OKLAHOMA STATE DEPARTMENT OF EDUCATION
Janet Barresi, State Superintendent of Public Instruction

DRAFT

Oklahoma School Testing Program

2011 Technical Report

Achieving Classroom Excellence

End-of-Instruction

Assessments

Submitted to
The Oklahoma State Department of Education
September 2011

PEARSON ALWAYS LEARNING

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 2011 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 2010-11 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 26
Section 4 29
4.1 Item Response Theory (IRT) models 29
4.2 Pre-Equating 30
4.3 Assessment of IRT Fit to the model 30
4.3.a Calibration and IRT Fit Results for Post-Equated Tests 31
4.4 Calibration and Equating 32
4.4.a Common Linking Items for Spring 2011 32
4.5 Item Stability Evaluation Methods 33
4.5.a Results of the Item Parameter Item Stability Check 34
4.6 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 66
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 2010-11 6
Table 1.5. Configuration of the OSTP-ACE/EOI Tests for Spring 2011 6
Table 1.6. Number of Common Linking Items per Subject for Spring 2011 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 2010-11 17
Table 3.2. Demographic Characteristics of Calibration and Equating Sample for Spring 2011. 17
Table 3.3. Test-Level Summaries of Classical Item Analyses for Winter/Trimester 2010-11 and Spring 2011 19
Table 3.4. DIF Flag Incidence Across All OSTP-ACE EOI Field Test Items for Winter/Trimester 2010-11 and Spring 2011 21
Table 3.5. Number of Items Per Subject Flagged and Rejected During Winter/Trimester 2010–2011 and Spring 2011 Field Test Data Review 23
Table 3.6. Cronbach's Alpha for Winter/Trimester 2010-11 and Spring 2011 Administrations by Subject 24
Table 3.7. Test Reliability by Subgroup for Spring 2011 25
Table 3.8.Inter-rater Reliability for English II Operational Writing Prompts for Winter/Trimester 2010-11 and Spring 2011 27
Table 3.9. Inter-rater Reliability for English III Operational Writing Prompts for Winter/Trimester 2010-11 and Spring 2011 28
Table 4.1. Number of Common Linking Items Per Subject for Spring 2011 32
Table 4.2. LOSS, HOSS, and Scaling Constants by Subject 34
Table 4.3. Performance-Level Cut Scores by Subject 35
Table 4.4. Raw Score to Scale Score Conversion Tables for Winter/Trimester 2010-11 36
Table 4.5.Raw Score to Scale Score Conversion Tables for Spring 2011 42
Table 5.1. Estimates of Accuracy and Consistency of Performance Classification for Winter/Trimester 2010-11 56
Table 5.2. Estimates of Accuracy and Consistency of Performance Classification for Spring 2011 56
Table 5.3. Accuracy and Consistency Estimates by Cut Score: False Positive- and False Negative Rates for Winter/Trimester 2010-11 58
Table 5.4. Accuracy and Consistency Estimates by Cut Score: False Positive- and False Negative Rates for Spring 2011 59
Table 6.1. Descriptive Statistics of the Scale Scores for Winter/Trimester 2010-11 - Overall. 60
Table 6.2. Descriptive Statistics of the Scale Scores for Winter/Trimester 2010-11 by Gender 60
Table 6.3. Descriptive Statistics of the Scale Scores for Winter/Trimester 2010-11 by Race/Ethnicity 60

Table 6.4. Descriptive Statistics of the Scale Scores for Winter/Trimester 2010-11 by Free/Reduced Lunch Status 61
Table 6.5. Descriptive Statistics of the Scale Scores for Spring 2011 - Overall 62
Table 6.6. Descriptive Statistics of the Scale Scores for Spring 2011 by Gender 62
Table 6.7. Descriptive Statistics of the Scale Scores for Spring 2011 by Race/Ethnicity 63
Table 6.8. Descriptive Statistics of the Scale Scores for Spring 2011 by Free/Reduced Lunch Status 64
Table 6.9. Percentage of Students by Performance Level for Winter/Trimester 2010-11 and Spring 2011 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 2010-11 and Spring 2011 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 2010-11 for each subject. In the Spring 2011 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 2010-11 test forms and then used again in the Spring 2011 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 2010-11 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 2010-11 and the Spring 2011 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 2010 administrations for Algebra I, Biology I, English II, and U.S. History and from the census Spring 2007 field test through the Spring 2010 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 2010-11 and Spring 2011 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 2010-11 and Spring 2011 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 2010-11	1	10-15	14.55	12.73	18.18	16.67
	2	60-70	65.45	65.45	63.64	63.33
	3/4	15-25	20.00	21.82	18.18	20.00
Spring 2011 Core A	1	10-15	12.73	9.09	14.55	13.33
	2	60-70	69.09	74.55	65.45	68.33
	3/4	15-25	18.18	16.36	20.00	18.33
Spring 2011 Core B	1	10-15	11.43	5.71	11.43	15.38
	2	60-70	65.71	74.29	77.14	58.97
	3/4	15-25	22.86	20.00	11.43	25.64

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 2010-11	1	10-15	14.75	6.35	10.00
	2	60-70	60.66	68.25	66.67
	3/4	15-25	24.59	25.40	23.33
Spring 2011 Core A	1	10-15	8.20	11.11	18.33
	2	60-70	73.77	69.84	65.00
	3/4	15-25	18.03	19.05	16.67
Spring 2011 Core B	1	10-15	10.26	5.00	10.00
	2	60-70	66.67	75.00	72.50
	3/4	15-25	23.07	20.00	17.50

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 2010-11 and Spring 2011 OSTP-ACE EOI assessments. The Spring 2011 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 2011 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 2010-11

					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 2011

					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	80/1*	60	6	20	0
English III	16	62/1*	20	82/1*	62	10	20	0
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 2011

Subject	No. of CL Items	Total No. of Items*
Algebra I	20	90
Algebra II	20	90
Biology I	21	99
English II	22	100
English III	23	103
Geometry	20	90
U.S. History	20	100

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 2010-11 Algebra I administration was comprised of one form with 55 operational items and 20 field test items. There were two core forms and 11 field test sets in the Spring 2011 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 2010-11								
Operational	15	15	31	31	9	9	55	55
FT Form 1	5	5	11	11	4	4	20	20
Spring 2011								
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	11	11	3	3	20	20
FT Form 2	6	6	12	12	2	2	20	20
FT Form 3	6	6	11	11	3	3	20	20
FT Form 4	6	6	11	11	3	3	20	20
FT Form 5	6	6	11	11	3	3	20	20
FT Form 6	6	6	11	11	3	3	20	20
FT Form 7	6	6	10	10	4	4	20	20
FT Form 8	6	6	10	10	4	4	20	20
FT Form 9	6	6	11	11	3	3	20	20
FT Form 10	6	6	11	11	3	3	20	20
FT Form 11	6	6	11	11	3	3	20	20

Note: FT = Field Test.

Algebra II. The Winter/Trimester 2010-11 Algebra II administration was comprised of one form with 55 operational items and 20 field test items. There were two core forms and 11 field test sets in the Spring 2011 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 2010-11								
Operational	15	15	31	31	9	9	55	55
FT Form 1	4	4	13	13	3	3	20	20
Spring 2011								
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	11	11	3	3	20	20
FT Form 2	5	5	12	12	3	3	20	20
FT Form 3	5	5	12	12	3	3	20	20
FT Form 4	4	4	13	13	3	3	20	20
FT Form 5	5	5	12	12	3	3	20	20
FT Form 6	6	6	11	11	3	3	20	20
FT Form 7	5	5	12	12	3	3	20	20
FT Form 8	5	5	11	11	4	4	20	20
FT Form 9	5	5	12	12	3	3	20	20
FT Form 10	5	5	12	12	3	3	20	20
FT Form 11	5	5	11	11	4	4	20	20

Note: FT = Field Test.

Geometry. The Winter/Trimester 2010-11 Geometry administration was comprised of one form with 55 operational items and 20 field test items. There were two core forms and 11 field test sets in the Spring 2011 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 2010-11												
Operational	6	6	20	20	12	12	10	10	7	7	55	55
FT Form 1	2	2	7	7	4	4	4	4	3	3	20	20
Spring 2011												
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	2	2	7	7	4	4	4	4	3	3	20	20
FT Form 2	2	2	7	7	4	4	4	4	3	3	20	20
FT Form 3	2	2	8	8	4	4	3	3	3	3	20	20
FT Form 4	2	2	7	7	4	4	4	4	3	3	20	20
FT Form 5	2	2	7	7	4	4	5	5	2	2	20	20
FT Form 6	2	2	7	7	4	4	5	5	2	2	20	20
FT Form 7	2	2	7	7	4	4	4	4	3	3	20	20
FT Form 8	2	2	7	7	4	4	5	5	2	2	20	20
FT Form 9	2	2	7	7	4	4	4	4	3	3	20	20
FT Form 10	2	2	7	7	4	4	5	5	2	2	20	20
FT Form 11	2	2	7	7	4	4	4	4	3	3	20	20

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

Biology I. The Winter/Trimester 2010-11 Biology I administration was comprised of one form with 60 operational items and 20 field test items. There were two core forms and 11 field test sets in the Spring 2011 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 2010-11														
Operational	9	9	9	9	9	9	12	12	9	9	8	8	56	56
FT Form 1	6	6	2	2	2	2	1	1	3	3	5	5	19	19
Spring 2011														
Core A	8	8	10	10	9	9	12	12	9	9	8	8	56	56
Core B	8	8	8	8	8	8	14	14	9	9	9	9	56	56
FT Form 1	2	2	3	3	2	2	5	5	2	2	4	4	18	18
FT Form 2	3	3	3	3	3	3	3	3	3	3	3	3	18	18
FT Form 3	3	3	4	4	4	4	3	3	2	2	3	3	19	19
FT Form 4	3	3	1	1	5	5	4	4	3	3	2	2	18	18
FT Form 5	4	4	3	3	2	2	3	3	4	4	2	2	18	18
FT Form 6	4	4	3	3	4	4	5	5	2	2	-	-	18	18
FT Form 7	5	5	3	3	4	4	3	3	3	3	1	1	19	19
FT Form 8	7	7	4	4	2	2	4	4	-	-	-	-	17	17
FT Form 9	4	4	4	4	4	4	5	5	1	1	-	-	18	18
FT Form 10	3	3	4	4	3	3	5	5	3	3	-	-	18	18
FT Form 11	5	5	4	4	2	2	5	5	1	1	-	-	17	17

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 2010-11 English II administration was comprised of one form with 60 operational MC items, 1 open-ended writing prompt, and 20 field test MC items. All multiple-choice 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 2011 administration. Each of the forms contained 60 operational MC items, 1 operational open-ended writing prompt, and 20 field test MC items, totaling 81 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 2010-11 and Spring 2011 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 2010-11														
Operational	6	6	18	18	18	18	6	6	1	6	12	12	61	66
FT Form 1	3	3	10	10	5	5	2	2	-	-	-	-	20	20
Spring 2011														
Core A	6	6	18	18	18	18	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	2	2	5	5	4	4	1	1	-	-	8	8	20	20
FT Form 2	2	2	7	7	6	6	5	5	-	-	-	-	20	20
FT Form 3	2	2	5	5	2	2	3	3	-	-	8	8	20	20
FT Form 4	2	2	7	7	8	8	3	3	-	-	-	-	20	20
FT Form 5	2	2	5	5	6	6	1	1	-	-	6	6	20	20
FT Form 6	2	2	7	7	8	8	3	3	-	-	-	-	20	20
FT Form 7	2	2	5	5	6	6	1	1	-	-	6	6	20	20
FT Form 8	2	2	10	10	6	6	2	2	-	-	-	-	20	20
FT Form 9	2	2	6	6	6	6	0	0	-	-	6	6	20	20
FT Form 10	3	3	6	6	6	6	5	5	-	-	-	-	20	20
FT Form 11	1	1	6	6	5	5	2	2	-	-	6	6	20	20
FT Form 12	3	3	6	6	8	8	3	3	-	-	-	-	20	20
FT Form 13	2	2	6	6	2	2	2	2	-	-	8	8	20	20
FT Form 14	2	2	8	8	8	8	2	2	-	-	-	-	20	20
FT Form 15	3	3	7	7	8	8	2	2	-	-	-	-	20	20
FT Form 16	1	1	6	6	10	10	3	3	-	-	-	-	20	20

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

English III. The Winter/Trimester 2010-11 English III administration was comprised of one form with 62 operational MC items, 1 open-ended writing prompt, and 20 field test MC items. All multiple-choice 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 2011 administration. Each of the forms contained a set of 62 operational MC items, 1 operational open-ended writing prompt, and 20 field test MC items, totaling 83 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 2010-11 and Spring 2011 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 2010-11														
Operational	6	6	18	18	18	18	6	6	1	10	14	14	63	72
FT Form 1	3	3	6	6	8	8	3	3	-	-	-	-	20	20
Spring 2011														
Core A	6	6	17	17	19	19	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	1	1	5	5	5	5	1	1	-	-	8	8	20	20
FT Form 2	2	2	7	7	6	6	1	1	-	-	4	4	20	20
FT Form 3	2	2	7	7	8	8	3	3	-	-	-	-	20	20
FT Form 4	1	1	3	3	3	3	1	1	-	-	12	12	20	20
FT Form 5	2	2	4	4	4	4	2	2	-	-	8	8	20	20
FT Form 6	2	2	9	9	4	4	1	1	-	-	4	4	20	20
FT Form 7	3	3	7	7	5	5	5	5	-	-	-	-	20	20
FT Form 8	3	3	2	2	8	8	3	3	-	-	4	4	20	20
FT Form 9	3	3	10	10	6	6	1	1	-	-	-	-	20	20
FT Form 10	3	3	10	10	5	5	2	2	-	-	-	-	20	20
FT Form 11	3	3	6	6	9	9	2	2	-	-	-	-	20	20
FT Form 12	2	2	6	6	2	2	2	2	-	-	8	8	20	20
FT Form 13	1	1	9	9	7	7	3	3	-	-	-	-	20	20
FT Form 14	2	2	6	6	5	5	1	1	-	-	6	6	20	20
FT Form 15	3	3	7	7	7	7	3	3	-	-	-	-	20	20
FT Form 16	2	2	6	6	4	4	2	2	-	-	6	6	20	20

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

U.S. History. The Winter/Trimester 2010-11 U.S. History administration was comprised of one form with 60 operational items and 20 field test items. There were two core forms and 11 field test sets in the Spring 2011 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 2010-11 and Spring 2011 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 2010-11														
Operational	6	6	9	9	9	9	12	12	9	9	15	15	60	60
FT Form 1	3	3	4	4	1	1	5	5	2	2	5	5	20	20
Spring 2011														
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	1	1	2	2	3	3	2	2	3	3	9	9	20	20
FT Form 2	3	3	2	2	4	4	4	4	1	1	6	6	20	20
FT Form 3	1	1	2	2	3	3	6	6	2	2	6	6	20	20
FT Form 4	1	1	2	2	2	2	7	7	3	3	5	5	20	20
FT Form 5	1	1	2	2	3	3	6	6	2	2	6	6	20	20
FT Form 6	3	3	4	4	2	2	4	4	3	3	4	4	20	20
FT Form 7	1	1	3	3	4	4	4	4	1	1	7	7	20	20
FT Form 8	1	1	3	3	3	3	5	5	5	5	3	3	20	20
FT Form 9	1	1	3	3	4	4	3	3	3	3	6	6	20	20
FT Form 10	2	2	5	5	3	3	4	4	3	3	3	3	20	20
FT Form 11	-	-	4	4	4	4	7	7	3	3	2	2	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 Access™ website, and optionally with data received directly from Oklahoma. Oklahoma educators also use the Access™ 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 2010-11 and Spring 2011 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 2010-11 and Spring 2011) 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 <u>exclusion from analysis</u> 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 2010-11 and Spring 2011 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 2010-11

Subject	Total	Male	Female	African American	Native American	Hispanic	Asian	Pacific Islander	White	Other
Algebra I	1,348	698	636	126	249	149	28	3	746	41
Algebra II	1,053	509	536	101	167	69	20	0	661	30
Biology I	1,376	651	725	143	205	117	14	1	864	26
English II	1,425	705	713	132	213	113	32	1	893	34
English III	1,420	703	717	123	182	85	50	5	919	51
Geometry	1,544	789	728	147	228	117	60	1	946	38
U.S. History	1,394	696	695	141	226	116	19	2	848	40

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 2011

Subject	Total	Male	Female	African American	Native American	Hispanic	Asian	Pacific Islander	White	Other
Algebra I	27,063	12,597	12,850	2,272	4,421	4,286	515	56	14,641	872
Algebra II	27,020	12,641	13,549	2,290	4,263	3,212	660	50	15,780	765
Biology I	32,606	15,838	16,287	2,720	5,589	3,790	683	73	18,704	1,047
English II	36,873	18,091	18,776	3,557	6,187	3,800	780	74	21,313	1,162
English III	37,220	18,792	18,420	3,694	6,369	3,468	815	58	21,588	1,228
Geometry	33,070	15,861	16,514	3,116	5,713	4,004	694	69	18,521	953
U.S. History	29,765	14,531	14,878	2,720	4,773	3,155	672	61	17,566	818

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 2010-11 or Spring 2011 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 2010-11 and Spring 2011 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 2010-11 and Spring 2011

Subject and Administration	Sample Size	Mean	Mean % of Max	Items / Points	Mean *p*	Mean r_{pb}	Omit Min	Omit Max
Algebra I-W10	1,348	29.92	0.54	55	0.54	0.38	0.00	0.37
Algebra I-S11 CA	14,873	35.16	0.64	55	0.64	0.39	0.01	0.09
Algebra I-S11 CB	12,190	35.44	0.64	55	0.64	0.39	0.01	0.08
Algebra II-W10	1,053	31.47	0.57	55	0.57	0.43	0.00	0.09
Algebra II-S11 CA	14,685	31.95	0.58	55	0.58	0.43	0.01	0.16
Algebra II-S11 CB	12,335	31.54	0.57	55	0.57	0.42	0.02	0.15
Biology I-W10	1,376	38.15	0.64	60	0.64	0.40	0.00	0.07
Biology I-S11 CA	17,798	40.98	0.68	60	0.68	0.38	0.01	0.07
Biology I-S11 CB	14,808	41.94	0.70	60	0.70	0.37	0.00	0.06
English II-W10	1,425	47.15	0.71	61/66	0.72	0.39	0.00	0.21
English II-S11 CAA	8,134	50.21	0.76	61/66	0.77	0.35	0.00	5.16
English II-S11 CAB	7,995	50.34	0.76	61/66	0.77	0.35	0.00	4.85
English II-S11 CBA	7,770	49.91	0.76	61/66	0.76	0.34	0.01	5.16
English II-S11 CBB	7,803	49.88	0.76	61/66	0.76	0.34	0.01	4.85
English III-W10	1,420	46.56	0.65	63/72	0.65	0.41	0.00	0.21
English III-S11 CAA	8,167	48.07	0.67	63/72	0.67	0.40	0.01	5.93
English III-S11 CAB	8,131	48.30	0.67	63/72	0.67	0.40	0.01	5.33
English III-S11 CBA	7,707	49.02	0.68	63/72	0.68	0.39	0.00	5.93
English III-S11 CBB	7,699	49.39	0.69	63/72	0.68	0.39	0.00	5.33
Geometry-W10	1,544	34.87	0.63	55	0.63	0.45	0.00	0.13
Geometry-S11 CA	17,947	37.86	0.69	55	0.69	0.41	0.01	0.06
Geometry-S11 CA	15,123	37.42	0.68	55	0.68	0.42	0.01	0.08
U.S. History-W10	1,394	37.83	0.63	60	0.63	0.40	0.00	0.14
U.S. History-S11 CA	16,228	40.77	0.68	60	0.68	0.38	0.01	0.04
U.S. History-S11 CB	13,537	41.02	0.68	60	0.68	0.38	0.00	0.06

Note: W10 = Winter/Trimester 2010-11; S11 CA = Spring 2011 Core A; S11 CB = Spring 2011 Core B; S11 CAA=Spring 11 MC form A +OE form A; S11 CAB=Spring 11 MC form A +OE form B; S11 CBA=Spring 11 MC form B +OE form A; S11 CBB=Spring 11 MC form B +OE form 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 2010-11 and Spring 2011 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 2011 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 2010-11 and Spring 2011

Subject	Total FT Items	Native American	Asian	African American	Hispanic	Female
Winter 2010-11						
Algebra I	20	1	0	2	0	1
Algebra II	20	0	0	0	1	1
Geometry	20	1	0	2	0	0
Biology I	20	0	0	2	1	2
English II	20	0	1	1	1	4
English III	20	0	1	1	1	0
U.S. History	20	0	0	1	2	2
Spring 2011						
Algebra I	220	4	6	28	4	14
Algebra II	220	3	7	17	3	1
Geometry	220	2	9	12	2	19
Biology I	320	5	11	40	42	27
English II	320	5	11	20	36	35
English III	220	1	5	15	5	12
U.S. History	220	0	7	21	8	19

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 2010-11 and Spring 2011 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 which met 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 2010–2011 and Spring 2011 Field Test Data Review

Subject	No. of FT Items	No. Flagged	Rejected	Accepted	Accepted with Edits
Winter 2010-11					
Algebra I	20	6	2	18	0
Algebra II	20	15	1	17	2
Geometry	20	14	6	13	1
Biology I	20	10	2	18	0
English II	20	9	6	14	0
English III	20	8	2	16	2
U.S. History	20	10	2	15	3
Spring 2011					
Algebra I	220	98	20	180	20
Algebra II	220	89	22	181	17
Geometry	220	82	23	177	20
Biology I	320	170	85	235	0
English II	320	152	53	267	0
English III	220	82	13	195	12
U.S. History	220	97	12	181	27

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:

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

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 2010-11 and Spring 2011 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 2010-11 and Spring 2011 Administrations by Subject

Subject	Administration and Form	Alpha
Algebra I	Winter 2010-11	0.89
	Spring 2011 – Core A	0.90
	Spring 2011 – Core B	0.89
Algebra II	Winter 2010-11	0.92
	Spring 2011 – Core A	0.92
	Spring 2011 – Core B	0.92
Biology I	Winter 2010-11	0.91
	Spring 2011 – Core A	0.90
	Spring 2011 – Core B	0.89
English II	Winter 2010-11	0.90
	Spring 2011 – Core AA	0.87
	Spring 2011 – Core AB	0.87
	Spring 2011 – Core BA	0.86
	Spring 2011 – Core BB	0.87
English III	Winter 2010-11	0.91
	Spring 2011 – Core AA	0.90
	Spring 2011 – Core AB	0.90
	Spring 2011 – Core BA	0.89
	Spring 2011 – Core BB	0.89
Geometry	Winter 2010-11	0.93
	Spring 2011 – Core A	0.91
	Spring 2011 – Core B	0.91
U.S. History	Winter 2010-11	0.91
	Spring 2011 – Core A	0.90
	Spring 2011 – Core B	0.90

Note: Core AA=Core MC form A+OE form A; Core AB=Core MC form A+OE form B; Core BA=Core MC form B+OE form A; Core BB=Core MC form B+OE form B.

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 2011 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 2011

Subject	Core	Male	Female	African-American	Native American	Hispanic	Asian	White
Algebra I	A	0.90	0.89	0.88	0.89	0.90	0.92	0.89
	B	0.90	0.89	0.89	0.88	0.89	0.92	0.89
Algebra II	A	0.92	0.91	0.90	0.91	0.91	0.93	0.92
	B	0.92	0.91	0.89	0.90	0.91	0.93	0.92
Biology I	A	0.90	0.89	0.89	0.89	0.89	0.92	0.89
	B	0.90	0.88	0.88	0.88	0.90	0.91	0.88
English II	AA	0.88	0.86	0.88	0.86	0.89	0.93	0.85
	AB	0.87	0.86	0.87	0.85	0.89	0.91	0.84
	BA	0.87	0.86	0.86	0.84	0.88	0.90	0.84
	BB	0.87	0.86	0.86	0.86	0.89	0.90	0.85
English III	AA	0.90	0.89	0.89	0.90	0.89	0.91	0.89
	AB	0.90	0.89	0.89	0.88	0.90	0.91	0.89
	BA	0.90	0.88	0.89	0.88	0.89	0.89	0.89
	BB	0.90	0.88	0.88	0.88	0.89	0.89	0.89
Geometry	A	0.91	0.91	0.90	0.90	0.91	0.91	0.90
	B	0.92	0.91	0.90	0.90	0.91	0.93	0.91
U.S. History	A	0.90	0.89	0.89	0.89	0.90	0.92	0.89
	B	0.90	0.89	0.90	0.88	0.89	0.93	0.89

Note: Core AA=Core MC form A+OE form A; Core AB=Core MC form A+OE form B; Core BA=Core MC form B+OE form A; Core BB=Core MC form B+OE form B.

