
U.S. History	1,597	694.0	81.5	699	407	737.8	87.6	735
Core B								
Algebra I	1,357	706.3	58.1	714	285	766.6	64.4	767
Algebra II	1,025	697.2	89.9	702	338	786.1	85.9	784
Biology I	1,515	689.5	81.2	693	302	762.3	85.6	767
English II – OE A	830	717.0	88.0	724	203	773.9	84.2	797
English II – OE B	843	723.1	79.5	725	217	771.6	86.8	781
English III – OE A	794	716.6	70.9	722	196	773.5	67.3	775
English III – OE B	761	719.9	70.0	726	196	766.3	75.8	769
Geometry	1,379	717.1	78.7	725	328	805.3	92.9	800
U.S. History	1,321	695.1	81.4	696	364	745.3	78.8	747

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.7. Descriptive Statistics of the Scale Scores for Spring 2010 by Race/Ethnicity (cont.)

	White			
Subject	***N***	**Mean**	***SD***	**Med.**
Core A	9,348	736.5	53.6	740
Algebra I	9,502	728.6	87.1	731
Algebra II	11,345	749.1	74.6	750
Biology I	5,245	774.2	74.3	773
English II – OE A	5,164	776.2	73.2	774
English II – OE B	5,432	759.3	69.6	765
English III – OE A	5,327	759.3	70.2	765
English III – OE B	10,663	754.7	77.2	758
Geometry	10,680	737.8	74.7	741
U.S. History	9,348	736.5	53.6	740
Core B				
Algebra I	7,777	736.4	55.9	738
Algebra II	7,604	727.9	85.3	735
Biology I	9,510	750.3	73.2	753
English II – OE A	5,283	769.8	73.8	771
English II – OE B	5,191	770.1	72.5	772
English III – OE A	5,373	758.1	68.3	764
English III – OE B	5,197	757.3	68.0	763
Geometry	8,774	756.6	77.1	758
U.S. History	8,800	738.6	75.1	740

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

Table 6.8. Descriptive Statistics of the Scale Scores for Spring 2010 by Free/Reduced Lunch Status

	Free/Reduced Lunch = Yes				Free/Reduced Lunch = No			
Subject	**N**	**Mean**	**SD**	**Med.**	**N**	**Mean**	**SD**	**Med.**
Core A								
Algebra I	7,850	711.4	56.7	715	8,389	741.4	54.6	744
Algebra II	6,091	689.8	88.6	700	9,365	736.2	84.7	741
Biology I	8,543	708.1	79.6	711	10,612	753.1	74.7	756
English II – OE A	3,919	736.5	76.4	741	4,962	779.9	73.9	782
English II – OE B	3,895	737.4	73.3	735	4,838	782.3	73.8	784
English III – OE A	3,808	722.5	72.1	727	5,156	764.7	68.5	765
English III – OE B	3,795	726.3	68.9	731	5,152	763.7	69.2	765
Geometry	7,749	718.3	77.0	720	10,098	760.0	77.2	764
U.S. History	7,280	701.2	78.6	704	10,326	741.2	74.6	747
Core B								
Algebra I	6,572	713.2	57.3	718	6,938	741.2	55.1	746
Algebra II	5,045	690.4	89.9	702	7,474	734.5	84.6	741
Biology I	6,890	708.7	77.4	712	9,043	753.4	74.1	753
English II – OE A	4,046	732.3	78.3	738	4,990	774.8	72.9	780
English II – OE B	3,875	732.9	74.4	739	4,936	775.2	73.1	781
English III – OE A	3,802	724.8	71.3	732	5,167	760.7	68.1	764
English III – OE B	3,675	724.8	71.6	731	5,001	761.4	66.7	763
Geometry	6,340	719.3	79.1	725	8,506	760.6	78.1	763
U.S. History	6,102	701.0	77.7	707	8,542	742.6	75.2	747

Note: N = Sample size; SD = Standard Deviation; Med. = Median.

6.2 Performance Level Distribution

The distributions of students in the four performance levels based on student performance in the Winter/Trimester 2009-10 and Spring 2010 administration are presented in Table 6.9 (please see Appendix B and Appendix C for distributions by scale score for Winter/Trimester 2009-10 and Spring 2010, respectively). As above, these percentages exclude invalid student data and second-time test-takers. The percentage distributions for each of the content areas are comparable to previous administrations (e.g., Winter/Trimester 2008-09 and Spring 2009).

Table 6.9. Percentage of Students by Performance Level for Winter/Trimester 2009-10 and Spring 2010

Subject	*N*	Unsatisfactory	Limited Knowledge	Proficient	Advanced
Winter 2009-10					
Algebra I	1,692	13.2%	19.7%	49.9%	17.2%
Algebra II	2,038	12.5%	14.1%	44.8%	28.6%
Biology I	2,379	7.6%	16.1%	41.4%	34.9%
English II	2,590	3.3%	16.5%	52.7%	27.6%
English III	2,766	9.9%	14.1%	51.2%	24.8%
Geometry	2,127	7.0%	14.1%	46.6%	32.4%
U.S. History	2,032	7.3%	23.2%	30.7%	38.8%
Spring 2010 Core A					
Algebra I	16,239	10.4%	14.4%	47.2%	28.0%
Algebra II	15,456	16.7%	17.9%	38.6%	26.7%
Biology I	19,155	8.3%	17.1%	41.1%	33.6%
English II – OE A	8,740	1.3%	12.5%	50.1%	36.1%
English II – OE B	8,733	1.3%	13.0%	50.6%	35.1%
English III – OE A	8,805	8.0%	12.2%	52.4%	27.4%
English III – OE B	8,947	7.6%	12.2%	53.2%	27.0%
Geometry	17,847	6.8%	15.5%	39.1%	38.6%
U.S. History	17,606	5.2%	23.6%	29.5%	41.7%
Spring 2010 Core B					
Algebra I	13,510	8.8%	17.7%	45.6%	27.9%
Algebra II	12,519	15.7%	19.1%	39.2%	26.0%
Biology I	15,933	7.3%	16.6%	44.4%	31.8%
English II – OE A	8,884	2.1%	14.3%	46.7%	36.9%
English II – OE B	8,811	2.2%	14.6%	48.1%	35.0%
English III – OE A	8,816	7.4%	11.6%	54.8%	26.2%
English III – OE B	8,676	8.5%	10.3%	55.2%	26.0%
Geometry	14,846	7.1%	14.6%	39.5%	38.9%
U.S. History	14,644	5.1%	21.0%	31.1%	42.7%

6.3 Conditional Standard Error of Measurement

The conditional standard error of measurement (*CSEM*) was computed for each reported scale score. *CSEM* was computed using an IRT-based approach based on the following formula:

$$CSEM(O_X \mid \theta) = \sqrt{\left[\sum_{X=0}^{MaxX} O_X^2\, p(X \mid \theta)\right] - \left[\sum_{X=0}^{MaxX} O_X \cdot p(X \mid \theta)\right]^2} \tag{9}$$

where O_X is the observed scaled score for a particular number-correct score X, θ is the IRT ability scale value conditioned on, and $p(\bullet)$ is the probability function. Pearson has implemented a computational approach for estimating $CSEM(O_x \mid \theta)$ in which $p(X \mid \theta)$ is computed using a recursive algorithm given by Thissen, Pommerich, Billeaud, and Williams (1995). This algorithm is a polytomous generalization of the algorithm for dichotomous items given by Lord and Wingersky (1984). The values of θ used with the algorithm are obtained through the true score equating process (i.e., by solving for θ through the test characteristic curve for each number-correct score, X). There is one *CSEM* per number-correct score. The CSEMs by subject appear Table 4.4 and Table 4.5 for Winter/Trimester 2009-10 and Spring 2010, respectively.

6.4 Standard Error of Measurement

Measurement error is associated with every test score. A student's true score is the hypothetical average score that would result if the student took the test repeatedly under similar conditions. The standard error of measurement (SEM), as an overall test-level measure of error, can be used to construct a range around any given observed test score that likely includes the student's true score. *SEM* is computed by taking the square root of the average value of the variances of the error of measurement associated with each of the raw score or scales scores:

$$SEM = \sqrt{\frac{\sum_j (CSEM_j^{\,2} \cdot N_j)}{N_T}} \tag{10}$$

where,

SEM = Standard Error of Measurement
CSEM = Conditional Standard of Measurement
N_j = number of examinees obtaining score j in the population
N_T = total number of students in test sample

SEM was computed for each of the content areas. Table 7.4 presents the overall estimates of *SEM* for each of the content areas for the Winter/Trimester 2009-2010 and Spring 2010 administrations.

Table 6.10. Overall Estimates of *SEM* by Subject

Subject	SEM
Winter 2009-10	
Algebra I	4.85
Algebra II	5.35
Biology I	5.23
English II	4.91
English III	5.04
Geometry	5.29
U.S. History	4.85
Spring 2010	
Algebra I – A	4.69
Algebra I – B	4.78
Algebra II – A	6.09
Algebra II – B	6.44
Biology I – A	5.35
Biology I – B	5.29
English II – A	4.77
English II – B	4.82
English III – A	5.14
English III – B	5.27
Geometry – A	5.26
Geometry – B	5.29
U.S. History - A	4.69
U.S. History - B	4.78

Note: *SEM* = Standard Error of Measurement.

References

American Educational Research Association (AERA), American Psychological Association (APA), & the National Council on Measurement in Education (NCME) (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum.

Holland, P. W., & Thayer, D.T. (1988). Differential Item Performance and the Mantel-Haenszel Procedure. (ETS RR-86-31). Princeton, NJ: Educational Testing Service.

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

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

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score "equatings." *Applied Psychological Measurement, 8,* 453-461.

Kim, S. & Kolen, M. J. (2004). STUIRT: A computer program. Iowa City, IA: The University of Iowa. (Available from the web address: http://www.uiowa.edu/~casma).

Kolen, M.J. (2004). POLYEQUATE: A computer program. Iowa City, IA: The University of Iowa. (Available from the web address: http://www.uiowa.edu/~casma).

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: methods and practices (2nd ed.)*. New York: Springer.

Kraemer, H. C. (1982). Kappa coefficient. Encyclopedia of Statistical Sciences. Wiley.

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

Michaelides, M. P. (2008). An Illustration of a Mantel-Haenszel Procedure to Flag Misbehaving Common Items in Test Equating. *Practical Assessment Research & Evaluation, 13(7).* Available online: http://pareonline.net/pdf/v13n7.pdf

Muraki, E. (1997). The generalized partial credit model. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 153-164). New York: Springer Verlag.

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

Thissen, D., Chen, W-H., & Bock, R. D. (2003). *MUTILOG for Windows, Version 7* [Computer Software]. Lincolnwood, IL: Scientific Software International.

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.

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement, 5,* 245–262.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8,* 125-145.

Appendix A

Standards, Objectives/Skills, and Processes Assessed by Subject

Algebra I	
Standard 1: Number Sense and Algebraic Operations	
Standard 1.1	Equations and Formulas
	1.1a Translate
	1.1b Literal Equations
	1.1c Problem Solving with Formulas
	1.1d Problem Solving
Standard 1.2	Expressions
	1.2a Simplify expressions…
	1.2b Compute with polynomials…
	1.2c Factor polynomials
Standard 2: Relations and Functions	
Standard 2.1	Relations/Functions
	2.1a Distinguish linear and nonlinear
	2.1b Distinguish between relations…
	2.1c Dependent, Independ, Domain, Range
	2.1d Evaluate a function…
Standard 2.2	Linear Equations and Graphs
	2.2a Solve linear equations
	2.2b Graph Transformations
	2.2c Slope
	2.2d Equation of a Line
	2.2e Match to a graph, table, etc.
Standard 2.3	Linear Inequalities and Graphs
	2.3a Solve linear inequalities
	2.3b Match to a table, graph, etc.
Standard 2.4	Systems of Equations
Standard 3: Data Analysis, Probability & Statistics	
Standard 3.1	Data Analysis
	3.1a Data Representations
	3.1b Data Predictions
	3.1c Problem Solving
Standard 3.2	Line of Best Fit

Algebra II	
Standard 1: Number Sense and Algebraic Operations	
Standard 1.1	Rational Exponents
	1.1a Convert expressions from radical notations to rational exponents and vice versa.
	1.1b Add, subtract, multiply, divide, and simplify radical expressions and expressions containing rational exponents.
Standard 1.2	Polynomial and Rational Expressions
	1.2a Divide polynomial expressions by lower degree polynomials.
	1.2b Add, subtract, multiply, divide, and simplify rational expressions, including complex fractions.
Standard 1.3	Complex Numbers
	1.3b Add, subtract, multiply, divide, and simplify expressions involving complex numbers.
Standard 2: Relations and Functions	
Standard 2.1	Functions and Function Notation
	2.1a Recognize the parent graphs of polynomial, exponential, and logarithmic functions and predict the effects of transformations on the parent graphs, using various methods and tools which may include graphing calculators.
	2.1b Use function notation to add, subtract, multiply, and divide functions.
	2.1c Combine functions by composition.
	2.1d Use algebraic, interval, and set notations to specify the domain and range of functions of various types.
	2.1e Find and graph the inverse of a function, if it exists.
Standard 2.2	Systems of Equations
	2.2a Model a situation that can be described by a system of equations and inequalities and use the model to answer questions about the situation.
	2.2b Solve systems of linear equations and inequalities using various methods and tools which may include substitution, elimination, matrices, graphing, and graphing calculators.
	2.2c Use either one quadratic equation and one linear equation or two quadratic equations to solve problems.
Standard 2.3	Quadratic Equations and Functions
	2.3a Solve quadratic equations by graphing, factoring, completing the square and quadratic formula.
	2.3b Graph a quadratic function and identify the x- and y-intercepts and maximum or minimum value, using various methods and tools which may include a graphing calculator.
	2.3c Model a situation that can be described by a quadratic function and use the model to answer questions about the situation.

