



OKLAHOMA STATE DEPARTMENT OF EDUCATION

Janet Barresi, State Superintendent of Public Instruction

Oklahoma School Testing Program

Oklahoma Core Curriculum Tests

Grades 3 to 8 Assessments

2011 Technical Report

Submitted to
The Oklahoma State Department of Education
October 2011

PEARSON

ALWAYS LEARNING

Executive Summary

Introduction

The Oklahoma Core Curriculum Tests (OCCT) is one component of the Oklahoma School Testing Program (OSTP). The OCCT is a state-wide criterion referenced assessment program that includes currently include tests of Mathematic and Reading in grades 3 through 8; Science in grades 5 and 8; Social Studies in grades 5, 7(Geography), and 8 (U.S. History, Constitution, and Government); and Writing in grades 5 and 8. Each test is designed as a measure of a student's knowledge relative to the *Priority Academic Student Skills (PASS)*, Oklahoma's content standards.

The OCCT tests of Writing were administered on March 9, 2011. Three tests—grade 7 Geography, grade 8 Mathematics, and grade 8 Reading are primarily computer delivery (a paper form is available), and were administered during the online test window from April 11, 2011 to May 6,2011. The remaining tests were administered as paper test between April 11, 2011and May 11, 2011. This report provides technical details of work accomplished through the end of 2011 on all of these tests.

Purpose

The purpose of this Technical Report is to provide objective information regarding technical aspects of the OSTP-OCCT 3-8 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 OCCT 3-8 assessments. Other sources of information regarding this battery of tests 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-OCCT 3-8 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-OCCT 3-8 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-OCCT 3-8 testing program.

Information provided in this report presents valuable information about the OSTP-OCCT 3-8 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-OCCT 3-8 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.

DRAFT

Table of Contents

Section 1.....	7
1.1 Skills Assessed by the OCCT	8
1.2 Summary of Test Development and Content Validity.....	10
1.2.a Aligning Test to <i>PASS</i> Content Standards	10
1.2.b Additional Considerations in Item Selection.....	11
1.2.c Configuration of Test Forms and Field-Test Design.....	12
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 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
3.4 Data Review	21
3.4.a Data Review Materials and Meetings.....	22
3.4.b Results of Data Review	22
3.5 Test Reliability	23
3.6 Analysis of the Writing Tests	25
3.6.a Prompt Scoring.....	25
3.6.b Adjustment for Rater-Year Effects.....	25
3.6.c Inter-rater reliability.....	27
Section 4.....	29
4.1 Item Response Theory (IRT) models.....	29
4.2 Assessment of IRT Fit to the model	29
4.3 Calibration and Equating	30
4.4 Anchor items and Anchor Stability Evaluation Methods	30
4.5 Scaling and Scoring Results.....	32
Section 5.....	46
5.1 Classification Consistency and Accuracy	46
Section 6.....	50
6.1 Descriptive Statistics	50
6.2 Performance Level Distribution.....	54
6.3 Conditional Standard Error of Measurement	55

6.4 Standard Error of Measurement..... 56
References 58
Appendix A 59
Appendix B 77

DRAFT

List of Tables

Table 1.1. Testable Standards for OCCT Grades 3 to 8	8
Table 1.2. Criteria for Aligning the Test with <i>PASS</i> Standards and Objectives.	10
Table 1.3. Percentage of Items by Depth of Knowledge Levels	12
Table 3.1. Demographic Characteristics of the Student Sample for Spring 2011	17
Table 3.3. Test-Level Summaries of Classical Item Analyses for Spring 2011	19
Table 3.4. DIF Flag Incidence Across All OSTP-OCCT 3-8 Field Test Items for Spring 2011	21
Table 3.5. Number of Items Per Subject Flagged and Rejected During Spring 2011 Field Test Data Review.....	23
Table 3.7. Test Reliability by Subgroup for Spring 2011	24
Table 3.9. Inter-rater Reliability for Grade 5 and 8 Writing Prompts for Spring 2011	28
Table 4.1. Number of Linking Items Per Subject for Spring 2011	31
Table 4.2. LOSS, HOSS, and Scaling Constants by Subject	33
Table 4.3. Raw Score to Scale Score Conversion Tables for Mathematics (Grades 3 to 5) Spring 2011	34
Table 4.4. Raw Score to Scale Score Conversion Tables for Mathematics (Grades 6 to 8) Spring 2011	36
Table 4.5. Raw Score to Scale Score Conversion Tables for Reading (Grades 3 to 5) Spring 2011	38
Table 4.6. Raw Score to Scale Score Conversion Tables for Reading (Grades 6 to 8) Spring 2011	40
Table 4.7. Raw Score to Scale Score Conversion Tables for Science Spring 2011.....	42
Table 4.8. Raw Score to Scale Score Conversion Tables for Social Studies Spring 2011	44
Table 5.1. Estimates of Accuracy and Consistency of Performance Classification.....	47
Table 5.3. Accuracy and Consistency Estimates and False Positive/False Negative Rates by Cut Score	49
Table 6.1. Descriptive Statistics of the Scale Scores for Spring 2011 - Overall	50
Table 6.2. Descriptive Statistics of the Scale Scores for Spring 2011 by Gender	51
Table 6.3. Descriptive Statistics of the Scale Scores for Spring 2011 by Race/Ethnicity	52
Table 6.4. Descriptive Statistics of the Scale Scores for Spring 2011 by Free/Reduced Lunch Status	54
Table 6.9. Percentage of Students by Performance Level for Spring 2011	55
Table 6.10. Overall Estimates of <i>SEM</i> by Test.....	57

Section 1

Overview of the Oklahoma School Testing Program (OSTP) Oklahoma Core Curriculum Tests (OCCT)of Grades 3 to 8

The Achieving Classroom Excellence End-of-Instruction assessment is a state-mandated, criterion-referenced testing program used to assess student proficiency. In the spring of 2011, the OCCT assessments were administered to all eligible public school students in Grades 3 through 8. Currently, this assessment program includes tests of Mathematics and Reading and the end of grades 3 through 8, tests of science and writing following grades 5 and 8, and social studies tests at the end of grades 5, 7 and 8. The 2011 administration of the OCCT was the 17th for students in Grades 5 and 8 and the 7th for students in Grades 3, 4, and grade 7 social studies (Geography). This was the 6th operational administration of the Reading and Mathematics tests in Grades 6 and 7.

All 19 assessments are designed to measure student performance relative to a specific set of academic skills established by committees of Oklahoma educators. This set of skills—the *Priority Academic Student Skills (PASS)*—represents skills that students are expected to master by the end of each grade for each subject. The OCCT are untimed tests, and with the exception of the writing assessment which is a single open-ended written response to a prompt, student performance is measured exclusively by multiple choice (MC) items. The MC Tests in grades 3 through five are administered in two sessions. All tests in grades 6 through 8th grade and the grade 5 writing test are administered in a single session. The grade 8 Mathematics and Reading tests as well as the grade 7 geography tests were primarily computer delivered (paper forms were available only for make-ups for test test-takers with accommodations requiring a paper form). All other tests were administered exclusively as paper and pencil tests.

Pearson content specialists and research scientists worked with the Oklahoma State Department of Education (SDE) to construct OCCT test forms aligned to the *PASS* standards. In each test, a form consisted of a set of operational items used to produce student test scores and a set of embedded field-test items. The two Writing assessments consisted of a single constructed-response (CR) item. The Reading, Mathematics, Science, and Social Studies assessments were composed of Multiple-Choice (MC) items only. For each content and grade, there were eight forms¹ consisting of a common set of operational items and a unique set of 10 field-test items. Responses to the operational items were used to produce student scores. Responses to the field-test items were used to evaluate the psychometric properties of these newly developed items for possible inclusion on future forms. In addition, to the regular operational form, an equivalent form was designated for all Mathematics and Reading tests as well as grade 7 Geography and a Braille version of each 2011 operational forms was created. 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

¹ The grade 6 math test was an exception to this format. The grade 6 math assessment contained a single operational core and 11 field-test forms. Due to changes in the OCCT mathematics test blueprints and a realignment of the *PASS* standards in 2010-11, some objectives and skills had no operational items available for use on the 2011 operational forms. Three field-test items from underrepresented objectives were placed on all forms. Each of the 11 forms also contained seven field-test items unique to the form.

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 Skills Assessed by the OCCT

The OCCT is developed with the expressed purpose of measuring the Oklahoma *PASS* content standards. In some cases, the *PASS* standards contain objectives that are not easily assessed in a large-scale and standardized format (e.g., English-Language Arts *PASS* standards include listening, reviewing). Standards that are not assessed using the OCCT must be assessed by school districts locally. A complete listing of all standards and objectives for all subjects and grades (measured and unmeasured) can be found on the SDE website (<http://sde.state.ok.us/Curriculum/PASS/default.html>).

A list of the testable standards for each subject are listed in Tables 1.1. For Math² and Reading³, the same testable standards appear in each grade level.

The Tables in Appendix A provide information drawn from the 2011 test *PASS* blueprints. These tables show the *PASS* standards and objectives, as well as the number of items required for standard and objective according to the blueprint and actual number of items number of items appearing on the 2011 operational form.

Table 1.1. Testable Standards for OCCT Grades 3 to 8

Mathematics Grades 3 to 8	
Standard 1.	Algebraic Reasoning: Patterns and Relationships
Standard 2.	Number Sense and Operation
Standard 3.	Geometry
Standard 4.	Measurement
Standard 5.	Data Analysis
Reading Grades 4 to 8 (Grade 3)	
Standard 1. (Standard 2.)	Vocabulary
Standard 3. (Standard 4.)	Comprehension/Critical Literacy
Standard 4. (Standard 5.)	Literature
Standard 5. (Standard 6.)	Research and Information

² The Mathematics *PASS* standards were revised in 2009-2010 and required significant changes to the test blueprints, and thus required significant changes to the OCCT Mathematics item bank. The test administered in 2011 was the first time the new standards were assessed, and Pearson assisted in the setting of new performance standards in the summer of 2011.

³ While the Reading *PASS* standards that are assessed by OCCT are the same, the enumeration of these standards is slightly different in grade 3.

Science Grades 5 & 8

PASS Process/Inquiry Standards and Objectives

- Process 1. Observe and Measure
- Process 2. Classify
- Process 3. Experiment
- Process 4. Interpret and Communicate

Grade 5 PASS Content Standards

- Standard 1. Properties of Matter and Energy
- Standard 2. Organisms and Environments
- Standard 3. Structures of the Earth and the Solar System

Grade 8 PASS Content Standards

- Standard 1. Properties and Chemical Changes in Matter
- Standard 2. Motion and Forces
- Standard 3. Diversity and Adaptations of Organisms
- Standard 4. Structures/Forces of the Earth/Solar System
- Standard 5. Earth's History

Social Studies Grade 5

- Standard 2. Early Exploration
- Standard 3. Colonial America
- Standard 4. American Revolution
- Standard 5. Early Federal Period
- Standard 7. Geographic Skills

Social Studies Grade 7 (Geography)

- Standard 1./6. Geographic Tools/Geography Skills
- Standard 2. Regions
- Standard 3. Physical Systems
- Standard 4. Human Systems
- Standard 5. Human/Environment Interaction

Social Studies Grade 7 (U.S. History)

- Standard 1. Social Studies Process Skills
- Standard 3. Causes of the American Revolution
- Standard 4. Results of the American Revolution
- Standard 5. Governing Documents/Early Federal Period
- Standard 6. Northern/Southern Economic Growth
- Standard 7. Jacksonian Era
- Standard 8. Cultural Growth and Reform
- Standard 9. Westward Movement
- Standard 10. Eve of War

1.2 Summary of Test Development and Content Validity

To ensure content validity of the OCCT 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 Spring 2011 assessments.

1.2.a Aligning Test to *PASS* Content Standards

In addition to the test Blueprints provided by SDE (see Appendix A), 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 <i>PASS</i> standard. The number of items 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. Depth-of Knowledge Consistency	The test is constructed using items from a variety of Depth of Knowledge levels that are consistent with the processes students need in order to demonstrate proficiency for each <i>PASS</i> objective.
3. Range of Knowledge Correspondence	The test is constructed so that at least 75% of the objectives for a <i>PASS</i> standard have at least one corresponding assessment item.
4. Balance-of-Representation	The test is constructed according to the Test Blueprint which reflects the degree of representation given on the test to each <i>PASS</i> standard and/or objective in terms of the percent of total test items measuring each standard and the number of test items measuring each standard and/or objective. The test construction shall yield a balance of representation with an index of 0.7 or higher of assessed objectives related to a standard.
5. Source-of-Challenge	Each test item is constructed in such a way that the major cognitive demand comes directly from the targeted <i>PASS</i> objective or concept being assessed, not from specialized knowledge or cultural background that the test-taker may bring to the testing situation.

1.2.b Additional Considerations in Item Selection

The source of the operational items eligible for inclusion on the Spring 2011 form is a pool of previously field-tested or operationally-administered items ranging from the Spring 2005 through the Spring 2010 administrations. In each case, items were calibrated using live data from the operational administrations to estimate parameters for these items.

To equate the forms across years, a 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 Spring 2011 operational assessments.

DRAFT

Table 1.3. Percentage of Items by Depth of Knowledge Levels

Subject	Grade	DOK Level					
		1		2		3	
		Target	Actual	Target	Actual	Target	Actual
Math	3	20-25	20	65-70	66	5-15	14
	4	20-25	22	65-70	64	5-15	14
	5	20-25	20	65-70	66	5-15	14
	6	10-15	10	65-70	72	15-25	18
	7	10-15	14	65-70	72	15-25	14
	8	10-15	18	65-70	66	15-25	16
Reading	3	20-25	14	65-70	68	5-15	18
	4	20-25	16	65-70	62	5-15	22
	5	20-25	10	65-70	74	5-15	16
	6	10-15	14	65-70	66	15-25	20
	7	10-15	10	65-70	70	15-25	20
	8	10-15	10	65-70	72	15-25	18
Science	5	20-25	20	65-70	62	5-15	18
	8	10-15	18	65-70	69	15-25	13
Social Studies	5	20-25	23	65-70	62	5-15	15
	7	10-15	18	65-70	67	15-25	16
	8	10-15	16	65-70	69	15-25	16

Note: All values are in percentages.

1.2.c Configuration of Test Forms and Field-Test Design

Error! Reference source not found. provides an overview of the number of operational and field test items for the Spring 2011 OSTP-OCCT 3-8 assessments. The Spring 2011 test is comprised of single core of operational items on each form. For each of the MC based tests, at least 20% of the operational items were designated as potential anchor items to be used in the equating process (the process for acceptance as an anchor items is detailed in section 4). For the 17 MC tests, eight to eleven field-test forms were created. Each field-test form included the operation core and 10 field-test items. These items are embedded in the operational test forms with the intent of building the item bank for future use. Each form of the assessment was spiraled within classrooms to obtain randomly-equivalent samples of examinees for the field test items.

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. Ten multiple choice field test items were placed in common positions on each forms of each assessment. Field test items were prioritized for inclusion on forms based on current item bank analyses which revealed which particular standards and objectives would benefit most from field testing. The tables in Appendix A contain the counts of field-test items aligned with each *PASS* objective.