Table 3.7. Test Reliability by Subgroup for Spring 2011 (cont.)

Subject	Core	English Language Learner	Individual Education Plan	Economically Disadvantaged
Algebra I	A	0.89	0.89	0.89
	B	0.89	0.89	0.88
Algebra II	A	0.91	0.89	0.90
	B	0.89	0.89	0.90
Biology I	A	0.85	0.90	0.89
	B	0.85	0.90	0.89
English II	AA	0.86	0.90	0.88
	AB	0.86	0.88	0.87
	BA	0.88	0.89	0.87
	BB	0.87	0.89	0.87
English III	AA	0.85	0.87	0.89
	AB	0.82	0.87	0.89
	BA	0.86	0.88	0.89
	BB	0.86	0.89	0.89
Geometry	A	0.89	0.89	0.90
	B	0.90	0.89	0.91
U.S. History	A	0.86	0.91	0.89
	B	0.86	0.92	0.89

Note: Core AA=Core MC form A+OE form A; Core AB=Core MC form A+OE form B; Core BA=Core MC form B+OE form A; Core BB=Core MC form B+OE form B.

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 2010-11 English II and English III tests contained one operational writing prompt each and the Spring 2011 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 2010-11 and Spring 2011 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 2010-11 and Spring 2011 operational writing prompts are close to or within the moderate range.

Table 3.8.Inter-rater Reliability for English II Operational Writing Prompts for Winter/Trimester 2010-11 and Spring 2011

			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 2010-11													
1	4	1,325	0.00	0.23	15.77	69.06	14.72	0.23	0.00	69.06	30.49	0.45	0.44
2	4	1,325	0.00	0.30	16.53	66.04	16.83	0.30	0.00	66.04	33.36	0.60	0.40
3	4	1,325	0.00	0.00	15.62	70.19	14.11	0.08	0.00	70.19	29.74	0.08	0.44
4	4	1,325	0.00	0.23	15.17	68.83	15.55	0.23	0.00	68.83	30.72	0.45	0.49
5	4	1,325	0.00	0.30	16.53	65.96	17.13	0.08	0.00	65.96	33.66	0.38	0.44
Spring 2011 Core Form A													
1	4	15,081	0.00	0.45	17.68	64.05	17.37	0.45	0.00	64.05	35.04	0.90	0.30
2	4	15,081	0.00	0.53	17.46	63.83	17.77	0.41	0.00	63.83	35.23	0.94	0.29
3	4	15,081	0.00	0.44	17.46	63.85	17.82	0.44	0.00	63.85	35.28	0.88	0.30
4	4	15,081	0.00	0.56	17.01	65.06	16.91	0.46	0.00	65.06	33.92	1.02	0.32
5	4	15,081	0.00	0.68	17.86	62.68	18.12	0.65	0.00	62.68	35.99	1.33	0.30
Spring 2011 Core Form B													
1	4	15,030	0.00	0.35	16.06	67.82	15.46	0.31	0.00	67.82	31.52	0.67	0.31
2	4	15,030	0.00	0.34	15.90	68.39	15.08	0.29	0.00	68.39	30.98	0.63	0.31
3	4	15,030	0.00	0.15	14.77	70.17	14.69	0.22	0.00	70.17	29.46	0.37	0.31
4	4	15,030	0.00	0.22	14.65	70.74	14.04	0.35	0.00	70.74	28.69	0.57	0.35
5	4	15,030	0.00	0.22	15.28	68.74	15.46	0.31	0.00	68.74	30.74	0.53	0.31

Table 3.9. Inter-rater Reliability for English III Operational Writing Prompts for Winter/Trimester 2010-11 and Spring 2011

			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 2010-11													
1	4	1,351	0.00	0.15	15.03	71.35	13.40	0.07	0.00	71.35	28.42	0.22	0.57
2	4	1,351	0.00	0.15	14.14	71.06	14.36	0.30	0.00	71.06	28.50	0.44	0.56
3	4	1,351	0.00	0.15	14.43	70.84	14.43	0.15	0.00	70.84	28.87	0.30	0.52
4	4	1,351	0.00	0.30	15.99	69.95	13.40	0.37	0.00	69.95	29.39	0.67	0.53
5	4	1,351	0.00	0.67	18.13	62.62	18.28	0.30	0.00	62.62	36.42	0.96	0.47
Spring 2011 Core Form A													
1	4	14,932	0.01	0.56	17.89	63.84	17.04	0.64	0.02	63.84	34.92	1.23	0.36
2	4	14,932	0.00	0.60	17.44	64.58	16.82	0.55	0.01	64.58	34.26	1.16	0.37
3	4	14,932	0.01	0.48	17.26	65.09	16.70	0.46	0.00	65.09	33.97	0.94	0.37
4	4	14,932	0.01	0.70	17.31	64.38	17.06	0.55	0.00	64.38	34.36	1.26	0.37
5	4	14,932	0.01	0.78	18.42	61.71	18.38	0.70	0.00	61.71	36.80	1.49	0.34
Spring 2011 Core Form B													
1	4	14,986	0.01	0.61	16.55	65.42	16.93	0.48	0.01	65.42	33.48	1.10	0.35
2	4	14,986	0.01	0.46	15.98	66.84	16.24	0.47	0.00	66.84	32.22	0.94	0.35
3	4	14,986	0.00	0.41	14.77	69.12	15.34	0.37	0.00	69.12	30.11	0.77	0.35
4	4	14,986	0.02	0.46	15.45	68.05	15.54	0.48	0.00	68.05	30.99	0.96	0.37
5	4	14,986	0.02	0.50	17.34	64.57	17.02	0.54	0.00	64.57	34.37	1.06	0.33

Section 4

Calibration, Equating, and Scaling

This section introduces the IRT models, methods, and processes that were used to calibrate, equate, and scale the OCCT EOI tests. The three-parameter logistic (3-PL) item response theory (IRT) model (Lord & Novick, 1968) was used for dichotomously-scored test items and the Generalized Partial Credit (GPC; Muraki, 1997) model was used for polytomously-scored test items. For 2010-2011 academic season, preequating procedures were applied to the subjects of Algebra I, Algebra II, Biology I, Geometry, and U.S. History, and post-equating procedures for the subjects of ELA II and ELA III.

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 Pre-Equating

Pre-equating procedures were applied to OCCT ACE EOI tests consisting entirely of dichotomously-scored multiple-choice items. These subjects included Algebra I, Algebra II, Biology I, Geometry, and U.S. History. ACE EOI tests English II and English III remained post-equated. All pre-equated forms were constructed using only previously-administered operational items and a set of unscored field-test items. Pearson Psychometric & Research Services staff created raw score to scale score (RSSS) tables using the freely-available program, POLYEQUATE (Kolen, 2004). Banked item parameter estimates for the forms' operational items were imported into POLYEQUATE as both the "new" and "old" forms to create a table of raw score to true score equivalents.

Scaling constants provided in Table 4.2 (M1 and M2) were used to rescale true score equivalents to the reported scale score metric. The lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS) for each subject also appear in Table 4.2.

Performance level cut scores appear in Table 4.3. Because the scale cut score may not always be present in the RSSS table, the scale scores that were closest to, but below the scale scores (thetas) set in standard setting were used as the "effective" cut scores. 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 the pre-equated tests, are shown in Table 4.4 and Table 4.5, respectively, for the Winter 2010 and Spring 2011 administrations. This next section outlines post-equating work completed for the ACE EOI English II and English III tests.

4.3 Assessment of IRT Fit to the model

For post-equated tests, 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 2010-11 or Spring 2011.

4.3.a Calibration and IRT Fit Results for Post-Equated Tests

4.3.a.i Winter/Trimester 2010-11

English II. For the Winter/Trimester 2010-11 English II assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. One English II items were flagged for further review based on their fit statistics.

English III. For the Winter/Trimester 2010-11 English III assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. Six English III 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, equating, or scoring. No items were dropped from any of the Winter/Trimester 2010-11 ACE EOI assessments for calibration, equating, or scoring as a result of the Q_1 results.

4.3.a.ii Spring 2011

English II. For the Spring 2011 English II assessment, based on the calibration sample, the Z-statistics for most operational items were smaller than the critical Z-statistic. One English II items were flagged for further review based on their fit statistics.

English III. For the Spring 2011 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.

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, equating, or scoring. No items were dropped from any of the Spring 2011 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 content specialists from Pearson and SDE (for more on data review, please see Section 3.4).

4.4 Calibration and Equating

The 3-PL model was used for calibration and equating of all items for the purposes of rescaling field test items to the bank metric for Algebra I, Algebra II, Geometry, Biology I, and U.S. History because 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 for the calibration and equating of the operational test form and for the field test items.

A common item, non-equivalent groups (CINEG) design was used for ACE EOI English II and English III tests to link the current test forms (i.e., Winter/Trimester 2010-11 and Spring 2011) 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 ACE EOI English II and English III Winter/Trimester 2010-11 and Spring 2011 tests, all operational items were used as common or anchor items to link to the base 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.

4.4.a Common Linking Items for Spring 2011

Table 4.1 presents the number and percentage of common linking items for all post-equated subject for the Spring 2011 administration. The common linking items were necessary as a result of two core operational forms for the Spring 2011 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 2011

Subject	Number of Items on Test	Number of Linking Items	Percent of Test
English II	61	22	36%
English III	63	23	37%

4.5 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 2010 and Spring 2011 administration. The process used in this evaluation is called an item parameter stability check[1].

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

[1] Note that the item stability check was applied only to post-equated tests.

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).

4.5.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 2010-11, no items were removed from English II or from English III. For Spring 2011, ten anchor items were removed from English II and English III each. Any item removed from the item parameter stability check set still contributed to student scores.

4.6 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:

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

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 English II and English III are shown in Table 4.4 and Table 4.5 for Winter/Trimester 2010-11 and Spring 2011, 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	654	700	783
Biology I	627	691	775
English II	588	693	797
English III	670	700	802
Geometry	635	700	777
U.S. History	603	689	747

Table 4.4. Raw Score to Scale Score Conversion Tables for Winter/Trimester 2010-11

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	49	1	440	48	1	440	49	1	440	41	1
1	490	49	1	440	48	1	440	49	1	440	41	1
2	490	49	1	440	48	1	440	49	1	440	41	1
3	490	49	1	440	48	1	440	49	1	440	41	1
4	490	49	1	440	48	1	440	49	1	440	41	1
5	490	49	1	440	48	1	440	49	1	440	41	1
6	490	49	1	440	48	1	440	49	1	440	41	1
7	490	49	1	440	48	1	440	49	1	440	41	1
8	490	49	1	440	48	1	440	49	1	440	41	1
9	490	49	1	440	48	1	440	49	1	440	41	1
10	490	49	1	440	48	1	440	49	1	440	41	1
11	490	49	1	440	48	1	440	49	1	440	41	1
12	537	53	1	440	48	1	440	49	1	440	41	1
13	569	56	1	440	48	1	440	49	1	440	41	1
14	590	55	1	442	49	1	473	52	1	440	41	1
15	606	53	1	481	54	1	506	56	1	440	41	1
16	619	49	1	510	58	1	530	58	1	440	41	1
17	631	44	1	532	59	1	549	57	1	454	43	1
18	640	39	1	551	58	1	564	55	1	482	47	1
19	649	34	1	567	56	1	577	51	1	504	50	1
20	662	31	2	581	52	1	589	47	1	522	51	1
21	663	27	2	594	48	1	603	43	2	537	50	1
22	670	25	2	606	44	1	610	40	2	551	48	1
23	676	23	2	616	41	1	620	36	2	563	46	1
24	682	21	2	627	38	2	629	34	2	574	43	1
25	687	20	2	636	35	2	637	31	2	588	40	2
26	693	19	2	645	33	2	645	30	2	593	37	2
27	700	19	3	653	31	2	653	28	2	602	34	2
28	703	18	3	661	29	2	660	27	2	610	32	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 2010-11 (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	708	17	3	669	28	2	667	25	2	618	30	2
30	713	17	3	676	26	2	674	24	2	625	29	2
31	717	16	3	683	25	2	680	24	2	632	27	2
32	722	16	3	691	24	3	689	23	3	639	26	2
33	727	16	3	697	24	3	693	22	3	645	25	2
34	731	16	3	704	23	3	699	22	3	652	24	2
35	736	15	3	710	22	3	704	21	3	658	23	2
36	740	15	3	716	22	3	710	21	3	664	23	2
37	745	15	3	722	21	3	716	20	3	670	22	2
38	749	15	3	728	21	3	721	20	3	676	22	2
39	754	15	3	734	20	3	727	20	3	681	21	2
40	762	15	4	740	20	3	733	20	3	687	21	2
41	764	15	4	746	20	3	738	20	3	693	21	3
42	769	15	4	752	20	3	747	20	4	699	21	3
43	774	16	4	758	20	3	750	20	4	704	21	3
44	779	16	4	764	20	3	756	20	4	710	21	3
45	785	17	4	775	20	4	762	20	4	716	21	3
46	791	17	4	776	20	4	768	21	4	722	21	3
47	797	19	4	782	20	4	775	21	4	728	21	3
48	804	20	4	789	21	4	782	22	4	734	21	3
49	812	23	4	796	21	4	789	23	4	740	21	3
50	822	26	4	803	22	4	797	24	4	746	21	3
51	833	32	4	811	23	4	806	25	4	753	22	3
52	847	40	4	819	25	4	815	27	4	760	22	3
53	868	50	4	829	28	4	826	30	4	767	23	3
54	904	57	4	839	31	4	838	33	4	775	24	3
55	999	39	4	852	35	4	852	38	4	783	25	3
56	-	-	-	868	40	4	870	43	4	797	27	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 2010-11 (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	-	-	-	888	45	4	893	47	4	802	28	4
58	-	-	-	918	46	4	925	47	4	813	30	4
59	-	-	-	975	37	4	982	35	4	825	33	4
60	-	-	-	999	32	4	999	31	4	839	37	4
61	-	-	-	-	-	-	-	-	-	855	41	4
62	-	-	-	-	-	-	-	-	-	875	46	4
63	-	-	-	-	-	-	-	-	-	900	50	4
64	-	-	-	-	-	-	-	-	-	936	48	4
65	-	-	-	-	-	-	-	-	-	997	34	4
66	-	-	-	-	-	-	-	-	-	999	34	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 2010-11 (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	66	1	440	64	1	440	46	1
1	440	66	1	440	64	1	440	46	1
2	440	66	1	440	64	1	440	46	1
3	440	66	1	440	64	1	440	46	1
4	440	66	1	440	64	1	440	46	1
5	440	66	1	440	64	1	440	46	1
6	440	66	1	440	64	1	440	46	1
7	440	66	1	440	64	1	440	46	1
8	440	66	1	440	64	1	440	46	1
9	440	66	1	440	64	1	440	46	1
10	440	66	1	440	64	1	440	46	1
11	440	66	1	440	64	1	440	46	1
12	440	66	1	503	69	1	440	46	1
13	504	72	1	544	71	1	440	46	1
14	545	75	1	571	70	1	477	51	1
15	574	75	1	591	66	1	506	55	1
16	596	71	1	606	60	1	527	57	1
17	614	65	1	619	54	1	543	56	1
18	629	59	1	635	47	2	558	54	1
19	642	52	1	640	41	2	571	51	1
20	654	45	2	649	36	2	583	48	1
21	663	40	2	657	32	2	595	44	1
22	672	35	2	665	28	2	605	41	1
23	680	31	2	672	26	2	614	38	1
24	688	29	2	679	24	2	623	35	1
25	700	26	3	685	23	2	631	33	1
26	702	24	3	691	22	2	639	31	1
27	708	23	3	700	21	3	646	29	1
28	714	22	3	703	20	3	653	28	1

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 2010-11 (cont.)

Raw Score	Algebra II			Geometry			English III		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
29	720	21	3	708	19	3	660	26	1
30	726	20	3	714	18	3	670	25	2
31	731	19	3	719	18	3	672	24	2
32	736	19	3	724	17	3	678	23	2
33	742	18	3	729	17	3	684	23	2
34	747	18	3	734	16	3	689	22	2
35	752	17	3	738	16	3	695	21	2
36	757	17	3	743	16	3	700	21	3
37	762	17	3	748	16	3	705	21	3
38	767	17	3	752	15	3	710	20	3
39	772	17	3	757	15	3	715	20	3
40	777	17	3	762	15	3	721	20	3
41	783	17	4	767	15	3	726	19	3
42	788	17	4	772	16	3	731	19	3
43	794	17	4	777	16	4	736	19	3
44	799	18	4	782	16	4	741	19	3
45	806	19	4	788	17	4	746	19	3
46	812	20	4	794	18	4	751	19	3
47	819	21	4	801	19	4	756	19	3
48	828	23	4	808	20	4	761	19	3
49	837	26	4	816	23	4	766	19	3
50	848	29	4	825	26	4	772	19	3
51	861	33	4	837	32	4	777	19	3
52	878	39	4	851	39	4	783	19	3
53	902	43	4	872	49	4	788	19	3
54	944	42	4	907	55	4	794	20	3
55	999	32	4	999	36	4	802	20	4
56							806	20	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 2010-11 (cont.)

Raw Score	Algebra II			Geometry			English III		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
57	-	-	-	-	-	-	813	21	4
58	-	-	-	-	-	-	819	21	4
59	-	-	-	-	-	-	826	22	4
60	-	-	-	-	-	-	833	22	4
61	-	-	-	-	-	-	841	23	4
62	-	-	-	-	-	-	849	24	4
63	-	-	-	-	-	-	858	25	4
64	-	-	-	-	-	-	867	27	4
65	-	-	-	-	-	-	877	29	4
66	-	-	-	-	-	-	888	32	4
67	-	-	-	-	-	-	901	35	4
68	-	-	-	-	-	-	917	37	4
69	-	-	-	-	-	-	938	37	4
70	-	-	-	-	-	-	968	32	4
71	-	-	-	-	-	-	999	25	4
72	-	-	-	-	-	-	999	25	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 2011

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	48	1	440	47	1	440	47	1	440	51	1
1	440	48	1	440	47	1	440	47	1	440	51	1
2	440	48	1	440	47	1	440	47	1	440	51	1
3	440	48	1	440	47	1	440	47	1	440	51	1
4	440	48	1	440	47	1	440	47	1	440	51	1
5	440	48	1	440	47	1	440	47	1	440	51	1
6	440	48	1	440	47	1	440	47	1	440	51	1
7	440	48	1	440	47	1	440	47	1	440	51	1
8	440	48	1	440	47	1	440	47	1	440	51	1
9	440	48	1	440	47	1	440	47	1	440	51	1
10	440	48	1	440	47	1	440	47	1	440	51	1
11	440	48	1	440	47	1	440	47	1	440	51	1
12	440	48	1	440	47	1	440	47	1	440	51	1
13	440	48	1	452	49	1	440	47	1	440	51	1
14	458	51	1	492	53	1	440	47	1	440	51	1
15	496	55	1	519	56	1	472	51	1	449	52	1
16	522	58	1	539	56	1	501	55	1	490	57	1
17	543	58	1	555	54	1	524	57	1	519	59	1
18	560	56	1	569	51	1	543	57	1	541	60	1
19	574	53	1	581	48	1	559	55	1	559	58	1
20	587	49	1	592	44	1	574	52	1	574	55	1
21	598	45	1	602	40	1	586	48	1	587	51	1
22	608	41	1	611	37	1	603	45	2	603	47	2
23	617	37	1	619	34	1	608	41	2	609	43	2
24	627	34	2	627	31	2	618	38	2	618	39	2
25	634	32	2	635	29	2	627	35	2	627	35	2
26	642	30	2	642	28	2	636	32	2	635	33	2
27	649	28	2	649	26	2	644	30	2	643	30	2
28	656	27	2	655	25	2	651	29	2	651	29	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 2011 (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	663	25	2	662	25	2	659	27	2	658	27	2
30	669	24	2	668	24	2	666	26	2	665	26	2
31	675	24	2	674	23	2	672	25	2	671	25	2
32	682	23	2	680	23	2	679	24	2	678	24	2
33	691	22	3	691	22	3	689	23	3	689	23	3
34	694	22	3	692	22	3	691	22	3	690	22	3
35	699	22	3	698	22	3	697	22	3	696	22	3
36	705	21	3	704	21	3	703	21	3	702	21	3
37	711	21	3	709	21	3	709	21	3	708	21	3
38	717	21	3	715	21	3	715	20	3	714	21	3
39	723	21	3	721	21	3	721	20	3	720	20	3
40	729	21	3	727	21	3	726	20	3	725	20	3
41	735	21	3	733	21	3	732	20	3	731	20	3
42	741	21	3	739	21	3	738	20	3	737	20	3
43	747	21	3	746	21	3	747	20	4	747	20	4
44	753	21	3	752	21	3	749	20	4	748	20	4
45	760	22	3	758	22	3	756	20	4	754	20	4
46	767	22	3	765	22	3	762	20	4	760	20	4
47	775	23	4	775	22	4	768	21	4	767	21	4
48	781	23	4	780	23	4	775	21	4	773	21	4
49	789	24	4	787	24	4	782	22	4	780	22	4
50	797	25	4	796	25	4	790	23	4	788	23	4
51	806	27	4	804	26	4	798	24	4	796	24	4
52	816	29	4	814	28	4	807	26	4	804	26	4
53	827	31	4	824	30	4	817	28	4	814	28	4
54	839	34	4	836	34	4	828	32	4	825	32	4
55	854	39	4	850	39	4	841	36	4	838	36	4
56	871	43	4	867	44	4	857	41	4	854	41	4
57	894	48	4	889	49	4	879	46	4	875	46	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 2011 (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	927	47	4	922	49	4	909	48	4	904	48	4
59	986	35	4	985	37	4	966	40	4	957	42	4
60	999	32	4	999	34	4	999	33	4	999	34	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 2011 (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	58	1	440	56	1	440	69	1	440	71	1
1	440	58	1	440	56	1	440	69	1	440	71	1
2	440	58	1	440	56	1	440	69	1	440	71	1
3	440	58	1	440	56	1	440	69	1	440	71	1
4	440	58	1	440	56	1	440	69	1	440	71	1
5	440	58	1	440	56	1	440	69	1	440	71	1
6	440	58	1	440	56	1	440	69	1	440	71	1
7	440	58	1	440	56	1	440	69	1	440	71	1
8	440	58	1	440	56	1	440	69	1	440	71	1
9	440	58	1	440	56	1	440	69	1	440	71	1
10	440	58	1	440	56	1	440	69	1	440	71	1
11	440	58	1	440	56	1	440	69	1	440	71	1
12	446	58	1	494	61	1	521	74	1	440	71	1
13	502	64	1	530	64	1	559	76	1	522	77	1
14	536	67	1	555	64	1	584	73	1	561	78	1
15	560	66	1	574	61	1	602	68	1	588	76	1
16	580	63	1	590	57	1	617	61	1	607	71	1
17	596	58	1	603	51	1	629	54	1	623	64	1
18	610	53	1	615	46	1	640	47	1	636	56	1
19	622	47	1	626	41	1	654	40	2	654	49	2
20	635	42	2	635	36	2	659	35	2	657	42	2
21	642	37	2	644	33	2	667	31	2	666	36	2
22	651	34	2	652	30	2	674	28	2	674	32	2
23	660	31	2	660	28	2	681	26	2	681	28	2
24	668	28	2	668	26	2	688	24	2	688	26	2
25	675	26	2	675	25	2	694	23	2	694	24	2
26	682	25	2	682	23	2	700	22	3	700	22	3
27	689	23	2	688	22	2	706	21	3	706	21	3
28	700	22	3	694	21	2	711	20	3	712	20	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 2011 (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	701	21	3	700	20	3	717	19	3	717	20	3
30	707	20	3	706	20	3	722	19	3	723	19	3
31	712	19	3	712	19	3	727	18	3	728	19	3
32	718	19	3	717	18	3	732	18	3	733	18	3
33	723	18	3	723	18	3	738	18	3	738	18	3
34	728	18	3	728	17	3	743	17	3	743	17	3
35	733	17	3	733	17	3	748	17	3	748	17	3
36	738	17	3	738	17	3	753	17	3	753	17	3
37	743	17	3	743	16	3	758	17	3	758	17	3
38	748	17	3	748	16	3	763	17	3	763	17	3
39	753	17	3	753	16	3	768	17	3	769	17	3
40	758	17	3	758	16	3	774	17	3	774	17	3
41	764	17	3	764	17	3	783	17	4	783	18	4
42	769	18	3	769	17	3	785	18	4	785	18	4
43	777	18	4	777	17	4	791	18	4	791	18	4
44	782	19	4	781	18	4	797	18	4	797	19	4
45	788	20	4	787	18	4	804	19	4	804	19	4
46	796	21	4	794	19	4	811	20	4	811	20	4
47	804	23	4	801	21	4	818	21	4	818	22	4
48	813	25	4	809	23	4	826	23	4	827	23	4
49	823	28	4	818	26	4	836	25	4	836	25	4
50	835	32	4	828	30	4	846	28	4	847	29	4
51	850	37	4	841	36	4	859	33	4	860	33	4
52	869	43	4	858	43	4	874	38	4	876	38	4
53	896	47	4	883	50	4	897	44	4	899	43	4
54	945	44	4	933	49	4	937	44	4	938	43	4
55	999	35	4	999	40	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 2011 (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	47	1	440	47	1	440	46	1	440	45	1
1	440	47	1	440	47	1	440	46	1	440	45	1
2	440	47	1	440	47	1	440	46	1	440	45	1
3	440	47	1	440	47	1	440	46	1	440	45	1
4	440	47	1	440	47	1	440	46	1	440	45	1
5	440	47	1	440	47	1	440	46	1	440	45	1
6	440	47	1	440	47	1	440	46	1	440	45	1
7	440	47	1	440	47	1	440	46	1	440	45	1
8	440	47	1	440	47	1	440	46	1	440	45	1
9	440	47	1	440	47	1	440	46	1	440	45	1
10	440	47	1	440	47	1	440	46	1	440	45	1
11	440	47	1	440	47	1	440	46	1	440	45	1
12	447	48	1	447	48	1	451	47	1	452	46	1
13	490	53	1	490	53	1	492	52	1	492	51	1
14	518	56	1	518	56	1	518	55	1	518	54	1
15	538	57	1	538	57	1	538	55	1	536	54	1
16	554	55	1	554	55	1	554	54	1	551	53	1
17	568	53	1	568	53	1	567	51	1	564	50	1
18	580	49	1	580	49	1	579	47	1	576	47	1
19	590	45	1	590	45	1	589	44	1	586	43	1
20	600	41	1	600	41	1	598	40	1	595	39	1
21	609	38	1	609	38	1	607	36	1	604	36	1
22	617	35	1	617	35	1	615	33	1	612	33	1
23	625	32	1	625	32	1	622	31	1	620	31	1
24	632	30	1	632	30	1	629	29	1	627	29	1
25	638	28	1	638	28	1	636	27	1	633	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 2011Table 4.5.Raw Score to Scale Score Conversion Tables for Spring 201(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	645	26	1	642	26	1	642	25	1	640	26	1
27	651	25	1	648	25	1	648	24	1	646	24	1
28	656	24	1	654	24	1	654	23	1	652	23	1
29	662	23	1	660	23	1	660	22	1	658	23	1
30	670	22	2	670	22	2	665	22	1	663	22	1
31	673	21	2	671	21	2	670	21	2	670	21	2
32	678	20	2	676	21	2	675	20	2	674	21	2
33	682	20	2	681	20	2	680	20	2	679	20	2
34	687	19	2	686	20	2	685	20	2	684	20	2
35	692	19	2	691	19	2	690	19	2	689	19	2
36	700	19	3	695	19	2	695	19	2	694	19	2
37	701	18	3	700	18	3	700	19	3	700	19	3
38	706	18	3	705	18	3	705	18	3	703	18	3
39	710	18	3	709	18	3	709	18	3	708	18	3
40	715	18	3	714	18	3	714	18	3	713	18	3
41	719	17	3	718	17	3	719	18	3	718	18	3
42	724	17	3	723	17	3	723	17	3	722	17	3
43	728	17	3	727	17	3	728	17	3	727	17	3
44	733	17	3	732	17	3	733	17	3	732	17	3
45	737	17	3	736	17	3	737	17	3	736	17	3
46	742	17	3	741	17	3	742	17	3	741	17	3
47	746	17	3	746	17	3	747	17	3	746	17	3
48	751	17	3	750	17	3	751	17	3	751	17	3
49	756	17	3	755	17	3	756	17	3	755	17	3
50	760	17	3	760	17	3	761	17	3	760	17	3
51	765	17	3	765	17	3	766	17	3	765	17	3
52	770	18	3	770	18	3	771	17	3	770	17	3
53	776	18	3	775	18	3	776	17	3	775	17	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 2011Table 4.5.Raw Score to Scale Score Conversion Tables for Spring 201(cont.)