Algebra II continued	
Standard 2.4	Identify, graph, and write the equations of the conic sections (circle, ellipse, parabola, and hyperbola).
Standard 2.5	Exponential and Logarithmic Functions
	2.5a Graph exponential and logarithmic functions.
	2.5b Apply the inverse relationship between exponential and logarithmic functions to convert from one form to another.
	2.5c Model a situation that can be described by an exponential or logarithmic function and use the model to answer questions about the situation.
Standard 2.6	Polynomial Equations and Functions
	2.6a Solve polynomial equations using various methods and tools which may include factoring and synthetic division.
	2.6b Sketch the graph of a polynomial function.
	2.6c Given the graph of a polynomial function, identify the x- and y-intercepts, relative maximums and relative minimums, using various methods and tools which may include a graphing calculator.
	2.6d Model a situation that can be described by a polynomial function and use the model to answer questions about the situation.
Standard 2.7	Rational Equations and Functions
	2.7a Solve rational equations.
	2.7b Sketch the graph of a rational function.
	2.7c Given the graph of a rational function, identify the x- and y-intercepts, asymptotes, using various methods and tools which may include a graphing calculator.
	2.7d Model a situation that can be described by a rational function and use the model to answer questions about the situation.
Standard 3: Data Analysis, Probability, & Statistics	
Standard 3.1	Analysis of Collected Data …
	3.1a Display data on a scatter plot.
	3.1b Interpret results using a linear, exponential or quadratic model/equation.
	3.1c Identify whether the model/equation is a curve of best fit for the data, using various methods and tools which may include a graphing calculator.
Standard 3.3	Identify and use arithmetic and geometric sequences

Geometry	
Standard 1: Logical Reasoning	
Standard 1.1	Identify and use logical reasoning skills (inductive and deductive) to make and test conjectures, formulate counter examples, and follow logical arguments.
Standard 1.2	State, use, and examine the validity of the converse, inverse, and contrapositive of "if-then" statements.
Standard 2: Properties of 2-Dimensional Figures	
Standard 2.2	Line and Angle Relationships
	2.2a Use the angle relationships formed by parallel lines cut by a transversal to solve problems.
	2.2b Use the angle relationships formed by two lines cut by a transversal to determine if the two lines are parallel and verify, using algebraic and deductive proofs.
	2.2c Use relationships between pairs of angles (for example, adjacent, complementary, vertical) to solve problems.
Standard 2.3	Polygons and Other Plane Figures
	2.3a Identify, describe, and analyze polygons (for example, convex, concave, regular, pentagonal, hexagonal, n-gonal).
	2.3b Apply the interior and exterior angle sum of convex polygons to solve problems, and verify using algebraic and deductive proofs.
	2.3c Develop and apply the properties of quadrilaterals to solve problems (for example, rectangles, parallelograms, rhombi, trapezoids, kites).
	2.3d Use properties of 2-dimensional figures and side length, perimeter or circumference, and area to determine unknown values and correctly identify the appropriate unit of measure of each.
Standard 2.4	Similarity
	2.4a Determine and verify the relationships of similarity of triangles, using algebraic and deductive proofs.
	2.4b Use ratios of similar 2-dimensional figures to determine unknown values, such as angles, side lengths, perimeter or circumference, and area.
Standard 2.5	Congruence
	2.5a Determine and verify the relationships of congruency of triangles, using algebraic and deductive proofs.
	2.5b Use the relationships of congruency of 2-dimensional figures to determine unknown values, such as angles, side lengths, perimeter or circumference, and area.
Standard 2.6	Circles
	2.6a Find angle measures and arc measures related to circles.
	2.6b Find angle measures and segment lengths using the relationships among radii, chords, secants, and tangents of a circle.

Geometry continued	
Standard 3: Triangles and Trigonometric Ratios	
Standard 3.1	Use the Pythagorean Theorem and its converse to find missing side lengths and to determine acute, right, and obtuse triangles, and verify using algebraic and deductive proofs.
Standard 3.2	Apply the 45-45-90 and 30-60-90 right triangle relationships to solve problems, and verify using algebraic and deductive proofs.
Standard 3.3	Express the trigonometric functions as ratios and use sine, cosine, and tangent ratios to solve real-world problems.
Standard 4: Properties of 3-Dimensional Figures	
Standard 4.1	Polyhedra and Other Solids
	4.1a Identify, describe, and analyze polyhedra (for example, regular, decahedral).
	4.1b Use properties of 3-dimensional figures; side lengths, perimeter or circumference, and area of a face; and volume, lateral area, and surface area to determine unknown values and correctly identify the appropriate unit of measure of each.
Standard 4.2	Similarity and Congruence
	4.2a Use ratios of similar 3-dimensional figures to determine unknown values, such as angles, side lengths, perimeter or circumference of a face, area of a face, and volume.
	4.2b Use the relationships of congruency of 3-dimensional figures to determine unknown values, such as angles, side lengths, perimeter or circumference of a face, area of a face, and volume.
4.3	Create a model of a 3-dimensional figure from a 2-dimensional drawing and make a 2-dimensional representation of a 3-dimensional object (for example, nets, blueprints, perspective drawings).
Standard 5: Coordinate Geometry	
Standard 5.1	Use coordinate geometry to find the distance between two points; the midpoint of a segment; and to calculate the slopes of a parallel, perpendicular, horizontal, and vertical lines.
Standard 5.2	Properties of Figures
	5.2a Given a set of points determine the type of figure formed based on its properties.
	5.2b Use transformations (reflection, rotation, translation)on geometric figures to solve problems within coordinate geometry.

Biology I	
***PASS* Process/Inquiry Standards and Objectives**	
Process 1 Observe and Measure	
P1.1	Qualitative/quantitative observations and changes
P1.2 P1.3	Use appropriate System International (SI) units and tools
Process 2 Classify	
P2.1	Use observable properties to classify
P2.2	Identify properties of a classification system
Process 3 Experiment	
P3.1	Evaluate the design of investigations
P3.2 P3.4	Identify a testable hypothesis, variables, and control in an experiment
P3.3	Use mathematics to show relationships
P3.5	Identify potential hazards and practice safety procedures in all science activities
Process 4 Interpret and Communicate	
P4.1	Select predictions based on observed patterns of evidence
P4.3	Interpret line, bar, trend, and circle graphs
P4.4	Accept or reject a hypothesis
P4.5	Make logical conclusions based on experimental data
P4.8	Identify an appropriate graph or chart
Process 5 Model	
P5.1	Interpret a model which explains a given set of observations
P5.2	Select predictions based on models
***PASS* Content Standards**	
Standard 1 The Cell	
1.1	Cell structures and functions
1.2	Differentiation of cells
Standard 2 The Molecular Basis of Heredity	
2.1	DNA structure and function in heredity
2.2	Sorting and recombination of genes
Standard 3 Biological Diversity	
3.1	Variation among organisms
3.2	Natural selection and biological adaptations
Standard 4 The Interdependence of Organisms	
4.1	Earth cycles including abiotic and biotic factors
4.2	Organisms both cooperate and compete
4.3	Population dynamics
Standard 5 Matter/Energy/Organization in Living Systems	
5.1	Complexity and organization used for survival
5.2	Matter and energy flow in living and nonliving systems

Biology I continued	
Standard 6 The Behavior of Organisms	
6.1	Specialized cells
6.2	Behavior patterns can be used to ensure reproductive success

English II	
Reading/Literature	
Standard 1 Vocabulary	
Standard 2 Comprehension	
2.1	Literal Understanding
2.2	Inferences and Interpretation
2.3	Summary and Generalization
2.4	Analysis and Evaluation
Standard 3 Literature	
3.1	Literary Genres
3.2	Literary Elements
3.3	Figurative Language
3.4	Literary Works
Standard 4 Research and Information	
Writing/Grammar/Usage and Mechanics	
Standard 1/2 Writing	
	Writing Prompt
Standard 3 Grammar/Usage and Mechanics	
3.1	Standard Usage
3.2	Mechanics and Spelling
3.3	Sentence Structure

English III	
Reading/Literature	
Standard 1 Vocabulary	
Standard 2 Comprehension	
2.1	Literal Understanding
2.2	Inference and Interpretation
2.3	Summary and Generalization
2.4	Analysis and Evaluation
Standard 3 Literature	
3.1	Literary Genres
3.2	Literary Elements
3.3	Figurative Language
3.4	Literary Works
Standard 4 Research and Information	
Writing/Grammar/Usage and Mechanics	
Standard 1/2 Writing	
	Writing Prompt
Standard 3 Grammar/Usage and Mechanics	
3.1	Standard English Usage
3.2	Mechanics and Spelling
3.3	Sentence Structure
3.4	Manuscript Conventions

U.S. History	
Standard 1 Civil War/Reconstruction Era	
Standard 2 Impact of Immigration and Industrialization	
2.1	Immigration and Impact on Native Americans
2.2	Industrialization
Standard 3 Imperialism, World War I, and Isolationism	
3.1	American Imperialism
3.2	World War I and Isolationism
Standard 4 United States During the 1920s and 1930s	
4.1	Cultural Life Between the Wars
4.2	Economic Destabilization
4.3	The Great Depression, the Dust Bowl, and the New Deal
Standard 5 World War II	
5.1	Preparing for War
5.2	World War II
Standard 6 United States Since World War II	
6.1	Post War Foreign Policies and Events
6.2	Events Changing Domestic and Foreign Policies and Events
6.3	Post War Domestic Policies and Events

Appendix B

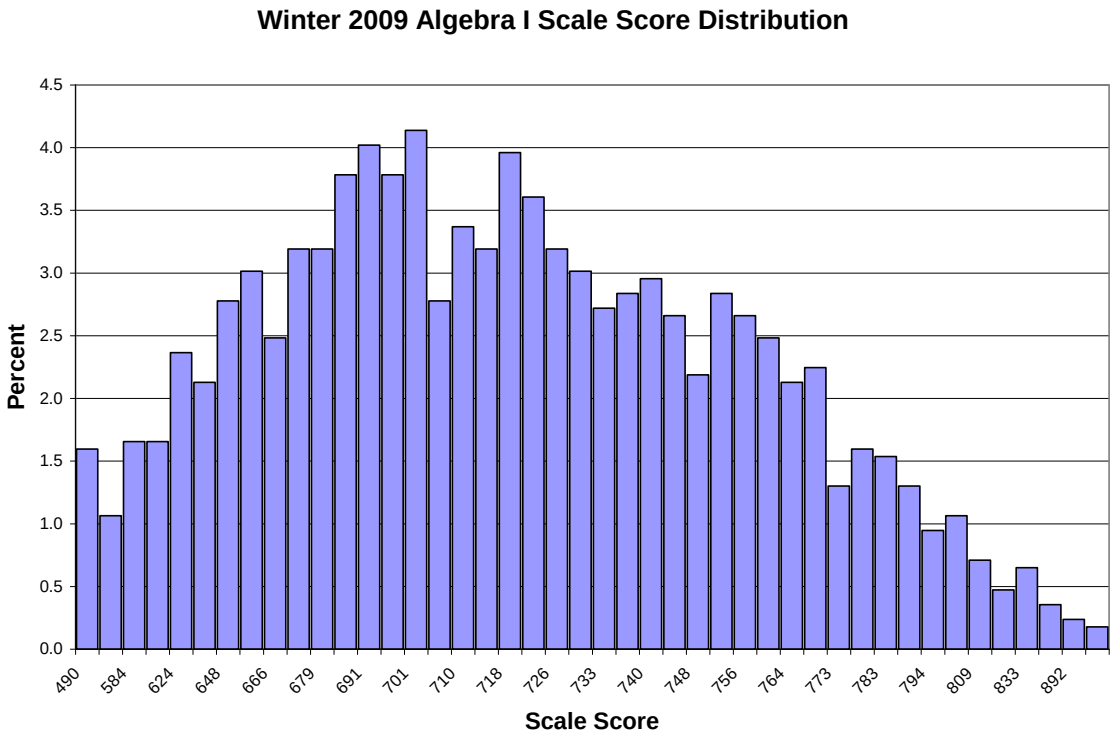
Scale Score Distributions for Winter/Trimester 2009-10