Table 1.4. Configuration of the OSTP-OCCT 3-8 Tests for Spring 2011

Subject	Grade	Total Number of					Maximum possible test items per FT form		
		Core Forms	FT Forms	OP items	FT items	Total items	OP	FT	Total
Math	3	1	8	50	80	130	50	10	60
	4	1	8	50	80	130	50	10	60
	5	1	8	50	80	130	50	10	60
	6	1	11	50	80*	130	50	10	60
	7	1	8	50	80	130	50	10	60
	8	1	8	50	96^	146	50	12^	60
Reading	3	1	8	50	80	130	50	10	60
	4	1	8	50	80	130	50	10	60
	5	1	8	50	80	130	50	10	60
	6	1	8	50	80	130	50	10	60
	7	1	8	50	80	130	50	10	60
	8	1	8	50	80	130	50	10	60
Science	5	1	8	45	80	125	45	10	55
	8	1	8	45	80	125	45	10	55
Social Studies	5	1	8	60	80	140	60	10	70
	7	1	8	45	80	125	45	10	55
	8	1	8	45	80	125	45	10	55

Note: OP = Operational; FT = Field Test; *three of these field test items were administered across all forms and became part of the scored set post-administration (replacing three operational items); ^two technologically-enhanced items were administered per form for a total of 16 items.

Section 2

Administration of the OCCT in Grades 3 to 8

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-OCCT 3-8 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 the OCCT.

2.1 Packaging and Shipping

To provide OSTP-OCCT 3-8 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

This section provides an overview of the initial statistical analyses carried out for the 2011 administration of the OCCT. Following the administration of the OCCT, student demographic and item response data were transmitted to Pearson research scientists, who are responsible for the all statistical analyses for the OCCT. The classical analyses described in this section (as well the calibration and equating of each test) were conducted of near population level data (each subject had at least 85% of the student population represented before processing).

3.1 Data Receipt Activities

After all tests were scored, a data clean-up process that removed invalid cases, ineligible responses, and absent students was preformed for each test. Additionally, 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 Pearson research scientist inspected several data fields to determine if the data met expectations. This included screening the following variables:

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

Exclusion Rules. Following data inspection and clean-up, exclusionary rules were applied to form the final sample that was used for classical item analyses, calibration, and equating. Any student who had attempted at least five responses was included in the data analyses. The demographic breakdown of the students in Spring 2011 item analysis and calibration sample appears in Table 3.1.

Table 3.1. Demographic Characteristics of the Student Sample for Spring 2011

Subject/Grade		Female	Male	African American	Native American	Hispanic	Asian	Pacific Islander	White	Other
MATH	3	22598	22442	4403	7422	6340	886	102	23909	1978
	4	21983	22323	4127	7438	6004	902	100	23821	1914
	5	21821	21936	4222	7491	5695	858	89	23582	1820
	6	21495	21633	4178	7432	5327	855	97	23491	1748
	7	21215	21433	4208	7336	5081	823	90	23455	1655
	8	20958	20347	3918	6969	4816	866	80	22994	1662
	All	130070	130114	25056	44088	33263	5190	558	141252	10777
READING	3	22447	21979	4344	7328	6261	876	101	23578	1938
	4	21828	21845	4066	7323	5890	896	99	23526	1873
	5	21767	21615	4195	7439	5602	856	89	23403	1798
	6	21486	21416	4153	7393	5263	852	97	23407	1737
	7	21243	21380	4207	7327	5057	827	90	23457	1658
	8	20934	20205	3924	6978	4759	849	77	22897	1655
	All	129705	128440	24889	43788	32832	5156	553	140268	10659
SCIENCE	5	22159	22246	4338	7611	5766	864	90	23889	1847
	8	21336	20916	4077	7176	4911	876	80	23418	1714
	All	43495	43162	8415	14787	10677	1740	170	47307	3561
SOCIAL STUDIES	5	23380	24439	4815	8274	6219	883	94	25546	1988
	7	22654	23896	4780	8123	5510	827	99	25428	1783
	8	22375	22799	4473	7830	5169	890	83	24903	1826
	All	68409	71134	14068	24227	16898	2600	276	75877	5597

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 0.25
- 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 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:

- Percentage of students endorsing each multiple choice option (overall and broken down by gender and ethnicity)
- Overall p -value for each item
- Point-biserial correlation (overall and broken down by gender and ethnicity)
- Point-biserial for non-key response options (overall and broken down by gender and ethnicity)
- Omit percentage per item
- Mean score by response option (overall and broken down by gender and ethnicity)

The classical analysis of operational items is used as an additional quality control step to ensure that operational items are not behaving in an unexpected or aberrant manner. The Item analysis results of the operational items are reviewed by Pearson research Scientists and in the case of unexpected item performance, a course of action (e.g., retain item, drop from operational scoring) regarding the item(s) are recommended to SDE. In the 2011 administration, all operational items performed adequately and were deemed appropriate 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 3.2. Note that students whose tests were invalidated were excluded. The operational test results indicate that the omit rates were small for all assessments (grade three, which is administered as a consumable booklet, is slightly higher). In all cases, the average p -value ranged from 0.54 to 0.73 and average point biserial correlations ranged from 0.31 to 0.37. In tandem, these summary statistics indicate sets of operational items that are functioning appropriately.

Table 3.2. Test-Level Summaries of Classical Item Analyses for Spring 2011

Subject	Grade	Sample Size	Mean	Mean % of Max	Items Points	Mean P	Mean r_{pb}	Omit Min	Omit Max
MATH	03	45290	37.55	0.75	50	0.73	0.37	0.13	3.19
	04	44562	36.46	0.73	50	0.73	0.36	0.00	0.46
	05	44112	34.87	0.70	50	0.67	0.36	0.00	0.40
	06	43339	31.47	0.63	50	0.59	0.36	0.00	0.29
	07	42850	30.83	0.62	50	0.57	0.35	0.00	0.26
	08	41332	32.00	0.64	50	0.59	0.36	0.00	0.10
READING	03	44692	34.58	0.69	50	0.66	0.37	0.13	1.75
	04	43948	35.60	0.71	50	0.67	0.33	0.00	0.32
	05	43751	37.94	0.76	50	0.71	0.35	0.00	0.23
	06	43107	33.58	0.67	50	0.67	0.35	0.00	0.86
	07	42813	34.81	0.70	50	0.69	0.36	0.00	0.22
	08	41166	36.68	0.73	50	0.72	0.32	0.00	0.08
SCIENCE	05	44754	31.10	0.69	45	0.63	0.33	0.02	0.31
	08	42512	28.78	0.64	45	0.58	0.31	0.00	0.31
SOCIAL	05	48087	35.31	0.59	60	0.54	0.33	0.00	0.31
	07	46560	29.00	0.64	45	0.60	0.31	0.00	0.24
	08	45365	27.66	0.61	45	0.58	0.36	0.00	0.25

r_{pb} = point biserial correlation.

3.3 Procedures for Detecting Item Bias

One of the goals of the OSTP-OCCT 3-8 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 OCCT 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.

The data in Table 3.3 summarizes the number of items in DIF categories for the 17 multiple choice tests for the OCCT 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 (described below in the Section 3.4.) and items that were determined to exhibit bias as a result of the content of the item were removed from the item bank excluding them from future use.

Table 3.3. DIF Flag Incidence Across All OSTP-OCCT 3-8 Field Test Items for Spring 2011

Subject/Grade		Total FT items	Female	African American	Native American	Hispanic	Asian
MATH	03	80	4	11	2	8	6
	04	80	3	9	0	7	7
	05	80	9	9	0	5	7
	06	80	7	9	0	3	2
	07	80	4	6	0	0	4
	08	80	4	3	0	1	7
READING	03	80	0	6	0	6	7
	04	80	0	7	0	7	7
	05	80	1	3	0	7	5
	06	80	3	8	0	9	4
	07	80	7	11	0	12	10
	08	80	7	8	1	3	7
SCIENCE	05	80	1	9	0	3	8
	08	80	6	7	0	5	6
SOCIAL STUDIES	05	80	2	1	0	4	3
STUDIES	07	80	4	5	0	8	9
	08	80	2	0	0	1	6

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 17 subjects based on the Spring 2011 field test administration. 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 OCCT 3-8 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.

3.4.a Data Review Materials and Meetings

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 OCCT assessments are as follows:

- p -value < .25 or > .90
- point-biserial correlation < .20
- 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.b Results of Data Review

The number of items inspected during data review as a result of the item meeting the statistical flagging criteria for the classical item analyses, DIF, and IRT procedures is presented in Table 3.4.

Table 3.4. Number of Items Per Subject Flagged and Rejected During Spring 2011 Field Test Data Review

Subject	Grade	FT Items	No. Flagged	Rejected	Accepted	Accepted with Edits
Mathematics	3	80	46	10	70	0
Mathematics	4	80	36	8	70	2
Mathematics	5	80	33	3	76	1
Mathematics	6	80	32	9	71	0
Mathematics	7	80	22	9	71	0
Mathematics	8	80	32	10	70	0
Reading	3	80	26	4	76	0
Reading	4	80	32	10	70	0
Reading	5	80	22	5	75	0
Reading	6	80	26	2	78	0
Reading	7	80	29	5	75	0
Reading	8	80	44	10	70	0
Science	5	80	28	8	71	1
Science	8	80	40	9	67	4
Social Studies	5	80	24	3	68	9
Social Studies	7	80	38	3	62	15
Social Studies	8	80	25	1	67	12

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 $i = 1, 2, \dots, k$, and s_{sum}^2 denotes the variance for the sum of all k items:

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

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

Table 3.5 presents Cronbach’s alpha for the operational tests by subject area for the Spring 2011 OCCT administration. These reliability coefficients indicate that the OSTP-OCCT assessments had strong internal consistency and that the tests produce relatively stable scores. Additionally, Table 3.5 shows the reliability analysis results by the different reporting subgroups for the OSTP-OCCT assessments for Spring 2011 for the operational items. In all instances, the reliability coefficients are well above the accepted lower limit of .70, with most values near .90.

Table 3.5. Test Reliability by Subgroup for Spring 2011

Subject	Grade	All	Female	Male	African American	Native American	Hispanic	Asian	Pacific Islander	White	Other
MATH	03	0.91	0.91	0.90	0.91	0.90	0.90	0.91	0.90	0.89	0.90
	04	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.92	0.89	0.90
	05	0.90	0.90	0.90	0.90	0.89	0.90	0.90	0.92	0.89	0.89
	06	0.90	0.89	0.90	0.89	0.89	0.89	0.91	0.91	0.90	0.90
	07	0.89	0.89	0.90	0.88	0.88	0.89	0.91	0.91	0.89	0.88
	08	0.90	0.89	0.90	0.89	0.89	0.89	0.89	0.92	0.89	0.89
READ	03	0.88	0.88	0.89	0.88	0.87	0.88	0.89	0.88	0.87	0.87
	04	0.89	0.88	0.89	0.88	0.88	0.89	0.88	0.89	0.87	0.88
	05	0.90	0.89	0.90	0.89	0.89	0.90	0.90	0.90	0.88	0.89
	06	0.88	0.88	0.89	0.88	0.87	0.88	0.88	0.89	0.88	0.87
	07	0.89	0.88	0.89	0.89	0.88	0.89	0.89	0.91	0.88	0.87
	08	0.86	0.86	0.86	0.88	0.85	0.88	0.88	0.93	0.83	0.86
SCIENCE	05	0.88	0.87	0.88	0.87	0.87	0.87	0.90	0.91	0.87	0.87
	08	0.86	0.84	0.87	0.82	0.84	0.83	0.86	0.86	0.85	0.84
SOCIAL STUDIES	05	0.88	0.87	0.89	0.86	0.87	0.86	0.89	0.89	0.88	0.87
	07	0.86	0.85	0.87	0.85	0.85	0.85	0.87	0.91	0.86	0.85
	08	0.90	0.89	0.91	0.88	0.89	0.89	0.90	0.89	0.90	0.89

3.6 Analysis of the Writing Tests

The administration of the spring 2011 Writing assessment took place on March 9, 2011. Students at Grades 5 and 8 were given one operational writing prompt. The following sections describe the statistical analyses conducted to place the 2011 operational writing prompts on the scale established in 2006.

3.6.a Prompt Scoring

The writing score is a weighted composite of five analytic scores that focus on specific domains of writing skills. These skills are listed in Table 3.6. Each student's response to a prompt is read by two independent raters; the raters' scores for each domain are averaged. The domain scores range from 1 (the lowest score) to 4 (the highest score).

Table 3.7. Test Reliability by Subgroup for Spring 2011

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

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

$$RCS = 15 * (0.30 * ID + 0.25 * OUC + 0.15 * WC + 0.15 * SP + 0.15 * GUM) \quad (2)$$

OSTP-OCCT 3-8 Spring 2011 operational writing prompts are within the moderate range.

3.6.b Adjustment for Rater-Year Effects

The baseline for each grade's operational writing scale was 2006. To place the 2011 operational prompts on the 2006 scale, transformation constants were obtained to adjust RCS scores for prompt difficulty and for rater-year effects relative to a target distribution. All calculations were performed on the RCS prior to rounding. For reporting, the scaled composite scores (SCS) were then rounded to the nearest integer between 15 and 60. For each of the 2007 field-test prompts, ETS provided a set of unique transformation constants to adjust for both prompt difficulty and rater-year effects. Based on ETS' report, *OCCT Writing: Scaling the 2007 Field-Test Prompts* (ETS, 2007), the following equation was used to adjust the 2010 raw composite scores:

$$SCS_{11} = B_{07}(RCS_{11}) + A_{07} \quad (3)$$

Where SCS_1 represents the scaled composite score after adjusting the 2011 prompt to the 2007 scale.

In 2011, Pearson also performed a rater drift to adjust for the difference in raters between the 2007 administration to the current administration. Pearson’s Performance Scoring Center (PSC) randomly scored approximately 500 student responses from 2007 for each grade’s prompt and distributed these into the current administration scoring throughout the entire scoring window. The student responses were pulled by lithocode and were only the valid scored responses (i.e., no condition codes such as off-topic present). The rescored prompts were then linked to their original 2007 scores and formed the basis for a second set of linear scaling constants.

The 2011 rater effect constants (C_{11} & D_{11}) were determined by using the means (M) and standard deviations (S) of the 2007 raw composite scores and the 2011 rescored raw composite scores as calculated below for each grade.

$$D_{11} = S_{07} / S_{11} \tag{4}$$

$$C_{11} = M_{07} - (M_{11} * D_{11}) \tag{5}$$

Because both are corrected due to raters and a rescaling to the 2007 scale is desired, a compound adjustment—using both sets of constants—is required. Final composite scores were created using the formula below:

$$SCS_{11} = B_{07} [(D_{11} * RCS_{11}) + C_{11}] + A_{07} \tag{6}$$

Table 3.8 provides the resulting score distribution statistics with no adjustment, only the ETS adjustment, and the compound adjustment. Transformation constants are provided at the bottom of the Tables.

Table 3.9. Results of Grades 5 and 8 Writing Prompt Scoring and scaling

Grade	Statistic	2011 No Adjustment	2011 Final	2010 Scores	2009 Scores
	N	46,057	46,057	44994	43665
	MIN	15	18	15	19
	MAX	60	60	60	60
	MEAN	40.49	46.21	43.67	44.57
	STD	8.12	7.99	8.25	8.54
Constants					
5	A	3.1308			
	B	0.9471			
	C	0.0898			
	D	1.0433			
	N	43,051	43,051	40962	42271
	MIN	15	60	19	18
	MAX	15	60	60	60
	MEAN	43.12	45.76	45.73	45.5
	STD	6.90	7.28	7.42	7.04
	Constants				
8	A	3.1309			
	B	0.9471			
	C	-4.3392			
	D	1.1458			

3.6.c 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). Raters for the grade 5 and 8 Writing assessments 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 **Error! Reference source not found.** The results show that exact and adjacent rater agreement on trait scores for both the grade 5 and 8 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 OCCT grade 5 and 8 Writing assessment's writing prompts are close to or within the moderate range.