	English III Core A Prompt A			English III Core A Prompt B			English III Core B Prompt A			English III Core B Prompt B		
Raw Score	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
54	781	18	3	780	18	3	781	17	3	781	17	3
55	786	19	3	786	19	3	787	18	3	786	18	3
56	792	19	3	792	19	3	792	18	3	792	18	3
57	802	20	4	802	20	4	802	18	4	802	18	4
58	805	21	4	804	20	4	804	19	4	803	19	4
59	811	21	4	811	21	4	810	20	4	809	19	4
60	819	22	4	818	22	4	817	20	4	816	20	4
61	826	24	4	825	23	4	824	21	4	823	21	4
62	835	25	4	834	25	4	832	23	4	831	22	4
63	844	27	4	843	26	4	840	24	4	839	24	4
64	855	29	4	853	28	4	849	27	4	848	26	4
65	866	32	4	865	31	4	860	30	4	858	29	4
66	880	35	4	877	34	4	872	33	4	870	32	4
67	895	38	4	892	37	4	886	37	4	884	37	4
68	914	40	4	910	40	4	904	41	4	900	41	4
69	937	40	4	932	41	4	926	42	4	921	43	4
70	968	35	4	962	36	4	957	39	4	950	41	4
71	999	27	4	999	27	4	999	30	4	999	29	4
72	999	27	4	999	27	4	999	30	4	999	29	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 2011 (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	35	1	440	35	1	440	35	1	440	34	1
1	440	35	1	440	35	1	440	35	1	440	34	1
2	440	35	1	440	35	1	440	35	1	440	34	1
3	440	35	1	440	35	1	440	35	1	440	34	1
4	440	35	1	440	35	1	440	35	1	440	34	1
5	440	35	1	440	35	1	440	35	1	440	34	1
6	440	35	1	440	35	1	440	35	1	440	34	1
7	440	35	1	440	35	1	440	35	1	440	34	1
8	440	35	1	440	35	1	440	35	1	440	34	1
9	440	35	1	440	35	1	440	35	1	440	34	1
10	440	35	1	440	35	1	440	35	1	440	34	1
11	440	35	1	440	35	1	440	35	1	440	34	1
12	440	35	1	440	35	1	440	35	1	440	34	1
13	440	35	1	440	35	1	440	35	1	440	34	1
14	440	35	1	440	35	1	440	35	1	440	34	1
15	457	38	1	456	37	1	445	35	1	444	35	1
16	480	41	1	479	41	1	469	40	1	468	40	1
17	499	44	1	498	43	1	489	43	1	488	43	1
18	515	44	1	514	44	1	507	45	1	506	45	1
19	528	44	1	527	44	1	521	45	1	520	45	1
20	541	42	1	540	42	1	535	44	1	534	44	1
21	551	40	1	550	40	1	547	42	1	546	42	1
22	561	38	1	560	38	1	557	40	1	556	40	1
23	570	35	1	569	36	1	567	38	1	566	38	1
24	579	33	1	578	33	1	577	36	1	576	36	1
25	588	31	2	588	31	2	588	34	2	588	34	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 2011 (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	594	29	2	593	30	2	594	32	2	593	32	2
27	601	28	2	600	28	2	601	30	2	600	30	2
28	608	27	2	607	27	2	609	29	2	608	29	2
29	614	26	2	613	26	2	616	28	2	615	28	2
30	621	25	2	620	25	2	623	27	2	622	27	2
31	627	24	2	626	24	2	630	26	2	629	26	2
32	633	23	2	632	24	2	636	25	2	635	25	2
33	639	23	2	638	23	2	642	24	2	642	24	2
34	644	22	2	644	23	2	649	24	2	648	24	2
35	650	22	2	650	22	2	655	23	2	654	23	2
36	656	22	2	655	22	2	661	23	2	660	23	2
37	662	22	2	661	22	2	667	23	2	666	23	2
38	667	22	2	667	22	2	673	22	2	672	22	2
39	673	22	2	673	22	2	679	22	2	678	22	2
40	679	22	2	679	22	2	685	22	2	684	22	2
41	685	22	2	684	22	2	693	22	3	693	22	3
42	693	22	3	693	22	3	697	22	3	696	22	3
43	697	22	3	696	22	3	703	22	3	703	22	3
44	703	22	3	703	22	3	709	22	3	709	22	3
45	709	22	3	709	23	3	716	22	3	715	23	3
46	716	23	3	716	23	3	722	23	3	722	23	3
47	723	23	3	722	23	3	729	23	3	729	23	3
48	730	24	3	729	24	3	736	23	3	736	23	3
49	737	24	3	737	24	3	743	24	3	743	24	3
50	744	25	3	744	25	3	751	24	3	750	24	3
51	752	26	3	752	26	3	758	25	3	758	25	3
52	761	26	3	760	27	3	767	26	3	767	26	3
53	769	28	3	769	28	3	775	27	3	775	27	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 2011 (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	779	29	3	779	29	3	785	28	3	785	28	3
55	797	31	4	797	31	4	797	29	4	797	29	4
56	800	32	4	801	33	4	805	31	4	805	31	4
57	813	35	4	813	35	4	817	33	4	817	33	4
58	827	37	4	827	38	4	830	35	4	830	36	4
59	842	41	4	844	42	4	844	38	4	845	39	4
60	860	44	4	862	45	4	861	42	4	863	43	4
61	882	48	4	885	49	4	881	45	4	883	46	4
62	908	50	4	912	50	4	905	47	4	909	48	4
63	942	47	4	947	47	4	937	46	4	942	45	4
64	990	36	4	996	35	4	983	37	4	988	36	4
65	999	34	4	999	35	4	999	33	4	999	33	4
66	999	34	4	999	35	4	999	33	4	999	33	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 2011 (cont.)

Raw Score	Algebra I Core A			Algebra I Core B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
0	490	54	1	490	54	1
1	490	54	1	490	54	1
2	490	54	1	490	54	1
3	490	54	1	490	54	1
4	490	54	1	490	54	1
5	490	54	1	490	54	1
6	490	54	1	490	54	1
7	490	54	1	490	54	1
8	490	54	1	490	54	1
9	490	54	1	490	54	1
10	490	54	1	490	54	1
11	490	54	1	490	54	1
12	490	54	1	504	54	1
13	537	57	1	560	59	1
14	575	60	1	587	59	1
15	598	59	1	606	58	1
16	615	56	1	620	54	1
17	628	52	1	632	48	1
18	640	46	1	642	43	1
19	649	41	1	651	37	1
20	662	35	2	662	32	2
21	665	31	2	666	28	2
22	672	27	2	672	25	2
23	679	24	2	678	23	2
24	685	22	2	684	21	2
25	690	20	2	689	19	2
26	695	19	2	694	18	2
27	700	17	3	700	17	3
28	705	16	3	704	17	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 2011 (cont.)

Raw Score	Algebra I Core A			Algebra I Core B		
	Scale Score	*CSEM*	Perf. Level	Scale Score	*CSEM*	Perf. Level
29	709	16	3	708	16	3
30	713	15	3	712	15	3
31	717	15	3	717	15	3
32	721	14	3	721	14	3
33	725	14	3	725	14	3
34	729	14	3	729	14	3
35	733	13	3	733	14	3
36	737	13	3	737	13	3
37	741	13	3	741	13	3
38	745	13	3	745	13	3
39	749	13	3	749	13	3
40	753	13	3	753	13	3
41	757	13	3	757	13	3
42	762	13	4	762	14	4
43	765	13	4	766	14	4
44	770	14	4	771	14	4
45	775	14	4	776	15	4
46	780	15	4	781	15	4
47	786	16	4	787	16	4
48	792	18	4	793	18	4
49	799	20	4	800	20	4
50	807	24	4	809	23	4
51	817	30	4	818	29	4
52	829	40	4	831	39	4
53	846	55	4	848	54	4
54	875	68	4	877	67	4
55	999	34	4	999	32	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 interpretation 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 2010-11 and Table 5.2 for Spring 2011 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 2010-11

Subject	Accuracy	Consistency	Kappa	False Positives	False Negatives
Algebra I	0.72	0.65	0.52	0.17	0.11
Algebra II	0.78	0.73	0.60	0.16	0.07
Biology I	0.77	0.71	0.59	0.11	0.11
English II	0.78	0.73	0.57	0.06	0.15
English III	0.76	0.73	0.58	0.20	0.04
Geometry	0.79	0.75	0.63	0.11	0.10
U.S. History	0.75	0.72	0.59	0.21	0.04

Table 5.2. Estimates of Accuracy and Consistency of Performance Classification for Spring 2011

Subject	Core	Accuracy	Consistency	Kappa	False Positives	False Negatives
Algebra I	A	0.71	0.73	0.58	0.02	0.27
	B	0.71	0.72	0.56	0.02	0.27
Algebra II	A	0.74	0.72	0.60	0.03	0.23
	B	0.73	0.71	0.59	0.03	0.23
Biology I	A	0.77	0.73	0.59	0.09	0.15
	B	0.77	0.73	0.58	0.09	0.13
English II	AA	0.75	0.70	0.52	0.04	0.21
	AB	0.77	0.70	0.52	0.06	0.17
	BA	0.78	0.70	0.50	0.10	0.12
	BB	0.79	0.71	0.53	0.08	0.14
English III	AA	0.80	0.74	0.55	0.13	0.08
	AB	0.79	0.73	0.55	0.13	0.07
	BA	0.81	0.74	0.55	0.09	0.10
	BB	0.80	0.74	0.55	0.12	0.07
Geometry	A	0.81	0.75	0.62	0.13	0.06
	B	0.79	0.74	0.61	0.09	0.11
U.S. History	A	0.78	0.73	0.59	0.05	0.17
	B	0.78	0.73	0.59	0.05	0.17

Note: Core AA=Core MC form A+OE form A; Core AB=Core MC form A+OE form B; Core BA=Core MC form B+OE form A; Core BB=Core MC form B+OE form B.

As shown in Table 5.1 and Table 5.2, the overall accuracy indices range between 72 and 79 percent for Winter/Trimester 2010-11 and 71 and 81 percent for Spring 2011 and overall consistency ranging between 65 and 75 percent for Winter/Trimester 2010-11 and 70 and 75 percent for Spring 2011. Kappa coefficients range from 0.52 and 0.63 for Winter/Trimester 2010-11 and 0.50 and 0.62 for Spring 2011. The rate of false positives range from 6 to 21 percent for Winter/Trimester 2010-11 and 2 to 13 percent for Spring 2011. The false negative rates range from 4 to 15 percent for Winter/Trimester 2010-11 and 6 to 27 percent for Spring 2011.

Table 5.3 and Table 5.4 provide the accuracy-, consistency-, false positive-, and false negative rates by cut score for Winter/Trimester 2010-11 and Spring 2011, respectively. The data in these tables reveal that the level of agreement for both accuracy and consistency is above 85 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 2010-11 and Spring 2011 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 2010-11 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 2010-11

	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	0.90	0.88	0.94	0.86	0.86	0.92	0.03	0.11	0.03	0.07	0.02	0.03
Algebra II	0.92	0.91	0.93	0.91	0.89	0.91	0.06	0.07	0.04	0.02	0.02	0.03
Biology I	0.94	0.91	0.92	0.92	0.89	0.89	0.05	0.02	0.05	0.02	0.06	0.03
English II	0.98	0.92	0.89	0.97	0.90	0.85	0.02	0.01	0.04	0.01	0.07	0.07
English III	0.94	0.92	0.88	0.94	0.91	0.86	0.05	0.06	0.10	0.01	0.01	0.02
Geometry	0.95	0.93	0.91	0.93	0.91	0.91	0.01	0.02	0.08	0.05	0.05	0.01
U.S. History	0.95	0.91	0.89	0.95	0.89	0.87	0.04	0.07	0.10	0.01	0.02	0.01

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 2011

		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	0.95	0.88	0.88	0.94	0.89	0.89	0.00	0.00	0.01	0.05	0.12	0.11
	B	0.95	0.88	0.87	0.94	0.89	0.88	0.00	0.00	0.01	0.05	0.12	0.11
Algebra II	A	0.93	0.89	0.92	0.91	0.89	0.91	0.01	0.01	0.01	0.06	0.11	0.06
	B	0.92	0.88	0.93	0.90	0.88	0.92	0.01	0.01	0.01	0.07	0.11	0.06
Biology I	A	0.95	0.92	0.90	0.95	0.90	0.87	0.00	0.01	0.08	0.05	0.08	0.02
	B	0.96	0.92	0.89	0.96	0.91	0.86	0.00	0.01	0.09	0.04	0.07	0.03
English II	AA	0.99	0.94	0.82	0.99	0.92	0.79	0.01	0.01	0.02	0.00	0.05	0.15
	AB	0.99	0.95	0.83	0.99	0.92	0.79	0.00	0.02	0.03	0.00	0.03	0.13
	BA	0.99	0.94	0.84	0.99	0.92	0.78	0.00	0.01	0.09	0.01	0.05	0.07
	BB	0.99	0.94	0.85	0.99	0.92	0.80	0.00	0.03	0.04	0.01	0.02	0.11
English III	AA	0.96	0.94	0.90	0.95	0.92	0.85	0.02	0.04	0.06	0.01	0.02	0.04
	AB	0.96	0.94	0.89	0.95	0.92	0.85	0.03	0.04	0.07	0.01	0.02	0.04
	BA	0.96	0.94	0.90	0.95	0.92	0.86	0.01	0.02	0.05	0.03	0.03	0.05
	BB	0.96	0.94	0.90	0.95	0.92	0.86	0.02	0.04	0.07	0.02	0.02	0.04
Geometry	A	0.96	0.94	0.91	0.96	0.91	0.88	0.03	0.03	0.07	0.01	0.04	0.02
	B	0.96	0.93	0.91	0.95	0.90	0.88	0.00	0.02	0.07	0.04	0.05	0.02
U.S. History	A	0.97	0.91	0.90	0.96	0.90	0.87	0.00	0.01	0.04	0.03	0.08	0.06
	B	0.97	0.91	0.90	0.96	0.90	0.87	0.00	0.01	0.04	0.03	0.08	0.06

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 2010-11 and Spring 2011 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 2010-11 - Overall

	Total			
Subject	***N***	**Mean**	***SD***	**Med.**
Algebra I	1,348	708.3	60.9	703
Algebra II	1,053	720.7	92.0	736
Biology I	1,376	725.9	84.2	728
English II	1,425	736.3	81.6	740
English III	1,420	754.9	81.4	766
Geometry	1,544	735.5	78.0	738
U.S. History	1,394	718.7	80.9	727

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

Table 6.2. Descriptive Statistics of the Scale Scores for Winter/Trimester 2010-11 by Gender

	Female				Male			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Algebra I	636	708.2	55.3	708	698	709	65.8	703
Algebra II	536	721.9	92.3	736	509	720.2	92.0	731
Biology I	725	724.7	80.3	728	651	727.2	88.4	734
English II	713	746.9	76.5	746	705	726.9	84.6	734
English III	717	765.8	72.9	772	703	743.7	87.9	756
Geometry	728	734.0	74.2	738	789	733.5	79.8	738
U.S. History	695	708.3	76.9	716	696	729.2	83.6	738

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

Table 6.3. Descriptive Statistics of the Scale Scores for Winter/Trimester 2010-11 by Race/Ethnicity

	African-American				Native American			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Algebra I	126	671.8	49.7	676	249	704.7	57.9	703
Algebra II	101	617.9	106.4	614	167	718.4	91.5	731
Biology I	143	665.2	84.3	661	205	722.1	77.8	710
English II	132	702.6	84.3	707	213	718.9	72.5	728
English III	123	677.1	87.1	684	182	741.3	81.7	756
Geometry	147	676.1	71.5	679	228	733.6	72.1	743
U.S. History	141	682.8	97.7	699	226	709.1	84.0	716

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

Table 6.3. Descriptive Statistics of the Scale Scores for Winter/Trimester 2010-11 by Race/Ethnicity (cont.)

	Hispanic				Asian			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Algebra I	149	692.2	58.7	687	28	718.3	71.3	700
Algebra II	69	689.5	78.7	688	20	751.6	98.4	752
Biology I	117	685.3	91.5	697	14	728.5	107.2	734
English II	113	687.3	87.6	699	32	728.2	109.9	747
English III	85	706.5	95.4	715	50	819.4	48.9	819
Geometry	117	712.0	83.1	714	60	809.0	69.8	825
U.S. History	116	694.5	77.5	696	19	725.5	88.9	710

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

Table 6.3. Descriptive Statistics of the Scale Scores for Winter/Trimester 2010-11 by Race/Ethnicity (cont.)

	White			
Subject	***N***	**Mean**	***SD***	**Med.**
Algebra I	746	716.9	58.9	713
Algebra II	661	739.7	79.0	747
Biology I	864	741.9	78.2	749
English II	893	752.2	76.4	753
English III	919	768.3	72.2	772
Geometry	946	743.7	73.2	750
U.S. History	848	729.2	74.4	738

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

Table 6.4. Descriptive Statistics of the Scale Scores for Winter/Trimester 2010-11 by Free/Reduced Lunch Status

	Free/Reduced Lunch = Yes				Free/Reduced Lunch = No			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Algebra I	599	695.4	57.2	700	735	719.4	61.9	713
Algebra II	355	677.6	100.9	700	690	743.4	78.5	752
Biology I	508	695.5	86.4	697	868	743.6	77.6	752
English II	599	715.7	79.5	722	819	752.5	79.0	760
English III	526	723.0	86.7	739	894	773.6	71.9	777
Geometry	606	706.4	78.6	708	911	751.9	70.6	757
U.S. History	618	699.4	85.0	710	773	734.3	74.1	738

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

Table 6.5. Descriptive Statistics of the Scale Scores for Spring 2011 - Overall

	Total			
Subject	***N***	**Mean**	***SD***	**Med.**
Core A				
Algebra I	14,873	732.2	54.6	737
Algebra II	14,685	726.5	83.3	732
Biology I	17,798	739.7	76.2	741
English II – OE A	8,134	760.4	75.2	761
English II – OE B	7,995	761.3	73.8	760
English III – OE A	8,167	756.2	68.8	760
English III – OE B	8,131	756.5	67.8	760
Geometry	17,947	752.1	74.4	753
U.S. History	16,228	733.1	73.8	738
Core B				
Algebra I	12,190	734.7	52.1	737
Algebra II	12,335	722.7	85.7	728
Biology I	14,808	744.9	72.9	746
English II – OE A	7,770	762.0	72.6	758
English II – OE B	7,803	761.7	73.4	767
English III – OE A	7,707	760.9	65.5	761
English III – OE B	7,699	762.0	65.2	765
Geometry	15,123	747.9	72.6	753
U.S. History	13,537	733.9	75.6	737

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

Table 6.6. Descriptive Statistics of the Scale Scores for Spring 2011 by Gender

	Female				Male			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Core A								
Algebra I	7,099	733.1	52.8	737	6,892	730.5	56.4	733
Algebra II	7,401	727.9	80.0	732	6,835	725.3	86.3	732
Biology I	8,897	734.9	72.6	735	8,649	745.2	79.4	747
English II – OE A	4,162	767.8	73.1	769	3,971	752.7	76.5	752
English II – OE B	4,012	769.4	72.8	769	3,983	753.2	74.0	752
English III – OE A	3,921	765.5	64.8	770	4,245	747.7	71.3	751
English III – OE B	4,019	764.1	64.5	765	4,112	749.1	70.1	755
Geometry	9,012	751.6	72.9	753	8,559	752.4	75.9	753
U.S. History	8,179	719.9	69.7	721	7,855	747.1	75.3	749
Core B								
Algebra I	5,751	736.8	49.1	737	5,705	733.0	54.9	733
Algebra II	6,148	724.1	81.6	728	5,806	721.3	89.3	728
Biology I	7,390	741.0	69.8	746	7,189	749.4	75.7	758
English II – OE A	3,976	768.2	71.5	767	3,794	755.5	73.1	758
English II – OE B	4,015	768.3	72.8	767	3,788	754.8	73.5	758
English III – OE A	3,955	766.0	62.8	766	3,752	755.4	67.9	761
English III – OE B	3,860	768.0	62.0	765	3,839	755.9	67.8	760
Geometry	7,502	747.9	70.1	748	7,302	747.8	75.2	753
U.S. History	6,699	721.3	71.9	725	6,676	746.9	77.0	748

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

Table 6.7. Descriptive Statistics of the Scale Scores for Spring 2011 by Race/Ethnicity

	African-American				Native American			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Core A								
Algebra I	1,262	711.3	55.8	717	2,384	722.1	53.5	725
Algebra II	1,233	686.0	88.2	694	2,331	711.2	79.5	717
Biology I	1,545	694.5	79.0	699	3,059	730.7	72.3	735
English II – OE A	864	726.5	75.3	730	1,296	756.6	71.7	752
English II – OE B	830	731.6	70.3	737	1,309	751.2	66.4	752
English III – OE A	834	727.9	73.8	733	1,331	744.5	67.9	751
English III – OE B	823	730.3	70.1	736	1,397	746.6	63.7	750
Geometry	1,674	711.4	76.1	718	3,082	742.6	69.0	743
U.S. History	1,488	694.2	75.7	697	2,621	723.8	72.4	726
Core B								
Algebra I	1,010	716.1	55.1	721	2,037	725.6	48.8	729
Algebra II	1,057	689.1	86.1	700	1,932	707.3	82.4	717
Biology I	1,175	703.6	72.1	709	2,530	736.3	68.3	739
English II – OE A	829	728.4	68.4	729	1,268	754.5	66.4	751
English II – OE B	819	730.2	69.9	736	1,288	751.8	69.7	758
English III – OE A	750	730.3	67.0	737	1,237	752.7	62.5	756
English III – OE B	744	732.6	64.8	741	1,196	752.6	62.6	755
Geometry	1,442	705.4	72.5	712	2,631	739.3	67.9	743
U.S. History	1,232	693.9	82.2	696	2,152	722.5	70.2	725

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

Table 6.7. Descriptive Statistics of the Scale Scores for Spring 2011 by Race/Ethnicity**Error! Reference source not found.** (cont.)