Algebra I Scale Score Distribution for Winter/Trimester 2009-10

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
490	27	1.6	27	1.6
547	18	1.1	45	2.7
584	28	1.7	73	4.3
607	28	1.7	101	6.0
624	40	2.4	141	8.3
637	36	2.1	177	10.5
648	47	2.8	224	13.2
662	51	3.0	275	16.3
666	42	2.5	317	18.7
673	54	3.2	371	21.9
679	54	3.2	425	25.1
685	64	3.8	489	28.9
691	68	4.0	557	32.9
700	64	3.8	621	36.7
701	70	4.1	691	40.8
705	47	2.8	738	43.6
710	57	3.4	795	47.0
714	54	3.2	849	50.2
718	67	4.0	916	54.1
722	61	3.6	977	57.7
726	54	3.2	1,031	60.9
729	51	3.0	1,082	63.9
733	46	2.7	1,128	66.7
737	48	2.8	1,176	69.5
740	50	3.0	1,226	72.5
744	45	2.7	1,271	75.1
748	37	2.2	1,308	77.3
752	48	2.8	1,356	80.1
756	45	2.7	1,401	82.8
762	42	2.5	1,443	85.3
764	36	2.1	1,479	87.4
768	38	2.2	1,517	89.7
773	22	1.3	1,539	91.0
777	27	1.6	1,566	92.6
783	26	1.5	1,592	94.1
788	22	1.3	1,614	95.4
794	16	0.9	1,630	96.3
801	18	1.1	1,648	97.4
809	12	0.7	1,660	98.1
819	8	0.5	1,668	98.6
833	11	0.7	1,679	99.2
853	6	0.4	1,685	99.6
892	4	0.2	1,689	99.8

Algebra I Scale Score Distribution for Winter/Trimester 2009-10 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
999	3	0.2	1,692	100.0

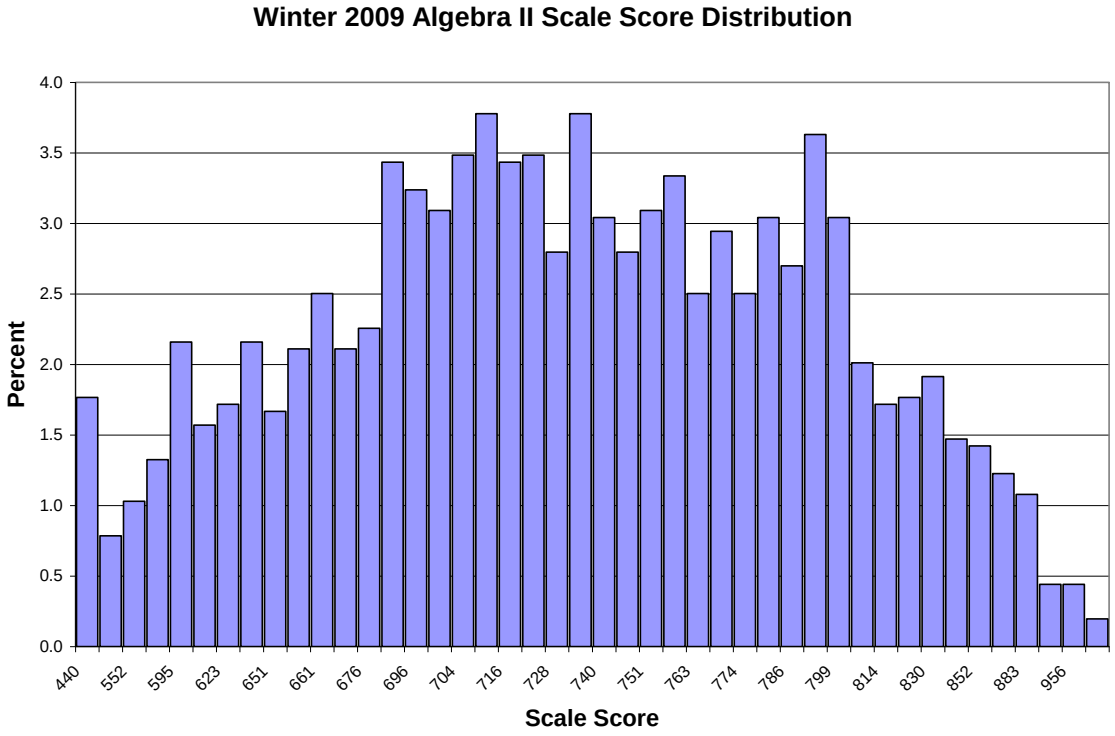


Algebra II Scale Score Distribution for Winter/Trimester 2009-10

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	36	1.8	36	1.8
512	16	0.8	52	2.6
552	21	1.0	73	3.6
577	27	1.3	100	4.9
595	44	2.2	144	7.1
610	32	1.6	176	8.6
623	35	1.7	211	10.4
634	44	2.2	255	12.5
651	34	1.7	289	14.2
652	43	2.1	332	16.3
661	51	2.5	383	18.8
669	43	2.1	426	20.9
676	46	2.3	472	23.2
683	70	3.4	542	26.6
696	66	3.2	608	29.8
697	63	3.1	671	32.9
704	71	3.5	742	36.4
710	77	3.8	819	40.2
716	70	3.4	889	43.6
722	71	3.5	960	47.1
728	57	2.8	1,017	49.9
734	77	3.8	1,094	53.7
740	62	3.0	1,156	56.7
746	57	2.8	1,213	59.5
751	63	3.1	1,276	62.6
757	68	3.3	1,344	65.9
763	51	2.5	1,395	68.4
768	60	2.9	1,455	71.4
774	51	2.5	1,506	73.9
780	62	3.0	1,568	76.9
786	55	2.7	1,623	79.6
793	74	3.6	1,697	83.3
799	62	3.0	1,759	86.3
806	41	2.0	1,800	88.3
814	35	1.7	1,835	90.0
822	36	1.8	1,871	91.8
830	39	1.9	1,910	93.7
840	30	1.5	1,940	95.2
852	29	1.4	1,969	96.6
866	25	1.2	1,994	97.8
883	22	1.1	2,016	98.9
909	9	0.4	2,025	99.4

Algebra II Scale Score Distribution for Winter/Trimester 2009-10 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
956	9	0.4	2,034	99.8
999	4	0.2	2,038	100.0

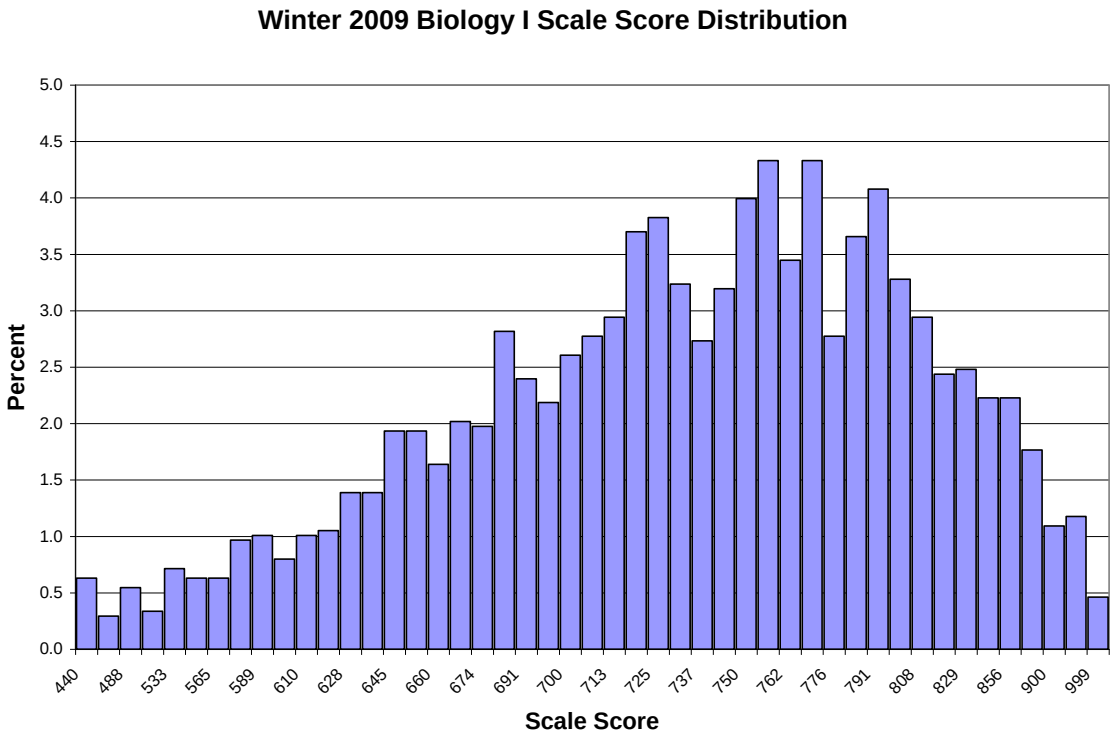


Biology I Scale Score Distribution for Winter/Trimester 2009-10

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	15	0.6	15	0.6
456	7	0.3	22	0.9
488	13	0.5	35	1.5
513	8	0.3	43	1.8
533	17	0.7	60	2.5
550	15	0.6	75	3.2
565	15	0.6	90	3.8
578	23	1.0	113	4.7
589	24	1.0	137	5.8
600	19	0.8	156	6.6
610	24	1.0	180	7.6
627	25	1.1	205	8.6
628	33	1.4	238	10.0
637	33	1.4	271	11.4
645	46	1.9	317	13.3
652	46	1.9	363	15.3
660	39	1.6	402	16.9
667	48	2.0	450	18.9
674	47	2.0	497	20.9
681	67	2.8	564	23.7
691	57	2.4	621	26.1
694	52	2.2	673	28.3
700	62	2.6	735	30.9
706	66	2.8	801	33.7
713	70	2.9	871	36.6
719	88	3.7	959	40.3
725	91	3.8	1,050	44.1
731	77	3.2	1,127	47.4
737	65	2.7	1,192	50.1
743	76	3.2	1,268	53.3
750	95	4.0	1,363	57.3
756	103	4.3	1,466	61.6
762	82	3.4	1,548	65.1
775	103	4.3	1,651	69.4
776	66	2.8	1,717	72.2
783	87	3.7	1,804	75.8
791	97	4.1	1,901	79.9
799	78	3.3	1,979	83.2
808	70	2.9	2,049	86.1
818	58	2.4	2,107	88.6
829	59	2.5	2,166	91.0
841	53	2.2	2,219	93.3
856	53	2.2	2,272	95.5

Biology I Scale Score Distribution for Winter/Trimester 2009-10 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
875	42	1.8	2,314	97.3
900	26	1.1	2,340	98.4
937	28	1.2	2,368	99.5
999	11	0.5	2,379	100.0