Table 3.10. Inter-rater Reliability for Grade 5 and 8 Writing Prompts for Spring 2011

Trait	Max Points	Valid N	Point Discrepancy Percentages							Agreement Percentages			Kappa
			-3	-2	-1	0	1	2	3	Exact	Adjacent	+/- 2 or more	
Grade 5													
1	4	46,047	0.00	0.56	18.51	61.93	18.42	0.58	0.00	61.93	36.93	1.14	0.40
2	4	46,047	0.01	0.63	18.76	61.18	18.77	0.65	0.00	61.18	37.53	1.29	0.39
3	4	46,047	0.00	0.62	19.16	60.91	18.63	0.67	0.00	60.91	37.79	1.3	0.40
4	4	46,047	0.00	0.73	19.60	59.88	18.98	0.81	0.00	59.88	38.58	1.54	0.40
5	4	46,047	0.00	0.78	19.55	59.73	19.13	0.80	0.00	59.73	38.68	1.59	0.41
Grade 8													
1	4	43,051	0.00	0.39	16.28	66.75	16.19	0.40	0.00	66.75	32.47	0.78	0.36
2	4	43,051	0.00	0.45	16.64	65.87	16.58	0.45	0.00	65.87	33.22	0.91	0.36
3	4	43,051	0.00	0.40	16.27	66.31	16.68	0.38	0.00	66.31	32.95	0.74	0.37
4	4	43,051	0.00	0.41	15.63	68.00	15.63	0.34	0.00	68.00	31.26	0.74	0.42
5	4	43,051	0.00	0.43	15.83	67.03	16.37	0.34	0.00	67.03	32.2	0.77	0.38

Section 4

Calibration, Equating, and Scaling

4.1 Item Response Theory (IRT) models

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

$$P_i(\theta) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta - b_i)}}, \quad (7)$$

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.

The fitting of the IRT models to the 2011 assessment data were implemented using MULTILOG 7.0 (Thissen, Chen, & Bock, 2003). MULTILOG estimates parameters simultaneously for dichotomous items via marginal maximum likelihood procedures. All item and student ability calibrations were independently conducted and verified by at two Pearson research scientists.

4.2 Assessment of IRT Fit to the model

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

$$Q_{i_i} = \sum_{r=1}^{10} \frac{N_r (O_{ir} - E_{ir})^2}{E_{ir} (1 - E_{ir})}, \quad (8)$$

where Q_{i_i} is the fit of the i th 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 (7) and summing over examinees in cell r :

$$E_{ir} = \frac{1}{N_{ir}} \sum_{k \in r} P_i(\hat{\theta}_k). \quad (9)$$

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

$$Z_j = \frac{Q_{i_i} - df}{\sqrt{(2df)}}. \quad (10)$$

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. A total of two operational items (one grade 3 Mathematics item and one Grade 8 Social Studies Item) were flagged as potentially misfitting, but showed no other evidence of aberrant behavior, and were not sent for further review.

Field Test Items. The field test items across all subjects were evaluated using the Q_1 statistic to evaluate the extent to which the obtained proportions of item scores are close to the proportions that would be expected based on the estimated thetas and item parameters. Any field test items flagged by Q_1 were included in the data review for review by contest specialists from Pearson and SDE (for more on data review, please see Section 0).

4.3 Calibration and Equating

The 3-PL model was used for calibration of all multiple choice items. A common item, non-equivalent groups (CINEG) design was used for all content areas to link the current test forms 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. 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 Anchor items and Anchor Stability Evaluation Methods

Table 4.1 presents the number and percentage of anchor items (before and after anchor stability checks) by subject and grade for the Spring 2011 administration. The anchor set was comprised of at least 20% of all operational items (counts vary by subject/grade). In addition, the 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 Linking Items Per Subject for Spring 2011

Subject	Grade	Operational Items	Initial Anchor Set		Final Anchor Set	
			Items	%	Item	%
Math	3	50	21	42%	20	40%
	4	50	20	40%	19	38%
	5	50	18	36%	15	30%
	6	50	18	36%	18	36%
	7	50	20	40%	19	38%
	8	50	20	40%	20	40%
Reading	3	50	19	38%	19	38%
	4	50	20	40%	19	38%
	5	50	18	36%	18	36%
	6	50	20	40%	20	40%
	7	50	18	36%	16	32%
	8	50	16	32%	16	32%
Science	5	45	16	36%	14	31%
	8	45	15	33%	15	33%
Social Studies	5	60	19	32%	18	30%
	7	45	15	33%	9	20%
	8	45	15	33%	15	33%

Despite the careful selection and placement of the operational items to be used as anchor items to link the 2011 test to the bank scale, 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 Spring 2011 administration. The process used in this evaluation is called an anchor stability check.

The anchor 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 item, calculate the weighted sum of the squared deviation (d^2) between the two Item Characteristic Curves—one ICC for each set of parameters.
 - a) For each item calculate a weighted sum of the squared deviation between the ICCs based on old (x) and new (y) parameters at each point in this theta distribution.

$$d_i^2 = \sum^k [P_{ix}(\theta_k) - P_{iy}(\theta_k)]^2 \cdot g(\theta_k) \quad (11)$$

- b) Review and sort the items in a descending (largest to smallest) fashion according to the d^2 estimate.
 - c) Drop the items with the largest d^2 item from inclusion in the anchor set.
- 5) Repeat steps 2 through 3 until 10 items are dropped. This will result in 11 raw score to scale score tables.
 - 6) Compare each RSSS table with the RSSS based on the use of one less anchor item. When two adjacent RSSS tables no longer differ in performance classification at each of the cut score points, the anchor set is considered stable. The constants used to generate the RSSS based on the largest number of anchor items when stability is achieved should be retained as the final SL constants.

Before removing any item from the item parameter stability check, the following additional characteristics were examined: 1) prior and current year p -values and point-biserial correlations, 2) prior and current year IRT parameter estimates, 3) prior and current year item sequence, 4) standard and objective/skill of the item, 5) impact on blueprint representation, 6) passage ID/title for items linked to a stimulus, and 7) content review of the actual item. Decisions about whether to keep or remove an item were evaluated on a per item basis. If an item (note, only one item can be removed at a time) was removed from the, the process (beginning at the equating step) was 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).

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. Table 4.1 show the final number of anchor items used for equating for OCCT each test. Any item removed from the anchor parameter stability check set still contributed to student scores.

4.5 Scaling and Scoring Results

The Lowest Obtainable Scale Score (LOSS), Highest Obtainable Scale Score (HOSS), and final scaling constants for each of the subjects are shown in Table 4.4. The scaling constants, $M1$ (multiplicative) and $M2$ (additive), place the true scores associated with each raw score point onto the reporting or operational scale using a straightforward linear transformation:

$$\text{Scale Score} = (\hat{\tau} \times M1) + M2 \quad (12)$$

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

The raw score to number-correct scale 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

OCCT 3-8 assessments, there are three cut scores that divide scores into four performance levels: Unsatisfactory, Limited Knowledge, Proficient (or Satisfactory), and Advanced. The cut scores for each of the tests appear in **Error! Reference source not found.**. 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 are shown in Table 4.2 to Table 4.3 for Spring 2011. RSSS tables for writing are not included in the list of tables as there no further transformation of the composite score beyond that described in section 3.6.

Table 4.4. LOSS, HOSS, and Scaling Constants by Subject

Content	Grade	M1	M2	LOSS	HOSS	Limited Cut	Proficient Cut	Advanced Cut
Mathematics	3	85	708.939	400	990	633	700	798
	4	85	702.339	400	990	639	700	805
	5	85	680.604	400	990	638	700	791
	6	85	729.793	400	990	664	700	795
	7	85	723.183	400	990	674	700	800
	8	85	672.0737	400	990	642	700	774
Reading	3	85	707.013	400	990	649	700	891
	4	85	702.672	400	990	658	700	845
	5	85	696.836	400	990	641	700	830
	6	85	744.586	400	990	647	700	828
	7	85	749.593	400	990	668	700	802
	8	85	714.419	400	990	655	700	833
Science	5	70	753.9	400	990	638	700	814
	8	70	745.5	400	990	647	700	829
Social Studies	5	70	713.81	400	990	645	700	786
	7	70	759.777	400	990	595	700	847
	8	70	709.94	400	990	622	700	821
Writing	5	NA	NA	15	60	26	36	54
	8	NA	NA	15	60	25	36	54

Table 4.5. Raw Score to Scale Score Conversion Tables for Mathematics (Grades 3 to 5) Spring 2011

Raw Score	Grade 3			Grade 4			Grade 5		
	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	43	400	1	44	400	1	54
1	400	1	43	400	1	44	400	1	54
2	400	1	43	400	1	44	400	1	54
3	400	1	43	400	1	44	400	1	54
4	400	1	43	400	1	44	400	1	54
5	400	1	43	400	1	44	400	1	54
6	400	1	43	400	1	44	400	1	54
7	400	1	43	400	1	44	400	1	54
8	400	1	43	400	1	44	400	1	54
9	400	1	43	406	1	45	400	1	54
10	434	1	48	447	1	51	411	1	55
11	464	1	52	475	1	55	464	1	61
12	486	1	53	497	1	55	496	1	63
13	505	1	52	515	1	54	520	1	63
14	521	1	50	531	1	51	539	1	59
15	535	1	46	545	1	47	556	1	54
16	548	1	43	557	1	43	570	1	49
17	559	1	39	569	1	40	582	1	43
18	570	1	36	580	1	37	593	1	39
19	581	1	34	591	1	35	604	1	35
20	590	1	32	601	1	33	613	1	32
21	600	1	30	610	1	31	622	1	30
22	609	1	29	620	1	30	630	1	28
23	617	1	28	629	1	29	638	2	27
24	625	1	27	637	1	28	646	2	25
25	633	2	26	645	2	27	654	2	25
26	641	2	25	653	2	26	661	2	24
27	648	2	24	661	2	25	668	2	23
28	656	2	24	669	2	25	675	2	23
29	663	2	23	676	2	24	682	2	22
30	670	2	23	684	2	24	689	2	22
31	677	2	22	691	2	23	696	2	22
32	684	2	22	698	2	23	703	3	22
33	691	2	22	705	3	23	710	3	22

34	698	2	22	713	3	23	717	3	22
35	706	3	22	720	3	23	724	3	22
36	713	3	22	728	3	23	732	3	23
37	721	3	22	735	3	24	740	3	23
38	728	3	23	744	3	24	748	3	24
39	737	3	23	752	3	25	757	3	24
40	745	3	24	761	3	26	766	3	25
41	754	3	25	771	3	27	776	3	27
42	764	3	26	781	3	29	787	3	28
43	775	3	28	793	3	31	799	4	30
44	786	3	31	806	4	34	812	4	33
45	800	4	35	821	4	38	827	4	37
46	816	4	40	839	4	43	845	4	42
47	836	4	47	862	4	48	869	4	47
48	865	4	53	893	4	50	902	4	49
49	915	4	53	948	4	43	961	4	39
50	990	4	40	990	4	34	990	4	33

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

Table 4.6. Raw Score to Scale Score Conversion Tables for Mathematics (Grades 6 to 8)Spring 2011

Raw Score	Grade 6			Grade 7			Grade 8		
	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	63	400	1	73	400	1	67
1	400	1	63	400	1	73	400	1	67
2	400	1	63	400	1	73	400	1	67
3	400	1	63	400	1	73	400	1	67
4	400	1	63	400	1	73	400	1	67
5	400	1	63	400	1	73	400	1	67
6	400	1	63	400	1	73	400	1	67
7	400	1	63	400	1	73	400	1	67
8	400	1	63	400	1	73	400	1	67
9	400	1	63	400	1	73	400	1	67
10	470	1	71	400	1	73	400	1	67
11	510	1	73	400	1	73	421	1	69
12	537	1	72	483	1	81	490	1	74
13	559	1	68	526	1	83	527	1	76
14	576	1	61	556	1	81	553	1	73
15	591	1	54	579	1	76	572	1	67
16	604	1	48	597	1	68	588	1	60
17	616	1	42	612	1	60	602	1	53
18	627	1	37	625	1	52	614	1	46
19	636	1	34	637	1	45	624	1	40
20	645	1	31	647	1	40	634	1	35
21	654	1	29	657	1	35	642	2	32
22	662	1	27	666	1	32	651	2	29
23	670	2	26	674	2	29	659	2	27
24	678	2	25	682	2	28	666	2	25
25	685	2	24	690	2	26	673	2	24
26	692	2	23	697	2	25	680	2	23
27	699	2	22	704	3	24	687	2	22
28	706	3	22	711	3	23	693	2	22
29	713	3	21	718	3	22	700	3	21
30	720	3	21	725	3	22	706	3	21
31	726	3	21	731	3	22	713	3	21
32	733	3	20	738	3	21	719	3	21

33	739	3	20	745	3	21	726	3	20
34	745	3	20	751	3	21	732	3	20
35	752	3	20	758	3	21	739	3	20
36	759	3	20	765	3	21	746	3	20
37	765	3	20	772	3	21	752	3	20
38	772	3	20	779	3	21	759	3	21
39	779	3	20	786	3	21	767	3	21
40	786	3	21	794	3	21	774	4	22
41	794	3	22	802	4	22	782	4	22
42	803	4	23	810	4	23	791	4	24
43	812	4	24	819	4	25	800	4	25
44	821	4	27	830	4	27	810	4	28
45	833	4	30	841	4	30	822	4	31
46	846	4	35	855	4	33	837	4	36
47	863	4	40	872	4	38	854	4	42
48	888	4	45	895	4	42	879	4	47
49	931	4	44	934	4	41	922	4	48
50	990	4	34	990	4	31	990	4	35

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

Table 4.7. Raw Score to Scale Score Conversion Tables for Reading (Grades 3to 5) Spring 2011

Raw Score	Grade 3			Grade 4			Grade 5		
	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	52	400	1	49	400	1	47
1	400	1	52	400	1	49	400	1	47
2	400	1	52	400	1	49	400	1	47
3	400	1	52	400	1	49	400	1	47
4	400	1	52	400	1	49	400	1	47
5	400	1	52	400	1	49	400	1	47
6	400	1	52	400	1	49	400	1	47
7	400	1	52	400	1	49	400	1	47
8	400	1	52	400	1	49	400	1	47
9	427	1	54	437	1	53	400	1	47
10	473	1	59	475	1	57	439	1	51
11	501	1	60	500	1	58	473	1	55
12	522	1	59	520	1	56	497	1	55
13	539	1	55	537	1	52	515	1	53
14	553	1	50	550	1	47	529	1	49
15	565	1	45	562	1	42	542	1	45
16	576	1	41	573	1	38	553	1	41
17	586	1	37	583	1	34	563	1	36
18	596	1	34	592	1	31	572	1	33
19	605	1	31	600	1	29	581	1	30
20	613	1	30	608	1	27	589	1	28
21	622	1	28	615	1	25	597	1	26
22	630	1	27	622	1	24	604	1	25
23	638	1	27	629	1	23	611	1	24
24	646	1	26	635	1	23	618	1	23
25	653	2	25	642	1	22	625	1	22
26	661	2	25	648	1	22	631	1	22
27	668	2	24	654	1	21	637	1	21
28	676	2	24	661	2	21	643	2	21
29	683	2	24	667	2	21	650	2	20
30	690	2	24	673	2	21	656	2	20
31	698	2	24	679	2	21	662	2	20
32	705	3	24	686	2	21	668	2	20
33	713	3	24	692	2	21	674	2	20

34	721	3	24	699	2	21	680	2	20
35	729	3	25	706	3	22	687	2	20
36	737	3	25	713	3	22	694	2	20
37	746	3	26	721	3	23	701	3	21
38	755	3	27	728	3	23	708	3	21
39	764	3	28	737	3	24	715	3	22
40	775	3	29	745	3	25	723	3	23
41	786	3	31	754	3	26	732	3	24
42	798	3	33	764	3	28	741	3	26
43	811	3	36	775	3	30	751	3	28
44	827	3	39	788	3	33	763	3	31
45	844	3	44	802	3	37	776	3	35
46	866	3	48	818	3	42	792	3	41
47	894	4	51	839	3	48	812	3	50
48	934	4	48	869	4	54	840	4	60
49	990	4	36	920	4	52	888	4	65
50	990	4	36	990	4	40	990	4	46