	Hispanic				Asian			
Subject	***N***	**Mean**	***SD***	**Med.**	***N***	**Mean**	***SD***	**Med.**
Core A								
Algebra I	2,361	724.2	57.5	729	269	772.5	62.1	775
Algebra II	1,738	706.6	86.6	711	368	791.3	86.2	791
Biology I	2,060	711.2	77.0	711	380	772.1	92.0	767
English II – OE A	859	725.3	78.5	723	176	763.3	99.2	769
English II – OE B	788	724.9	75.6	729	200	777.3	91.7	779
English III – OE A	786	738.8	66.9	742	184	776.5	74.6	779
English III – OE B	794	732.9	68.3	736	187	775.9	71.3	786
Geometry	2,135	732.1	78.0	733	368	808.7	77.4	813
U.S. History	1,707	708.1	78.7	715	354	740.2	83.0	749
Core B								
Algebra I	1,925	724.4	52.2	729	246	770.5	62.8	774
Algebra II	1,474	702.6	89.8	712	292	788.2	89.9	785
Biology I	1,730	714.5	76.6	721	303	777.2	75.9	787
English II – OE A	824	726.7	75.7	729	184	789.3	85.5	797
English II – OE B	821	726.5	79.1	729	178	780.3	87.2	785
English III – OE A	728	740.2	66.1	742	191	785.5	66.5	792
English III – OE B	715	743.2	64.3	746	170	786.4	68.1	792
Geometry	1,869	726.5	71.8	728	326	801.6	79.1	809
U.S. History	1,448	708.2	76.5	714	318	752.4	89.1	754

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

Table 6.7. Descriptive Statistics of the Scale Scores for Spring 2011 by Race/Ethnicity (cont.)

	White			
Subject	***N***	**Mean**	***SD***	**Med.**
Core A				
Algebra I	8,098	739.3	51.8	741
Algebra II	8,550	738.3	78.5	743
Biology I	10,147	753.8	71.5	753
English II – OE A	4,674	773.8	69.9	769
English II – OE B	4,605	775.1	71.0	769
English III – OE A	4,768	766.5	66.0	770
English III – OE B	4,642	767.6	65.6	770
Geometry	10,133	763.6	70.6	764
U.S. History	9,576	745.3	69.2	747
Core B				
Algebra I	6,543	741.9	50.6	745
Algebra II	7,230	733.2	82.3	738
Biology I	8,557	757.9	69.1	758
English II – OE A	4,431	775.9	69.1	775
English II – OE B	4,468	775.8	69.2	775
English III – OE A	4,566	770.2	63.7	771
English III – OE B	4,631	771.2	63.7	775
Geometry	8,388	760.2	69.4	764
U.S. History	7,990	747.0	71.6	748

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

Table 6.8. Descriptive Statistics of the Scale Scores for Spring 2011 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	6,458	717.5	55.1	721	7,533	744.1	51.0	745
Algebra II	5,482	700.1	83.4	706	8,754	743.3	78.5	748
Biology I	7,682	717.1	75.8	723	9,864	757.8	71.6	760
English II – OE A	3,626	737.1	74.5	737	4,508	779.2	70.3	779
English II – OE B	3,563	737.7	72.5	737	4,432	780.3	69.3	779
English III – OE A	3,551	733.4	69.1	737	4,616	773.8	63.2	776
English III – OE B	3,587	733.5	67.8	741	4,544	774.7	62.0	775
Geometry	7,684	731.3	73.2	733	9,887	768.0	71.2	769
U.S. History	6,352	710.7	74.6	715	9,682	748.0	69.4	749
Core B								
Algebra I	5,280	721.5	50.7	725	6,176	746.3	50.6	749
Algebra II	4,602	697.5	84.2	706	7,352	738.5	82.3	743
Biology I	6,301	724.0	72.2	727	8,278	761.2	69.3	765
English II – OE A	3,502	739.1	70.7	743	4,268	780.8	68.7	775
English II – OE B	3,501	739.1	73.1	743	4,302	780.2	68.4	785
English III – OE A	3,200	740.3	66.6	747	4,507	775.5	60.7	776
English III – OE B	3,206	742.2	64.6	746	4,493	776.1	61.9	775
Geometry	6,492	726.2	70.7	728	8,312	764.7	69.6	764
U.S. History	5,406	712.1	75.3	720	7,969	749.0	72.0	748

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 2010-11 and Spring 2011 administration are presented in Table 6.9 (please see Appendix B and Appendix C for distributions by scale score for Winter/Trimester 2010-11 and Spring 2011, 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 2009-10 and Spring 2010).

Table 6.9. Percentage of Students by Performance Level for Winter/Trimester 2010-11 and Spring 2011

Subject	*N*	Unsatisfactory	Limited Knowledge	Proficient	Advanced
Winter 2010-11					
Algebra I	1,348	14.9%	26.7%	38.7%	19.7%
Algebra II	1,053	18.3%	13.1%	44.8%	23.7%
Biology I	1,376	11.4%	20.2%	34.2%	34.2%
English II	1,425	3.7%	21.5%	50.9%	23.9%
English III	1,420	12.9%	7.2%	50.4%	29.5%
Geometry	1,544	7.1%	20.8%	39.8%	32.3%
U.S. History	1,394	6.5%	23.7%	28.3%	41.6%
Spring 2011 Core A					
Algebra I	14,873	6.8%	14.6%	48.8%	29.8%
Algebra II	14,685	13.8%	18.2%	42.0%	26.0%
Biology I	17,798	5.9%	15.3%	44.4%	34.3%
English II – OE A	8,134	1.6%	12.9%	49.3%	36.1%
English II – OE B	7,995	1.3%	11.9%	50.8%	36.0%
English III – OE A	8,167	9.4%	6.7%	56.3%	27.6%
English III – OE B	8,131	8.8%	7.7%	56.0%	27.6%
Geometry	17,947	4.7%	12.8%	44.5%	38.0%
U.S. History	16,228	3.3%	18.9%	30.4%	47.4%
Spring 2011 Core B					
Algebra I	12,190	5.4%	14.7%	49.5%	30.4%
Algebra II	12,335	13.8%	20.2%	41.5%	24.5%
Biology I	14,808	5.4%	12.4%	43.9%	38.3%
English II – OE A	7,770	1.3%	11.8%	52.8%	34.1%
English II – OE B	7,803	1.5%	11.9%	52.4%	34.3%
English III – OE A	7,707	7.6%	6.6%	56.7%	29.1%
English III – OE B	7,699	7.0%	6.9%	56.4%	29.7%
Geometry	15,123	5.6%	15.8%	41.8%	36.8%
U.S. History	13,537	3.4%	18.2%	30.0%	48.4%

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 the Winter/Trimester 2010-11 and Spring 2011, 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 6.10 presents the overall estimates of *SEM* for each of the content areas for the Winter/Trimester 2010-2011 and Spring 2011 administrations.

Table 6.10. Overall Estimates of *SEM* by Subject

Subject	SEM*
Winter 2010-11	
Algebra I	25.67
Algebra II	36.00
Biology I	29.39
English II	27.48
English III	24.77
Geometry	29.04
U.S. History	28.32
Spring 2011	
Algebra I – A	22.96
Algebra I – B	21.52
Algebra II – A	32.00
Algebra II – B	34.54
Biology I – A	27.77
Biology I – B	27.17
English II – AA	30.77
English II – AB	30.96
English II – BA	29.22
English II – BB	29.48
English III – AA	22.46
English III – AB	22.10
English III – BA	21.35
English III – BB	21.21
Geometry – A	27.18
Geometry – B	26.37
U.S. History - A	27.04
U.S. History - B	27.68

Note: *SEM = Standard Error of Measurement; SEM values are on the reportable scale metric; AA=Core MC form A+OE form A; AB=Core MC form A+OE form B; BA=Core MC form B+OE form A; BB=Core MC form B+OE form B.

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

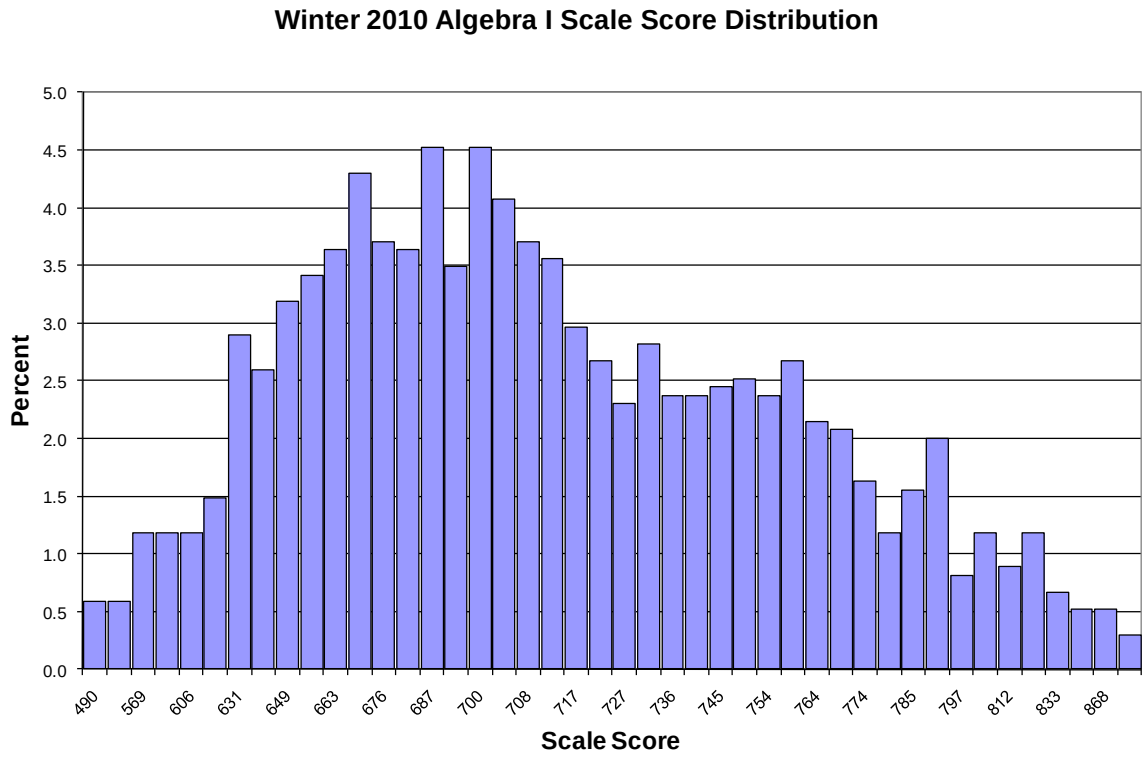
Scale Score Distributions for Winter/Trimester 2010-11

Algebra I Scale Score Distribution for Winter/Trimester 2010-11

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
490	8	0.6	8	0.6
537	8	0.6	16	1.2
569	16	1.2	32	2.4
590	16	1.2	48	3.6
606	16	1.2	64	4.7
619	20	1.5	84	6.2
631	39	2.9	123	9.1
640	35	2.6	158	11.7
649	43	3.2	201	14.9
662	46	3.4	247	18.3
663	49	3.6	296	22.0
670	58	4.3	354	26.3
676	50	3.7	404	30.0
682	49	3.6	453	33.6
687	61	4.5	514	38.1
693	47	3.5	561	41.6
700	61	4.5	622	46.1
703	55	4.1	677	50.2
708	50	3.7	727	53.9
713	48	3.6	775	57.5
717	40	3.0	815	60.5
722	36	2.7	851	63.1
727	31	2.3	882	65.4
731	38	2.8	920	68.2
736	32	2.4	952	70.6
740	32	2.4	984	73.0
745	33	2.4	1,017	75.4
749	34	2.5	1,051	78.0
754	32	2.4	1,083	80.3
762	36	2.7	1,119	83.0
764	29	2.2	1,148	85.2
769	28	2.1	1,176	87.2
774	22	1.6	1,198	88.9
779	16	1.2	1,214	90.1
785	21	1.6	1,235	91.6
791	27	2.0	1,262	93.6
797	11	0.8	1,273	94.4
804	16	1.2	1,289	95.6
812	12	0.9	1,301	96.5
822	16	1.2	1,317	97.7
833	9	0.7	1,326	98.4
847	7	0.5	1,333	98.9
868	7	0.5	1,340	99.4

Algebra I Scale Score Distribution for Winter/Trimester 2010-11 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
904	4	0.3	1,344	99.7
999	4	0.3	1,348	100.0

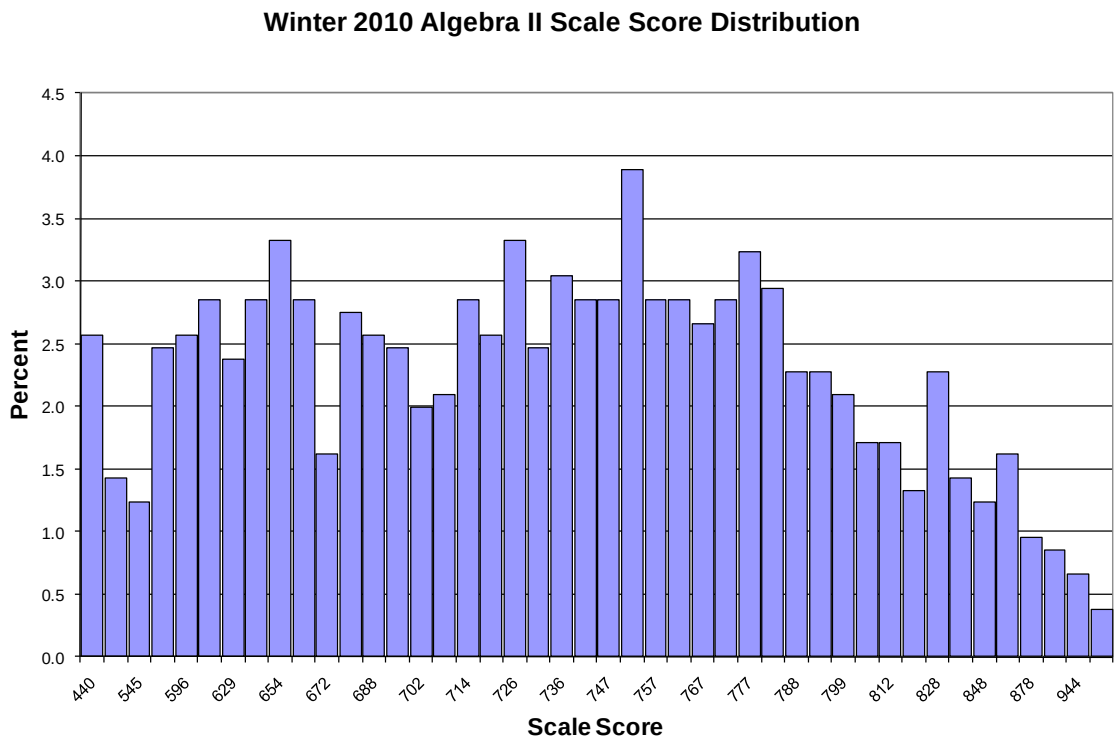


Algebra II Scale Score Distribution for Winter/Trimester 2010-11

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	27	2.6	27	2.6
504	15	1.4	42	4.0
545	13	1.2	55	5.2
574	26	2.5	81	7.7
596	27	2.6	108	10.3
614	30	2.8	138	13.1
629	25	2.4	163	15.5
642	30	2.8	193	18.3
654	35	3.3	228	21.7
663	30	2.8	258	24.5
672	17	1.6	275	26.1
680	29	2.8	304	28.9
688	27	2.6	331	31.4
700	26	2.5	357	33.9
702	21	2.0	378	35.9
708	22	2.1	400	38.0
714	30	2.8	430	40.8
720	27	2.6	457	43.4
726	35	3.3	492	46.7
731	26	2.5	518	49.2
736	32	3.0	550	52.2
742	30	2.8	580	55.1
747	30	2.8	610	57.9
752	41	3.9	651	61.8
757	30	2.8	681	64.7
762	30	2.8	711	67.5
767	28	2.7	739	70.2
772	30	2.8	769	73.0
777	34	3.2	803	76.3
783	31	2.9	834	79.2
788	24	2.3	858	81.5
794	24	2.3	882	83.8
799	22	2.1	904	85.8
806	18	1.7	922	87.6
812	18	1.7	940	89.3
819	14	1.3	954	90.6
828	24	2.3	978	92.9
837	15	1.4	993	94.3
848	13	1.2	1,006	95.5
861	17	1.6	1,023	97.2
878	10	0.9	1,033	98.1
902	9	0.9	1,042	99.0

Algebra II Scale Score Distribution for Winter/Trimester 2010-11 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
944	7	0.7	1,049	99.6
999	4	0.4	1,053	100.0

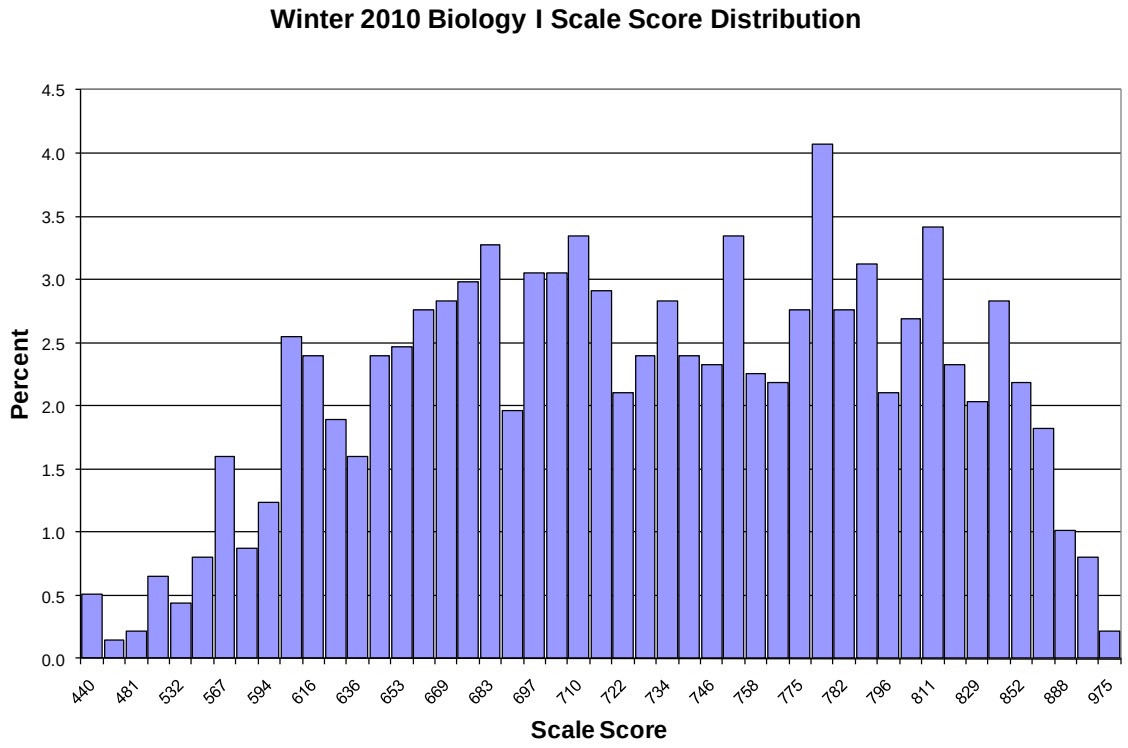


Biology I Scale Score Distribution for Winter/Trimester 2010-11

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	7	0.5	7	0.5
442	2	0.1	9	0.7
481	3	0.2	12	0.9
510	9	0.7	21	1.5
532	6	0.4	27	2.0
551	11	0.8	38	2.8
567	22	1.6	60	4.4
581	12	0.9	72	5.2
594	17	1.2	89	6.5
606	35	2.5	124	9.0
616	33	2.4	157	11.4
627	26	1.9	183	13.3
636	22	1.6	205	14.9
645	33	2.4	238	17.3
653	34	2.5	272	19.8
661	38	2.8	310	22.5
669	39	2.8	349	25.4
676	41	3.0	390	28.3
683	45	3.3	435	31.6
691	27	2.0	462	33.6
697	42	3.1	504	36.6
704	42	3.1	546	39.7
710	46	3.3	592	43.0
716	40	2.9	632	45.9
722	29	2.1	661	48.0
728	33	2.4	694	50.4
734	39	2.8	733	53.3
740	33	2.4	766	55.7
746	32	2.3	798	58.0
752	46	3.3	844	61.3
758	31	2.3	875	63.6
764	30	2.2	905	65.8
775	38	2.8	943	68.5
776	56	4.1	999	72.6
782	38	2.8	1,037	75.4
789	43	3.1	1,080	78.5
796	29	2.1	1,109	80.6
803	37	2.7	1,146	83.3
811	47	3.4	1,193	86.7
819	32	2.3	1,225	89.0
829	28	2.0	1,253	91.1
839	39	2.8	1,292	93.9
852	30	2.2	1,322	96.1

Biology I Scale Score Distribution for Winter/Trimester 2010-11 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
868	25	1.8	1,347	97.9
888	14	1.0	1,361	98.9
918	11	0.8	1,372	99.7
975	3	0.2	1,375	99.9
999	1	0.1	1,376	100.0

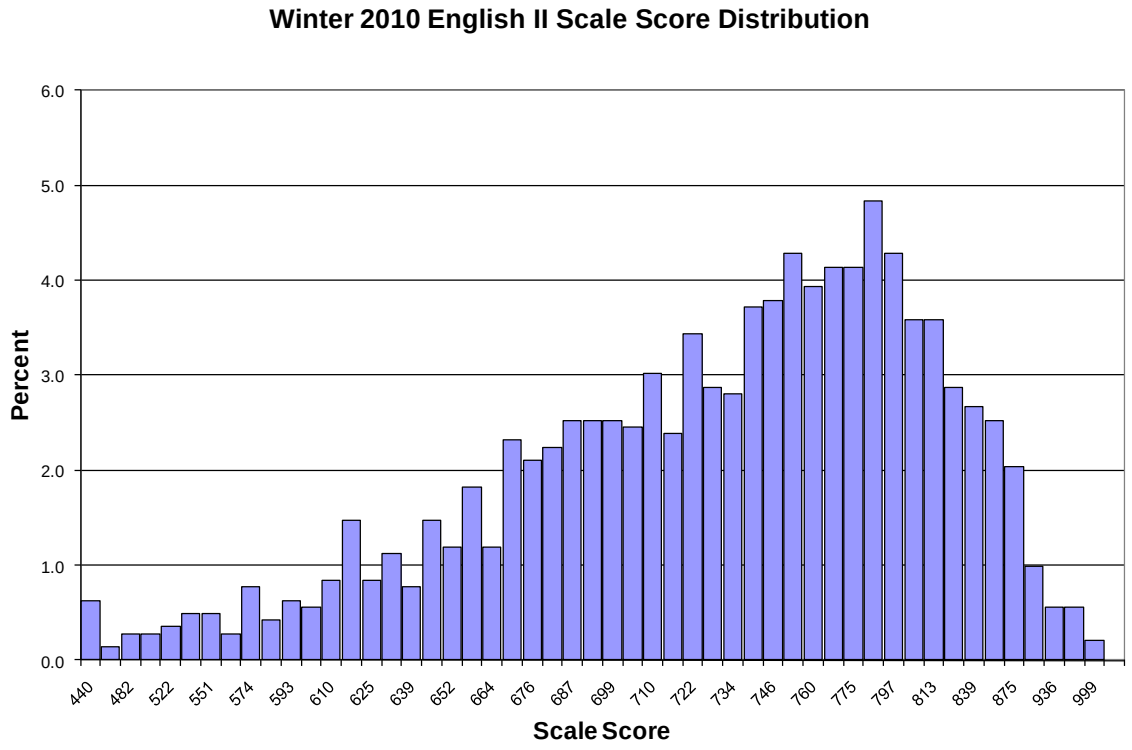


English II Scale Score Distribution for Winter/Trimester 2010-11

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	9	0.6	9	0.6
454	2	0.1	11	0.8
482	4	0.3	15	1.1
504	4	0.3	19	1.3
522	5	0.4	24	1.7
537	7	0.5	31	2.2
551	7	0.5	38	2.7
563	4	0.3	42	2.9
574	11	0.8	53	3.7
588	6	0.4	59	4.1
593	9	0.6	68	4.8
602	8	0.6	76	5.3
610	12	0.8	88	6.2
618	21	1.5	109	7.6
625	12	0.8	121	8.5
632	16	1.1	137	9.6
639	11	0.8	148	10.4
645	21	1.5	169	11.9
652	17	1.2	186	13.1
658	26	1.8	212	14.9
664	17	1.2	229	16.1
670	33	2.3	262	18.4
676	30	2.1	292	20.5
681	32	2.2	324	22.7
687	36	2.5	360	25.3
693	36	2.5	396	27.8
699	36	2.5	432	30.3
704	35	2.5	467	32.8
710	43	3.0	510	35.8
716	34	2.4	544	38.2
722	49	3.4	593	41.6
728	41	2.9	634	44.5
734	40	2.8	674	47.3
740	53	3.7	727	51.0
746	54	3.8	781	54.8
753	61	4.3	842	59.1
760	56	3.9	898	63.0
767	59	4.1	957	67.2
775	59	4.1	1,016	71.3
783	69	4.8	1,085	76.1
797	61	4.3	1,146	80.4
802	51	3.6	1,197	84.0
813	51	3.6	1,248	87.6