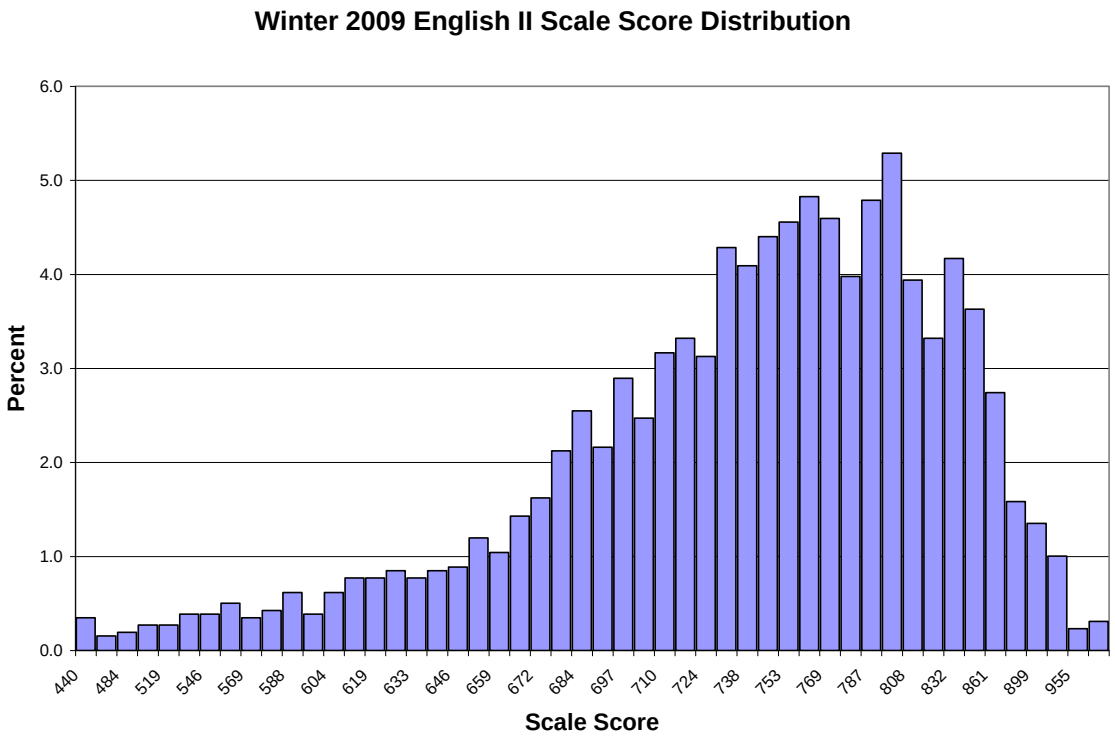


English II Scale Score Distribution for Winter/Trimester 2009-10

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	9	0.3	9	0.3
461	4	0.2	13	0.5
484	5	0.2	18	0.7
503	7	0.3	25	1.0
519	7	0.3	32	1.2
534	10	0.4	42	1.6
546	10	0.4	52	2.0
558	13	0.5	65	2.5
569	9	0.3	74	2.9
578	11	0.4	85	3.3
588	16	0.6	101	3.9
596	10	0.4	111	4.3
604	16	0.6	127	4.9
612	20	0.8	147	5.7
619	20	0.8	167	6.4
626	22	0.8	189	7.3
633	20	0.8	209	8.1
640	22	0.8	231	8.9
646	23	0.9	254	9.8
653	31	1.2	285	11.0
659	27	1.0	312	12.0
666	37	1.4	349	13.5
672	42	1.6	391	15.1
678	55	2.1	446	17.2
684	66	2.5	512	19.8
693	56	2.2	568	21.9
697	75	2.9	643	24.8
704	64	2.5	707	27.3
710	82	3.2	789	30.5
717	86	3.3	875	33.8
724	81	3.1	956	36.9
731	111	4.3	1,067	41.2
738	106	4.1	1,173	45.3
745	114	4.4	1,287	49.7
753	118	4.6	1,405	54.2
761	125	4.8	1,530	59.1
769	119	4.6	1,649	63.7
778	103	4.0	1,752	67.6
787	124	4.8	1,876	72.4
797	137	5.3	2,013	77.7
808	102	3.9	2,115	81.7
819	86	3.3	2,201	85.0
832	108	4.2	2,309	89.2

English II Scale Score Distribution for Winter/Trimester 2009-10 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
846	94	3.6	2,403	92.8
861	71	2.7	2,474	95.5
878	41	1.6	2,515	97.1
899	35	1.4	2,550	98.5
923	26	1.0	2,576	99.5
955	6	0.2	2,582	99.7
999	8	0.3	2,590	100.0

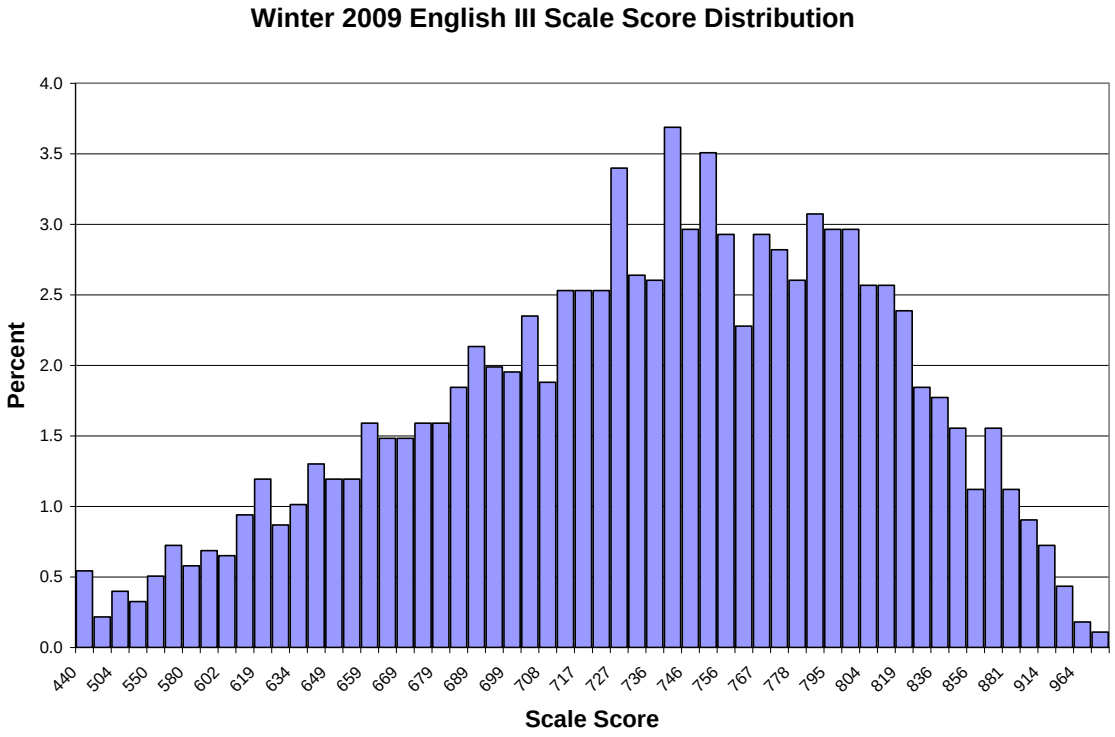


English III Scale Score Distribution for Winter/Trimester 2009-10

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	15	0.5	15	0.5
461	6	0.2	21	0.8
504	11	0.4	32	1.2
531	9	0.3	41	1.5
550	14	0.5	55	2.0
566	20	0.7	75	2.7
580	16	0.6	91	3.3
591	19	0.7	110	4.0
602	18	0.7	128	4.6
611	26	0.9	154	5.6
619	33	1.2	187	6.8
627	24	0.9	211	7.6
634	28	1.0	239	8.6
641	36	1.3	275	9.9
649	33	1.2	308	11.1
653	33	1.2	341	12.3
659	44	1.6	385	13.9
664	41	1.5	426	15.4
669	41	1.5	467	16.9
674	44	1.6	511	18.5
679	44	1.6	555	20.1
684	51	1.8	606	21.9
689	59	2.1	665	24.0
695	55	2.0	720	26.0
699	54	2.0	774	28.0
703	65	2.3	839	30.3
708	52	1.9	891	32.2
713	70	2.5	961	34.7
717	70	2.5	1,031	37.3
722	70	2.5	1,101	39.8
727	94	3.4	1,195	43.2
731	73	2.6	1,268	45.8
736	72	2.6	1,340	48.4
741	102	3.7	1,442	52.1
746	82	3.0	1,524	55.1
751	97	3.5	1,621	58.6
756	81	2.9	1,702	61.5
761	63	2.3	1,765	63.8
767	81	2.9	1,846	66.7
772	78	2.8	1,924	69.6
778	72	2.6	1,996	72.2
784	85	3.1	2,081	75.2
795	82	3.0	2,163	78.2

English III Scale Score Distribution for Winter/Trimester 2009-10 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
797	82	3.0	2,245	81.2
804	71	2.6	2,316	83.7
811	71	2.6	2,387	86.3
819	66	2.4	2,453	88.7
827	51	1.8	2,504	90.5
836	49	1.8	2,553	92.3
846	43	1.6	2,596	93.9
856	31	1.1	2,627	95.0
868	43	1.6	2,670	96.5
881	31	1.1	2,701	97.7
896	25	0.9	2,726	98.6
914	20	0.7	2,746	99.3
936	12	0.4	2,758	99.7
964	5	0.2	2,763	99.9
999	3	0.1	2,766	100.0

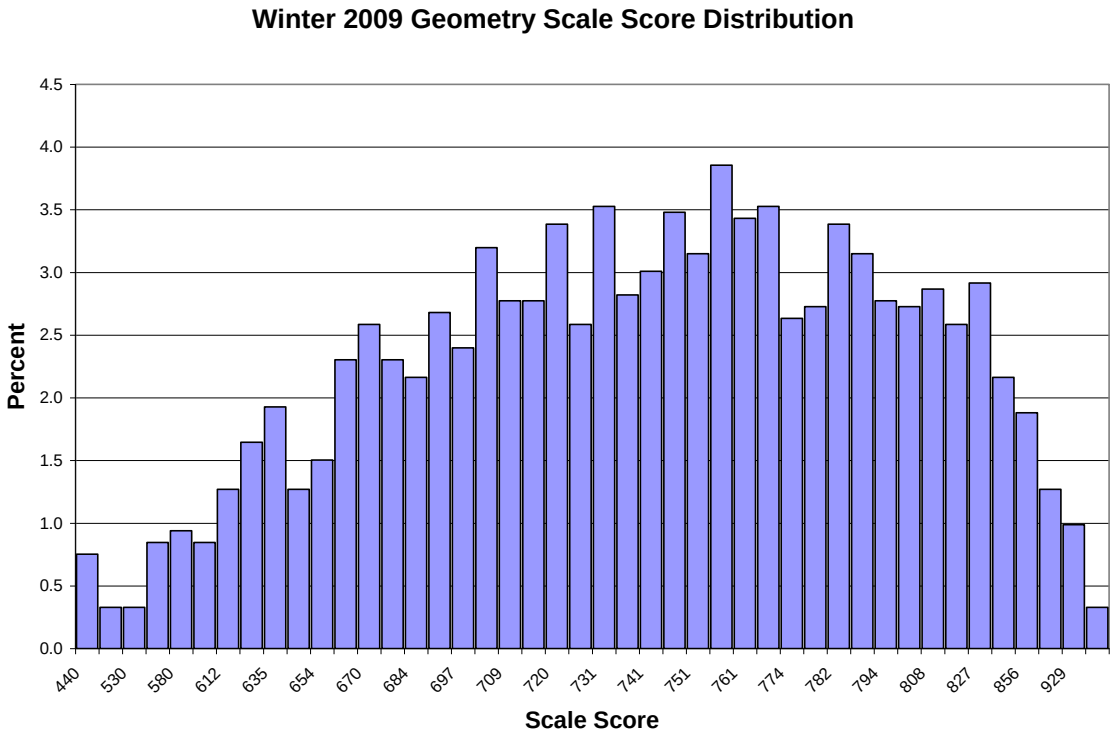


Geometry Scale Score Distribution for Winter/Trimester 2009-10

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	16	0.8	16	0.8
486	7	0.3	23	1.1
530	7	0.3	30	1.4
559	18	0.8	48	2.3
580	20	0.9	68	3.2
597	18	0.8	86	4.0
612	27	1.3	113	5.3
624	35	1.6	148	7.0
635	41	1.9	189	8.9
645	27	1.3	216	10.2
654	32	1.5	248	11.7
662	49	2.3	297	14.0
670	55	2.6	352	16.5
677	49	2.3	401	18.9
684	46	2.2	447	21.0
695	57	2.7	504	23.7
697	51	2.4	555	26.1
703	68	3.2	623	29.3
709	59	2.8	682	32.1
715	59	2.8	741	34.8
720	72	3.4	813	38.2
725	55	2.6	868	40.8
731	75	3.5	943	44.3
736	60	2.8	1,003	47.2
741	64	3.0	1,067	50.2
746	74	3.5	1,141	53.6
751	67	3.1	1,208	56.8
756	82	3.9	1,290	60.6
761	73	3.4	1,363	64.1
766	75	3.5	1,438	67.6
774	56	2.6	1,494	70.2
777	58	2.7	1,552	73.0
782	72	3.4	1,624	76.4
788	67	3.1	1,691	79.5
794	59	2.8	1,750	82.3
801	58	2.7	1,808	85.0
808	61	2.9	1,869	87.9
817	55	2.6	1,924	90.5
827	62	2.9	1,986	93.4
839	46	2.2	2,032	95.5
856	40	1.9	2,072	97.4
881	27	1.3	2,099	98.7

Geometry Scale Score Distribution for Winter/Trimester 2009-10 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
929	21	1.0	2,120	99.7
999	7	0.3	2,127	100.0