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

Table 4.8. Raw Score to Scale Score Conversion Tables for Reading (Grades 6 to 8) Spring 2011

Raw Score	Grade 6			Grade 7			Grade 8		
	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	54	400	1	53	400	1	48
1	400	1	54	400	1	53	400	1	48
2	400	1	54	400	1	53	400	1	48
3	400	1	54	400	1	53	400	1	48
4	400	1	54	400	1	53	400	1	48
5	400	1	54	400	1	53	400	1	48
6	400	1	54	400	1	53	400	1	48
7	400	1	54	400	1	53	400	1	48
8	400	1	54	400	1	53	400	1	48
9	400	1	54	455	1	59	400	1	48
10	461	1	61	491	1	62	431	1	53
11	495	1	64	516	1	62	469	1	57
12	519	1	63	535	1	58	497	1	59
13	538	1	59	551	1	53	518	1	58
14	553	1	54	564	1	48	535	1	55
15	566	1	48	576	1	43	550	1	50
16	577	1	43	586	1	38	563	1	46
17	588	1	39	596	1	34	575	1	42
18	597	1	35	605	1	31	585	1	38
19	606	1	32	613	1	29	595	1	35
20	615	1	30	621	1	27	604	1	33
21	623	1	28	628	1	25	613	1	31
22	631	1	27	635	1	24	621	1	29
23	638	1	26	642	1	23	629	1	28
24	646	1	25	649	1	23	637	1	27
25	653	2	25	656	1	22	645	1	26
26	660	2	24	662	1	22	652	1	26
27	667	2	24	668	2	21	660	2	26
28	674	2	24	675	2	21	667	2	25
29	681	2	23	681	2	21	675	2	25
30	689	2	23	688	2	21	682	2	25
31	696	2	23	694	2	21	689	2	25
32	703	3	23	701	3	21	697	2	25
33	710	3	23	708	3	21	705	3	25
34	718	3	23	715	3	21	712	3	26

35	725	3	24	722	3	22	721	3	26
36	733	3	24	729	3	22	729	3	27
37	741	3	24	737	3	22	738	3	27
38	750	3	25	745	3	23	747	3	29
39	759	3	26	753	3	23	757	3	30
40	768	3	27	762	3	24	768	3	31
41	778	3	28	771	3	25	779	3	33
42	789	3	30	781	3	27	792	3	36
43	800	3	33	792	3	28	806	3	39
44	814	3	37	804	4	31	822	3	43
45	829	4	43	817	4	35	842	4	48
46	848	4	49	834	4	40	865	4	52
47	873	4	55	855	4	46	896	4	53
48	911	4	56	885	4	50	940	4	48
49	989	4	41	938	4	45	990	4	37
50	990	4	40	990	4	36	990	4	37

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

DRAFT

Table 4.9. Raw Score to Scale Score Conversion Tables for Science Spring 2011

Raw Score	Grade 5			Grade 8		
	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	74	400	1	87
1	400	1	74	400	1	87
2	400	1	74	400	1	87
3	400	1	74	400	1	87
4	400	1	74	400	1	87
5	400	1	74	400	1	87
6	400	1	74	400	1	87
7	400	1	74	400	1	87
8	476	1	78	400	1	87
9	530	1	81	430	1	89
10	560	1	79	533	1	95
11	583	1	72	575	1	94
12	600	1	64	601	1	88
13	615	1	55	621	1	78
14	628	1	48	637	1	67
15	639	2	41	651	2	56
16	650	2	36	663	2	47
17	660	2	33	674	2	40
18	669	2	30	684	2	35
19	677	2	28	694	2	32
20	686	2	27	702	3	29
21	694	2	26	711	3	27
22	701	3	25	719	3	26
23	709	3	24	727	3	25
24	717	3	24	734	3	24
25	724	3	23	742	3	23
26	731	3	23	749	3	23
27	739	3	23	757	3	23
28	746	3	23	764	3	22
29	754	3	23	771	3	22
30	761	3	23	778	3	22
31	769	3	23	786	3	22
32	777	3	23	793	3	22
33	785	3	23	801	3	22
34	794	3	24	809	3	23

35	803	3	24	817	3	23
36	812	3	25	826	3	25
37	822	4	26	835	4	26
38	833	4	28	845	4	28
39	845	4	30	857	4	31
40	858	4	32	870	4	35
41	874	4	35	886	4	39
42	893	4	38	907	4	42
43	919	4	39	936	4	41
44	962	4	32	990	4	29
45	990	4	26	990	4	29

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

DRAFT

Table 4.10. Raw Score to Scale Score Conversion Tables for Social Studies Spring 2011

Raw Score	Grade 5			Grade 7			Grade 8		
	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	74	400	1	59	400	1	81
1	400	1	74	400	1	59	400	1	81
2	400	1	74	400	1	59	400	1	81
3	400	1	74	400	1	59	400	1	81
4	400	1	74	400	1	59	400	1	81
5	400	1	74	400	1	59	400	1	81
6	400	1	74	400	1	59	400	1	81
7	400	1	74	400	1	59	400	1	81
8	400	1	74	413	1	60	400	1	81
9	400	1	74	474	1	68	400	1	81
10	400	1	74	510	1	71	467	1	84
11	400	1	74	537	1	70	535	1	88
12	400	1	74	559	1	67	569	1	86
13	482	1	80	577	1	62	592	1	80
14	525	1	82	594	1	56	610	1	71
15	552	1	81	609	2	51	625	2	61
16	572	1	76	623	2	47	638	2	53
17	588	1	70	636	2	43	650	2	45
18	602	1	62	649	2	41	661	2	39
19	614	1	55	662	2	39	671	2	34
20	625	1	48	673	2	37	680	2	31
21	635	1	42	685	2	36	688	2	29
22	644	1	38	696	2	35	697	2	27
23	653	2	34	707	3	34	704	3	25
24	660	2	31	718	3	33	712	3	24
25	668	2	29	728	3	33	719	3	23
26	675	2	27	738	3	32	726	3	22
27	682	2	26	749	3	31	732	3	22
28	689	2	25	759	3	31	739	3	21
29	695	2	24	769	3	31	746	3	21
30	702	3	23	779	3	30	752	3	21
31	708	3	22	789	3	30	759	3	20
32	714	3	22	800	3	30	766	3	21
33	720	3	21	810	3	31	772	3	21
34	726	3	21	822	3	31	779	3	21

35	731	3	20	833	3	32	787	3	22
36	737	3	20	845	3	33	795	3	23
37	743	3	20	858	4	34	804	3	24
38	748	3	20	872	4	36	813	3	26
39	754	3	19	887	4	37	824	4	29
40	760	3	19	904	4	39	836	4	33
41	765	3	19	924	4	39	851	4	37
42	771	3	19	950	4	36	870	4	42
43	777	3	19	988	4	27	897	4	45
44	783	3	19	990	4	27	944	4	40
45	789	4	19	990	4	27	990	4	31
46	795	4	19						
47	801	4	19						
48	807	4	19						
49	814	4	20						
50	821	4	21						
50	828	4	22						
50	836	4	23						
50	845	4	25						
50	855	4	27						
50	866	4	31						
50	880	4	35						
50	898	4	38						
50	924	4	39						
50	969	4	30						
50	990	4	26						

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 **Error! Reference source not found.** 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 a 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 Spring 2011 present the overall accuracy and consistency indices for all of the OCCT 3-8tests.

Table 5.1. Estimates of Accuracy and Consistency of Performance Classification

Subject	Grade	Accuracy	Consistency	Kappa(K)	False Positive	False Negative
MATH	3	0.74	0.64	0.48	0.14	0.13
	4	0.74	0.65	0.48	0.13	0.13
	5	0.73	0.64	0.47	0.14	0.13
	6	0.77	0.70	0.55	0.12	0.11
	7	0.77	0.69	0.54	0.12	0.11
	8	0.77	0.68	0.55	0.12	0.12
READING	3	0.83	0.77	0.55	0.10	0.08
	4	0.82	0.76	0.57	0.10	0.08
	5	0.79	0.73	0.52	0.13	0.07
	6	0.76	0.69	0.48	0.16	0.08
	7	0.71	0.65	0.43	0.21	0.07
	8	0.74	0.67	0.39	0.19	0.07
SCIENCE	5	0.76	0.68	0.43	0.11	0.13
	8	0.83	0.77	0.52	0.07	0.10
SOCIAL STUDIES	5	0.76	0.67	0.52	0.12	0.12
	7	0.77	0.69	0.44	0.13	0.10
	8	0.81	0.73	0.55	0.10	0.09

As shown in Table 5.1 the overall accuracy indices range between 73 and 83 percent and overall consistency ranging between 64 and 77 for the Spring 2011 OCCT administration. Kappa coefficients range from 0.43 and 0.55 in Spring 2011. The rate of false positives range from 7 to 21 and false negative rates range from 7 to 13 percent for the Spring 2011 administration.

Table 5.2 provides the accuracy, consistency, false positive, and false negative rates by cut score for Spring 2011. The data in these tables reveal that the level of agreement for both accuracy and consistency is above 80 percent in all cases, with most above 90 percent. In general, the high rates of accuracy and consistency support the cut decisions made using these assessments. Similar to Table 5.1, the false positive and false negative rates were comparable for the Spring 2011 administrations and are quite low.

The importance of the dichotomous categorization is particularly notable when they map onto proficient/not proficient decisions for the assessments. For the OCCT 3-8 tests, the U+L/P+A is the important dichotomization, because it directly translates to the proficient/not proficient decision point, which is important in computing Adequate Yearly Progress (AYP). 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 is proficient and should have been proficient); 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 5.2, for example, 93 percent of grade 3 Mathematics students were correctly classified as proficient or not proficient based on their performance (scenario 1), 3 percent were considered proficient but their true performance is below the standard (scenario 2), and 4 percent were not considered proficient although their true performance is above the standard (scenario 3). Overall, the rates for accurate classification are above 85% for the administrations for all subjects - students are appropriately (more than 85% of the time) categorized into proficient/not proficient classifications based on their true ability using their observed score (raw score) as their classification score.

Table 5.2. Accuracy and Consistency Estimates and False Positive/False Negative Rates by Cut Score

Subject	Grade	Accuracy			Consistency			False Positive			False Negative		
		U / (L+P+A)	(U+L) / (P+A)	(U+L+P) / A									
MATH	3	0.97	0.93	0.84	0.95	0.90	0.79	0.02	0.03	0.09	0.02	0.04	0.07
	4	0.96	0.93	0.86	0.94	0.90	0.81	0.02	0.03	0.08	0.02	0.04	0.06
	5	0.96	0.92	0.85	0.95	0.89	0.8	0.02	0.03	0.09	0.02	0.04	0.06
	6	0.94	0.91	0.92	0.91	0.88	0.9	0.03	0.04	0.05	0.04	0.05	0.03
	7	0.93	0.91	0.92	0.9	0.87	0.89	0.03	0.04	0.05	0.04	0.05	0.03
	8	0.95	0.91	0.91	0.93	0.88	0.87	0.02	0.04	0.05	0.03	0.05	0.04
READING	3	0.96	0.92	0.96	0.94	0.89	0.95	0.02	0.03	0.04	0.02	0.05	0.00
	4	0.95	0.91	0.96	0.93	0.88	0.96	0.02	0.04	0.04	0.03	0.05	0.00
	5	0.97	0.92	0.91	0.95	0.89	0.9	0.02	0.03	0.09	0.02	0.05	0.00
	6	0.95	0.91	0.91	0.92	0.87	0.89	0.02	0.04	0.09	0.03	0.05	0.00
	7	0.95	0.93	0.84	0.93	0.9	0.81	0.02	0.03	0.16	0.03	0.04	0.00
	8	0.96	0.93	0.86	0.95	0.9	0.82	0.02	0.03	0.14	0.02	0.04	0.00
SCIENCE	5	0.98	0.96	0.83	0.97	0.94	0.77	0.01	0.02	0.08	0.01	0.02	0.09
	8	0.98	0.95	0.91	0.96	0.93	0.88	0.01	0.02	0.04	0.01	0.02	0.05
SOCIAL STUDIES	5	0.94	0.91	0.91	0.92	0.87	0.87	0.02	0.04	0.05	0.03	0.05	0.04
	7	0.98	0.93	0.87	0.96	0.9	0.83	0.01	0.03	0.09	0.01	0.04	0.04
	8	0.96	0.92	0.93	0.94	0.89	0.91	0.02	0.04	0.05	0.02	0.04	0.02

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 Spring 2011 appears in Table 6.1 through **Error! Reference source not found.**. The scales scores presented exclude invalid student cases and second-time testers.

Table 6.1. Descriptive Statistics of the Scale Scores for Spring 2011 - Overall

Subject/Grade	Scale Score				
	N	Mean	SD	Med.	
MATH	3	45290	736	115	745
	4	44562	741	116	744
	5	44112	728	120	740
	6	43339	722	109	733
	7	42850	721	118	731
	8	41332	714	114	726
READ	3	44692	730	116	737
	4	43948	715	110	721
	5	43751	716	120	723
	6	43107	716	113	725
	7	42813	722	112	737
	8	41166	744	113	757
SCIENCE	5	44754	769	120	785
	8	42512	760	127	778
SOCIAL STUDIES	5	48087	722	113	737
	7	46560	766	120	779
	8	45365	725	125	739

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

Table 6.2. Descriptive Statistics of the Scale Scores for Spring 2011 by Gender

Subject	Grade	Female				Male			
		N	Mean	SD	Med.	N	Mean	SD	Med.
MATH	3	22691	733	111	737	22591	739	119	745
	4	22076	738	112	744	22466	743	120	752
	5	21942	727	112	732	22160	728	127	740
	6	21575	723	102	733	21752	723	116	733
	7	21296	721	113	731	21546	720	123	731
	8	20968	714	111	726	20363	714	118	726
READING	3	22544	736	111	746	22140	724	120	737
	4	21918	719	105	728	22009	711	114	721
	5	21892	723	112	732	21850	710	127	723
	6	21564	722	107	725	21530	710	118	718
	7	21317	729	107	737	21487	715	116	729
	8	20942	753	111	757	20221	735	115	747
SCIENCE	5	22275	769	111	777	22470	769	129	785
	8	21428	761	118	771	21082	759	136	778
SOCIAL STUDIES	5	23466	719	105	731	24616	725	121	743
	7	22656	761	112	769	23904	771	127	789
	8	22439	722	116	732	22925	728	132	746

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

Table 6.3. Descriptive Statistics of the Scale Scores for Spring 2011 by Race/Ethnicity

Subject	Grade	African American				Native American				Hispanic				Asian			
		N	Mean	SD	Med.	N	Mean	SD	Med.	N	Mean	SD	Med.	N	Mean	SD	Med.
MATH	3	4403	692	104	698	7422	733	102	737	6340	716	99	721	886	781	109	786
	4	4128	698	103	705	7438	735	104	744	6005	724	93	728	902	788	103	793
	5	4222	697	106	710	7491	721	106	724	5695	719	96	724	858	784	94	787
	6	4179	692	103	699	7433	714	100	720	5329	706	94	713	855	769	102	779
	7	4208	691	116	704	7336	714	106	725	5081	699	113	711	823	786	101	786
	8	3918	677	120	693	6969	706	105	719	4816	693	105	706	866	774	99	767
READING	3	4344	699	101	705	7328	729	97	737	6261	703	109	713	876	735	160	755
	4	4067	690	88	692	7323	712	93	721	5891	694	101	699	896	720	154	745
	5	4195	690	98	694	7439	714	102	723	5602	694	107	701	856	727	154	751
	6	4154	688	101	696	7394	710	103	718	5265	690	103	696	852	729	152	750
	7	4207	696	103	708	7327	721	95	729	5057	695	119	708	827	741	143	762
	8	3924	708	115	721	6978	739	107	747	4759	711	114	721	849	760	139	768
SCIENCE	5	4338	734	102	746	7611	767	108	777	5766	750	96	754	864	802	90	812
	8	4077	730	114	742	7176	754	120	771	4911	742	102	749	876	794	103	801
SOCIAL STUDIES	5	4815	686	103	695	8274	720	106	731	6219	702	97	714	883	760	85	765
	7	4780	713	136	728	8123	761	112	769	5510	737	120	749	827	810	116	822
	8	4473	688	122	704	7830	718	116	732	5169	702	110	712	890	766	105	772

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

Table 6.3. Descriptive Statistics of the Scale Scores for Spring 2011 by Race/Ethnicity (cont.)