English II Scale Score Distribution for Winter/Trimester 2010-11 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
825	41	2.9	1,289	90.5
839	38	2.7	1,327	93.1
855	36	2.5	1,363	95.6
875	29	2.0	1,392	97.7
900	14	1.0	1,406	98.7
936	8	0.6	1,414	99.2
997	8	0.6	1,422	99.8
999	3	0.2	1,425	100.0

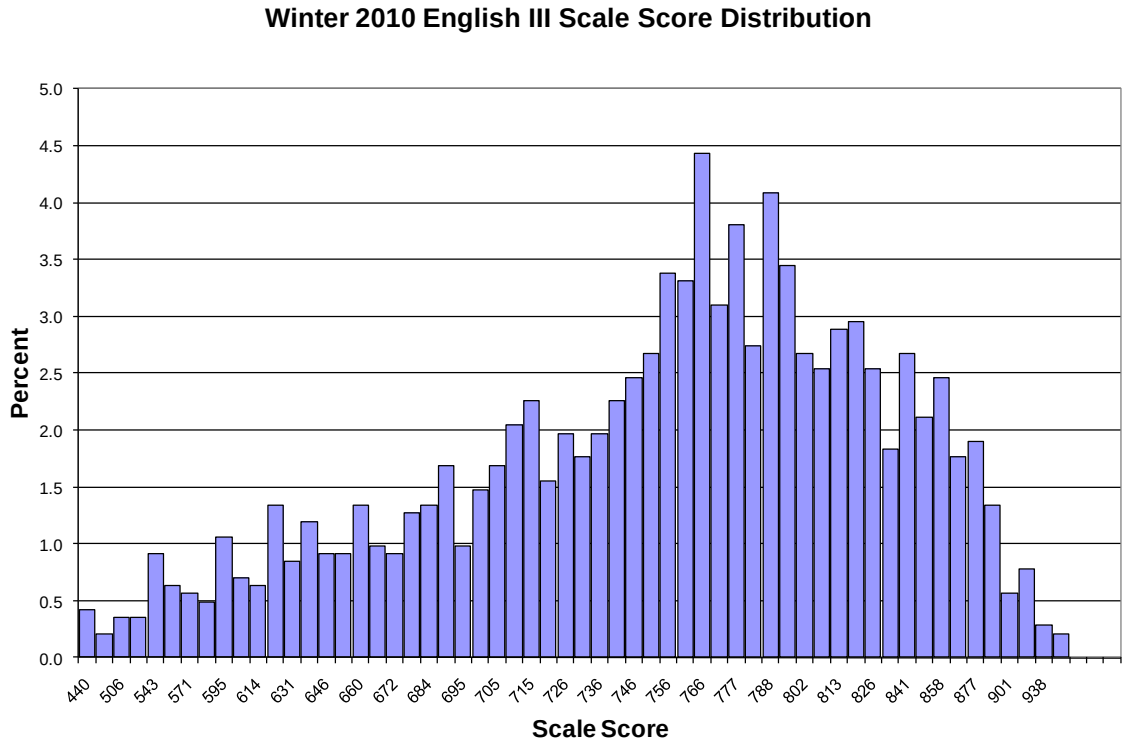


English III Scale Score Distribution for Winter/Trimester 2010-11

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	6	0.4	6	0.4
477	3	0.2	9	0.6
506	5	0.4	14	1.0
527	5	0.4	19	1.3
543	13	0.9	32	2.3
558	9	0.6	41	2.9
571	8	0.6	49	3.5
583	7	0.5	56	3.9
595	15	1.1	71	5.0
605	10	0.7	81	5.7
614	9	0.6	90	6.3
623	19	1.3	109	7.7
631	12	0.8	121	8.5
639	17	1.2	138	9.7
646	13	0.9	151	10.6
653	13	0.9	164	11.5
660	19	1.3	183	12.9
670	14	1.0	197	13.9
672	13	0.9	210	14.8
678	18	1.3	228	16.1
684	19	1.3	247	17.4
689	24	1.7	271	19.1
695	14	1.0	285	20.1
700	21	1.5	306	21.5
705	24	1.7	330	23.2
710	29	2.0	359	25.3
715	32	2.3	391	27.5
721	22	1.5	413	29.1
726	28	2.0	441	31.1
731	25	1.8	466	32.8
736	28	2.0	494	34.8
741	32	2.3	526	37.0
746	35	2.5	561	39.5
751	38	2.7	599	42.2
756	48	3.4	647	45.6
761	47	3.3	694	48.9
766	63	4.4	757	53.3
772	44	3.1	801	56.4
777	54	3.8	855	60.2
783	39	2.7	894	63.0
788	58	4.1	952	67.0
794	49	3.5	1,001	70.5
802	38	2.7	1,039	73.2

English III Scale Score Distribution for Winter/Trimester 2010-11 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
806	36	2.5	1,075	75.7
813	41	2.9	1,116	78.6
819	42	3.0	1,158	81.5
826	36	2.5	1,194	84.1
833	26	1.8	1,220	85.9
841	38	2.7	1,258	88.6
849	30	2.1	1,288	90.7
858	35	2.5	1,323	93.2
867	25	1.8	1,348	94.9
877	27	1.9	1,375	96.8
888	19	1.3	1,394	98.2
901	8	0.6	1,402	98.7
917	11	0.8	1,413	99.5
938	4	0.3	1,417	99.8
968	3	0.2	1,420	100.0

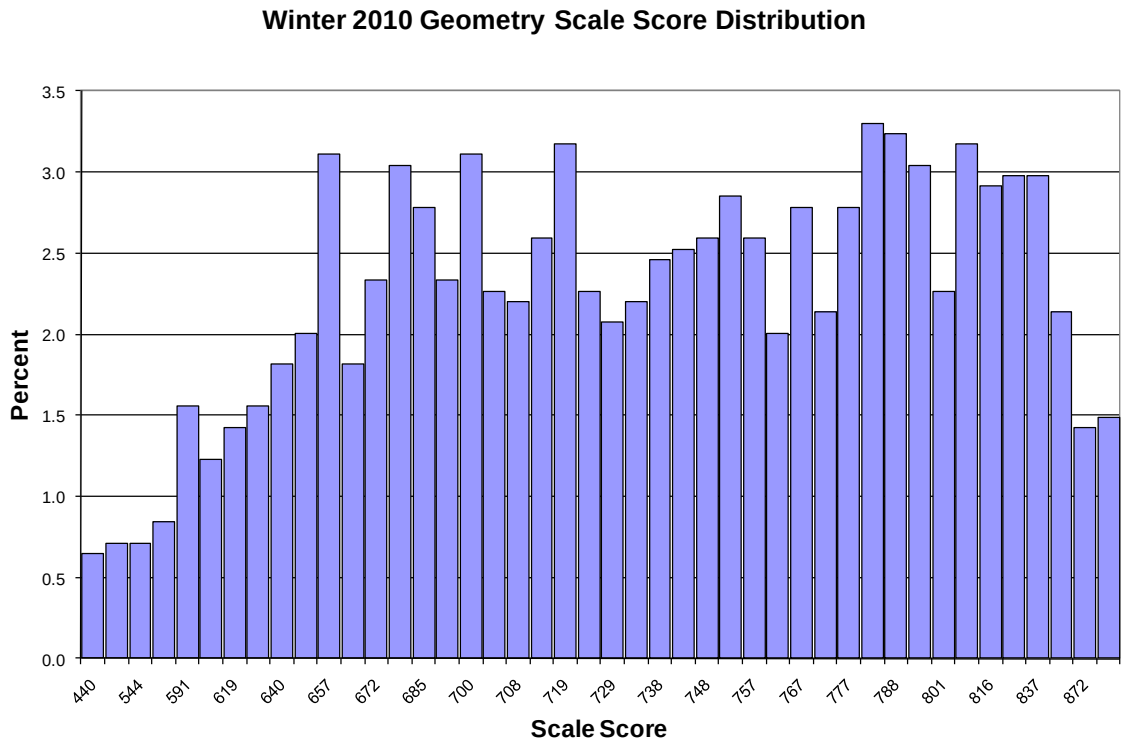


Geometry Scale Score Distribution for Winter/Trimester 2010-11

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	10	0.6	10	0.6
503	11	0.7	21	1.4
544	11	0.7	32	2.1
571	13	0.8	45	2.9
591	24	1.6	69	4.5
606	19	1.2	88	5.7
619	22	1.4	110	7.1
635	24	1.6	134	8.7
640	28	1.8	162	10.5
649	31	2.0	193	12.5
657	48	3.1	241	15.6
665	28	1.8	269	17.4
672	36	2.3	305	19.8
679	47	3.0	352	22.8
685	43	2.8	395	25.6
691	36	2.3	431	27.9
700	48	3.1	479	31.0
703	35	2.3	514	33.3
708	34	2.2	548	35.5
714	40	2.6	588	38.1
719	49	3.2	637	41.3
724	35	2.3	672	43.5
729	32	2.1	704	45.6
734	34	2.2	738	47.8
738	38	2.5	776	50.3
743	39	2.5	815	52.8
748	40	2.6	855	55.4
752	44	2.8	899	58.2
757	40	2.6	939	60.8
762	31	2.0	970	62.8
767	43	2.8	1,013	65.6
772	33	2.1	1,046	67.7
777	43	2.8	1,089	70.5
782	51	3.3	1,140	73.8
788	50	3.2	1,190	77.1
794	47	3.0	1,237	80.1
801	35	2.3	1,272	82.4
808	49	3.2	1,321	85.6
816	45	2.9	1,366	88.5
825	46	3.0	1,412	91.5
837	46	3.0	1,458	94.4
851	33	2.1	1,491	96.6

Geometry Scale Score Distribution for Winter/Trimester 2010-11 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
872	22	1.4	1,513	98.0
907	23	1.5	1,536	99.5
999	8	0.5	1,544	100.0

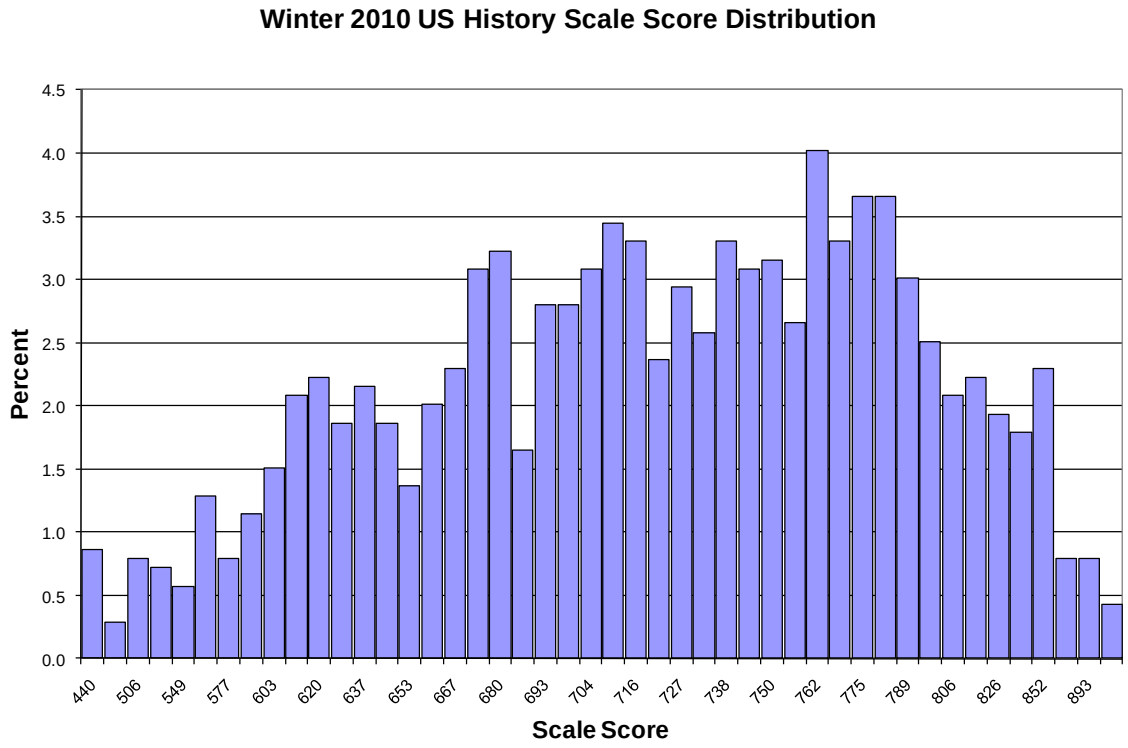


U.S. History Scale Score Distribution for Winter/Trimester 2010-11

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	12	0.9	12	0.9
473	4	0.3	16	1.1
506	11	0.8	27	1.9
530	10	0.7	37	2.7
549	8	0.6	45	3.2
564	18	1.3	63	4.5
577	11	0.8	74	5.3
589	16	1.1	90	6.5
603	21	1.5	111	8.0
610	29	2.1	140	10.0
620	31	2.2	171	12.3
629	26	1.9	197	14.1
637	30	2.2	227	16.3
645	26	1.9	253	18.1
653	19	1.4	272	19.5
660	28	2.0	300	21.5
667	32	2.3	332	23.8
674	43	3.1	375	26.9
680	45	3.2	420	30.1
689	23	1.6	443	31.8
693	39	2.8	482	34.6
699	39	2.8	521	37.4
704	43	3.1	564	40.5
710	48	3.4	612	43.9
716	46	3.3	658	47.2
721	33	2.4	691	49.6
727	41	2.9	732	52.5
733	36	2.6	768	55.1
738	46	3.3	814	58.4
747	43	3.1	857	61.5
750	44	3.2	901	64.6
756	37	2.7	938	67.3
762	56	4.0	994	71.3
768	46	3.3	1,040	74.6
775	51	3.7	1,091	78.3
782	51	3.7	1,142	81.9
789	42	3.0	1,184	84.9
797	35	2.5	1,219	87.4
806	29	2.1	1,248	89.5
815	31	2.2	1,279	91.8
826	27	1.9	1,306	93.7
838	25	1.8	1,331	95.5
852	32	2.3	1,363	97.8

U.S. History Scale Score Distribution for Winter/Trimester 2010-11 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
870	11	0.8	1,374	98.6
893	11	0.8	1,385	99.4
925	6	0.4	1,391	99.8
982	2	0.1	1,393	99.9
999	1	0.1	1,394	100.0



Appendix C

Scale Score Distributions for Spring 2011

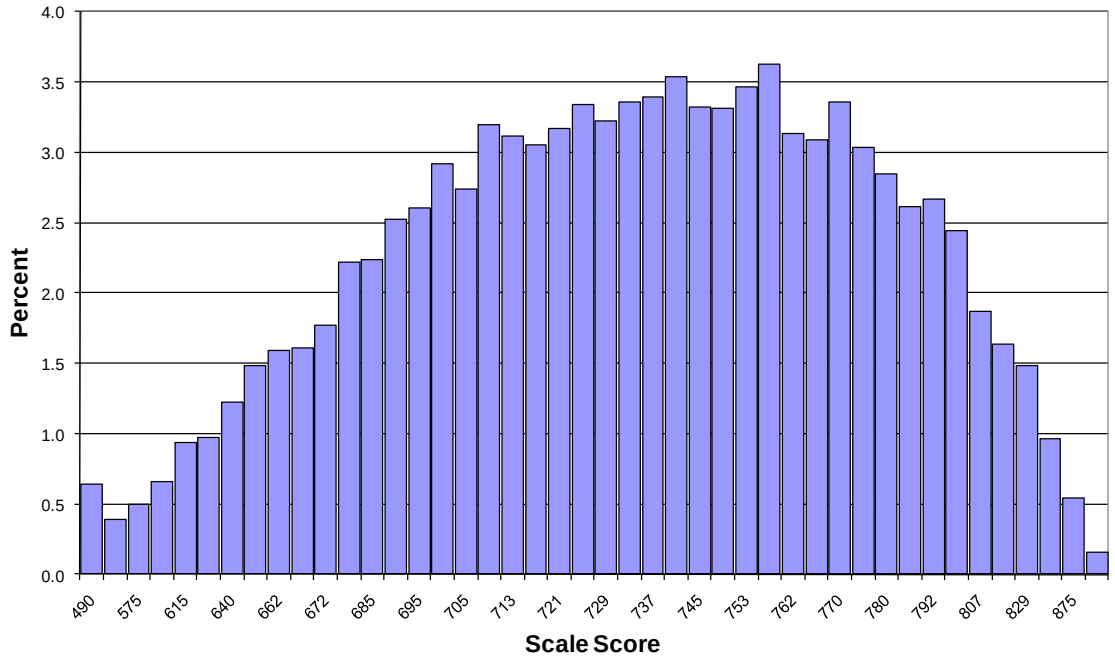
Algebra I Core A Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
490	95	0.6	95	0.6
537	58	0.4	153	1.0
575	74	0.5	227	1.5
598	98	0.7	325	2.2
615	140	0.9	465	3.1
628	145	1.0	610	4.1
640	182	1.2	792	5.3
649	221	1.5	1,013	6.8
662	237	1.6	1,250	8.4
665	240	1.6	1,490	10.0
672	263	1.8	1,753	11.8
679	330	2.2	2,083	14.0
685	333	2.2	2,416	16.2
690	376	2.5	2,792	18.8
695	387	2.6	3,179	21.4
700	434	2.9	3,613	24.3
705	407	2.7	4,020	27.0
709	476	3.2	4,496	30.2
713	464	3.1	4,960	33.3
717	454	3.1	5,414	36.4
721	471	3.2	5,885	39.6
725	497	3.3	6,382	42.9
729	480	3.2	6,862	46.1
733	499	3.4	7,361	49.5
737	505	3.4	7,866	52.9
741	526	3.5	8,392	56.4
745	494	3.3	8,886	59.7
749	493	3.3	9,379	63.1
753	516	3.5	9,895	66.5
757	539	3.6	10,434	70.2
762	466	3.1	10,900	73.3
765	459	3.1	11,359	76.4
770	499	3.4	11,858	79.7
775	452	3.0	12,310	82.8
780	423	2.8	12,733	85.6
786	389	2.6	13,122	88.2
792	397	2.7	13,519	90.9
799	363	2.4	13,882	93.3
807	278	1.9	14,160	95.2
817	244	1.6	14,404	96.8
829	221	1.5	14,625	98.3
846	143	1.0	14,768	99.3

Algebra I Core A Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
875	81	0.5	14,849	99.8
999	24	0.2	14,873	100.0

Spring 2011 Algebra I Core A Scale Score Distribution

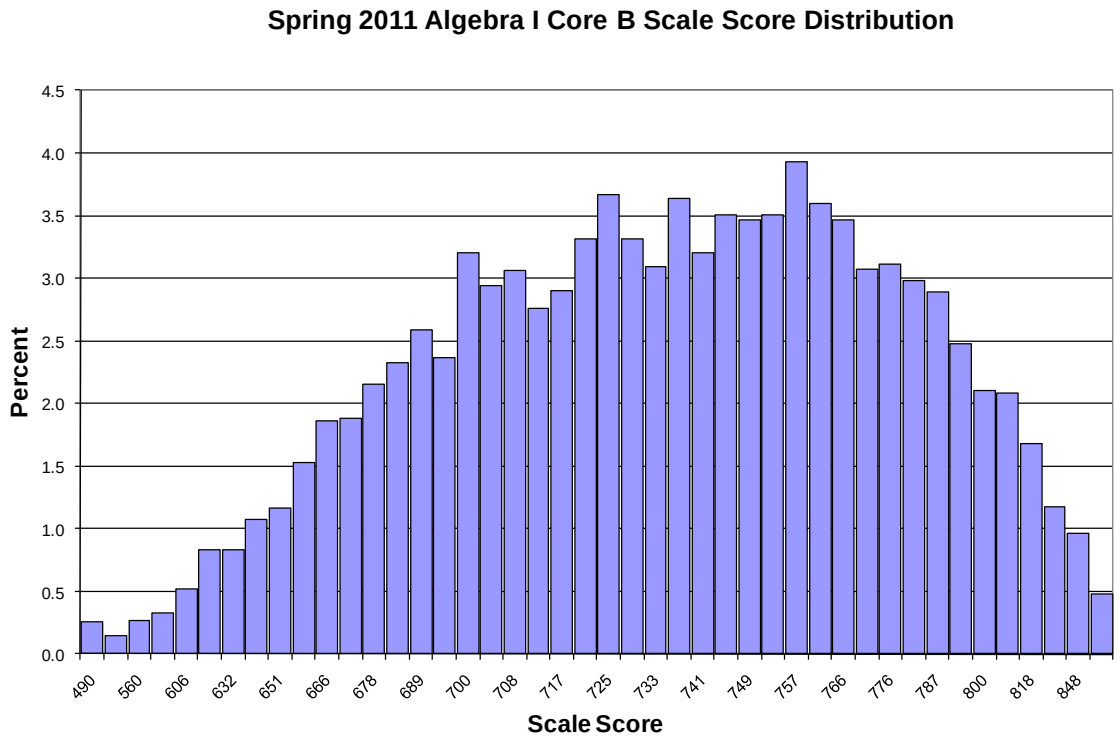


Algebra I Core B Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
490	31	0.3	31	0.3
504	18	0.1	49	0.4
560	33	0.3	82	0.7
587	40	0.3	122	1.0
606	63	0.5	185	1.5
620	101	0.8	286	2.3
632	101	0.8	387	3.2
642	131	1.1	518	4.2
651	142	1.2	660	5.4
662	186	1.5	846	6.9
666	227	1.9	1,073	8.8
672	229	1.9	1,302	10.7
678	262	2.1	1,564	12.8
684	284	2.3	1,848	15.2
689	316	2.6	2,164	17.8
694	288	2.4	2,452	20.1
700	390	3.2	2,842	23.3
704	358	2.9	3,200	26.3
708	373	3.1	3,573	29.3
712	337	2.8	3,910	32.1
717	354	2.9	4,264	35.0
721	404	3.3	4,668	38.3
725	447	3.7	5,115	42.0
729	404	3.3	5,519	45.3
733	377	3.1	5,896	48.4
737	444	3.6	6,340	52.0
741	390	3.2	6,730	55.2
745	427	3.5	7,157	58.7
749	423	3.5	7,580	62.2
753	427	3.5	8,007	65.7
757	479	3.9	8,486	69.6
762	439	3.6	8,925	73.2
766	422	3.5	9,347	76.7
771	375	3.1	9,722	79.8
776	379	3.1	10,101	82.9
781	364	3.0	10,465	85.8
787	352	2.9	10,817	88.7
793	302	2.5	11,119	91.2
800	256	2.1	11,375	93.3
809	254	2.1	11,629	95.4
818	205	1.7	11,834	97.1
831	143	1.2	11,977	98.3

Algebra I Core B Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
848	118	1.0	12,095	99.2
877	58	0.5	12,153	99.7
999	37	0.3	12,190	100.0