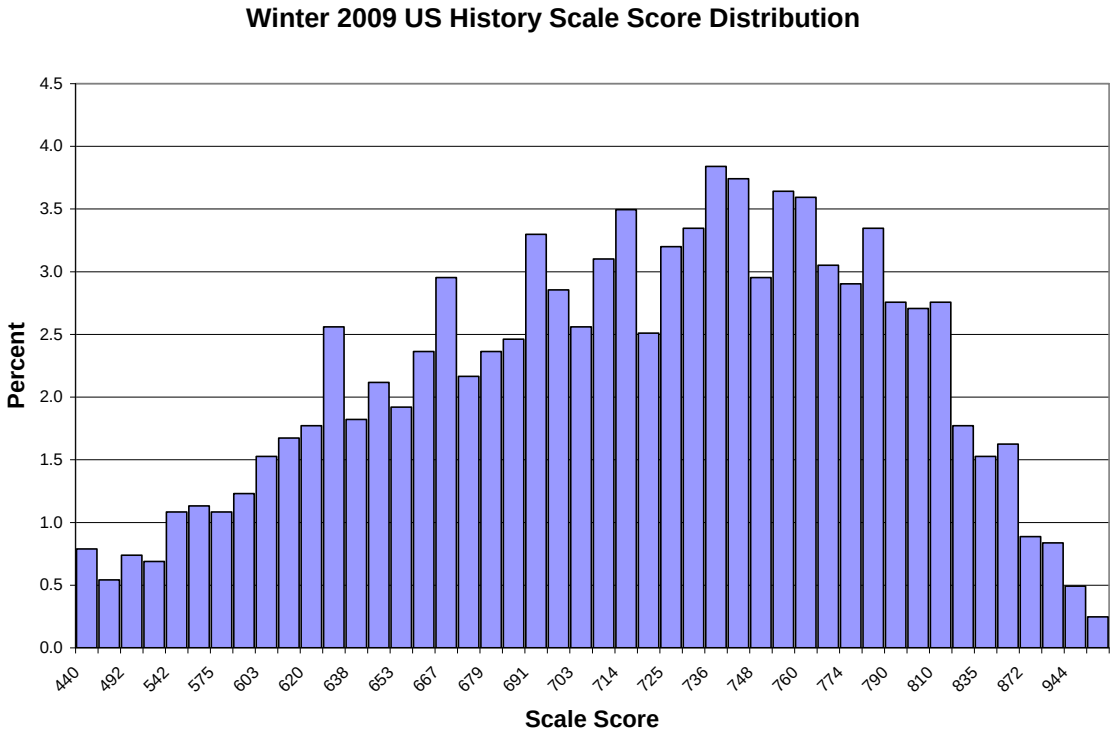


U.S. History Scale Score Distribution for Winter/Trimester 2009-10

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	16	0.8	16	0.8
452	11	0.5	27	1.3
492	15	0.7	42	2.1
520	14	0.7	56	2.8
542	22	1.1	78	3.8
559	23	1.1	101	5.0
575	22	1.1	123	6.1
588	25	1.2	148	7.3
603	31	1.5	179	8.8
611	34	1.7	213	10.5
620	36	1.8	249	12.3
629	52	2.6	301	14.8
638	37	1.8	338	16.6
646	43	2.1	381	18.8
653	39	1.9	420	20.7
660	48	2.4	468	23.0
667	60	3.0	528	26.0
673	44	2.2	572	28.1
679	48	2.4	620	30.5
689	50	2.5	670	33.0
691	67	3.3	737	36.3
697	58	2.9	795	39.1
703	52	2.6	847	41.7
708	63	3.1	910	44.8
714	71	3.5	981	48.3
719	51	2.5	1,032	50.8
725	65	3.2	1,097	54.0
730	68	3.3	1,165	57.3
736	78	3.8	1,243	61.2
747	76	3.7	1,319	64.9
748	60	3.0	1,379	67.9
754	74	3.6	1,453	71.5
760	73	3.6	1,526	75.1
767	62	3.1	1,588	78.1
774	59	2.9	1,647	81.1
782	68	3.3	1,715	84.4
790	56	2.8	1,771	87.2
799	55	2.7	1,826	89.9
810	56	2.8	1,882	92.6
821	36	1.8	1,918	94.4
835	31	1.5	1,949	95.9
851	33	1.6	1,982	97.5
872	18	0.9	2,000	98.4

U.S. History Scale Score Distribution for Winter/Trimester 2009-10 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
900	17	0.8	2,017	99.3
944	10	0.5	2,027	99.8
999	5	0.2	2,032	100.0



Appendix C

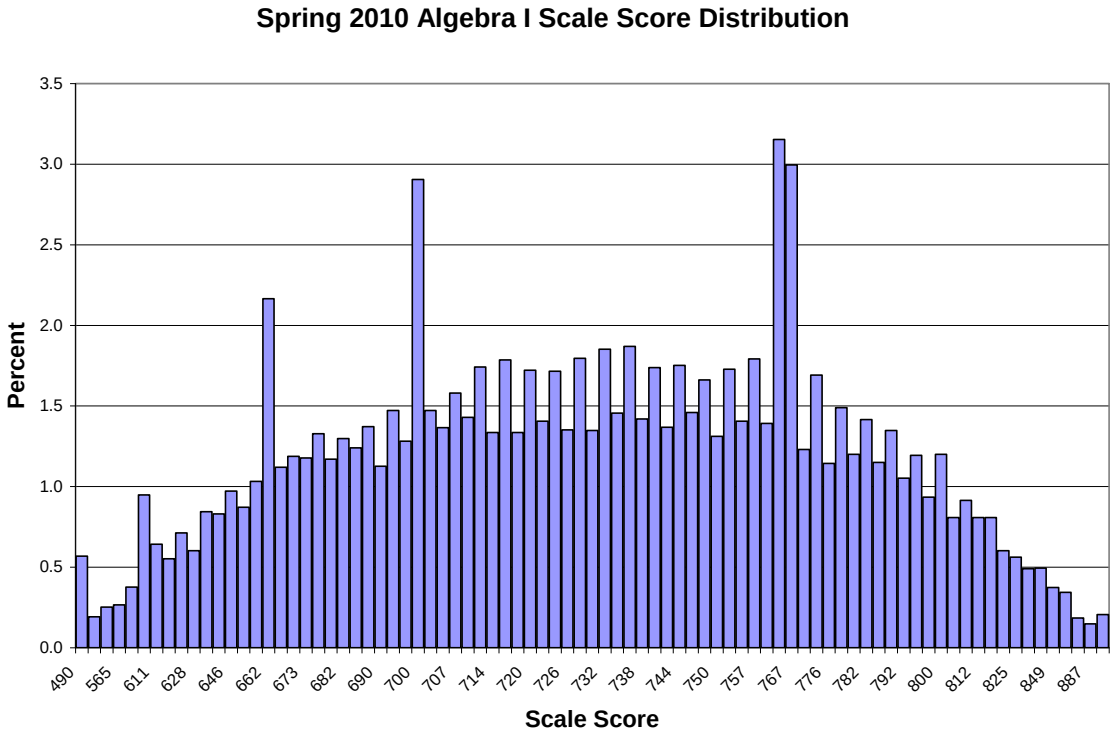
Scale Score Distributions for Spring 2010

Algebra I Score Distribution for Spring 2010

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
490	169	0.6	169	0.6
513	57	0.2	226	0.8
538	75	0.3	301	1.0
565	79	0.3	380	1.3
572	112	0.4	492	1.7
594	282	0.9	774	2.6
611	191	0.6	965	3.2
613	164	0.6	1,129	3.8
625	212	0.7	1,341	4.5
628	179	0.6	1,520	5.1
636	251	0.8	1,771	6.0
639	247	0.8	2,018	6.8
646	289	1.0	2,307	7.8
649	259	0.9	2,566	8.6
655	307	1.0	2,873	9.7
662	644	2.2	3,517	11.8
666	333	1.1	3,850	12.9
669	353	1.2	4,203	14.1
673	350	1.2	4,553	15.3
676	395	1.3	4,948	16.6
679	348	1.2	5,296	17.8
682	386	1.3	5,682	19.1
685	369	1.2	6,051	20.3
687	408	1.4	6,459	21.7
690	335	1.1	6,794	22.8
692	438	1.5	7,232	24.3
696	381	1.3	7,613	25.6
700	864	2.9	8,477	28.5
702	438	1.5	8,915	30.0
705	406	1.4	9,321	31.3
707	470	1.6	9,791	32.9
709	425	1.4	10,216	34.3
711	518	1.7	10,734	36.1
714	397	1.3	11,131	37.4
715	531	1.8	11,662	39.2
718	397	1.3	12,059	40.5
720	512	1.7	12,571	42.3
722	418	1.4	12,989	43.7
724	510	1.7	13,499	45.4
726	402	1.4	13,901	46.7
728	534	1.8	14,435	48.5
730	401	1.3	14,836	49.9

Algebra I Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
732	551	1.9	15,387	51.7
734	433	1.5	15,820	53.2
736	556	1.9	16,376	55.0
738	422	1.4	16,798	56.5
740	517	1.7	17,315	58.2
742	407	1.4	17,722	59.6
744	521	1.8	18,243	61.3
746	434	1.5	18,677	62.8
749	494	1.7	19,171	64.4
750	390	1.3	19,561	65.8
753	514	1.7	20,075	67.5
754	418	1.4	20,493	68.9
757	533	1.8	21,026	70.7
758	414	1.4	21,440	72.1
762	938	3.2	22,378	75.2
767	891	3.0	23,269	78.2
771	366	1.2	23,635	79.4
772	503	1.7	24,138	81.1
776	340	1.1	24,478	82.3
777	443	1.5	24,921	83.8
781	357	1.2	25,278	85.0
782	421	1.4	25,699	86.4
786	342	1.1	26,041	87.5
788	401	1.3	26,442	88.9
792	313	1.1	26,755	89.9
794	355	1.2	27,110	91.1
798	278	0.9	27,388	92.1
800	357	1.2	27,745	93.3
805	240	0.8	27,985	94.1
807	272	0.9	28,257	95.0
812	240	0.8	28,497	95.8
815	240	0.8	28,737	96.6
821	179	0.6	28,916	97.2
825	167	0.6	29,083	97.8
833	146	0.5	29,229	98.3
838	147	0.5	29,376	98.7
849	111	0.4	29,487	99.1
855	102	0.3	29,589	99.5
877	55	0.2	29,644	99.6
887	44	0.1	29,688	99.8
999	61	0.2	29,749	100.0

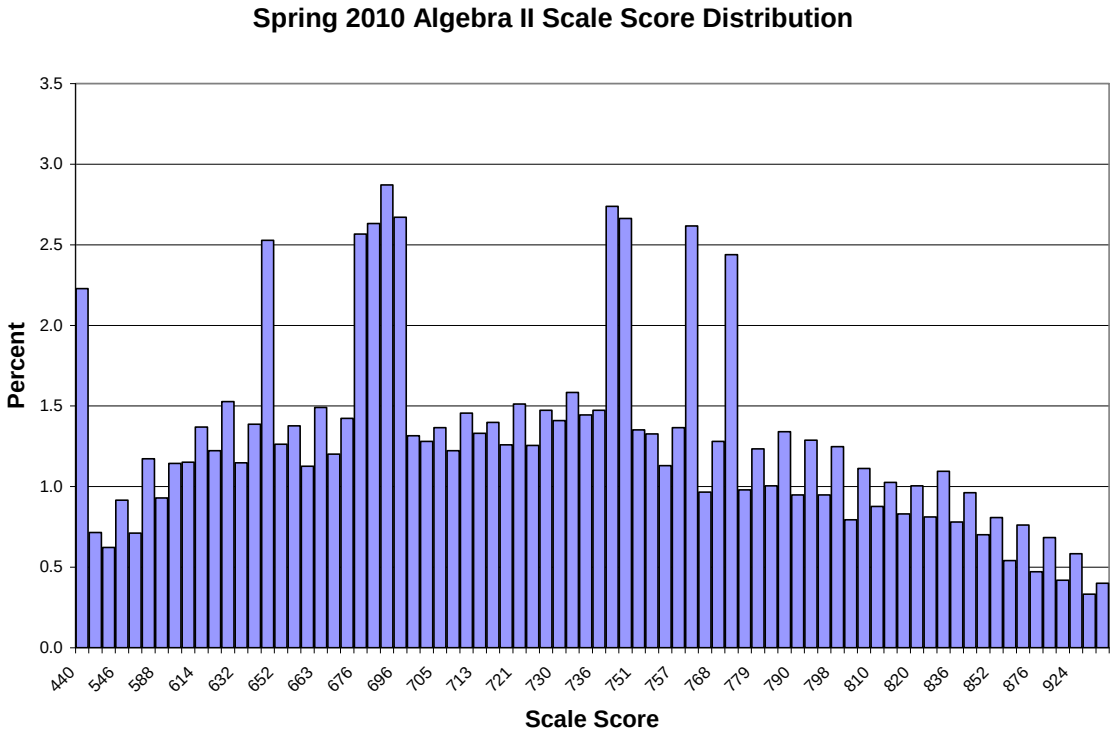

Spring 2010 Algebra I Scale Score Distribution
Percent
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
490
565
611
628
646
662
673
682
690
700
707
714
720
726
732
738
744
750
757
767
776
782
792
800
812
825
849
887
Scale Score

Algebra II Score Distribution for Spring 2010

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	623	2.2	623	2.2
480	200	0.7	823	2.9
526	174	0.6	997	3.6
546	256	0.9	1,253	4.5
564	199	0.7	1,452	5.2
578	328	1.2	1,780	6.4
588	260	0.9	2,040	7.3
599	320	1.1	2,360	8.4
606	322	1.2	2,682	9.6
614	383	1.4	3,065	11.0
620	342	1.2	3,407	12.2
627	427	1.5	3,834	13.7
632	321	1.1	4,155	14.9
638	388	1.4	4,543	16.2
651	707	2.5	5,250	18.8
652	353	1.3	5,603	20.0
655	385	1.4	5,988	21.4
660	315	1.1	6,303	22.5
663	417	1.5	6,720	24.0
668	336	1.2	7,056	25.2
670	398	1.4	7,454	26.6
676	718	2.6	8,172	29.2
683	736	2.6	8,908	31.8
689	803	2.9	9,711	34.7
696	747	2.7	10,458	37.4
700	368	1.3	10,826	38.7
702	358	1.3	11,184	40.0
705	382	1.4	11,566	41.3
708	342	1.2	11,908	42.6
710	407	1.5	12,315	44.0
713	372	1.3	12,687	45.4
716	391	1.4	13,078	46.7
719	352	1.3	13,430	48.0
721	423	1.5	13,853	49.5
724	351	1.3	14,204	50.8
726	412	1.5	14,616	52.2
730	394	1.4	15,010	53.7
731	443	1.6	15,453	55.2
735	404	1.4	15,857	56.7
736	412	1.5	16,269	58.2
741	766	2.7	17,035	60.9
746	745	2.7	17,780	63.6