Subject	Grade	Pacific Islander				White				Other			
		N	Mean	SD	Med.	N	Mean	SD	Med.	N	Mean	SD	Med.
MATH	3	102	708	86	710	23910	755	101	754	1983	738	113	745
	4	100	743	98	735	23822	759	103	761	1922	741	108	744
	5	89	694	139	710	23582	745	101	748	1823	727	114	732
	6	97	711	110	726	23492	738	97	745	1753	724	110	739
	7	90	693	127	704	23455	735	107	745	1657	723	103	731
	8	80	707	97	719	22994	726	114	739	1663	711	114	719
READING	3	101	709	149	721	23579	750	98	755	1943	737	108	746
	4	99	708	106	713	23527	733	92	737	1880	721	93	728
	5	89	674	183	701	23403	736	97	741	1801	724	105	732
	6	97	682	165	718	23409	733	101	741	1742	721	107	725
	7	90	674	150	701	23457	736	99	745	1660	726	95	737
	8	77	714	114	721	22897	760	107	768	1656	743	113	747
SCIENCE	5	90	731	155	761	23889	788	99	794	1850	775	104	777
	8	80	742	82	749	23418	776	114	786	1714	759	124	771
SOCIAL STUDIES	5	94	693	149	714	25546	740	99	748	1991	722	110	731
	7	99	745	128	759	25428	783	114	789	1783	767	116	779
	8	83	721	87	726	24903	742	116	752	1826	724	125	739

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

Table 6.4. Descriptive Statistics of the Scale Scores for Spring 2011 by Free/Reduced Lunch Status

Subject	Grade	Free/Reduced Lunch = No				Free/Reduced Lunch = Yes			
		N	Mean	SD	Median	N	Mean	SD	Median
MATH	3	17102	764	114	764	28188	719	113	728
	4	17515	771	111	771	27047	721	115	728
	5	17479	754	117	766	26633	710	119	724
	6	18668	747	106	759	24671	704	108	713
	7	18976	743	122	751	23874	703	112	718
	8	19120	739	106	746	22212	692	117	706
READING	3	16843	759	113	764	27849	712	114	721
	4	17278	742	104	745	26670	698	110	706
	5	17373	744	115	751	26378	698	120	708
	6	18578	742	110	750	24529	696	111	710
	7	18990	741	116	753	23823	706	106	715
	8	19026	771	102	779	22140	722	117	729
SCIENCE	5	17838	794	117	803	26916	752	119	769
	8	19633	779	127	793	22879	743	125	757
SOCIAL STUDIES	5	19318	747	110	760	28769	706	113	720
	7	19861	798	111	810	26699	742	121	759
	8	21103	747	124	759	24262	706	122	726

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 Spring 2011 administration are presented in Table 6.5 (please see Appendix B). As above, these percentages exclude invalid student data. The percentage distributions for each of the content areas are comparable to previous administrations (e.g., Spring 2010).

Table 6.5. Percentage of Students by Performance Level for Spring 2011

Subject/Grade	N	Unsatisfactory	Limited Knowledge	Proficient	Advanced	
MATH	3	44825	8.9%	21.3%	43.4%	26.5%
	4	44107	10.4%	17.9%	45.8%	25.9%
	5	43493	9.6%	21.0%	45.5%	23.9%
	6	42828	17.9%	15.4%	49.2%	17.5%
	7	42269	19.4%	13.0%	48.7%	19.0%
	8	40808	12.2%	21.4%	40.9%	25.4%
	All	258330	13.0%	18.3%	45.6%	23.1%
READ	3	44126	12.2%	17.9%	65.7%	4.2%
	4	43372	16.0%	20.6%	59.4%	3.9%
	5	43041	11.3%	21.4%	58.4%	8.9%
	6	42539	14.9%	20.5%	55.2%	9.4%
	7	42192	15.9%	13.0%	54.6%	16.5%
	8	40699	10.1%	13.3%	62.0%	14.6%
	All	255969	13.4%	17.8%	59.2%	9.5%
SCIENCE	5	44048	2.2%	9.0%	58.7%	30.0%
	8	41656	2.6%	7.9%	71.4%	18.1%
	All	85704	2.4%	8.5%	64.9%	24.2%
SOCIAL STUDIES	5	47453	12.5%	17.6%	46.5%	23.4%
	7	46039	3.2%	16.2%	62.2%	18.4%
	8	44629	6.7%	22.3%	57.3%	13.8%
	All	138121	7.5%	18.6%	55.2%	18.6%

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:

$$CSEM(O_x | \theta) = \sqrt{\left[\sum_{X=0}^{MaxX} O_x^2 p(X | \theta) \right] - \left[\sum_{X=0}^{MaxX} O_x \cdot p(X | \theta) \right]^2} \quad (13)$$

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 $CSEM(O_x | \theta)$ in which $p(X | \theta)$ 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 Tables 6.6 to 6.7 for the Spring 2011 administration of the OCCT.

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:

$$SEM = \sqrt{\frac{\sum_j (CSEM_j^2 \cdot N_j)}{N_T}} \quad (14)$$

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 6.8. presents the overall estimates of *SEM* for each of the content areas for the Spring 2011 administration.

Table 6.9. Overall Estimates of SEM by Test

Subject	Grade	SEM in OPI Units
MATHEMATICS	03	32
	04	32
	05	30
	06	28
	07	33
	08	30
	READING	03
04		29
05		33
06		30
07		27
08		35
SCIENCE		05
	08	29
SOCIAL STUDIES	05	30
	07	35
	08	35

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

DRAFT

Table 6.10. OCCT Test Blueprint and Actual Item Counts: Grade 3 Mathematics

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field- Tested in 2011
Algebraic Reasoning: Patterns and Relationships	7	7	13
Algebra Patterns (1.1)	2	2	0
Equations (1.2)	2	2	4
Number Properties (1.3)	3	3	9
Number Sense and Operation	20	20	27
Number Sense (2.1)	10	10	12
Number Operations (2.2)	10	10	15
Geometry	7	7	11
Properties of shapes (3.1)	3	3	6
Spatial Reasoning (3.2)	2	2	4
Coordinate Geometry (3.3)	2	2	1
Measurement	9	9	17
Measurement (4.1)	4	4	8
Time and Temperature (4.2)	2	2	3
Money (4.3)	3	3	6
Data Analysis	7	7	12
Data Analysis (5.1)	4	4	7
Probability (5.2)	3	3	5
Total Test	50	50	80

Table 6.11. OCCT Test Blueprint and Actual Item Counts: Grade 4 Mathematics

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field- Tested in 2011
Algebraic Reasoning: Patterns and Relationships	7	7	9
Algebra Patterns (1.1)	3	3	0
Equations (1.2)	2	2	5
Number Properties (1.3)	2	2	4
Number Sense and Operation	18	18	23
Number Sense (2.1)	8	8	13
Number Operations (2.2)	10	10	10
Geometry	9	9	18
Lines (3.1)	2	2	5
Angles (3.2)	2	2	2
Polygons (3.3)	3	3	7
Transformations (3.4)	2	2	4
Measurement	9	9	21
Measurement (4.1)	5	5	12
Time and Temperature (4.2)	2	2	5
Money (4.3)	2	2	4
Data Analysis	7	7	9
Data Analysis (5.1)	2	2	0
Probability (5.2)	2	2	4
Central Tendency (5.3)	3	3	5
Total Test	50	50	80

Table 6.12. OCCT Test Blueprint and Actual Item Counts: Grade 5 Mathematics

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field- Tested in 2011
Algebraic Reasoning: Patterns and Relationships	13	13	18
Algebra Patterns (1.1)	5	5	5
Equations (1.2)	4	4	3
Number Properties (1.3)	4	4	10
Number Sense and Operation	16	16	26
Number Sense (2.1)	8	8	13
Number Operations (2.2)	8	8	13
Geometry	7	7	17
Circles and Polygons (3.1)	4	4	10
Angles (3.2)	3	3	7
Measurement	7	7	12
Measurement (4.1)	5	5	10
Money (4.2)	2	2	2
Data Analysis	7	7	7
Data Analysis (5.1)	3	3	0
Probability (5.2)	2	2	1
Central Tendency (5.3)	2	2	6
Total Test	50	50	80

Table 6.13. OCCT Test Blueprint and Actual Item Counts: Grade 6 Mathematics

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field- Tested in 2011
Algebraic Reasoning: Patterns and Relationships	13	13	21
Algebra Patterns (1.1)	4	4	6
Expressions and Equations (1.2)	4	4	10
Number Properties (1.3)	3	3	1
Solving Equations (1.4)	2	2	4
Number Sense and Operation	15	15	27
Number Sense (2.1)	5	5	12
Number Operations (2.2)	10	10	15
Geometry	8	8	10
Three Dimensional Figures (3.1)	2	2	4
Congruent and Similar Figures (3.2)	2	2	2
Coordinate Geometry (3.3)	4	4	4
Measurement	7	7	18
Circles (4.1)	4	2	12
Conversions (4.2)	3	5	6
Data Analysis	7	7	4
Data Analysis (5.1)	3	3	1
Probability (5.2)	2	2	2
Central Tendency (5.3)	2	2	1
Total Test	50	50	80

Table 6.14. OCCT Test Blueprint and Actual Item Counts: Grade 7 Mathematics

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field- Tested in 2011
Algebraic Reasoning: Patterns and Relationships	15	15	32
Linear Relationships (1.1)	5	5	10
Solving Equations (1.2)	5	5	11
Solving and Graphing Inequalities (1.3)	5	5	11
Number Sense and Operation	11	11	10
Number Sense (2.1)	5	5	4
Number Operations (2.2)	6	6	6
Geometry	8	8	5
Classifying Figures (3.1)	2-3	3	1
Lines and Angles (3.2)	2-3	3	2
Transformations (3.3)	2-3	2	2
Measurement	9	9	20
Perimeter and Area (4.1)	5	5	9
Circles (4.2)	2	2	8
Composite Figures (4.3)	2	2	3
Data Analysis	7	7	13
Data Analysis (5.1)	2	2	5
Probability (5.2)	2	2	1
Central Tendency (5.3)	3	3	7
Total Test	50	50	80

Table 6.15. OCCT Test Blueprint and Actual Item Counts: Grade 8 Mathematics

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field- Tested in 2011
Algebraic Reasoning: Patterns and Relationships	16	16	35
Equations (1.1)	10-12	11	26
Inequalities (1.2)	4-6	5	9
Number Sense and Operation	11	11	23
Number Sense (2.1)	3-4	4	6
Number Operations (2.2)	7-8	7	17
Geometry	9	9	10
Three Dimensional Figures (3.1)	5	5	7
Pythagorean Theorem (3.2)	4	4	3
Measurement	7	7	2
Surface Area and Volume (4.1)	3	3	0
Ratio and Proportions (4.2)	2	2	0
Composite Figures (4.3)	2	2	2
Data Analysis	7	7	10
Data Analysis (5.1)	3	3	0
Central Tendency (5.3)	4	4	10
Total Test	50	50	80

Table 6.16. OCCT Test Blueprint and Actual Item Counts: Grade 3 Reading

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Vocabulary	12	12	17
Words in Context (2.1)	2-4	3	5
Affixes, Roots, and Stems (2.2)	2-4	3	5
Synonyms, Antonyms, and Homonyms (2.3)	2-4	2	2
Using Resource Materials (2.4)	2-4	4	5
Comprehension/Critical Literacy	24	24	46
Literal Understanding (4.1)	5	5	7
Inferences and Interpretation (4.2)	7	8	14
Summary and Generalization (4.3)	6	5	14
Analysis and Evaluation (4.4)	6	6	11
Literature	8	8	9
Literary Elements (5.2)	4	4	3
Figurative Language/Sound Devices (5.3)	4	4	6
Research and Information	6	6	8
Accessing Information (6.1)	6	6	8
Total	50	50	80

Table 6.17. OCCT Test Blueprint and Actual Item Counts: Grade 4 Reading

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Vocabulary	12	12	16
Words in Context (1.1)	4	3	3
Affixes, Roots, and Stems (1.2)	4	5	6
Synonyms, Antonyms, and Homonyms (1.3)	4	4	7
Comprehension/Critical Literacy	23	23	41
Literal Understanding (3.1)	4	4	9
Inferences and Interpretation (3.2)	6	6	9
Summary and Generalization (3.3)	7	6	14
Analysis and Evaluation (3.4)	6	7	9
Literature	9	9	15
Literary Elements (4.2)	5	5	8
Figurative Language/Sound Devices (4.3)	4	4	7
Research and Information	6	6	8
Accessing Information (5.1)	6	6	8
Total Test	50	50	80

Table 6.18. OCCT Test Blueprint and Actual Item Counts: Grade 5 Reading

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Vocabulary	12	12	20
Words in Context (1.1)	4	4	5
Affixes, Roots, and Stems (1.2)	4	4	7
Synonyms, Antonyms, and Homonyms (1.3)	4	4	8
Comprehension/Critical Literacy	20	20	33
Literal Understanding (3.1)	4	4	7
Inferences and Interpretation (3.2)	4-6	6	11
Summary and Generalization (3.3)	4-6	5	7
Analysis and Evaluation (3.4)	4-6	5	8
Literature	12	12	19
Literary Genre (4.1)	4	4	5
Literary Elements (4.2)	4	4	6
Figurative Language/Sound Devices (4.3)	4	4	8
Research and Information	6	6	8
Accessing Information (5.1)	2-4	4	5
Interpreting Information (5.2)	2-4	2	3
Total	50	50	80

Table 6.19. OCCT Test Blueprint and Actual Item Counts: Grade 6 Reading

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Vocabulary	8	8	13
Words in Context (1.1)	4	4	6
Word Origins (1.2)	4	4	7
Comprehension/Critical Literacy	20	20	33
Literal Understanding (3.1)	4	4	11
Inferences and Interpretation (3.2)	4-6	6	5
Summary and Generalization (3.3)	4-6	6	9
Analysis and Evaluation (3.4)	4-6	4	8
Literature	14	14	22
Literary Genres (4.1)	4	4	8
Literary Elements (4.2)	4-6	6	5
Figurative Language/Sound Devices (4.3)	4-6	4	9
Research and Information	8	8	12
Accessing Information (5.1)	4	4	3
Interpreting Information (5.2)	4	4	9
Total	50	50	80

Table 6.20. OCCT Test Blueprint and Actual Item Counts: Grade 7 Reading

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Vocabulary	10	10	17
Words in Context (1.1)	3-4	4	7
Word Origins (1.2)	3-4	3	8
Idioms and Comparisons (1.3)	3-4	3	2
Comprehension/Critical Literacy	20	20	39
Literal Understanding (3.1)	4	4	10
Inferences and Interpretation (3.2)	4-6	5	8
Summary and Generalization (3.3)	4-6	5	12
Analysis and Evaluation (3.4)	4-6	6	9
Literature	12	12	14
Literary Genres (4.1)	4	4	6
Literary Elements (4.2)	4	4	5
Figurative Language/Sound Devices (4.3)	4	4	3
Research and Information	8	8	10
Accessing Information (5.1)	4	4	4
Interpreting Information (5.2)	4	4	6
Total	50	50	80