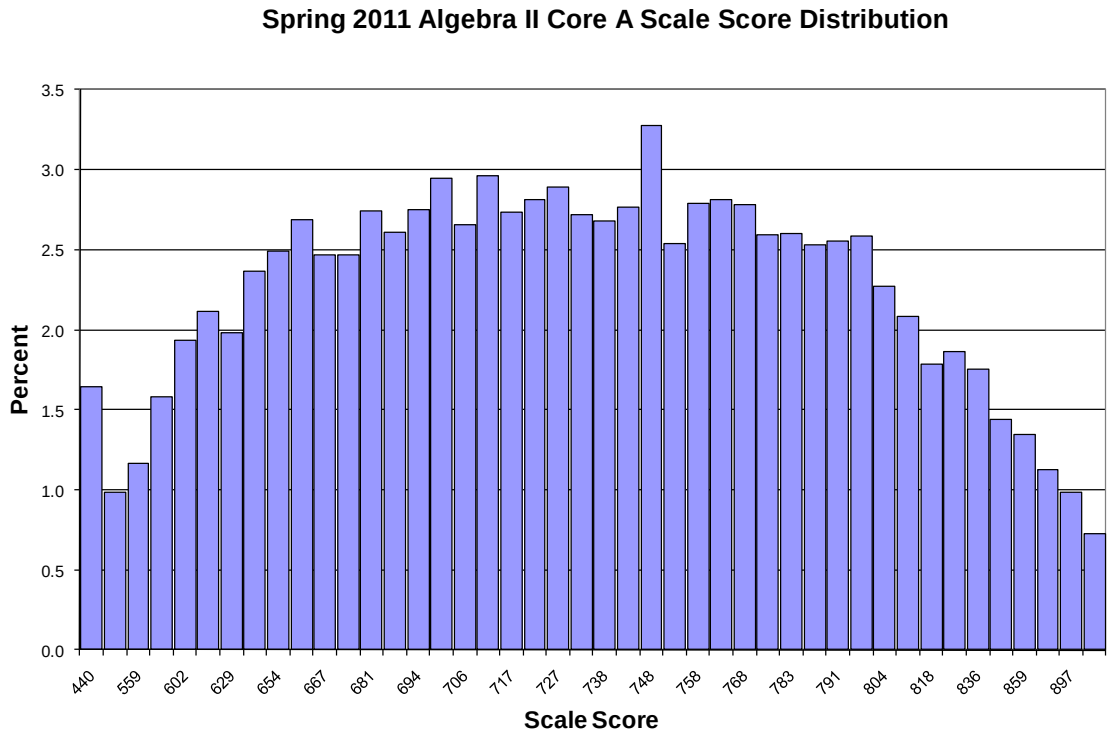


Algebra II Core A Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	241	1.6	241	1.6
521	145	1.0	386	2.6
559	171	1.2	557	3.8
584	232	1.6	789	5.4
602	284	1.9	1,073	7.3
617	310	2.1	1,383	9.4
629	291	2.0	1,674	11.4
640	348	2.4	2,022	13.8
654	366	2.5	2,388	16.3
659	395	2.7	2,783	19.0
667	362	2.5	3,145	21.4
674	363	2.5	3,508	23.9
681	403	2.7	3,911	26.6
688	383	2.6	4,294	29.2
694	404	2.8	4,698	32.0
700	433	2.9	5,131	34.9
706	390	2.7	5,521	37.6
711	435	3.0	5,956	40.6
717	402	2.7	6,358	43.3
722	413	2.8	6,771	46.1
727	425	2.9	7,196	49.0
732	399	2.7	7,595	51.7
738	393	2.7	7,988	54.4
743	406	2.8	8,394	57.2
748	481	3.3	8,875	60.4
753	373	2.5	9,248	63.0
758	410	2.8	9,658	65.8
763	413	2.8	10,071	68.6
768	409	2.8	10,480	71.4
774	381	2.6	10,861	74.0
783	382	2.6	11,243	76.6
785	372	2.5	11,615	79.1
791	375	2.6	11,990	81.6
797	380	2.6	12,370	84.2
804	334	2.3	12,704	86.5
811	306	2.1	13,010	88.6
818	262	1.8	13,272	90.4
826	274	1.9	13,546	92.2
836	257	1.8	13,803	94.0
846	212	1.4	14,015	95.4
859	198	1.3	14,213	96.8
874	165	1.1	14,378	97.9

Algebra II Core A Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
897	145	1.0	14,523	98.9
937	107	0.7	14,630	99.6
999	55	0.4	14,685	100.0

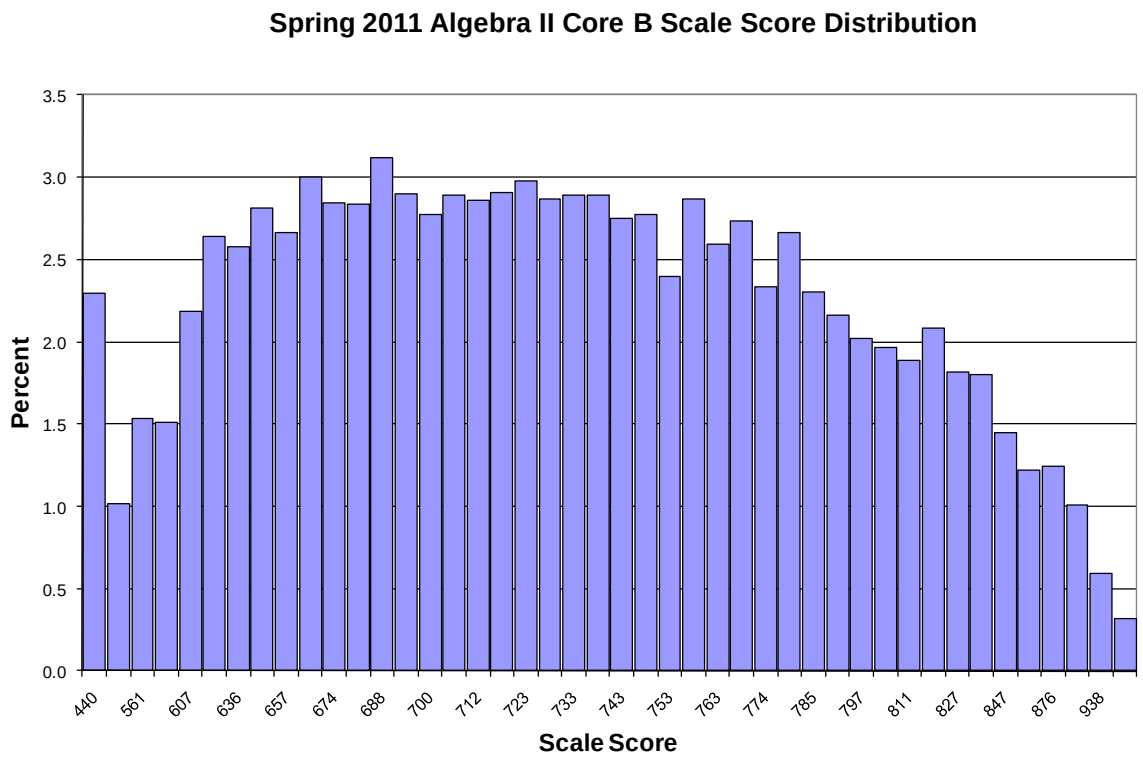


Algebra II Core B Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	283	2.3	283	2.3
522	125	1.0	408	3.3
561	189	1.5	597	4.8
588	186	1.5	783	6.3
607	270	2.2	1,053	8.5
623	326	2.6	1,379	11.2
636	318	2.6	1,697	13.8
654	347	2.8	2,044	16.6
657	329	2.7	2,373	19.2
666	370	3.0	2,743	22.2
674	351	2.8	3,094	25.1
681	350	2.8	3,444	27.9
688	385	3.1	3,829	31.0
694	358	2.9	4,187	33.9
700	342	2.8	4,529	36.7
706	357	2.9	4,886	39.6
712	353	2.9	5,239	42.5
717	359	2.9	5,598	45.4
723	367	3.0	5,965	48.4
728	354	2.9	6,319	51.2
733	357	2.9	6,676	54.1
738	357	2.9	7,033	57.0
743	339	2.7	7,372	59.8
748	342	2.8	7,714	62.5
753	296	2.4	8,010	64.9
758	354	2.9	8,364	67.8
763	320	2.6	8,684	70.4
769	337	2.7	9,021	73.1
774	288	2.3	9,309	75.5
783	329	2.7	9,638	78.1
785	284	2.3	9,922	80.4
791	267	2.2	10,189	82.6
797	249	2.0	10,438	84.6
804	242	2.0	10,680	86.6
811	233	1.9	10,913	88.5
818	257	2.1	11,170	90.6
827	224	1.8	11,394	92.4
836	222	1.8	11,616	94.2
847	179	1.5	11,795	95.6
860	151	1.2	11,946	96.8
876	153	1.2	12,099	98.1
899	124	1.0	12,223	99.1

Algebra II Core B Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
938	73	0.6	12,296	99.7
999	39	0.3	12,335	100.0

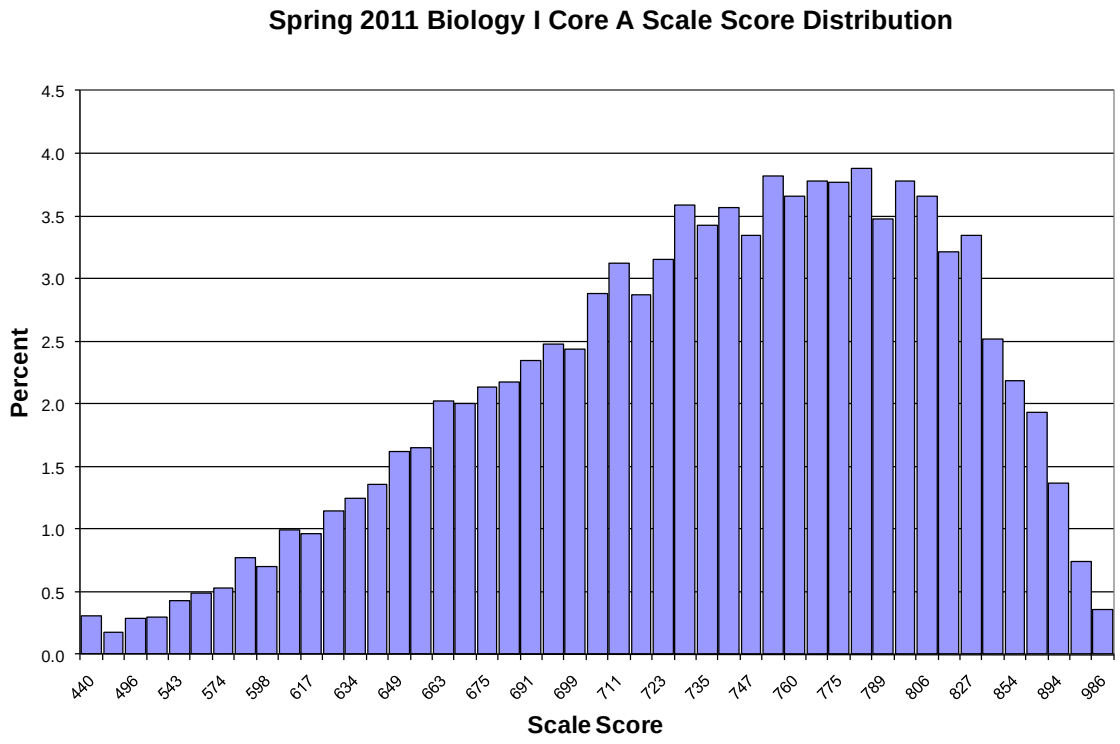


Biology I Core A Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	55	0.3	55	0.3
458	31	0.2	86	0.5
496	51	0.3	137	0.8
522	53	0.3	190	1.1
543	77	0.4	267	1.5
560	87	0.5	354	2.0
574	94	0.5	448	2.5
587	137	0.8	585	3.3
598	125	0.7	710	4.0
608	176	1.0	886	5.0
617	172	1.0	1,058	5.9
627	203	1.1	1,261	7.1
634	221	1.2	1,482	8.3
642	241	1.4	1,723	9.7
649	289	1.6	2,012	11.3
656	293	1.6	2,305	13.0
663	360	2.0	2,665	15.0
669	356	2.0	3,021	17.0
675	379	2.1	3,400	19.1
682	387	2.2	3,787	21.3
691	418	2.3	4,205	23.6
694	440	2.5	4,645	26.1
699	434	2.4	5,079	28.5
705	512	2.9	5,591	31.4
711	556	3.1	6,147	34.5
717	510	2.9	6,657	37.4
723	561	3.2	7,218	40.6
729	638	3.6	7,856	44.1
735	609	3.4	8,465	47.6
741	635	3.6	9,100	51.1
747	595	3.3	9,695	54.5
753	679	3.8	10,374	58.3
760	651	3.7	11,025	61.9
767	673	3.8	11,698	65.7
775	670	3.8	12,368	69.5
781	690	3.9	13,058	73.4
789	619	3.5	13,677	76.8
797	672	3.8	14,349	80.6
806	651	3.7	15,000	84.3
816	572	3.2	15,572	87.5
827	596	3.3	16,168	90.8
839	448	2.5	16,616	93.4

Biology I Core A Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
854	388	2.2	17,004	95.5
871	343	1.9	17,347	97.5
894	243	1.4	17,590	98.8
927	132	0.7	17,722	99.6
986	63	0.4	17,785	99.9
999	13	0.1	17,798	100.0

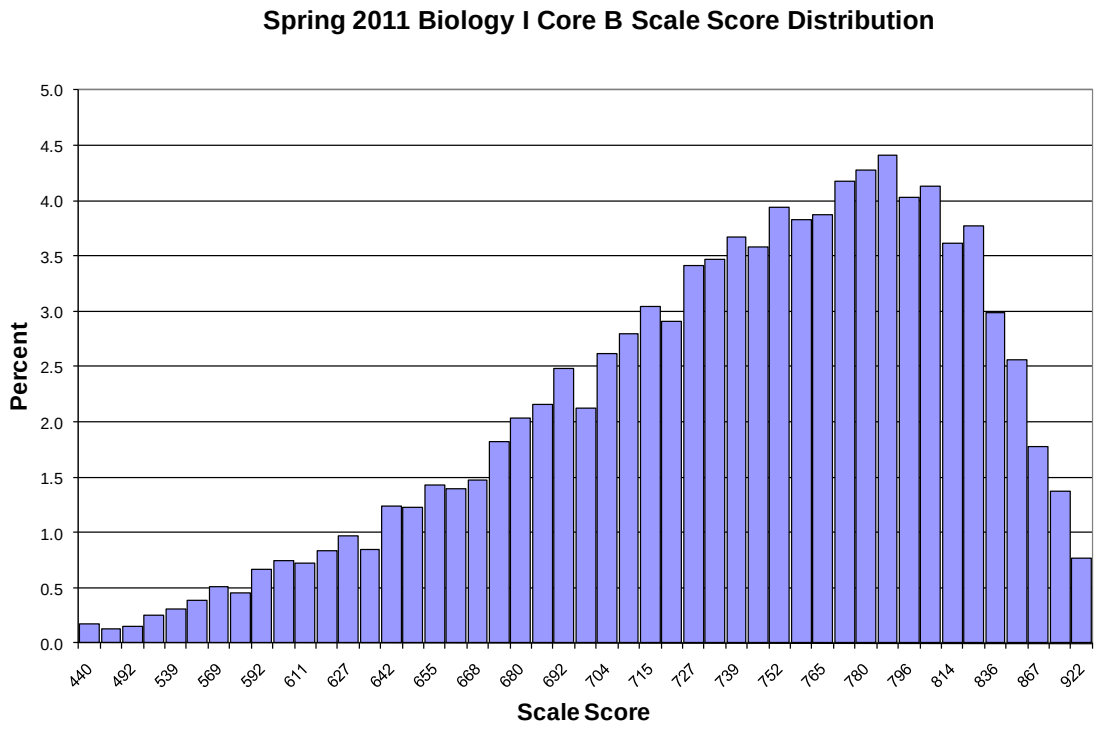


Biology I Core B Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	26	0.2	26	0.2
452	19	0.1	45	0.3
492	22	0.1	67	0.5
519	38	0.3	105	0.7
539	46	0.3	151	1.0
555	58	0.4	209	1.4
569	76	0.5	285	1.9
581	68	0.5	353	2.4
592	98	0.7	451	3.0
602	111	0.7	562	3.8
611	107	0.7	669	4.5
619	124	0.8	793	5.4
627	143	1.0	936	6.3
635	125	0.8	1,061	7.2
642	184	1.2	1,245	8.4
649	182	1.2	1,427	9.6
655	212	1.4	1,639	11.1
662	207	1.4	1,846	12.5
668	218	1.5	2,064	13.9
674	269	1.8	2,333	15.8
680	301	2.0	2,634	17.8
691	320	2.2	2,954	19.9
692	367	2.5	3,321	22.4
698	315	2.1	3,636	24.6
704	388	2.6	4,024	27.2
709	415	2.8	4,439	30.0
715	450	3.0	4,889	33.0
721	431	2.9	5,320	35.9
727	505	3.4	5,825	39.3
733	514	3.5	6,339	42.8
739	544	3.7	6,883	46.5
746	530	3.6	7,413	50.1
752	584	3.9	7,997	54.0
758	567	3.8	8,564	57.8
765	573	3.9	9,137	61.7
775	618	4.2	9,755	65.9
780	633	4.3	10,388	70.2
787	653	4.4	11,041	74.6
796	597	4.0	11,638	78.6
804	612	4.1	12,250	82.7
814	535	3.6	12,785	86.3
824	558	3.8	13,343	90.1

Biology I Core B Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
836	443	3.0	13,786	93.1
850	380	2.6	14,166	95.7
867	263	1.8	14,429	97.4
889	203	1.4	14,632	98.8
922	114	0.8	14,746	99.6
985	51	0.3	14,797	99.9
999	11	0.1	14,808	100.0

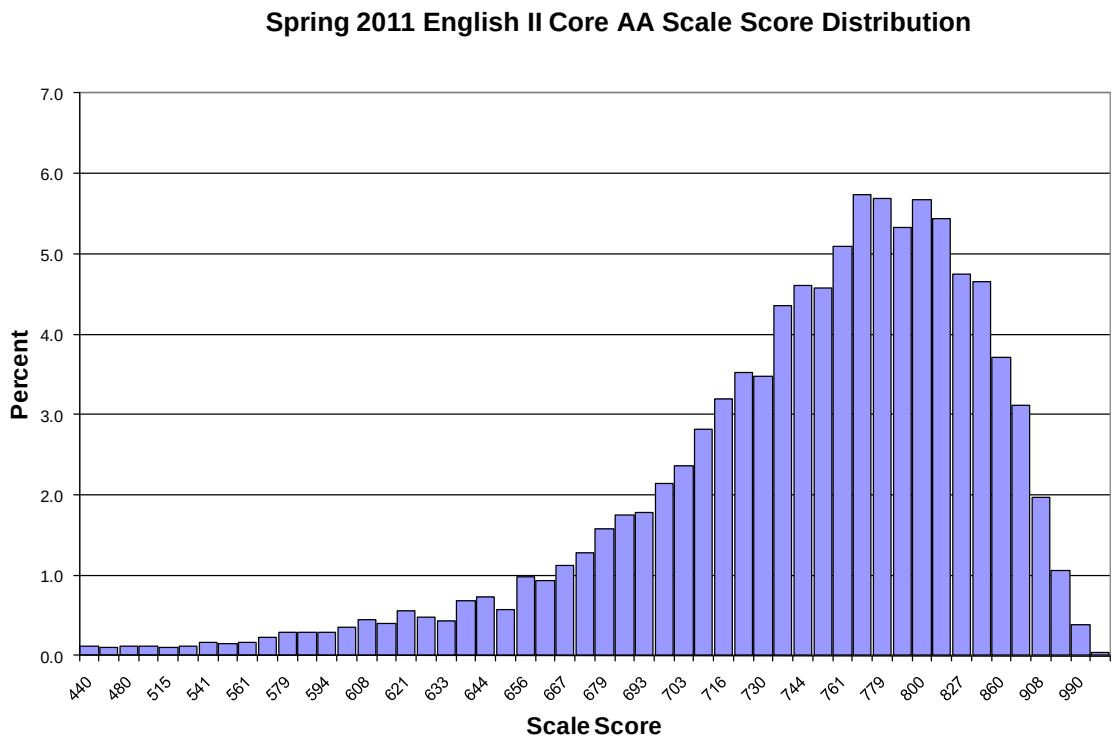


English II Core AA Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	9	0.1	9	0.1
457	8	0.1	17	0.2
480	9	0.1	26	0.3
499	10	0.1	36	0.4
515	8	0.1	44	0.5
528	10	0.1	54	0.7
541	13	0.2	67	0.8
551	12	0.1	79	1.0
561	13	0.2	92	1.1
570	19	0.2	111	1.4
579	23	0.3	134	1.6
588	24	0.3	158	1.9
594	23	0.3	181	2.2
601	29	0.4	210	2.6
608	37	0.5	247	3.0
614	32	0.4	279	3.4
621	45	0.6	324	4.0
627	39	0.5	363	4.5
633	35	0.4	398	4.9
639	55	0.7	453	5.6
644	60	0.7	513	6.3
650	47	0.6	560	6.9
656	80	1.0	640	7.9
662	76	0.9	716	8.8
667	91	1.1	807	9.9
673	104	1.3	911	11.2
679	128	1.6	1,039	12.8
685	142	1.7	1,181	14.5
693	145	1.8	1,326	16.3
697	174	2.1	1,500	18.4
703	192	2.4	1,692	20.8
709	229	2.8	1,921	23.6
716	260	3.2	2,181	26.8
723	287	3.5	2,468	30.3
730	283	3.5	2,751	33.8
737	354	4.4	3,105	38.2
744	375	4.6	3,480	42.8
752	372	4.6	3,852	47.4
761	414	5.1	4,266	52.4
769	466	5.7	4,732	58.2
779	463	5.7	5,195	63.9
797	434	5.3	5,629	69.2

English II Core AA Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
800	462	5.7	6,091	74.9
813	442	5.4	6,533	80.3
827	386	4.7	6,919	85.1
842	379	4.7	7,298	89.7
860	302	3.7	7,600	93.4
882	254	3.1	7,854	96.6
908	160	2.0	8,014	98.5
942	86	1.1	8,100	99.6
990	31	0.4	8,131	100.0
999	3	0.0	8,134	100.0

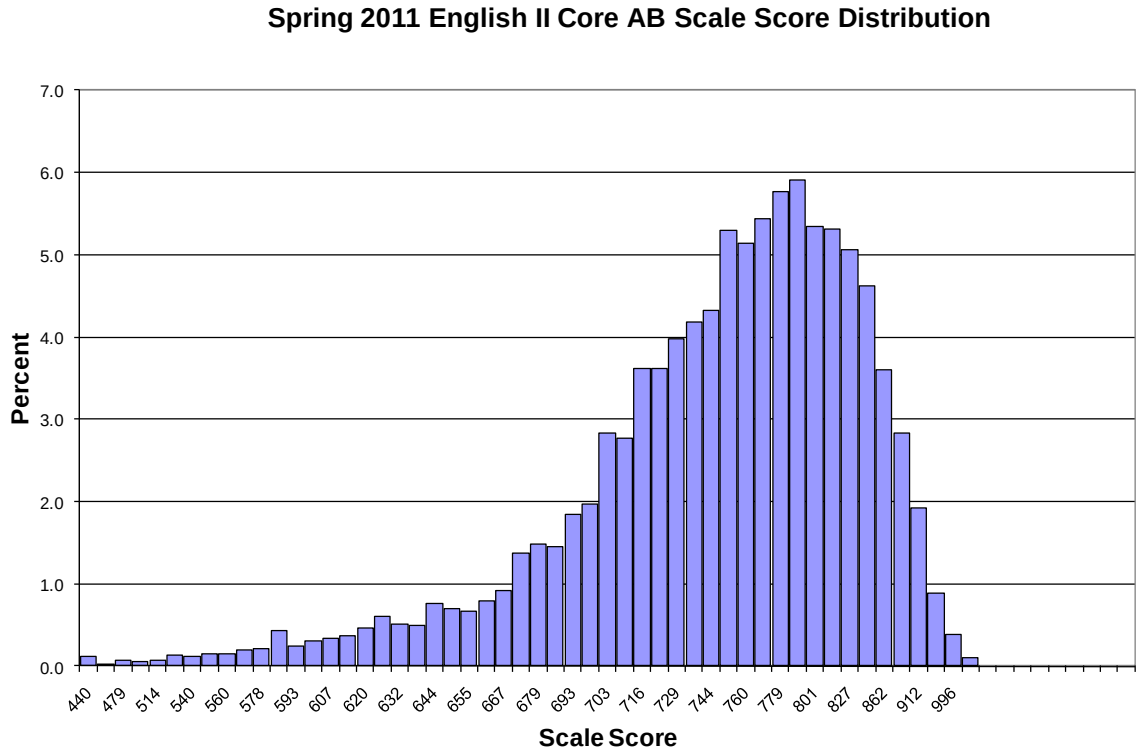


English II Core AB Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	9	0.1	9	0.1
456	2	0.0	11	0.1
479	6	0.1	17	0.2
498	5	0.1	22	0.3
514	6	0.1	28	0.4
527	11	0.1	39	0.5
540	10	0.1	49	0.6
550	12	0.2	61	0.8
560	12	0.2	73	0.9
569	16	0.2	89	1.1
578	17	0.2	106	1.3
588	34	0.4	140	1.8
593	20	0.3	160	2.0
600	24	0.3	184	2.3
607	27	0.3	211	2.6
613	30	0.4	241	3.0
620	37	0.5	278	3.5
626	48	0.6	326	4.1
632	41	0.5	367	4.6
638	40	0.5	407	5.1
644	61	0.8	468	5.9
650	56	0.7	524	6.6
655	53	0.7	577	7.2
661	63	0.8	640	8.0
667	74	0.9	714	8.9
673	110	1.4	824	10.3
679	118	1.5	942	11.8
684	116	1.5	1,058	13.2
693	147	1.8	1,205	15.1
696	158	2.0	1,363	17.0
703	227	2.8	1,590	19.9
709	222	2.8	1,812	22.7
716	289	3.6	2,101	26.3
722	289	3.6	2,390	29.9
729	318	4.0	2,708	33.9
737	334	4.2	3,042	38.0
744	346	4.3	3,388	42.4
752	423	5.3	3,811	47.7
760	411	5.1	4,222	52.8
769	435	5.4	4,657	58.2
779	461	5.8	5,118	64.0
797	472	5.9	5,590	69.9