Algebra II Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
751	378	1.4	18,158	64.9
752	371	1.3	18,529	66.2
756	316	1.1	18,845	67.4
757	382	1.4	19,227	68.7
762	732	2.6	19,959	71.3
767	270	1.0	20,229	72.3
768	358	1.3	20,587	73.6
774	682	2.4	21,269	76.0
778	274	1.0	21,543	77.0
779	345	1.2	21,888	78.2
784	281	1.0	22,169	79.2
785	375	1.3	22,544	80.6
790	265	0.9	22,809	81.5
791	360	1.3	23,169	82.8
796	265	0.9	23,434	83.8
798	349	1.2	23,783	85.0
803	222	0.8	24,005	85.8
804	311	1.1	24,316	86.9
810	245	0.9	24,561	87.8
812	287	1.0	24,848	88.8
818	232	0.8	25,080	89.7
820	281	1.0	25,361	90.7
826	227	0.8	25,588	91.5
829	306	1.1	25,894	92.6
836	218	0.8	26,112	93.3
839	269	1.0	26,381	94.3
847	196	0.7	26,577	95.0
852	226	0.8	26,803	95.8
860	151	0.5	26,954	96.4
867	213	0.8	27,167	97.1
876	132	0.5	27,299	97.6
888	191	0.7	27,490	98.3
898	117	0.4	27,607	98.7
924	163	0.6	27,770	99.3
936	93	0.3	27,863	99.6
999	112	0.4	27,975	100.0


Spring 2010 Algebra II Scale Score Distribution
Percent
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
440
546
588
614
632
652
663
676
696
705
713
721
730
736
751
757
768
779
790
798
810
820
836
852
876
924
Scale Score

Biology I Score Distribution for Spring 2010

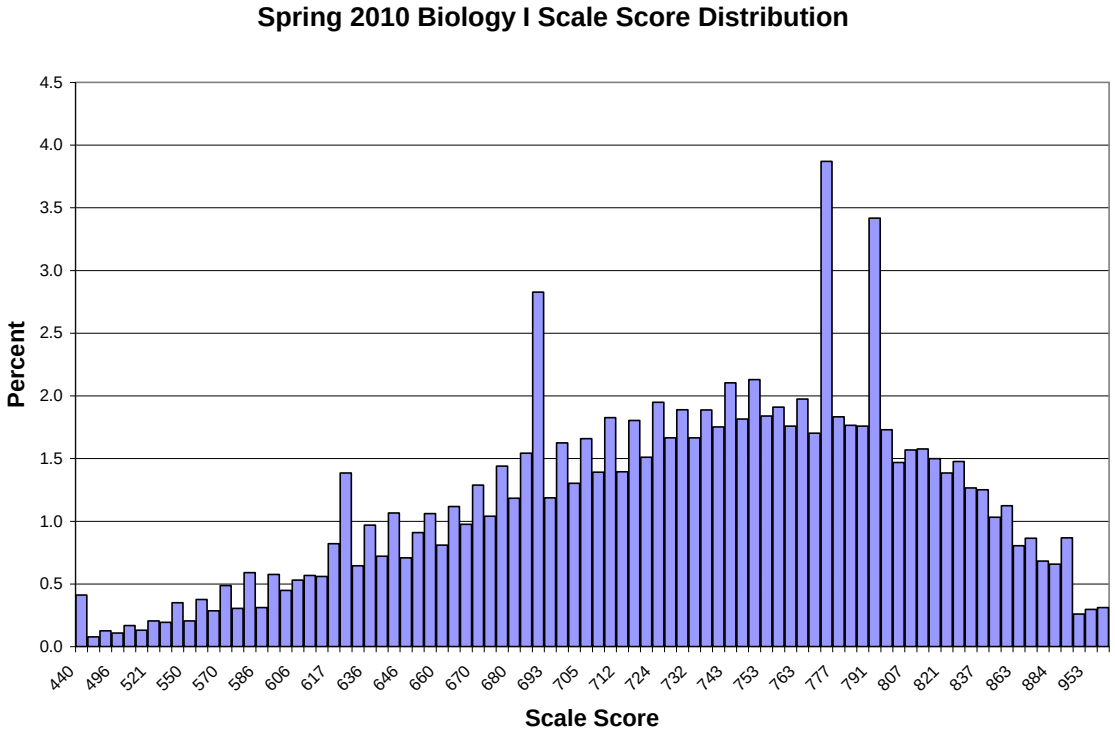
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	144	0.4	144	0.4
469	27	0.1	171	0.5
472	44	0.1	215	0.6
496	38	0.1	253	0.7
499	59	0.2	312	0.9
517	46	0.1	358	1.0
521	72	0.2	430	1.2
534	68	0.2	498	1.4
540	123	0.4	621	1.8
550	72	0.2	693	2.0
556	132	0.4	825	2.4
563	100	0.3	925	2.6
570	171	0.5	1,096	3.1
575	107	0.3	1,203	3.4
583	207	0.6	1,410	4.0
586	109	0.3	1,519	4.3
595	202	0.6	1,721	4.9
596	157	0.4	1,878	5.4
606	186	0.5	2,064	5.9
607	199	0.6	2,263	6.4
614	196	0.6	2,459	7.0
617	288	0.8	2,747	7.8
627	486	1.4	3,233	9.2
631	226	0.6	3,459	9.9
636	340	1.0	3,799	10.8
638	253	0.7	4,052	11.5
645	374	1.1	4,426	12.6
646	248	0.7	4,674	13.3
653	319	0.9	4,993	14.2
654	372	1.1	5,365	15.3
660	284	0.8	5,649	16.1
662	392	1.1	6,041	17.2
667	342	1.0	6,383	18.2
670	452	1.3	6,835	19.5
673	365	1.0	7,200	20.5
677	505	1.4	7,705	22.0
680	415	1.2	8,120	23.1
684	541	1.5	8,661	24.7
691	992	2.8	9,653	27.5
693	416	1.2	10,069	28.7
698	570	1.6	10,639	30.3
699	457	1.3	11,096	31.6

Biology I Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
705	582	1.7	11,678	33.3
706	488	1.4	12,166	34.7
711	641	1.8	12,807	36.5
712	489	1.4	13,296	37.9
718	633	1.8	13,929	39.7
719	530	1.5	14,459	41.2
724	684	1.9	15,143	43.2
726	584	1.7	15,727	44.8
731	663	1.9	16,390	46.7
732	584	1.7	16,974	48.4
737	662	1.9	17,636	50.3
739	615	1.8	18,251	52.0
743	738	2.1	18,989	54.1
746	637	1.8	19,626	55.9
750	747	2.1	20,373	58.1
753	645	1.8	21,018	59.9
756	670	1.9	21,688	61.8
760	617	1.8	22,305	63.6
763	693	2.0	22,998	65.5
767	597	1.7	23,595	67.2
775	1358	3.9	24,953	71.1
777	643	1.8	25,596	72.9
783	619	1.8	26,215	74.7
784	617	1.8	26,832	76.5
791	1199	3.4	28,031	79.9
799	607	1.7	28,638	81.6
800	515	1.5	29,153	83.1
807	550	1.6	29,703	84.7
810	553	1.6	30,256	86.2
816	526	1.5	30,782	87.7
821	486	1.4	31,268	89.1
826	518	1.5	31,786	90.6
833	444	1.3	32,230	91.9
837	439	1.3	32,669	93.1
847	362	1.0	33,031	94.1
850	394	1.1	33,425	95.3
863	282	0.8	33,707	96.1
865	303	0.9	34,010	96.9
883	239	0.7	34,249	97.6
884	231	0.7	34,480	98.3
910	304	0.9	34,784	99.1
952	91	0.3	34,875	99.4

Biology I Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
953	104	0.3	34,979	99.7
999	109	0.3	35,088	100.0



English II Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	39	0.1	39	0.1
450	8	0.0	47	0.1
451	8	0.0	55	0.2
475	17	0.0	72	0.2
476	14	0.0	86	0.2
494	9	0.0	95	0.3
495	17	0.0	112	0.3
496	3	0.0	115	0.3
510	19	0.1	134	0.4
511	8	0.0	142	0.4
512	12	0.0	154	0.4
523	14	0.0	168	0.5
524	18	0.1	186	0.5
525	19	0.1	205	0.6
535	18	0.1	223	0.6
536	17	0.0	240	0.7
537	27	0.1	267	0.8
546	43	0.1	310	0.9
547	16	0.0	326	0.9
548	11	0.0	337	1.0
555	25	0.1	362	1.0
556	20	0.1	382	1.1
557	32	0.1	414	1.2
564	27	0.1	441	1.2
565	26	0.1	467	1.3
566	37	0.1	504	1.4
572	15	0.0	519	1.5
573	37	0.1	556	1.6
574	36	0.1	592	1.7
580	30	0.1	622	1.8
581	36	0.1	658	1.9
588	119	0.3	777	2.2
589	30	0.1	807	2.3
590	27	0.1	834	2.4
595	75	0.2	909	2.6
596	36	0.1	945	2.7
597	24	0.1	969	2.7
602	89	0.3	1,058	3.0
603	28	0.1	1,086	3.1
604	23	0.1	1,109	3.1
608	42	0.1	1,151	3.2
609	75	0.2	1,226	3.5

English II Score Distribution for Spring 2010 (cont.)

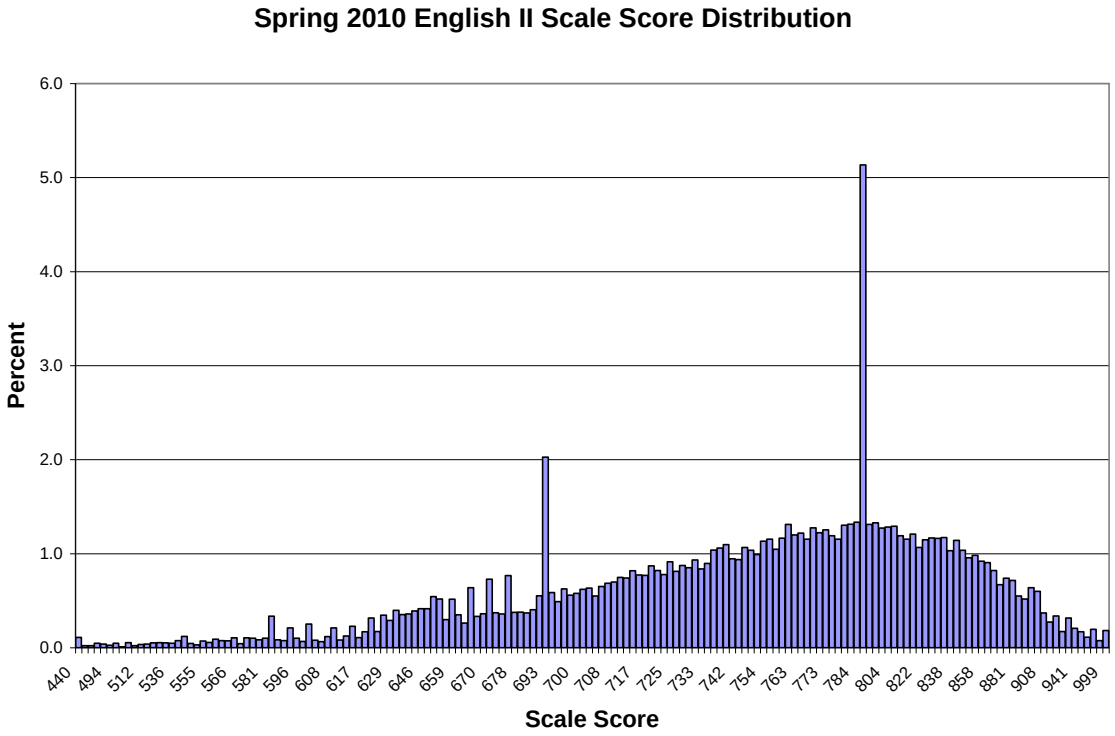
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
610	29	0.1	1,255	3.5
615	44	0.1	1,299	3.7
616	81	0.2	1,380	3.9
617	38	0.1	1,418	4.0
621	60	0.2	1,478	4.2
622	112	0.3	1,590	4.5
623	61	0.2	1,651	4.7
628	123	0.3	1,774	5.0
629	103	0.3	1,877	5.3
634	141	0.4	2,018	5.7
635	125	0.4	2,143	6.0
640	127	0.4	2,270	6.4
641	139	0.4	2,409	6.8
646	147	0.4	2,556	7.2
647	147	0.4	2,703	7.6
652	193	0.5	2,896	8.2
653	184	0.5	3,080	8.7
658	106	0.3	3,186	9.0
659	183	0.5	3,369	9.5
660	124	0.3	3,493	9.8
664	93	0.3	3,586	10.1
665	226	0.6	3,812	10.7
666	118	0.3	3,930	11.1
670	128	0.4	4,058	11.4
671	258	0.7	4,316	12.2
672	132	0.4	4,448	12.5
676	127	0.4	4,575	12.9
677	272	0.8	4,847	13.7
678	133	0.4	4,980	14.0
682	134	0.4	5,114	14.4
683	131	0.4	5,245	14.8
684	143	0.4	5,388	15.2
685	196	0.6	5,584	15.7
693	719	2.0	6,303	17.8
694	208	0.6	6,511	18.4
695	174	0.5	6,685	18.9
697	222	0.6	6,907	19.5
698	198	0.6	7,105	20.0
700	205	0.6	7,310	20.6
701	220	0.6	7,530	21.2
703	225	0.6	7,755	21.9
704	195	0.5	7,950	22.4