Table 6.21. OCCT Test Blueprint and Actual Item Counts: Grade 8 Reading

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Vocabulary	6	6	10
Words in Context (1.1)	2	2	4
Word Origins (1.2)	2	2	3
Idioms and Comparisons (1.3)	2	2	3
Comprehension/Critical Literacy	21	21	35
Literal Understanding (3.1)	4	4	9
Inferences and Interpretation (3.2)	4-6	5	7
Summary and Generalization (3.3)	5-7	5	8
Analysis and Evaluation (3.4)	6-8	7	11
Literature	15	15	24
Literary Genre (4.1)	4	4	7
Literary Elements (4.2)	5-7	6	5
Figurative Language/Sound Devices (4.3)	4-6	5	12
Research and Information	8	8	11
Accessing Information (5.1)	4	4	5
Interpreting Information (5.2)	4	4	6
Total	50	50	80

Table 6.22. OCCT Test Blueprint and Actual Item Counts: Grade 5 Science

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Process Standards			
Observe and Measure	10	10	13
SI Metric (P1.1)	5	5	5
Similar/different characteristics (P1.2)	5	5	8
Classify	10	10	24
Observable properties (P2.1)	5	5	12
Serial order (P2.2)	5	5	12
Experiment	11	11	14
Experimental design (P3.2)	7	7	9
Hazards/practice safety (P3.4)	4	4	5
Interpret and Communicate	14	14	29
Data tables/line/bar/trend and circle graphs (P4.2)	6	6	12
Prediction based on data (P4.3)	4	4	8
Explanations based on data (P4.4)	4	4	9
Total Test	45	45	80
Content Standards			
Properties of Matter and Energy	18	18	31
Matter has physical properties (1.1)	6	6	11
Physical properties can be measured (1.2)	6	6	8
Energy can be transferred (1.3)	6	6	12
Organisms and Environments	12	12	21
Dependence upon community (2.1)	6	6	12
Individual organism and species survival (2.2)	6	6	9
Structures of the Earth and the Solar System	11	11	23
Weather patterns (3.2)	6	6	13
Earth as a planet (3.3)	5	5	10
Total Test	41	41	75

* Items from the Safety Objective (P3.4) are not dual aligned to a content standard

Table 6.23. OCCT Test Blueprint and Actual Item Counts: Grade 8 Science

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Process Standards			
Observe and Measure	8	8	15
Qualitative/quantitative observations/changes (P1.1)	4	4	9
SI (metrics) units/appropriate tools (P1.2 and P1.3)	4	4	6
Classify	8	8	18
Classification system (P2.1)	4	4	7
Properties ordered (P2.2)	4	4	11
Experiment	16	16	26
Experimental design (P3.2)	6	6	12
Identify variables (P3.3)	6	6	8
Hazards/practice safety (P3.6)	4	4	6
Interpret and Communicate	13	13	21
Data tables/line/bar/trend and circle graphs (P4.2)	7	7	10
Explanations/prediction (P4.3)	6	6	11
Total Test	45	45	80
Content Standards			
Properties and Chemical Changes in Matter	7-8	8	9
Chemical reactions (1.1)	3-4	4	5
Conservation of matter (1.2)	3-4	4	4
Motion and Forces	8	8	19
Motion of an object (2.1)	4	4	10
Object subjected to a force (2.2)	4	4	9
Diversity and Adaptations of Organisms	9	9	14
Classification (3.1)	5	5	9
Internal and external structures (3.2)	4	4	5
Structures/Forces of the Earth/Solar System	8	8	17
Landforms result from constructive and destructive forces (4.1)	4	4	9
Rock cycle (4.2)	4	4	8
Earth's History	7-8	8	15
Catastrophic events (5.1)	3-4	4	7
Fossil evidence (5.2)	3-4	4	8
Total Test	41	41	74

* Items from the Safety Objective (P3.4) are not dual aligned to a content standard

Table 6.24. OCCT Test Blueprint and Actual Item Counts: Grade 5 Social Studies

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Early Exploration	8	8	10
Expeditions (2.1)	4	4	5
Native American Reaction (2.2)	4	4	5
Colonial America	12	12	13
Settlements and Migration (3.1)	4	4	4
Colonial Life (3.2)	4	4	3
Individuals and Groups (3.3)	4	4	6
American Revolution	12	12	16
Causes and Results (4.1)	4	4	4
Declaration of Independence (4.3)	4	4	7
Individuals (4.4)	4	4	5
Early Federal Period	8	8	13
Constitutional Provisions (5.2)	4	4	7
Ratification and Rights (5.3)	4	4	6
Geographic Skills	20	20	28
Maps/Charts/Graphs Usage (7.1)	7	7	10
Human/Environment Interaction (7.2)	5	5	5
Historical Places (7.3)	4	4	7
Westward Movement (7.4)	4	4	6
Total Test	60	60	80

Table 6.25. OCCT Test Blueprint and Actual Item Counts: Grade 7 Social Studies (Geography)

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Geographic Tools/Geography Skills	9	9	18
Map Concepts (1.2)	4	4	3
Maps/Charts/Graphs (6.1)	5	5	15
Regions	12	12	20
Regional Characteristics (2.1)	4	4	6
Conflict/Cooperation (2.2)	4	4	4
Locations (2.4)	4	4	10
Physical Systems	8	8	11
Climate/Weather (3.2)	4	4	7
Natural Disasters (3.3)	4	4	4
Human Systems	8	8	14
World Cultures (4.1)	4	4	8
Population Issues (4.5)	4	4	6
Human/Environment Interaction	8	8	17
Natural Resources (5.1)	4	4	6
Human Modification (5.2)	4	4	11
Total Test	45	45	80

Table 6.26. OCCT Test Blueprint and Actual Item Counts: Grade 8 Social Studies (U.S. History)

Pass Standard and Objective	Ideal Number of Items for Alignment to PASS*	Actual Number of Items on 2011 Test	Number of Item Field-Tested in 2011
Social Studies Process Skills (1.0)	6	6	15
Causes and Results of the American Revolution (3.0/4.0)	10	10	23
Causes of the American Revolution (3.0)	5	5	12
Results of the American Revolution (4.0)	5	5	11
Governing Documents/Early Federal Period (5.0)	6	6	13
Northern/Southern Economic Growth (6.0)	4	4	4
Jacksonian Era (7.0)	4	4	3
Cultural Growth and Reform (8.0)	4	4	7
Westward Movement (9.0)	6	6	9
Eve of War (10.0)	5	5	6
Total Test	45	45	80

Appendix B

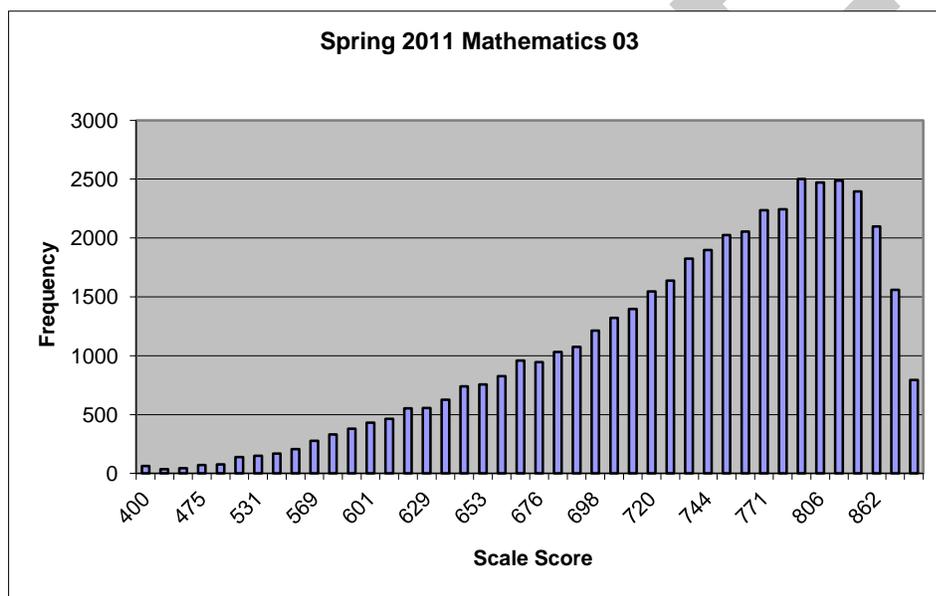
Scale Score Distributions for Spring 2011

DRAFT

Mathematics Grade 3 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	64	0.14	413	0.92
434	35	0.08	448	1
464	44	0.1	492	1.09
486	72	0.16	564	1.25
505	77	0.17	641	1.43
521	139	0.31	780	1.73
535	151	0.34	931	2.07
548	169	0.38	1100	2.45
559	208	0.46	1308	2.91
570	276	0.61	1584	3.52
581	330	0.73	1914	4.26
590	380	0.85	2294	5.1
600	430	0.96	2724	6.06
609	464	1.03	3188	7.09
617	554	1.23	3742	8.32
625	556	1.24	4298	9.56
633	627	1.39	4925	10.95
641	741	1.65	5666	12.6
648	756	1.68	6422	14.28
656	827	1.84	7249	16.12
663	959	2.13	8208	18.25
670	946	2.1	9154	20.36
677	1032	2.3	10186	22.65
684	1075	2.39	11261	25.04
691	1212	2.7	12473	27.74
698	1322	2.94	13795	30.68
706	1397	3.11	15192	33.78
713	1547	3.44	16739	37.23
721	1639	3.64	18378	40.87
728	1825	4.06	20203	44.93
737	1898	4.22	22101	49.15
745	2026	4.51	24127	53.65
754	2054	4.57	26181	58.22
764	2237	4.97	28418	63.2

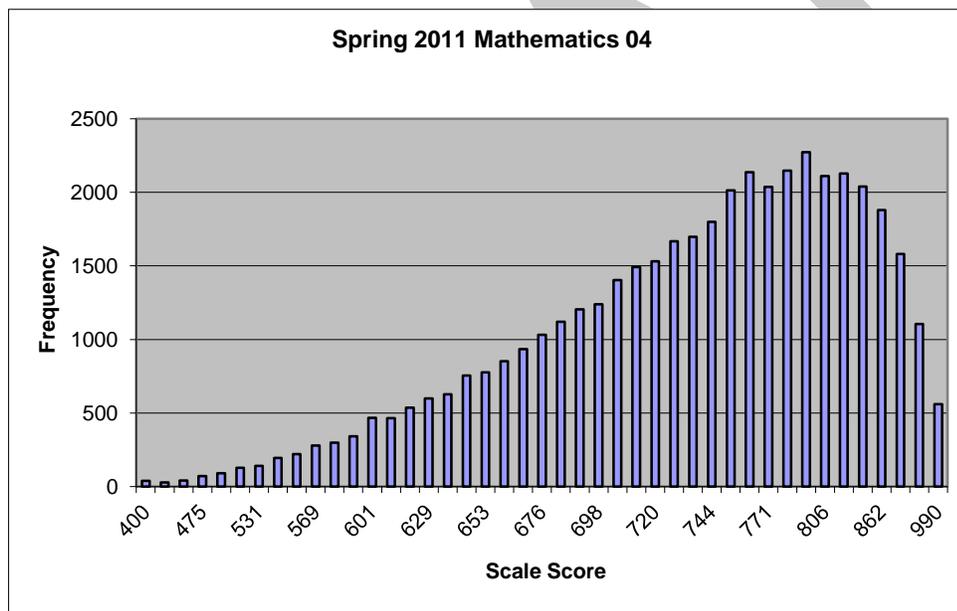
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
775	2243	4.99	30661	68.19
786	2500	5.56	33161	73.75
800	2471	5.5	35632	79.24
816	2487	5.53	38119	84.77
836	2396	5.33	40515	90.1
865	2099	4.67	42614	94.77
915	1559	3.47	44173	98.23
990	794	1.77	44967	100



Mathematics Grade 4 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	40	0.09	339	0.76
406	28	0.06	367	0.83
447	41	0.09	408	0.92
475	72	0.16	480	1.08
497	91	0.21	571	1.29
515	127	0.29	698	1.57
531	141	0.32	839	1.89
545	194	0.44	1033	2.33
557	220	0.5	1253	2.82
569	279	0.63	1532	3.45
580	299	0.67	1831	4.13
591	342	0.77	2173	4.9
601	468	1.05	2641	5.95
610	464	1.05	3105	7
620	537	1.21	3642	8.21
629	599	1.35	4241	9.56
637	628	1.42	4869	10.97
645	755	1.7	5624	12.67
653	777	1.75	6401	14.42
661	851	1.92	7252	16.34
669	934	2.1	8186	18.44
676	1031	2.32	9217	20.77
684	1120	2.52	10337	23.29
691	1205	2.72	11542	26.01
698	1240	2.79	12782	28.8
705	1404	3.16	14186	31.96
713	1492	3.36	15678	35.33
720	1530	3.45	17208	38.77
728	1668	3.76	18876	42.53
735	1698	3.83	20574	46.36
744	1800	4.06	22374	50.41
752	2013	4.54	24387	54.95
761	2136	4.81	26523	59.76
771	2037	4.59	28560	64.35

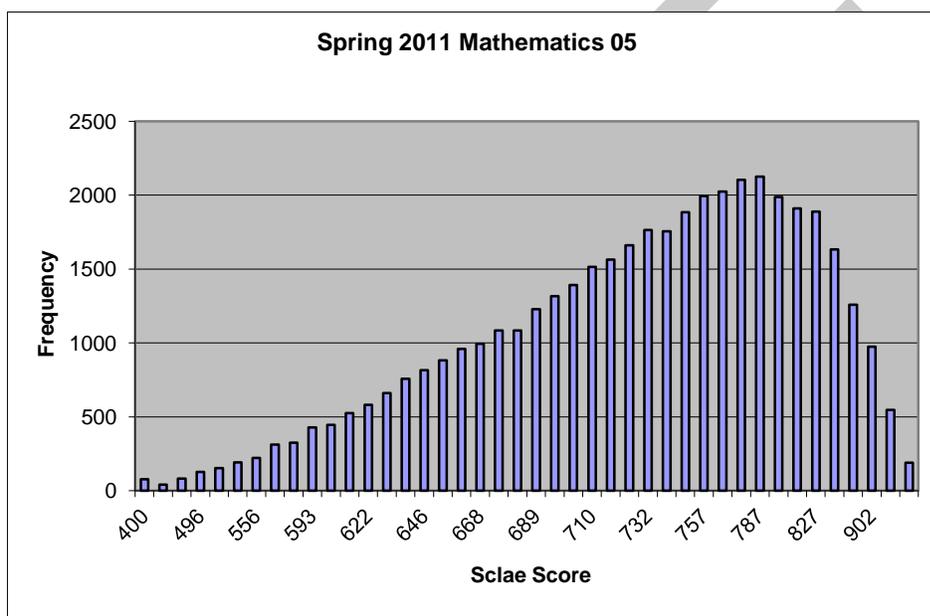
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
781	2147	4.84	30707	69.19
793	2273	5.12	32980	74.31
806	2110	4.75	35090	79.07
821	2128	4.79	37218	83.86
839	2040	4.6	39258	88.46
862	1880	4.24	41138	92.69
893	1580	3.56	42718	96.25
948	1104	2.49	43822	98.74
990	559	1.26	44381	100