English II Core AB Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
801	427	5.3	6,017	75.3
813	425	5.3	6,442	80.6
827	405	5.1	6,847	85.6
844	370	4.6	7,217	90.3
862	288	3.6	7,505	93.9
885	226	2.8	7,731	96.7
912	154	1.9	7,885	98.6
947	71	0.9	7,956	99.5
996	31	0.4	7,987	99.9
999	8	0.1	7,995	100.0

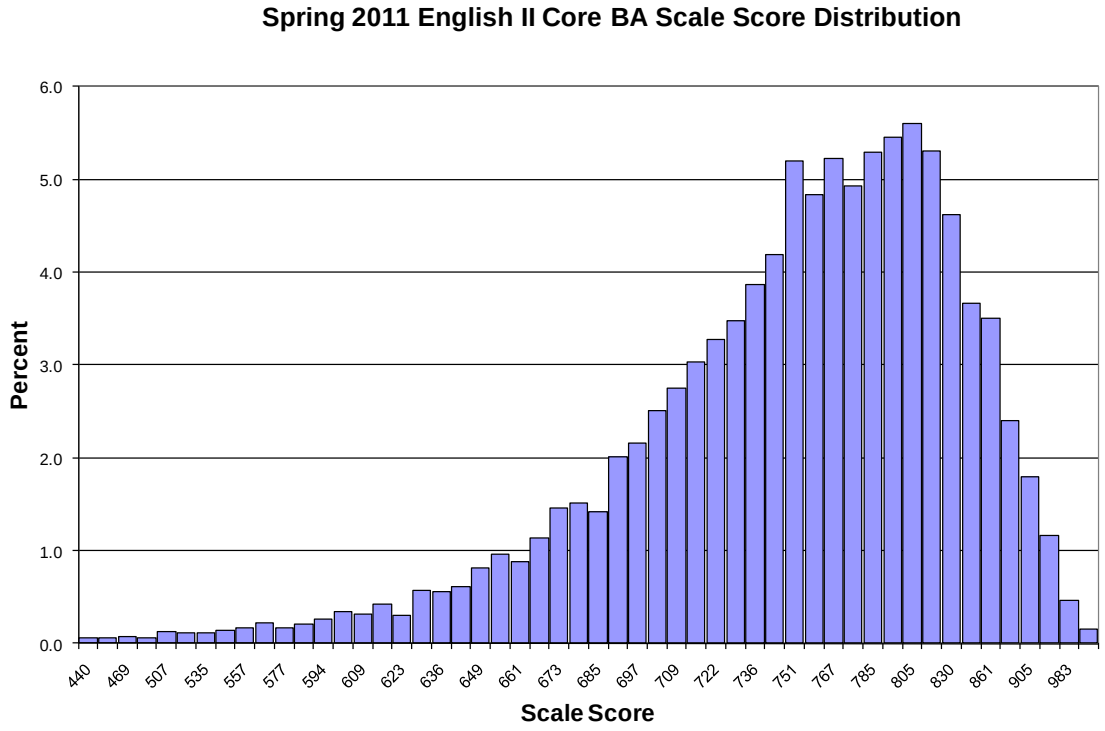


English II Core BA Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	5	0.1	5	0.1
445	5	0.1	10	0.1
469	6	0.1	16	0.2
489	5	0.1	21	0.3
507	10	0.1	31	0.4
521	9	0.1	40	0.5
535	9	0.1	49	0.6
547	11	0.1	60	0.8
557	13	0.2	73	0.9
567	17	0.2	90	1.2
577	13	0.2	103	1.3
588	16	0.2	119	1.5
594	20	0.3	139	1.8
601	27	0.3	166	2.1
609	25	0.3	191	2.5
616	33	0.4	224	2.9
623	23	0.3	247	3.2
630	44	0.6	291	3.7
636	43	0.6	334	4.3
642	48	0.6	382	4.9
649	63	0.8	445	5.7
655	75	1.0	520	6.7
661	68	0.9	588	7.6
667	88	1.1	676	8.7
673	113	1.5	789	10.2
679	118	1.5	907	11.7
685	110	1.4	1,017	13.1
693	156	2.0	1,173	15.1
697	168	2.2	1,341	17.3
703	195	2.5	1,536	19.8
709	214	2.8	1,750	22.5
716	236	3.0	1,986	25.6
722	255	3.3	2,241	28.8
729	270	3.5	2,511	32.3
736	301	3.9	2,812	36.2
743	326	4.2	3,138	40.4
751	404	5.2	3,542	45.6
758	376	4.8	3,918	50.4
767	406	5.2	4,324	55.6
775	383	4.9	4,707	60.6
785	411	5.3	5,118	65.9
797	424	5.5	5,542	71.3

English II Core BA Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
805	435	5.6	5,977	76.9
817	412	5.3	6,389	82.2
830	359	4.6	6,748	86.8
844	285	3.7	7,033	90.5
861	272	3.5	7,305	94.0
881	187	2.4	7,492	96.4
905	140	1.8	7,632	98.2
937	90	1.2	7,722	99.4
983	36	0.5	7,758	99.8
999	12	0.2	7,770	100.0

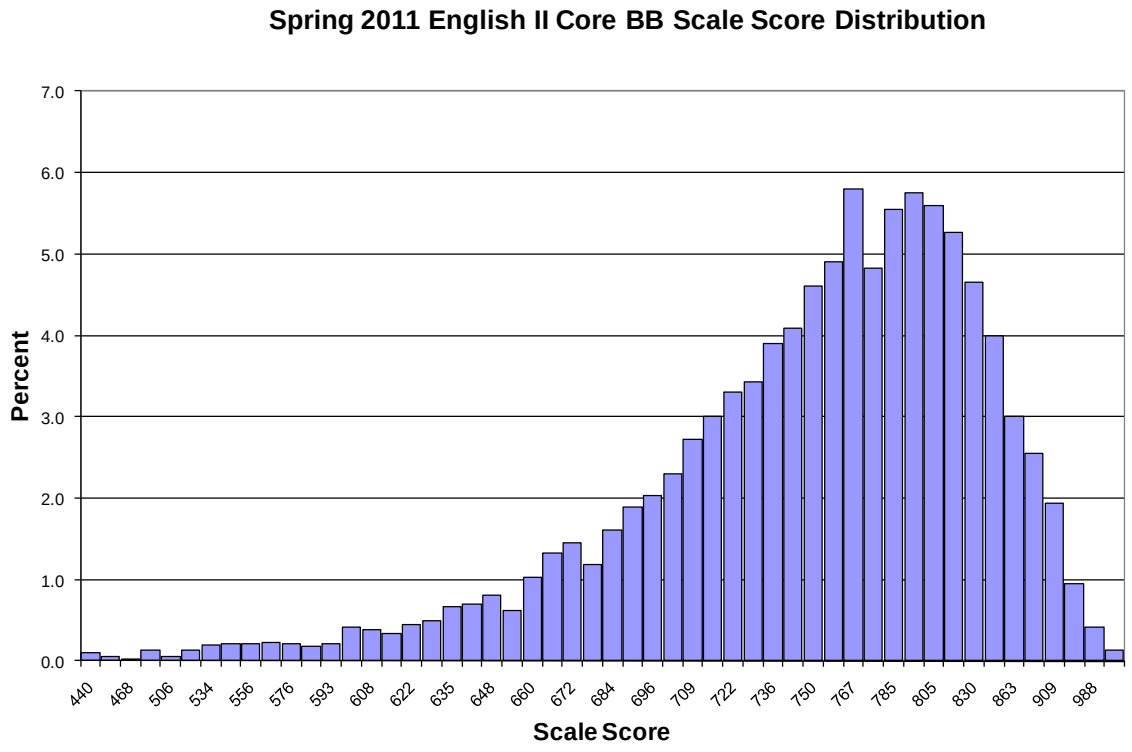


English II Core BB Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	8	0.1	8	0.1
444	4	0.1	12	0.2
468	1	0.0	13	0.2
488	10	0.1	23	0.3
506	4	0.1	27	0.3
520	10	0.1	37	0.5
534	15	0.2	52	0.7
546	16	0.2	68	0.9
556	16	0.2	84	1.1
566	18	0.2	102	1.3
576	17	0.2	119	1.5
588	14	0.2	133	1.7
593	16	0.2	149	1.9
600	32	0.4	181	2.3
608	30	0.4	211	2.7
615	26	0.3	237	3.0
622	35	0.4	272	3.5
629	39	0.5	311	4.0
635	52	0.7	363	4.7
642	55	0.7	418	5.4
648	63	0.8	481	6.2
654	48	0.6	529	6.8
660	80	1.0	609	7.8
666	103	1.3	712	9.1
672	113	1.4	825	10.6
678	93	1.2	918	11.8
684	126	1.6	1,044	13.4
693	148	1.9	1,192	15.3
696	159	2.0	1,351	17.3
703	180	2.3	1,531	19.6
709	213	2.7	1,744	22.4
715	234	3.0	1,978	25.3
722	258	3.3	2,236	28.7
729	268	3.4	2,504	32.1
736	304	3.9	2,808	36.0
743	319	4.1	3,127	40.1
750	359	4.6	3,486	44.7
758	383	4.9	3,869	49.6
767	452	5.8	4,321	55.4
775	376	4.8	4,697	60.2
785	433	5.5	5,130	65.7
797	449	5.8	5,579	71.5

English II Core BB Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
805	437	5.6	6,016	77.1
817	411	5.3	6,427	82.4
830	363	4.7	6,790	87.0
845	312	4.0	7,102	91.0
863	234	3.0	7,336	94.0
883	199	2.6	7,535	96.6
909	151	1.9	7,686	98.5
942	74	0.9	7,760	99.4
988	32	0.4	7,792	99.9
999	11	0.1	7,803	100.0

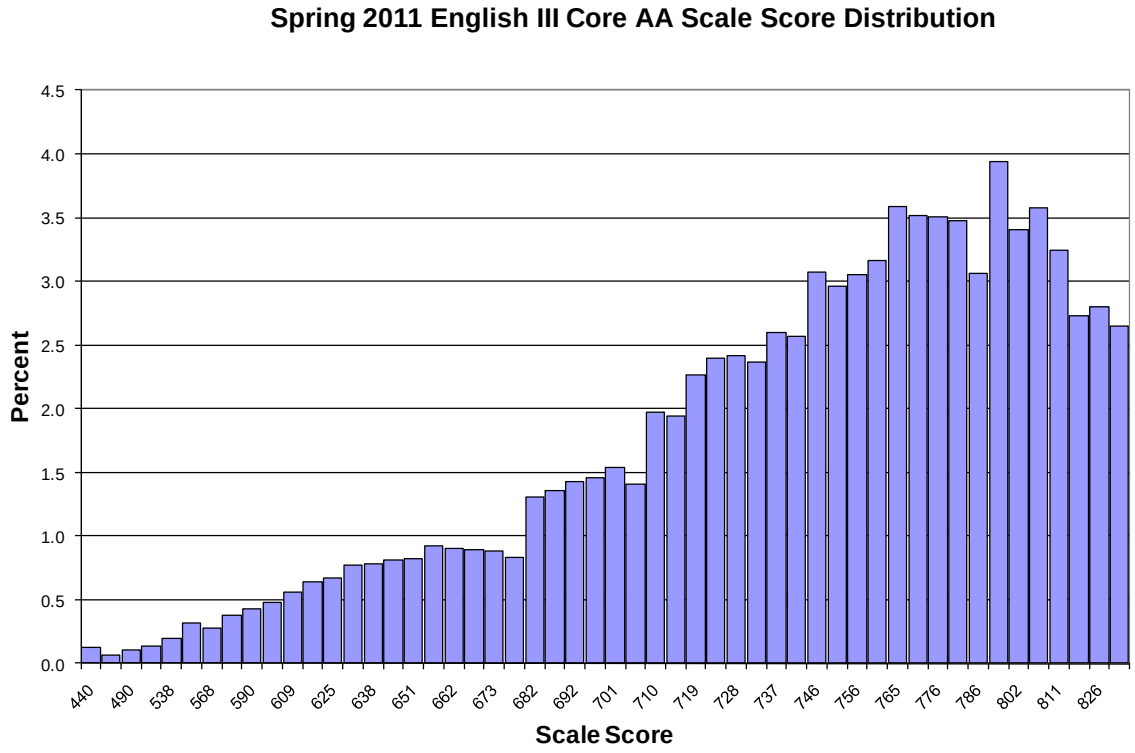


English III Core AA Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	10	0.1	10	0.1
447	5	0.1	15	0.2
490	9	0.1	24	0.3
518	11	0.1	35	0.4
538	16	0.2	51	0.6
554	26	0.3	77	0.9
568	23	0.3	100	1.2
580	31	0.4	131	1.6
590	35	0.4	166	2.0
600	39	0.5	205	2.5
609	46	0.6	251	3.1
617	52	0.6	303	3.7
625	55	0.7	358	4.4
632	63	0.8	421	5.2
638	64	0.8	485	5.9
645	66	0.8	551	6.7
651	67	0.8	618	7.6
656	75	0.9	693	8.5
662	74	0.9	767	9.4
670	73	0.9	840	10.3
673	72	0.9	912	11.2
678	68	0.8	980	12.0
682	107	1.3	1,087	13.3
687	111	1.4	1,198	14.7
692	117	1.4	1,315	16.1
700	119	1.5	1,434	17.6
701	126	1.5	1,560	19.1
706	115	1.4	1,675	20.5
710	161	2.0	1,836	22.5
715	159	1.9	1,995	24.4
719	185	2.3	2,180	26.7
724	196	2.4	2,376	29.1
728	197	2.4	2,573	31.5
733	193	2.4	2,766	33.9
737	212	2.6	2,978	36.5
742	210	2.6	3,188	39.0
746	251	3.1	3,439	42.1
751	242	3.0	3,681	45.1
756	249	3.0	3,930	48.1
760	258	3.2	4,188	51.3
765	293	3.6	4,481	54.9

English III Core AA Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
770	287	3.5	4,768	58.4
776	286	3.5	5,054	61.9
781	284	3.5	5,338	65.4
786	250	3.1	5,588	68.4
792	322	3.9	5,910	72.4
802	278	3.4	6,188	75.8
805	292	3.6	6,480	79.3
811	265	3.2	6,745	82.6
819	223	2.7	6,968	85.3
826	229	2.8	7,197	88.1
835	216	2.6	7,413	90.8
844	181	2.2	7,594	93.0
855	173	2.1	7,767	95.1
866	135	1.7	7,902	96.8
880	102	1.2	8,004	98.0
895	72	0.9	8,076	98.9
914	51	0.6	8,127	99.5
937	18	0.2	8,145	99.7
968	16	0.2	8,161	99.9
999	6	0.1	8,167	100.0

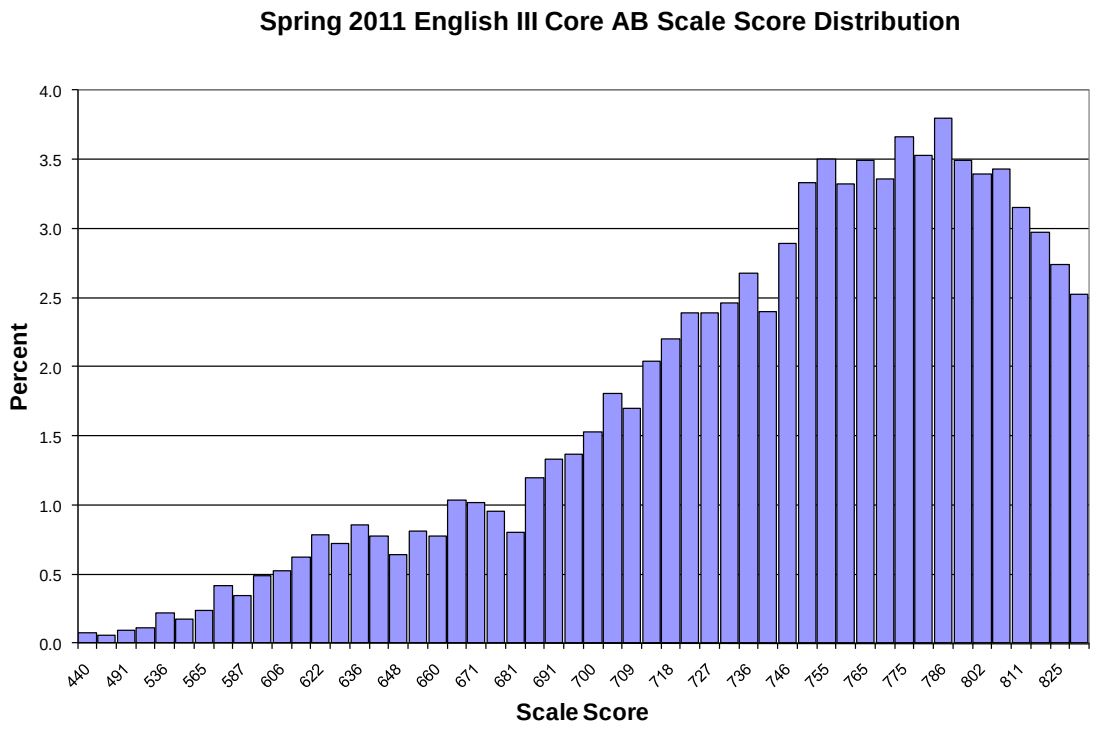

Spring 2011 English III Core AA Scale Score Distribution
Percent
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
440
490
538
568
590
609
625
638
651
662
673
682
692
701
710
719
728
737
746
756
765
776
786
802
811
826
Scale Score

English III Core AB Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	6	0.1	6	0.1
449	5	0.1	11	0.1
491	8	0.1	19	0.2
517	9	0.1	28	0.3
536	18	0.2	46	0.6
552	14	0.2	60	0.7
565	19	0.2	79	1.0
577	34	0.4	113	1.4
587	28	0.3	141	1.7
597	40	0.5	181	2.2
606	43	0.5	224	2.8
614	51	0.6	275	3.4
622	64	0.8	339	4.2
629	59	0.7	398	4.9
636	70	0.9	468	5.8
642	63	0.8	531	6.5
648	52	0.6	583	7.2
654	66	0.8	649	8.0
660	63	0.8	712	8.8
670	84	1.0	796	9.8
671	83	1.0	879	10.8
676	78	1.0	957	11.8
681	65	0.8	1,022	12.6
686	97	1.2	1,119	13.8
691	108	1.3	1,227	15.1
695	111	1.4	1,338	16.5
700	124	1.5	1,462	18.0
705	147	1.8	1,609	19.8
709	138	1.7	1,747	21.5
714	166	2.0	1,913	23.5
718	179	2.2	2,092	25.7
723	194	2.4	2,286	28.1
727	194	2.4	2,480	30.5
732	200	2.5	2,680	33.0
736	218	2.7	2,898	35.6
741	195	2.4	3,093	38.0
746	235	2.9	3,328	40.9
750	271	3.3	3,599	44.3
755	285	3.5	3,884	47.8
760	270	3.3	4,154	51.1
765	284	3.5	4,438	54.6

English III Core AB Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
770	273	3.4	4,711	57.9
775	298	3.7	5,009	61.6
780	287	3.5	5,296	65.1
786	309	3.8	5,605	68.9
792	284	3.5	5,889	72.4
802	276	3.4	6,165	75.8
804	279	3.4	6,444	79.3
811	256	3.1	6,700	82.4
818	242	3.0	6,942	85.4
825	223	2.7	7,165	88.1
834	205	2.5	7,370	90.6
843	199	2.4	7,569	93.1
853	146	1.8	7,715	94.9
865	124	1.5	7,839	96.4
877	118	1.5	7,957	97.9
892	87	1.1	8,044	98.9
910	43	0.5	8,087	99.5
932	24	0.3	8,111	99.8
962	13	0.2	8,124	99.9
999	7	0.1	8,131	100.0

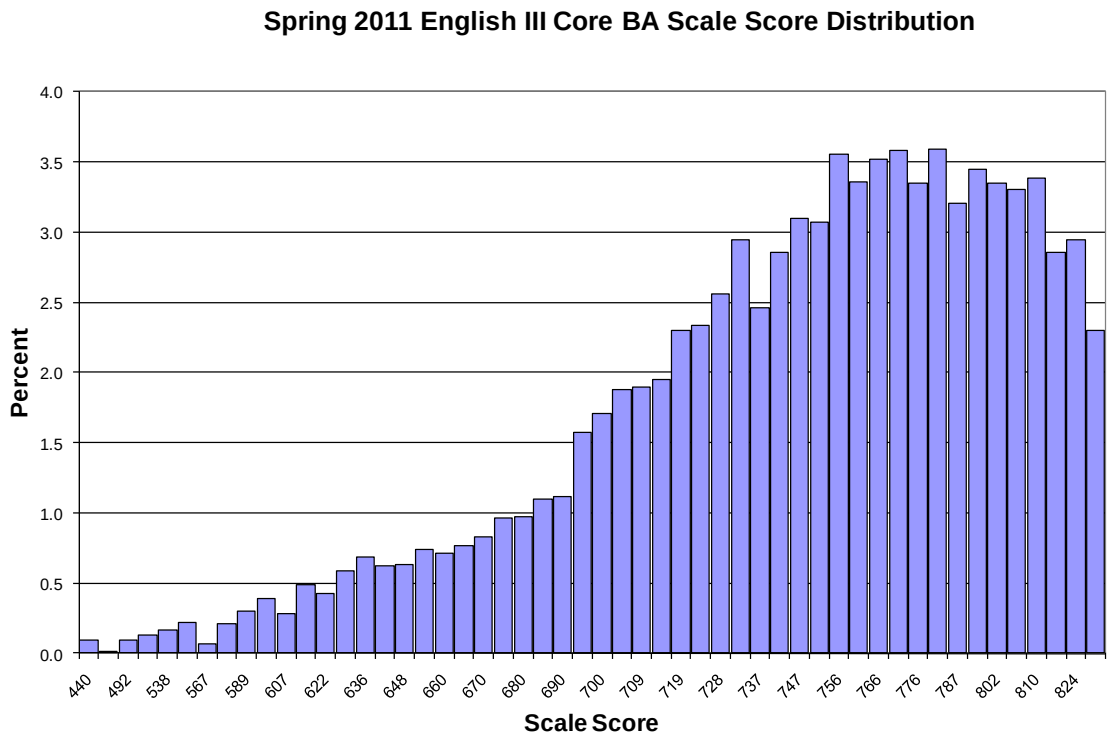

Spring 2011 English III Core AB Scale Score Distribution
Percent
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
440
491
536
565
587
606
622
636
648
660
671
681
691
700
709
718
727
736
746
755
765
775
786
802
811
825
Scale Score