English II Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
707	231	0.7	8,181	23.1
708	243	0.7	8,424	23.8
710	248	0.7	8,672	24.5
711	265	0.7	8,937	25.2
713	263	0.7	9,200	25.9
714	290	0.8	9,490	26.8
717	274	0.8	9,764	27.5
718	273	0.8	10,037	28.3
720	309	0.9	10,346	29.2
721	291	0.8	10,637	30.0
724	276	0.8	10,913	30.8
725	324	0.9	11,237	31.7
726	288	0.8	11,525	32.5
728	310	0.9	11,835	33.4
731	302	0.9	12,137	34.2
732	331	0.9	12,468	35.2
733	297	0.8	12,765	36.0
735	318	0.9	13,083	36.9
738	368	1.0	13,451	37.9
739	376	1.1	13,827	39.0
741	389	1.1	14,216	40.1
742	335	0.9	14,551	41.0
746	332	0.9	14,883	42.0
747	378	1.1	15,261	43.0
748	367	1.0	15,628	44.1
749	351	1.0	15,979	45.1
754	402	1.1	16,381	46.2
755	409	1.2	16,790	47.3
756	371	1.0	17,161	48.4
757	413	1.2	17,574	49.6
762	465	1.3	18,039	50.9
763	425	1.2	18,464	52.1
764	432	1.2	18,896	53.3
766	409	1.2	19,305	54.4
771	452	1.3	19,757	55.7
772	434	1.2	20,191	56.9
773	444	1.3	20,635	58.2
774	422	1.2	21,057	59.4
780	409	1.2	21,466	60.5
781	462	1.3	21,928	61.8
782	466	1.3	22,394	63.1
784	473	1.3	22,867	64.5

English II Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
797	1821	5.1	24,688	69.6
799	465	1.3	25,153	70.9
801	471	1.3	25,624	72.3
803	451	1.3	26,075	73.5
804	455	1.3	26,530	74.8
810	458	1.3	26,988	76.1
812	422	1.2	27,410	77.3
814	409	1.2	27,819	78.4
816	428	1.2	28,247	79.7
822	378	1.1	28,625	80.7
824	407	1.1	29,032	81.9
827	414	1.2	29,446	83.0
829	412	1.2	29,858	84.2
835	415	1.2	30,273	85.4
838	366	1.0	30,639	86.4
842	405	1.1	31,044	87.5
844	367	1.0	31,411	88.6
850	339	1.0	31,750	89.5
853	348	1.0	32,098	90.5
858	326	0.9	32,424	91.4
861	321	0.9	32,745	92.3
866	291	0.8	33,036	93.2
870	238	0.7	33,274	93.8
878	262	0.7	33,536	94.6
881	254	0.7	33,790	95.3
886	195	0.5	33,985	95.8
889	184	0.5	34,169	96.4
901	226	0.6	34,395	97.0
906	213	0.6	34,608	97.6
908	131	0.4	34,739	98.0
913	97	0.3	34,836	98.2
931	120	0.3	34,956	98.6
936	61	0.2	35,017	98.7
937	112	0.3	35,129	99.1
941	73	0.2	35,202	99.3
971	60	0.2	35,262	99.4
973	40	0.1	35,302	99.5
977	69	0.2	35,371	99.7
978	27	0.1	35,398	99.8
999	65	0.2	35,463	100.0


Spring 2010 English II Scale Score Distribution
Percent
6.0
5.0
4.0
3.0
2.0
1.0
0.0
440 494 512 536 555 566 581 596 608 617 629 646 659 670 678 693 700 708 717 725 733 742 754 763 773 784 804 822 838 858 881 908 941 999
Scale Score

English III Score Distribution for Spring 2010

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	91	0.3	91	0.3
454	6	0.0	97	0.3
458	7	0.0	104	0.3
487	14	0.0	118	0.3
489	11	0.0	129	0.4
502	18	0.1	147	0.4
504	7	0.0	154	0.4
515	14	0.0	168	0.5
517	26	0.1	194	0.5
529	41	0.1	235	0.7
536	35	0.1	270	0.8
547	30	0.1	300	0.8
548	27	0.1	327	0.9
551	43	0.1	370	1.0
561	29	0.1	399	1.1
562	42	0.1	441	1.2
563	37	0.1	478	1.3
564	29	0.1	507	1.4
573	31	0.1	538	1.5
574	43	0.1	581	1.6
575	88	0.2	669	1.9
584	76	0.2	745	2.1
585	36	0.1	781	2.2
586	51	0.1	832	2.3
593	52	0.1	884	2.5
594	97	0.3	981	2.8
596	49	0.1	1,030	2.9
601	64	0.2	1,094	3.1
603	112	0.3	1,206	3.4
605	58	0.2	1,264	3.6
609	66	0.2	1,330	3.7
611	135	0.4	1,465	4.1
613	66	0.2	1,531	4.3
616	67	0.2	1,598	4.5
618	81	0.2	1,679	4.7
619	72	0.2	1,751	4.9
621	85	0.2	1,836	5.2
623	73	0.2	1,909	5.4
625	63	0.2	1,972	5.5
626	81	0.2	2,053	5.8
628	77	0.2	2,130	6.0

English III Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
630	78	0.2	2,208	6.2
632	81	0.2	2,289	6.4
633	101	0.3	2,390	6.7
635	99	0.3	2,489	7.0
636	97	0.3	2,586	7.3
638	67	0.2	2,653	7.5
640	77	0.2	2,730	7.7
642	172	0.5	2,902	8.2
649	387	1.1	3,289	9.3
650	95	0.3	3,384	9.5
652	104	0.3	3,488	9.8
654	209	0.6	3,697	10.4
656	94	0.3	3,791	10.7
658	105	0.3	3,896	11.0
660	211	0.6	4,107	11.6
661	107	0.3	4,214	11.9
664	114	0.3	4,328	12.2
665	208	0.6	4,536	12.8
667	117	0.3	4,653	13.1
669	122	0.3	4,775	13.4
670	110	0.3	4,885	13.7
671	114	0.3	4,999	14.1
672	120	0.3	5,119	14.4
674	131	0.4	5,250	14.8
676	274	0.8	5,524	15.5
677	130	0.4	5,654	15.9
679	152	0.4	5,806	16.3
681	252	0.7	6,058	17.0
682	119	0.3	6,177	17.4
684	131	0.4	6,308	17.7
686	283	0.8	6,591	18.5
687	141	0.4	6,732	18.9
689	134	0.4	6,866	19.3
691	163	0.5	7,029	19.8
695	619	1.7	7,648	21.5
696	168	0.5	7,816	22.0
697	150	0.4	7,966	22.4
699	155	0.4	8,121	22.8
700	167	0.5	8,288	23.3
701	204	0.6	8,492	23.9
702	172	0.5	8,664	24.4

English III Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
703	144	0.4	8,808	24.8
705	173	0.5	8,981	25.3
706	181	0.5	9,162	25.8
707	185	0.5	9,347	26.3
708	205	0.6	9,552	26.9
709	176	0.5	9,728	27.4
711	178	0.5	9,906	27.9
712	209	0.6	10,115	28.4
713	181	0.5	10,296	29.0
714	206	0.6	10,502	29.5
716	218	0.6	10,720	30.1
717	397	1.1	11,117	31.3
718	199	0.6	11,316	31.8
721	215	0.6	11,531	32.4
722	408	1.1	11,939	33.6
723	210	0.6	12,149	34.2
726	450	1.3	12,599	35.4
727	466	1.3	13,065	36.7
731	480	1.3	13,545	38.1
732	459	1.3	14,004	39.4
735	257	0.7	14,261	40.1
736	502	1.4	14,763	41.5
737	269	0.8	15,032	42.3
740	251	0.7	15,283	43.0
741	547	1.5	15,830	44.5
742	290	0.8	16,120	45.3
745	498	1.4	16,618	46.7
746	273	0.8	16,891	47.5
747	275	0.8	17,166	48.3
749	261	0.7	17,427	49.0
750	273	0.8	17,700	49.8
752	327	0.9	18,027	50.7
753	294	0.8	18,321	51.5
754	296	0.8	18,617	52.4
755	244	0.7	18,861	53.0
757	316	0.9	19,177	53.9
758	306	0.9	19,483	54.8
759	299	0.8	19,782	55.6
760	282	0.8	20,064	56.4
763	314	0.9	20,378	57.3
764	323	0.9	20,701	58.2

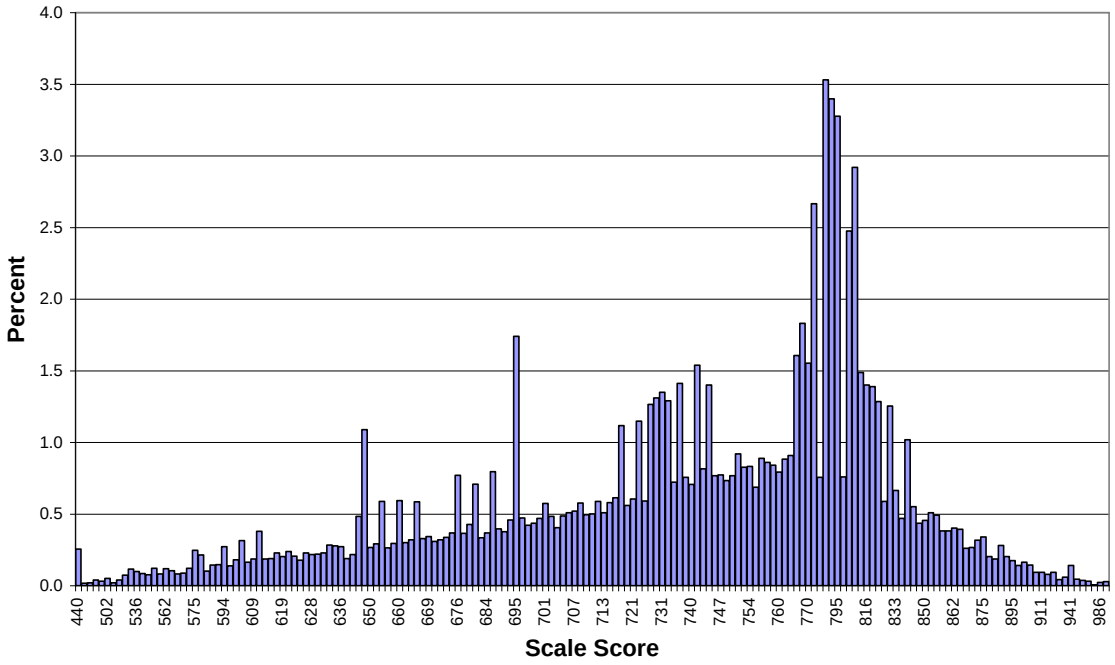
English III Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
765	571	1.6	21,272	59.8
769	651	1.8	21,923	61.7
770	552	1.6	22,475	63.2
775	948	2.7	23,423	65.9
776	269	0.8	23,692	66.6
781	1255	3.5	24,947	70.2
787	1208	3.4	26,155	73.6
795	1165	3.3	27,320	76.8
800	270	0.8	27,590	77.6
801	880	2.5	28,470	80.1
808	1038	2.9	29,508	83.0
815	529	1.5	30,037	84.5
816	498	1.4	30,535	85.9
823	494	1.4	31,029	87.3
824	457	1.3	31,486	88.6
831	209	0.6	31,695	89.1
832	446	1.3	32,141	90.4
833	236	0.7	32,377	91.1
840	167	0.5	32,544	91.5
841	362	1.0	32,906	92.5
842	196	0.6	33,102	93.1
849	155	0.4	33,257	93.5
850	162	0.5	33,419	94.0
851	181	0.5	33,600	94.5
852	175	0.5	33,775	95.0
859	136	0.4	33,911	95.4
860	136	0.4	34,047	95.8
862	143	0.4	34,190	96.2
863	140	0.4	34,330	96.6
870	93	0.3	34,423	96.8
871	95	0.3	34,518	97.1
874	113	0.3	34,631	97.4
875	121	0.3	34,752	97.7
882	72	0.2	34,824	97.9
883	66	0.2	34,890	98.1
887	100	0.3	34,990	98.4
888	72	0.2	35,062	98.6
895	62	0.2	35,124	98.8
896	50	0.1	35,174	98.9
902	58	0.2	35,232	99.1
903	51	0.1	35,283	99.2