Mathematics Grade 5 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	79	0.18	506	1.15
411	42	0.1	548	1.25
464	83	0.19	631	1.44
496	128	0.29	759	1.73
520	154	0.35	913	2.08
539	193	0.44	1106	2.52
556	222	0.51	1328	3.03
570	312	0.71	1640	3.74
582	326	0.74	1966	4.48
593	428	0.97	2394	5.45
604	445	1.01	2839	6.47
613	525	1.2	3364	7.66
622	581	1.32	3945	8.99
630	661	1.51	4606	10.49
638	758	1.73	5364	12.22
646	815	1.86	6179	14.08
654	882	2.01	7061	16.08
661	959	2.18	8020	18.27
668	995	2.27	9015	20.54
675	1085	2.47	10100	23.01
682	1085	2.47	11185	25.48
689	1229	2.8	12414	28.28
696	1316	3	13730	31.28
703	1392	3.17	15122	34.45
710	1514	3.45	16636	37.9
717	1564	3.56	18200	41.46
724	1661	3.78	19861	45.24
732	1764	4.02	21625	49.26
740	1755	4	23380	53.26
748	1884	4.29	25264	57.55
757	1994	4.54	27258	62.09
766	2025	4.61	29283	66.71
776	2104	4.79	31387	71.5
787	2125	4.84	33512	76.34
799	1987	4.53	35499	80.87

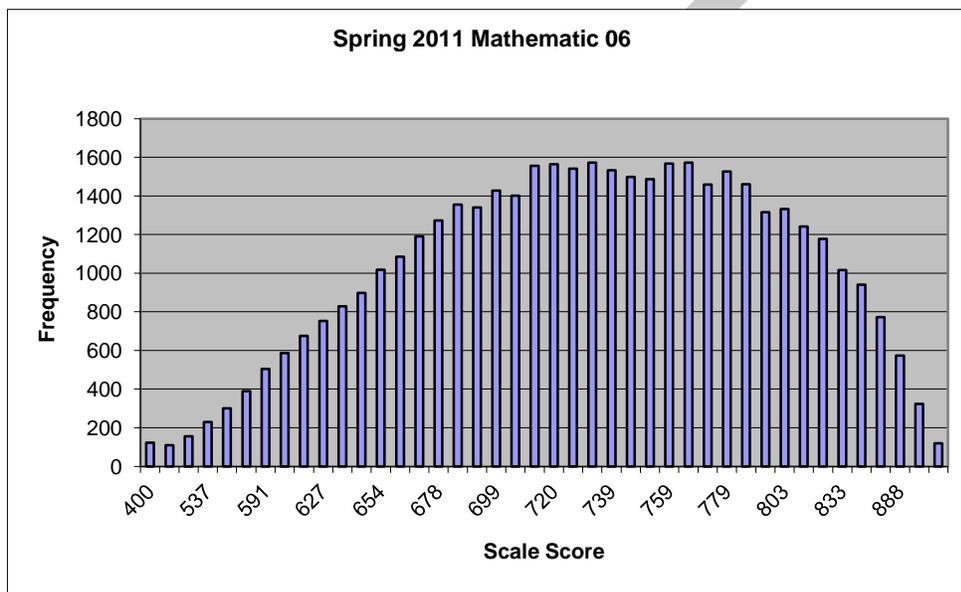
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
812	1910	4.35	37409	85.22
827	1888	4.3	39297	89.52
845	1632	3.72	40929	93.23
869	1259	2.87	42188	96.1
902	974	2.22	43162	98.32
961	547	1.25	43709	99.57
990	190	0.43	43899	100



Mathematics Grade 6 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	122	0.28	553	1.28
470	109	0.25	662	1.53
510	156	0.36	818	1.89
537	230	0.53	1048	2.42
559	301	0.7	1349	3.12
576	389	0.9	1738	4.02
591	505	1.17	2243	5.19
604	586	1.36	2829	6.54
616	675	1.56	3504	8.11
627	753	1.74	4257	9.85
636	829	1.92	5086	11.77
645	898	2.08	5984	13.84
654	1018	2.36	7002	16.2
662	1086	2.51	8088	18.71
670	1191	2.76	9279	21.47
678	1273	2.94	10552	24.41
685	1356	3.14	11908	27.55
692	1341	3.1	13249	30.65
699	1427	3.3	14676	33.95
706	1401	3.24	16077	37.19
713	1556	3.6	17633	40.79
720	1564	3.62	19197	44.41
726	1541	3.56	20738	47.98
733	1573	3.64	22311	51.61
739	1533	3.55	23844	55.16
745	1498	3.47	25342	58.63
752	1487	3.44	26829	62.07
759	1568	3.63	28397	65.69
765	1573	3.64	29970	69.33
772	1459	3.38	31429	72.71
779	1526	3.53	32955	76.24
786	1460	3.38	34415	79.62
794	1315	3.04	35730	82.66
803	1333	3.08	37063	85.74
812	1242	2.87	38305	88.62

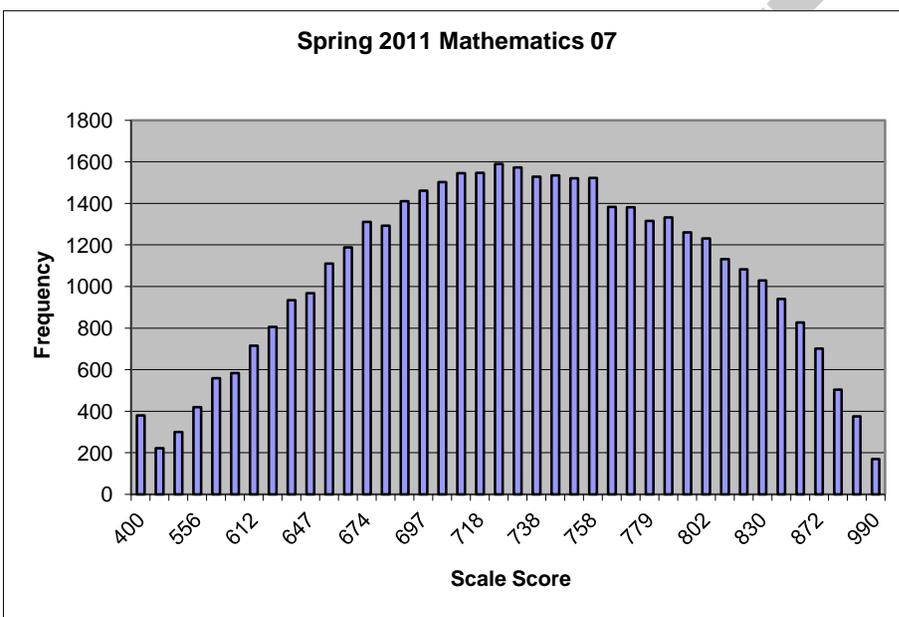
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
821	1177	2.72	39482	91.34
833	1016	2.35	40498	93.69
846	941	2.18	41439	95.87
863	772	1.79	42211	97.65
888	573	1.33	42784	98.98
931	323	0.75	43107	99.72
990	119	0.28	43226	100



Mathematics Grade 7 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	379	0.89	883	2.07
483	222	0.52	1105	2.59
526	300	0.7	1405	3.29
556	420	0.98	1825	4.28
579	559	1.31	2384	5.58
597	584	1.37	2968	6.95
612	715	1.67	3683	8.63
625	805	1.89	4488	10.51
637	934	2.19	5422	12.7
647	968	2.27	6390	14.97
657	1110	2.6	7500	17.57
666	1188	2.78	8688	20.35
674	1311	3.07	9999	23.42
682	1293	3.03	11292	26.45
690	1410	3.3	12702	29.75
697	1461	3.42	14163	33.18
704	1503	3.52	15666	36.7
711	1545	3.62	17211	40.32
718	1547	3.62	18758	43.94
725	1589	3.72	20347	47.66
731	1573	3.68	21920	51.35
738	1528	3.58	23448	54.93
745	1534	3.59	24982	58.52
751	1521	3.56	26503	62.08
758	1522	3.57	28025	65.65
765	1383	3.24	29408	68.89
772	1382	3.24	30790	72.13
779	1315	3.08	32105	75.21
786	1332	3.12	33437	78.33
794	1261	2.95	34698	81.28
802	1231	2.88	35929	84.16
810	1132	2.65	37061	86.82
819	1082	2.53	38143	89.35
830	1029	2.41	39172	91.76
841	940	2.2	40112	93.96

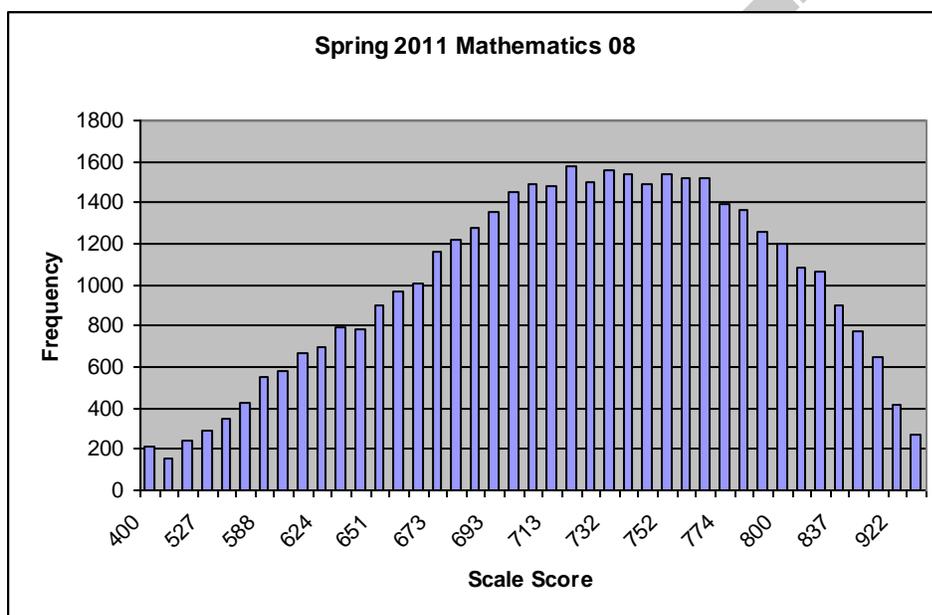
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
855	827	1.94	40939	95.9
872	701	1.64	41640	97.54
895	504	1.18	42144	98.72
934	375	0.88	42519	99.6
990	170	0.4	42689	100



Mathematics Grade 8 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	212	0.51	724	1.76
421	159	0.39	883	2.14
490	239	0.58	1122	2.72
527	292	0.71	1414	3.43
553	352	0.85	1766	4.29
572	423	1.03	2189	5.31
588	548	1.33	2737	6.64
602	581	1.41	3318	8.05
614	664	1.61	3982	9.67
624	697	1.69	4679	11.36
634	798	1.94	5477	13.29
642	788	1.91	6265	15.21
651	901	2.19	7166	17.39
659	970	2.35	8136	19.75
666	1009	2.45	9145	22.2
673	1164	2.83	10309	25.02
680	1222	2.97	11531	27.99
687	1280	3.11	12811	31.1
693	1352	3.28	14163	34.38
700	1454	3.53	15617	37.91
706	1487	3.61	17104	41.52
713	1483	3.6	18587	45.12
719	1582	3.84	20169	48.96
726	1498	3.64	21667	52.59
732	1554	3.77	23221	56.37
739	1539	3.74	24760	60.1
746	1490	3.62	26250	63.72
752	1535	3.73	27785	67.45
759	1520	3.69	29305	71.14
767	1523	3.7	30828	74.83
774	1397	3.39	32225	78.22
782	1368	3.32	33593	81.54
791	1258	3.05	34851	84.6
800	1200	2.91	36051	87.51
810	1080	2.62	37131	90.13

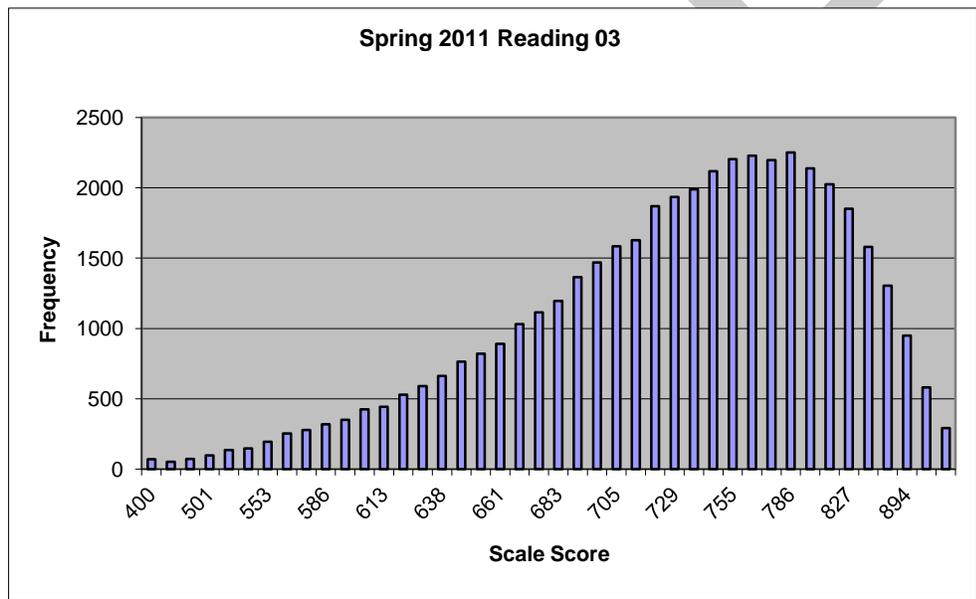
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
822	1061	2.58	38192	92.71
837	900	2.18	39092	94.89
854	773	1.88	39865	96.77
879	645	1.57	40510	98.33
922	419	1.02	40929	99.35
990	267	0.65	41196	100



Reading Grade 3 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	71	0.16	502	1.13
427	52	0.12	554	1.25
473	72	0.16	626	1.41
501	98	0.22	724	1.63
522	135	0.3	859	1.93
539	147	0.33	1006	2.27
553	195	0.44	1201	2.7
565	254	0.57	1455	3.28
576	278	0.63	1733	3.9
586	318	0.72	2051	4.62
596	350	0.79	2401	5.41
605	425	0.96	2826	6.36
613	442	1	3268	7.36
622	529	1.19	3797	8.55
630	589	1.33	4386	9.88
638	663	1.49	5049	11.37
646	764	1.72	5813	13.09
653	820	1.85	6633	14.93
661	891	2.01	7524	16.94
668	1031	2.32	8555	19.26
676	1114	2.51	9669	21.77
683	1196	2.69	10865	24.46
690	1366	3.08	12231	27.54
698	1470	3.31	13701	30.85
705	1584	3.57	15285	34.41
713	1628	3.67	16913	38.08
721	1868	4.21	18781	42.29
729	1935	4.36	20716	46.64
737	1989	4.48	22705	51.12
746	2117	4.77	24822	55.89
755	2203	4.96	27025	60.85
764	2229	5.02	29254	65.87
775	2196	4.94	31450	70.81
786	2250	5.07	33700	75.88