English III Core BA Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	7	0.1	7	0.1
451	1	0.0	8	0.1
492	7	0.1	15	0.2
518	10	0.1	25	0.3
538	13	0.2	38	0.5
554	17	0.2	55	0.7
567	5	0.1	60	0.8
579	16	0.2	76	1.0
589	23	0.3	99	1.3
598	30	0.4	129	1.7
607	22	0.3	151	2.0
615	38	0.5	189	2.5
622	33	0.4	222	2.9
629	45	0.6	267	3.5
636	53	0.7	320	4.2
642	48	0.6	368	4.8
648	49	0.6	417	5.4
654	57	0.7	474	6.2
660	55	0.7	529	6.9
665	59	0.8	588	7.6
670	64	0.8	652	8.5
675	74	1.0	726	9.4
680	75	1.0	801	10.4
685	85	1.1	886	11.5
690	86	1.1	972	12.6
695	121	1.6	1,093	14.2
700	132	1.7	1,225	15.9
705	145	1.9	1,370	17.8
709	146	1.9	1,516	19.7
714	150	1.9	1,666	21.6
719	177	2.3	1,843	23.9
723	180	2.3	2,023	26.2
728	197	2.6	2,220	28.8
733	227	2.9	2,447	31.8
737	190	2.5	2,637	34.2
742	220	2.9	2,857	37.1
747	239	3.1	3,096	40.2
751	237	3.1	3,333	43.2
756	274	3.6	3,607	46.8
761	259	3.4	3,866	50.2
766	271	3.5	4,137	53.7

English III Core BA Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
771	276	3.6	4,413	57.3
776	258	3.3	4,671	60.6
781	277	3.6	4,948	64.2
787	247	3.2	5,195	67.4
792	266	3.5	5,461	70.9
802	258	3.3	5,719	74.2
804	255	3.3	5,974	77.5
810	261	3.4	6,235	80.9
817	220	2.9	6,455	83.8
824	227	2.9	6,682	86.7
832	177	2.3	6,859	89.0
840	187	2.4	7,046	91.4
849	176	2.3	7,222	93.7
860	147	1.9	7,369	95.6
872	123	1.6	7,492	97.2
886	95	1.2	7,587	98.4
904	50	0.6	7,637	99.1
926	36	0.5	7,673	99.6
957	27	0.4	7,700	99.9
999	7	0.1	7,707	100.0

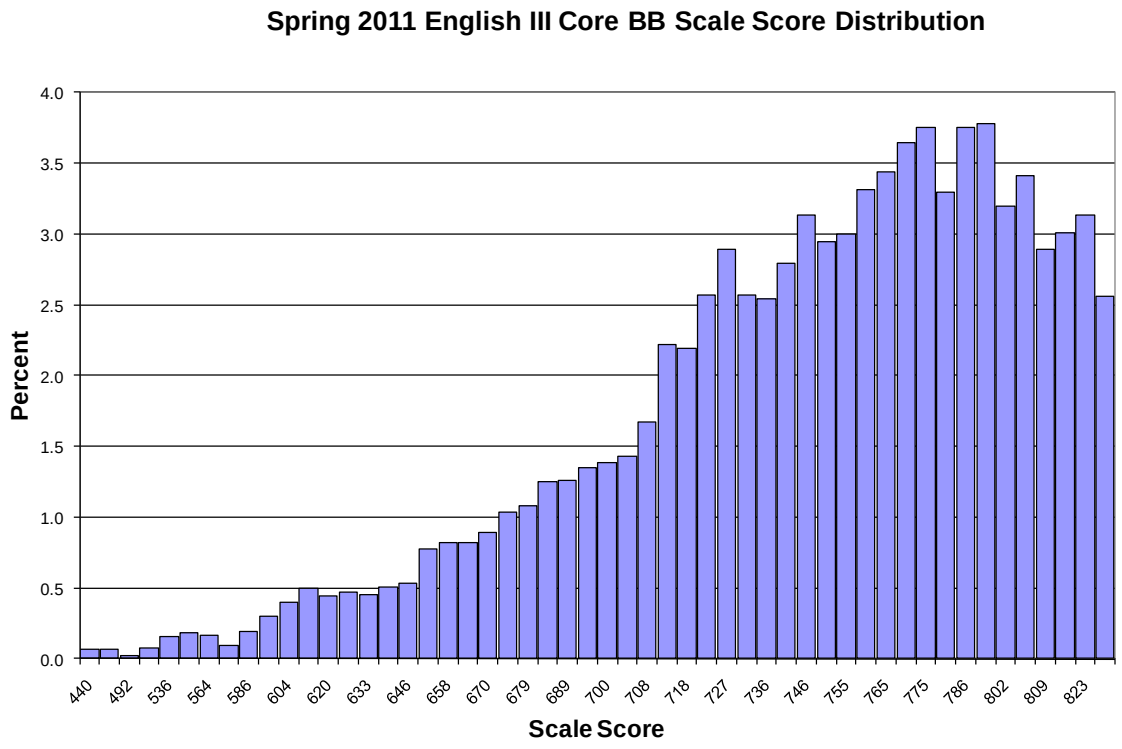

Spring 2011 English III Core BA Scale Score Distribution
Percent
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
440
492
538
567
589
607
622
636
648
660
670
680
690
700
709
719
728
737
747
756
766
776
787
802
810
824
Scale Score

English III Core BB Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	5	0.1	5	0.1
452	5	0.1	10	0.1
492	2	0.0	12	0.2
518	6	0.1	18	0.2
536	12	0.2	30	0.4
551	14	0.2	44	0.6
564	13	0.2	57	0.7
576	7	0.1	64	0.8
586	15	0.2	79	1.0
595	23	0.3	102	1.3
604	31	0.4	133	1.7
612	38	0.5	171	2.2
620	34	0.4	205	2.7
627	36	0.5	241	3.1
633	35	0.5	276	3.6
640	39	0.5	315	4.1
646	41	0.5	356	4.6
652	60	0.8	416	5.4
658	63	0.8	479	6.2
663	63	0.8	542	7.0
670	69	0.9	611	7.9
674	80	1.0	691	9.0
679	83	1.1	774	10.1
684	96	1.2	870	11.3
689	97	1.3	967	12.6
694	104	1.4	1,071	13.9
700	107	1.4	1,178	15.3
703	110	1.4	1,288	16.7
708	129	1.7	1,417	18.4
713	171	2.2	1,588	20.6
718	169	2.2	1,757	22.8
722	198	2.6	1,955	25.4
727	223	2.9	2,178	28.3
732	198	2.6	2,376	30.9
736	196	2.5	2,572	33.4
741	215	2.8	2,787	36.2
746	241	3.1	3,028	39.3
751	227	2.9	3,255	42.3
755	231	3.0	3,486	45.3
760	255	3.3	3,741	48.6
765	265	3.4	4,006	52.0

English III Core BB Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
770	281	3.6	4,287	55.7
775	289	3.8	4,576	59.4
781	254	3.3	4,830	62.7
786	289	3.8	5,119	66.5
792	291	3.8	5,410	70.3
802	246	3.2	5,656	73.5
803	263	3.4	5,919	76.9
809	223	2.9	6,142	79.8
816	232	3.0	6,374	82.8
823	241	3.1	6,615	85.9
831	197	2.6	6,812	88.5
839	189	2.5	7,001	90.9
848	191	2.5	7,192	93.4
858	150	1.9	7,342	95.4
870	111	1.4	7,453	96.8
884	88	1.1	7,541	97.9
900	85	1.1	7,626	99.1
921	41	0.5	7,667	99.6
950	16	0.2	7,683	99.8
999	16	0.2	7,699	100.0

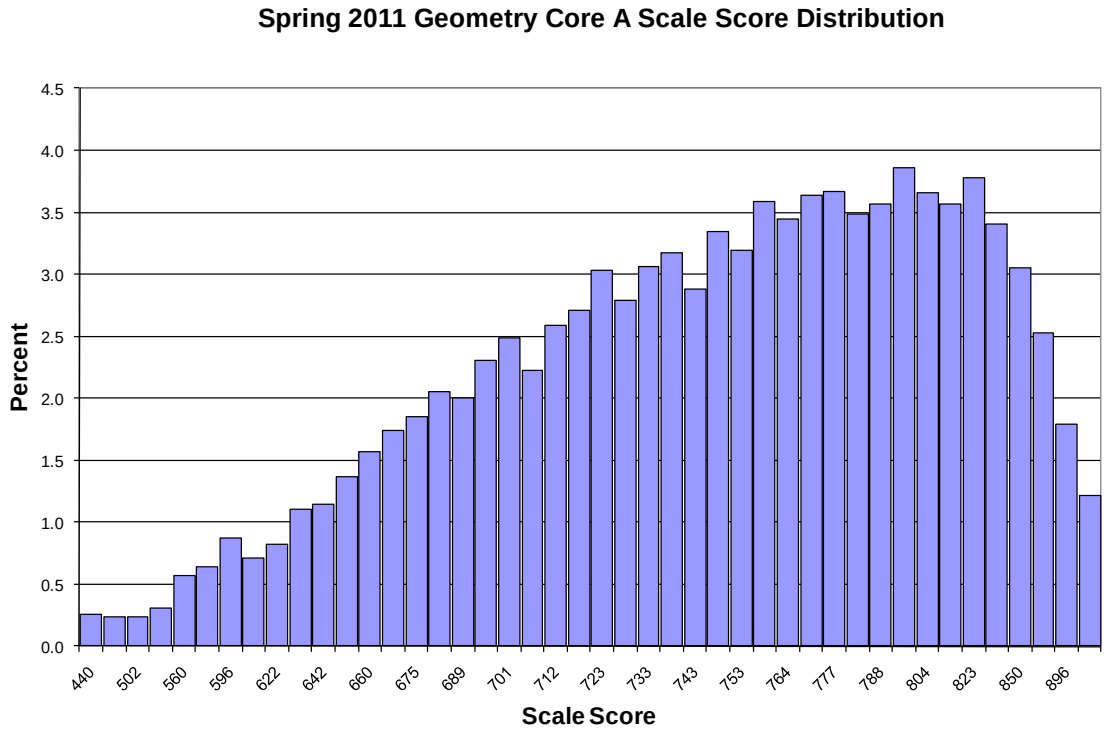

Spring 2011 English III Core BB Scale Score Distribution
Percent
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
440
492
536
564
586
604
620
633
646
658
670
679
689
700
708
718
727
736
746
755
765
775
786
802
809
823
Scale Score

Geometry Core A Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	46	0.3	46	0.3
446	42	0.2	88	0.5
502	43	0.2	131	0.7
536	56	0.3	187	1.0
560	102	0.6	289	1.6
580	115	0.6	404	2.3
596	156	0.9	560	3.1
610	127	0.7	687	3.8
622	148	0.8	835	4.7
635	199	1.1	1,034	5.8
642	205	1.1	1,239	6.9
651	245	1.4	1,484	8.3
660	281	1.6	1,765	9.8
668	312	1.7	2,077	11.6
675	333	1.9	2,410	13.4
682	368	2.1	2,778	15.5
689	360	2.0	3,138	17.5
700	413	2.3	3,551	19.8
701	446	2.5	3,997	22.3
707	399	2.2	4,396	24.5
712	465	2.6	4,861	27.1
718	487	2.7	5,348	29.8
723	544	3.0	5,892	32.8
728	501	2.8	6,393	35.6
733	550	3.1	6,943	38.7
738	570	3.2	7,513	41.9
743	517	2.9	8,030	44.7
748	601	3.3	8,631	48.1
753	574	3.2	9,205	51.3
758	644	3.6	9,849	54.9
764	618	3.4	10,467	58.3
769	653	3.6	11,120	62.0
777	659	3.7	11,779	65.6
782	625	3.5	12,404	69.1
788	640	3.6	13,044	72.7
796	693	3.9	13,737	76.5
804	656	3.7	14,393	80.2
813	641	3.6	15,034	83.8
823	679	3.8	15,713	87.6
835	612	3.4	16,325	91.0
850	547	3.0	16,872	94.0
869	453	2.5	17,325	96.5

Geometry Core A Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
896	321	1.8	17,646	98.3
945	218	1.2	17,864	99.5
999	83	0.5	17,947	100.0

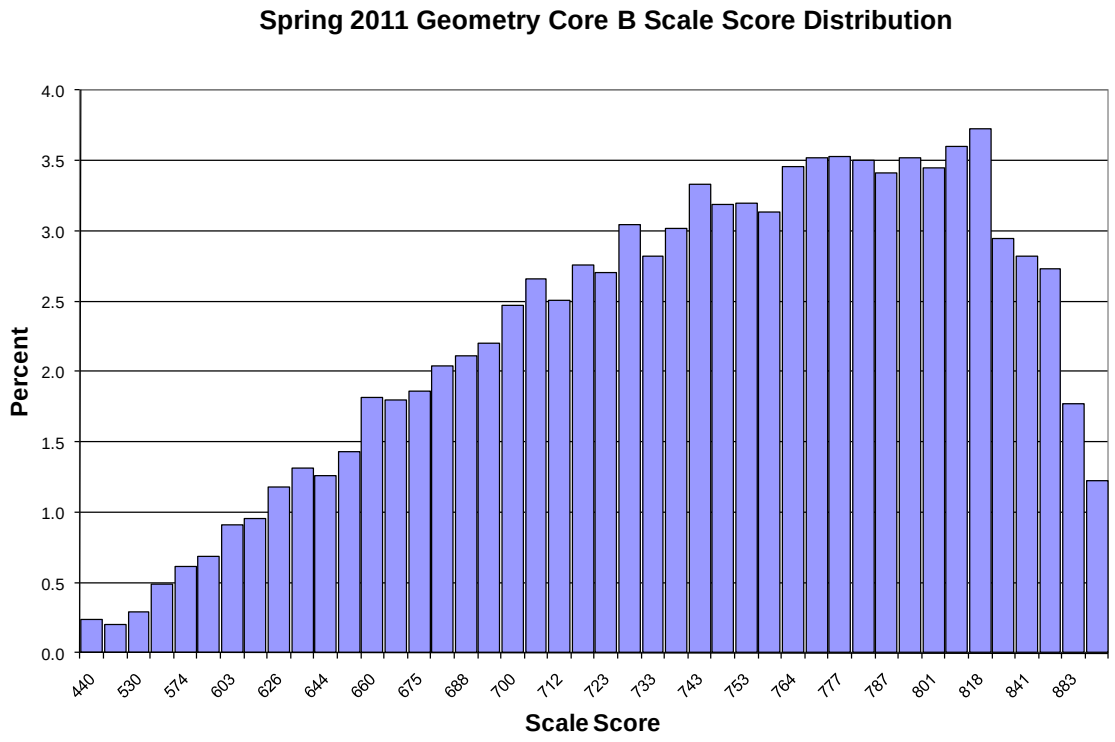


Geometry Core B Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	36	0.2	36	0.2
494	31	0.2	67	0.4
530	44	0.3	111	0.7
555	74	0.5	185	1.2
574	93	0.6	278	1.8
590	104	0.7	382	2.5
603	138	0.9	520	3.4
615	144	1.0	664	4.4
626	179	1.2	843	5.6
635	199	1.3	1,042	6.9
644	190	1.3	1,232	8.1
652	217	1.4	1,449	9.6
660	275	1.8	1,724	11.4
668	272	1.8	1,996	13.2
675	281	1.9	2,277	15.1
682	309	2.0	2,586	17.1
688	320	2.1	2,906	19.2
694	333	2.2	3,239	21.4
700	374	2.5	3,613	23.9
706	402	2.7	4,015	26.5
712	379	2.5	4,394	29.1
717	417	2.8	4,811	31.8
723	409	2.7	5,220	34.5
728	460	3.0	5,680	37.6
733	427	2.8	6,107	40.4
738	456	3.0	6,563	43.4
743	504	3.3	7,067	46.7
748	482	3.2	7,549	49.9
753	483	3.2	8,032	53.1
758	474	3.1	8,506	56.2
764	523	3.5	9,029	59.7
769	533	3.5	9,562	63.2
777	534	3.5	10,096	66.8
781	530	3.5	10,626	70.3
787	516	3.4	11,142	73.7
794	532	3.5	11,674	77.2
801	522	3.5	12,196	80.6
809	544	3.6	12,740	84.2
818	563	3.7	13,303	88.0
828	445	2.9	13,748	90.9
841	427	2.8	14,175	93.7
858	413	2.7	14,588	96.5

Geometry Core B Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
883	268	1.8	14,856	98.2
933	185	1.2	15,041	99.5
999	82	0.5	15,123	100.0

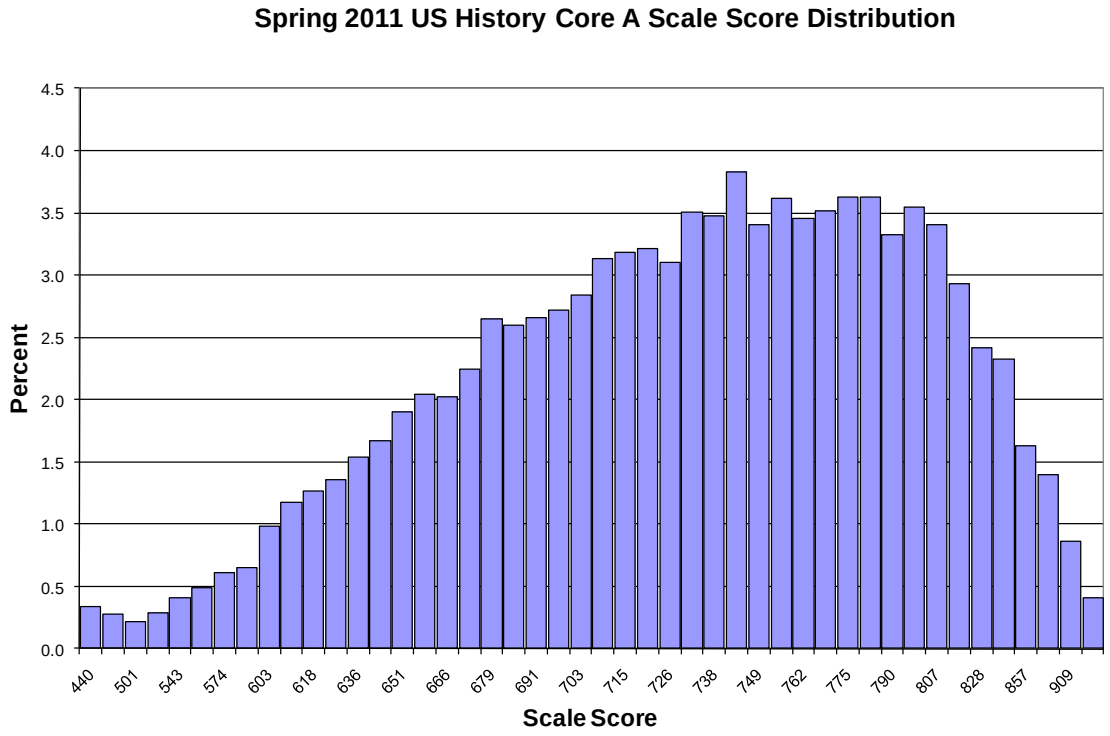


U.S. History Core A Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	55	0.3	55	0.3
472	45	0.3	100	0.6
501	36	0.2	136	0.8
524	47	0.3	183	1.1
543	66	0.4	249	1.5
559	79	0.5	328	2.0
574	99	0.6	427	2.6
586	106	0.7	533	3.3
603	159	1.0	692	4.3
608	191	1.2	883	5.4
618	206	1.3	1,089	6.7
627	221	1.4	1,310	8.1
636	249	1.5	1,559	9.6
644	271	1.7	1,830	11.3
651	309	1.9	2,139	13.2
659	332	2.0	2,471	15.2
666	329	2.0	2,800	17.3
672	365	2.2	3,165	19.5
679	430	2.6	3,595	22.2
689	421	2.6	4,016	24.7
691	431	2.7	4,447	27.4
697	441	2.7	4,888	30.1
703	461	2.8	5,349	33.0
709	509	3.1	5,858	36.1
715	516	3.2	6,374	39.3
721	522	3.2	6,896	42.5
726	503	3.1	7,399	45.6
732	569	3.5	7,968	49.1
738	564	3.5	8,532	52.6
747	622	3.8	9,154	56.4
749	553	3.4	9,707	59.8
756	587	3.6	10,294	63.4
762	561	3.5	10,855	66.9
768	571	3.5	11,426	70.4
775	588	3.6	12,014	74.0
782	589	3.6	12,603	77.7
790	539	3.3	13,142	81.0
798	576	3.5	13,718	84.5
807	553	3.4	14,271	87.9
817	475	2.9	14,746	90.9
828	392	2.4	15,138	93.3
841	377	2.3	15,515	95.6

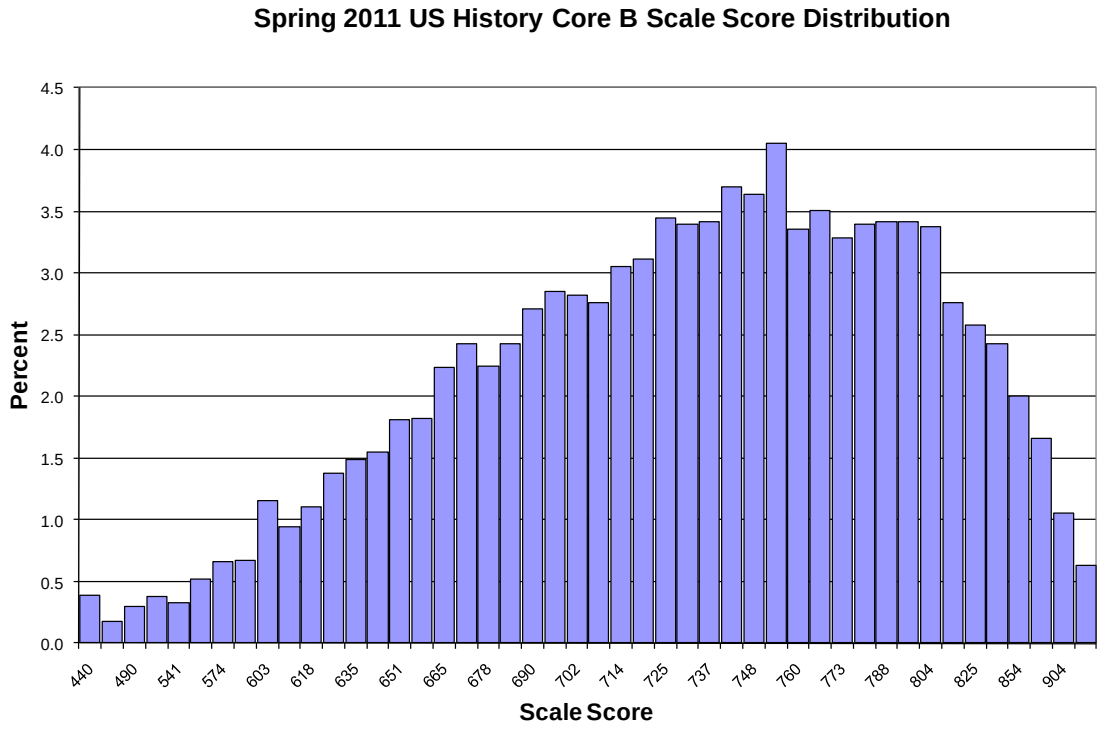
U.S. History Core A Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
857	265	1.6	15,780	97.2
879	227	1.4	16,007	98.6
909	140	0.9	16,147	99.5
966	67	0.4	16,214	99.9
999	14	0.1	16,228	100.0



U.S. History Core B Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
440	52	0.4	52	0.4
449	24	0.2	76	0.6
490	40	0.3	116	0.9
519	51	0.4	167	1.2
541	45	0.3	212	1.6
559	71	0.5	283	2.1
574	89	0.7	372	2.7
587	91	0.7	463	3.4
603	157	1.2	620	4.6
609	128	0.9	748	5.5
618	150	1.1	898	6.6
627	186	1.4	1,084	8.0
635	202	1.5	1,286	9.5
643	209	1.5	1,495	11.0
651	245	1.8	1,740	12.9
658	246	1.8	1,986	14.7
665	302	2.2	2,288	16.9
671	329	2.4	2,617	19.3
678	304	2.2	2,921	21.6
689	329	2.4	3,250	24.0
690	367	2.7	3,617	26.7
696	386	2.9	4,003	29.6
702	382	2.8	4,385	32.4
708	373	2.8	4,758	35.1
714	413	3.1	5,171	38.2
720	421	3.1	5,592	41.3
725	466	3.4	6,058	44.8
731	460	3.4	6,518	48.1
737	462	3.4	6,980	51.6
747	500	3.7	7,480	55.3
748	492	3.6	7,972	58.9
754	549	4.1	8,521	62.9
760	454	3.4	8,975	66.3
767	474	3.5	9,449	69.8
773	444	3.3	9,893	73.1
780	459	3.4	10,352	76.5
788	462	3.4	10,814	79.9
796	462	3.4	11,276	83.3
804	457	3.4	11,733	86.7
814	374	2.8	12,107	89.4
825	349	2.6	12,456	92.0
838	329	2.4	12,785	94.4

U.S. History Core B Score Distribution for Spring 2011 (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
854	271	2.0	13,056	96.4
875	224	1.7	13,280	98.1
904	143	1.1	13,423	99.2
957	85	0.6	13,508	99.8
999	29	0.2	13,537	100.0