English III Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
910	33	0.1	35,316	99.3
911	33	0.1	35,349	99.4
919	28	0.1	35,377	99.5
920	33	0.1	35,410	99.6
928	15	0.0	35,425	99.6
929	21	0.1	35,446	99.7
941	50	0.1	35,496	99.8
952	16	0.0	35,512	99.9
970	13	0.0	35,525	99.9
972	11	0.0	35,536	99.9
984	2	0.0	35,538	99.9
986	8	0.0	35,546	100.0
999	10	0.0	35,556	100.0

Spring 2010 English III Scale Score Distribution

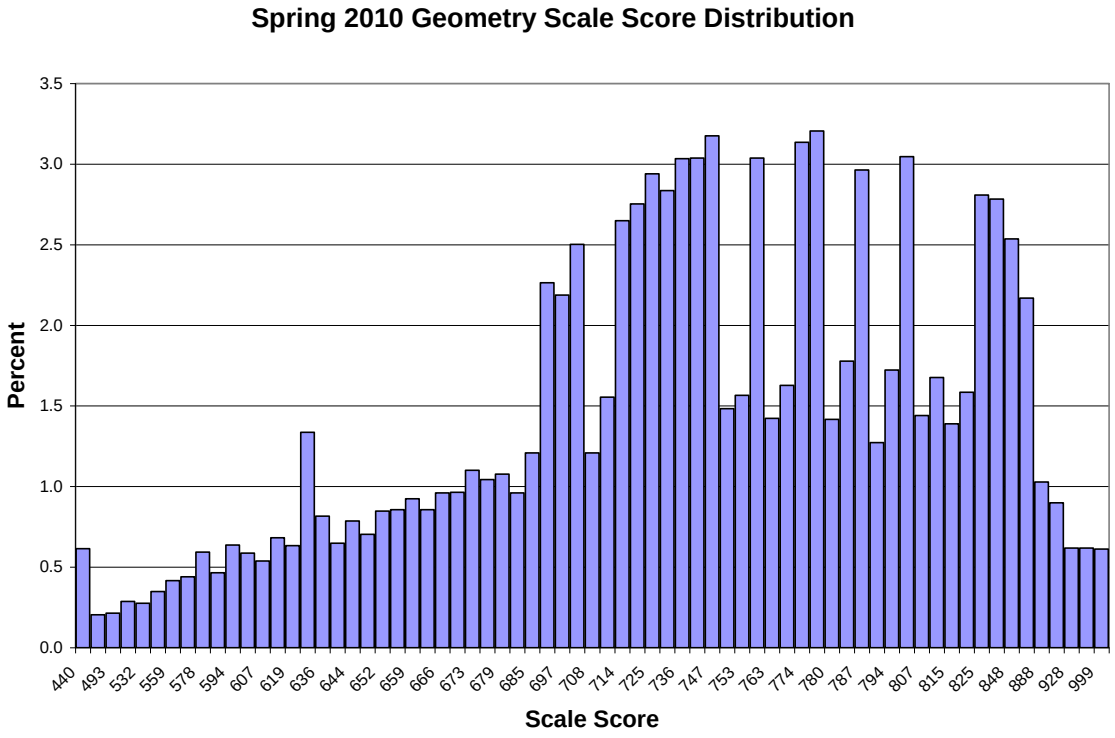


Geometry Score Distribution for Spring 2010

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	201	0.6	201	0.6
483	67	0.2	268	0.8
493	70	0.2	338	1.0
531	94	0.3	432	1.3
532	90	0.3	522	1.6
558	114	0.3	636	1.9
559	136	0.4	772	2.4
577	144	0.4	916	2.8
578	194	0.6	1,110	3.4
593	152	0.5	1,262	3.9
594	208	0.6	1,470	4.5
606	192	0.6	1,662	5.1
607	176	0.5	1,838	5.6
617	223	0.7	2,061	6.3
619	207	0.6	2,268	6.9
635	437	1.3	2,705	8.3
636	267	0.8	2,972	9.1
639	212	0.6	3,184	9.7
644	257	0.8	3,441	10.5
648	230	0.7	3,671	11.2
652	277	0.8	3,948	12.1
656	280	0.9	4,228	12.9
659	302	0.9	4,530	13.9
663	280	0.9	4,810	14.7
666	314	1.0	5,124	15.7
671	315	1.0	5,439	16.6
673	360	1.1	5,799	17.7
678	341	1.0	6,140	18.8
679	352	1.1	6,492	19.9
684	314	1.0	6,806	20.8
685	395	1.2	7,201	22.0
695	740	2.3	7,941	24.3
697	715	2.2	8,656	26.5
703	818	2.5	9,474	29.0
708	395	1.2	9,869	30.2
709	508	1.6	10,377	31.7
714	866	2.6	11,243	34.4
720	900	2.8	12,143	37.1
725	961	2.9	13,104	40.1
731	927	2.8	14,031	42.9
736	992	3.0	15,023	46.0
742	993	3.0	16,016	49.0

Geometry Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
747	1038	3.2	17,054	52.2
752	485	1.5	17,539	53.6
753	512	1.6	18,051	55.2
758	993	3.0	19,044	58.3
763	465	1.4	19,509	59.7
764	532	1.6	20,041	61.3
774	1025	3.1	21,066	64.4
775	1048	3.2	22,114	67.6
780	463	1.4	22,577	69.1
781	581	1.8	23,158	70.8
787	969	3.0	24,127	73.8
793	416	1.3	24,543	75.1
794	563	1.7	25,106	76.8
800	996	3.0	26,102	79.8
807	471	1.4	26,573	81.3
808	548	1.7	27,121	83.0
815	454	1.4	27,575	84.3
816	518	1.6	28,093	85.9
825	918	2.8	29,011	88.7
835	910	2.8	29,921	91.5
848	829	2.5	30,750	94.1
865	709	2.2	31,459	96.2
888	336	1.0	31,795	97.3
889	294	0.9	32,089	98.2
928	202	0.6	32,291	98.8
931	202	0.6	32,493	99.4
999	200	0.6	32,693	100.0

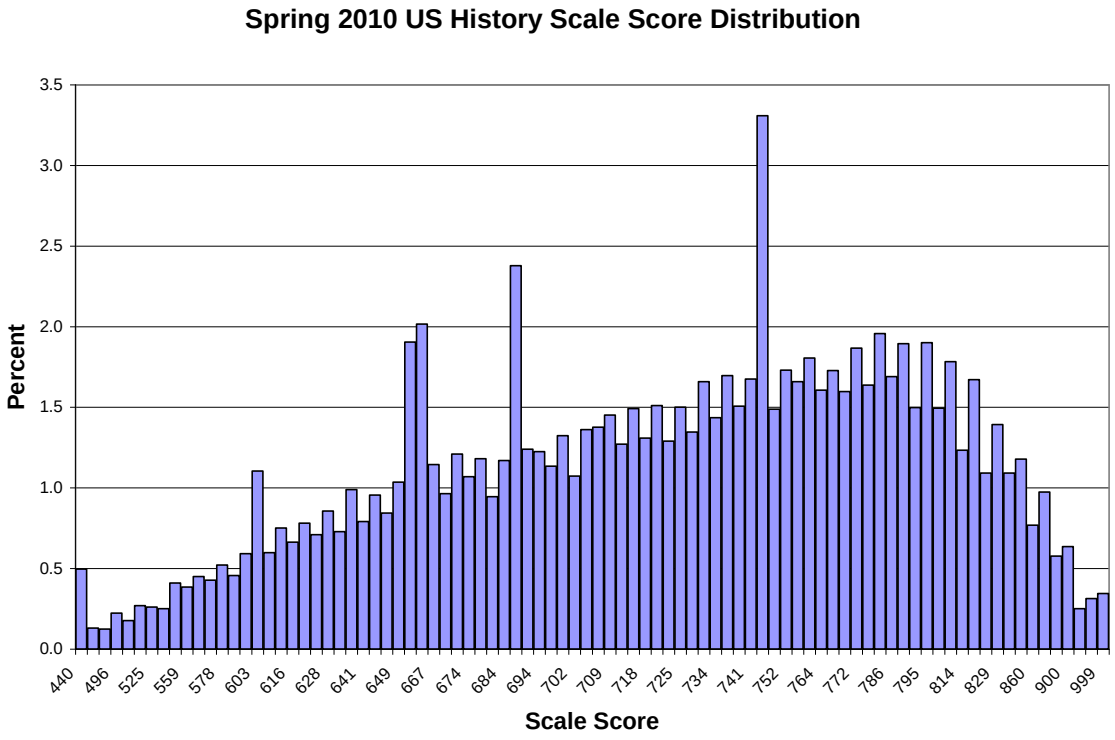

Spring 2010 Geometry Scale Score Distribution
Percent
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
440 493 532 559 578 594 607 619 636 644 652 659 666 673 679 685 697 708 714 725 736 747 753 763 774 780 787 794 807 815 825 848 888 928 999
Scale Score

U.S. History Score Distribution for Spring 2010

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	160	0.5	160	0.5
452	42	0.1	202	0.6
465	40	0.1	242	0.8
496	72	0.2	314	1.0
498	57	0.2	371	1.2
523	87	0.3	458	1.4
525	84	0.3	542	1.7
543	81	0.3	623	1.9
547	132	0.4	755	2.3
559	124	0.4	879	2.7
564	145	0.4	1,024	3.2
573	138	0.4	1,162	3.6
578	168	0.5	1,330	4.1
586	147	0.5	1,477	4.6
590	191	0.6	1,668	5.2
603	356	1.1	2,024	6.3
607	193	0.6	2,217	6.9
611	242	0.8	2,459	7.6
616	214	0.7	2,673	8.3
620	252	0.8	2,925	9.1
625	229	0.7	3,154	9.8
628	276	0.9	3,430	10.6
633	235	0.7	3,665	11.4
635	319	1.0	3,984	12.4
641	255	0.8	4,239	13.1
642	308	1.0	4,547	14.1
648	272	0.8	4,819	14.9
649	334	1.0	5,153	16.0
655	614	1.9	5,767	17.9
661	650	2.0	6,417	19.9
667	369	1.1	6,786	21.0
668	311	1.0	7,097	22.0
673	390	1.2	7,487	23.2
674	345	1.1	7,832	24.3
678	381	1.2	8,213	25.5
679	305	0.9	8,518	26.4
684	377	1.2	8,895	27.6
689	767	2.4	9,662	30.0
691	400	1.2	10,062	31.2
694	395	1.2	10,457	32.4
696	366	1.1	10,823	33.6
699	427	1.3	11,250	34.9

U.S. History Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
702	346	1.1	11,596	36.0
704	439	1.4	12,035	37.3
707	444	1.4	12,479	38.7
709	468	1.5	12,947	40.1
712	410	1.3	13,357	41.4
714	481	1.5	13,838	42.9
718	422	1.3	14,260	44.2
720	487	1.5	14,747	45.7
723	416	1.3	15,163	47.0
725	484	1.5	15,647	48.5
728	434	1.3	16,081	49.9
730	535	1.7	16,616	51.5
734	463	1.4	17,079	53.0
735	547	1.7	17,626	54.7
740	486	1.5	18,112	56.2
741	540	1.7	18,652	57.8
747	1067	3.3	19,719	61.1
751	480	1.5	20,199	62.6
752	558	1.7	20,757	64.4
757	535	1.7	21,292	66.0
759	582	1.8	21,874	67.8
764	518	1.6	22,392	69.4
765	557	1.7	22,949	71.2
771	515	1.6	23,464	72.8
772	602	1.9	24,066	74.6
778	528	1.6	24,594	76.3
779	631	2.0	25,225	78.2
786	545	1.7	25,770	79.9
787	611	1.9	26,381	81.8
794	483	1.5	26,864	83.3
795	613	1.9	27,477	85.2
804	482	1.5	27,959	86.7
805	575	1.8	28,534	88.5
814	398	1.2	28,932	89.7
816	539	1.7	29,471	91.4
827	352	1.1	29,823	92.5
829	449	1.4	30,272	93.9
842	352	1.1	30,624	95.0
846	380	1.2	31,004	96.1
860	248	0.8	31,252	96.9
867	314	1.0	31,566	97.9
885	186	0.6	31,752	98.5

U.S. History Score Distribution for Spring 2010 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
900	205	0.6	31,957	99.1
925	81	0.3	32,038	99.3
961	101	0.3	32,139	99.7
999	111	0.3	32,250	100.0