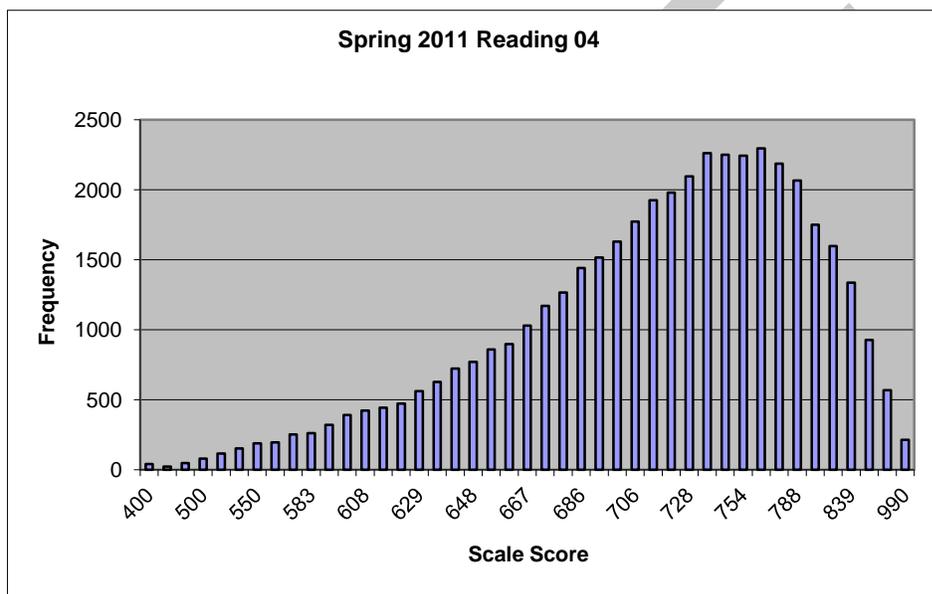
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
798	2137	4.81	35837	80.69
811	2024	4.56	37861	85.24
827	1850	4.17	39711	89.41
844	1580	3.56	41291	92.97
866	1303	2.93	42594	95.9
894	949	2.14	43543	98.04
934	581	1.31	44124	99.34
990	291	0.66	44415	100



Reading Grade 4 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	40	0.09	453	1.04
437	23	0.05	476	1.09
475	47	0.11	523	1.2
500	78	0.18	601	1.37
520	116	0.27	717	1.64
537	151	0.35	868	1.98
550	187	0.43	1055	2.41
562	196	0.45	1251	2.86
573	251	0.57	1502	3.43
583	260	0.59	1762	4.03
592	321	0.73	2083	4.76
600	391	0.89	2474	5.65
608	422	0.96	2896	6.62
615	443	1.01	3339	7.63
622	472	1.08	3811	8.71
629	561	1.28	4372	9.99
635	626	1.43	4998	11.42
642	722	1.65	5720	13.07
648	769	1.76	6489	14.83
654	858	1.96	7347	16.79
661	897	2.05	8244	18.84
667	1030	2.35	9274	21.2
673	1169	2.67	10443	23.87
679	1266	2.89	11709	26.76
686	1441	3.29	13150	30.06
692	1515	3.46	14665	33.52
699	1628	3.72	16293	37.24
706	1772	4.05	18065	41.29
713	1925	4.4	19990	45.69
721	1978	4.52	21968	50.21
728	2094	4.79	24062	55
737	2261	5.17	26323	60.17
745	2250	5.14	28573	65.31
754	2242	5.12	30815	70.43
764	2295	5.25	33110	75.68

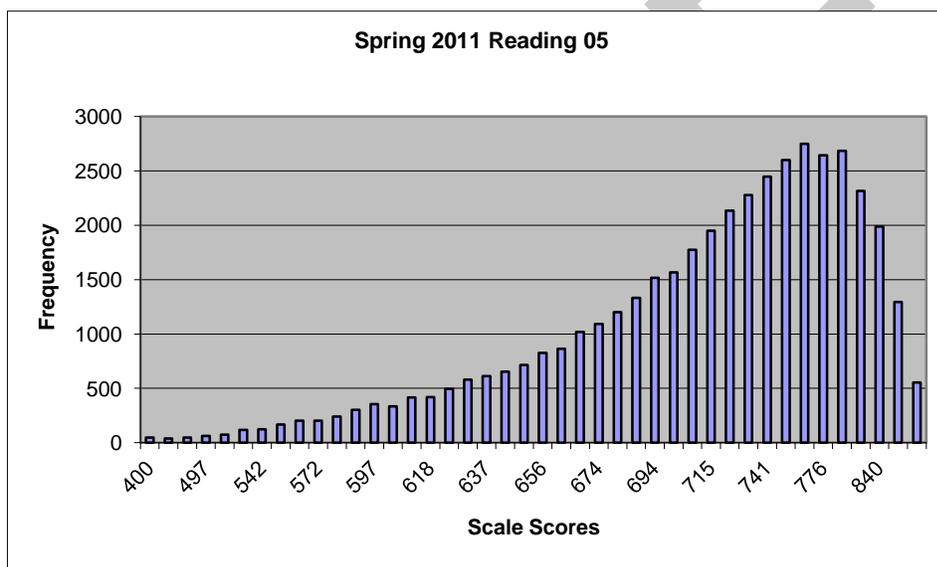
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
775	2186	5	35296	80.67
788	2065	4.72	37361	85.39
802	1750	4	39111	89.39
818	1598	3.65	40709	93.05
839	1335	3.05	42044	96.1
869	927	2.12	42971	98.22
920	567	1.3	43538	99.51
990	213	0.49	43751	100



Reading Grade 5 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	49	0.11	561	1.29
439	39	0.09	600	1.38
473	48	0.11	648	1.49
497	62	0.14	710	1.63
515	75	0.17	785	1.8
529	117	0.27	902	2.07
542	125	0.29	1027	2.36
553	168	0.39	1195	2.74
563	202	0.46	1397	3.21
572	203	0.47	1600	3.67
581	240	0.55	1840	4.23
589	303	0.7	2143	4.92
597	356	0.82	2499	5.74
604	335	0.77	2834	6.51
611	416	0.96	3250	7.46
618	419	0.96	3669	8.42
625	494	1.13	4163	9.56
631	579	1.33	4742	10.89
637	612	1.41	5354	12.29
643	653	1.5	6007	13.79
650	716	1.64	6723	15.44
656	825	1.89	7548	17.33
662	863	1.98	8411	19.31
668	1019	2.34	9430	21.65
674	1091	2.51	10521	24.16
680	1200	2.76	11721	26.91
687	1333	3.06	13054	29.98
694	1518	3.49	14572	33.46
701	1566	3.6	16138	37.06
708	1775	4.08	17913	41.13
715	1951	4.48	19864	45.61
723	2134	4.9	21998	50.51
732	2278	5.23	24276	55.74
741	2447	5.62	26723	61.36
751	2599	5.97	29322	67.33

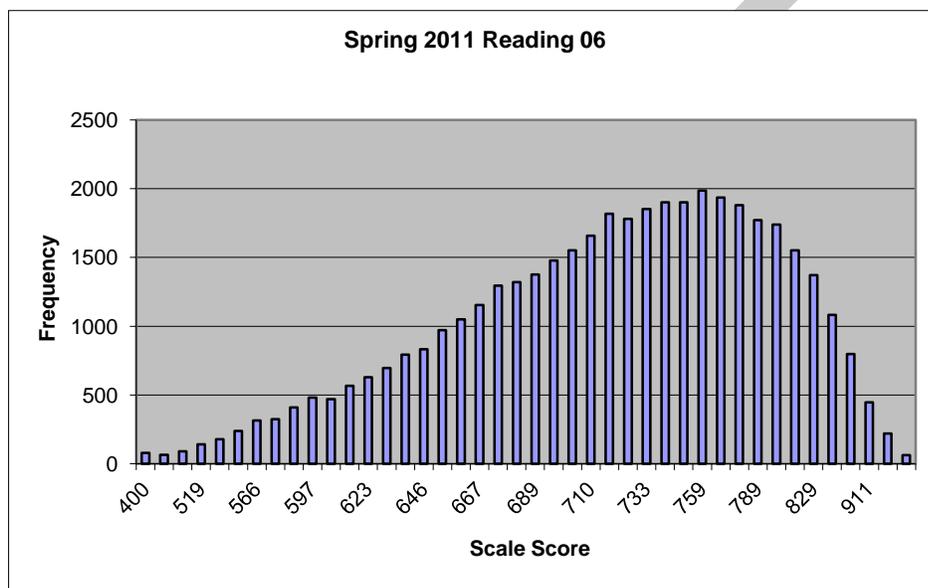
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
763	2748	6.31	32070	73.64
776	2644	6.07	34714	79.71
792	2684	6.16	37398	85.88
812	2317	5.32	39715	91.2
840	1988	4.56	41703	95.76
888	1293	2.97	42996	98.73
990	553	1.27	43549	100



Reading Grade 6 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	78	0.18	563	1.32
461	65	0.15	628	1.47
495	90	0.21	718	1.68
519	141	0.33	859	2.01
538	178	0.42	1037	2.43
553	237	0.55	1274	2.98
566	314	0.73	1588	3.72
577	323	0.76	1911	4.47
588	410	0.96	2321	5.43
597	480	1.12	2801	6.56
606	470	1.1	3271	7.66
615	567	1.33	3838	8.98
623	628	1.47	4466	10.45
631	697	1.63	5163	12.08
638	794	1.86	5957	13.94
646	832	1.95	6789	15.89
653	972	2.28	7761	18.17
660	1049	2.46	8810	20.62
667	1153	2.7	9963	23.32
674	1295	3.03	11258	26.35
681	1320	3.09	12578	29.44
689	1375	3.22	13953	32.66
696	1477	3.46	15430	36.12
703	1552	3.63	16982	39.75
710	1658	3.88	18640	43.63
718	1817	4.25	20457	47.88
725	1781	4.17	22238	52.05
733	1851	4.33	24089	56.38
741	1900	4.45	25989	60.83
750	1900	4.45	27889	65.28
759	1986	4.65	29875	69.93
768	1935	4.53	31810	74.45
778	1879	4.4	33689	78.85
789	1770	4.14	35459	83
800	1738	4.07	37197	87.06

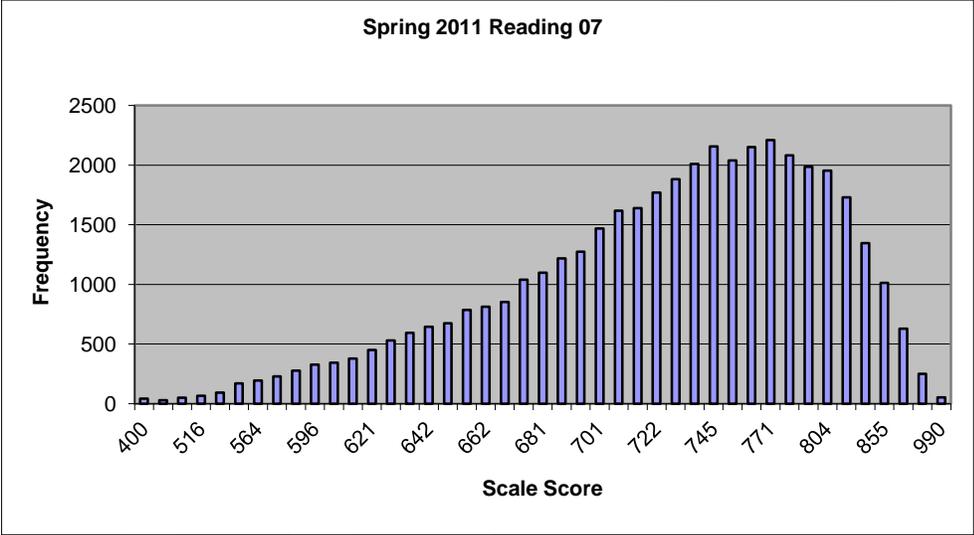
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
814	1552	3.63	38749	90.7
829	1370	3.21	40119	93.9
848	1081	2.53	41200	96.43
873	797	1.87	41997	98.3
911	446	1.04	42443	99.34
989	219	0.51	42662	99.85
990	62	0.15	42724	100



Reading Grade 7 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	43	0.1	578	1.35
455	30	0.07	608	1.42
491	51	0.12	659	1.54
516	67	0.16	726	1.7
535	92	0.22	818	1.92
551	169	0.4	987	2.31
564	193	0.45	1180	2.76
576	229	0.54	1409	3.3
586	276	0.65	1685	3.95
596	327	0.77	2012	4.71
605	343	0.8	2355	5.51
613	379	0.89	2734	6.4
621	450	1.05	3184	7.46
628	531	1.24	3715	8.7
635	595	1.39	4310	10.09
642	644	1.51	4954	11.6
649	674	1.58	5628	13.18
656	787	1.84	6415	15.02
662	812	1.9	7227	16.92
668	854	2	8081	18.92
675	1039	2.43	9120	21.36
681	1098	2.57	10218	23.93
688	1219	2.85	11437	26.78
694	1274	2.98	12711	29.76
701	1470	3.44	14181	33.21
708	1619	3.79	15800	37
715	1640	3.84	17440	40.84
722	1770	4.14	19210	44.98
729	1881	4.4	21091	49.39
737	2011	4.71	23102	54.1
745	2158	5.05	25260	59.15
753	2040	4.78	27300	63.93
762	2152	5.04	29452	68.97
771	2210	5.18	31662	74.14

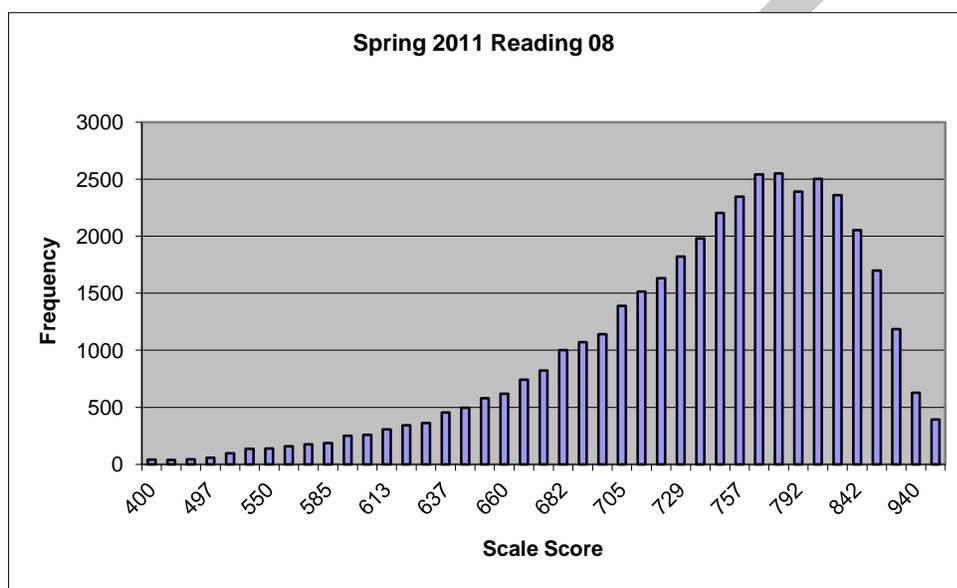
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
781	2081	4.87	33743	79.01
792	1986	4.65	35729	83.66
804	1953	4.57	37682	88.24
817	1730	4.05	39412	92.29
834	1347	3.15	40759	95.44
855	1013	2.37	41772	97.82
885	630	1.48	42402	99.29
938	250	0.59	42652	99.88
990	53	0.12	42705	100



Reading Grade 8 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	41	0.1	493	1.2
431	39	0.09	532	1.29
469	43	0.1	575	1.4
497	59	0.14	634	1.54
518	98	0.24	732	1.78
535	136	0.33	868	2.11
550	138	0.34	1006	2.44
563	159	0.39	1165	2.83
575	174	0.42	1339	3.25
585	186	0.45	1525	3.71
595	250	0.61	1775	4.31
604	258	0.63	2033	4.94
613	305	0.74	2338	5.68
621	342	0.83	2680	6.51
629	361	0.88	3041	7.39
637	453	1.1	3494	8.49
645	496	1.21	3990	9.69
652	579	1.41	4569	11.1
660	620	1.51	5189	12.61
667	741	1.8	5930	14.41
675	823	2	6753	16.41
682	1002	2.43	7755	18.84
689	1071	2.6	8826	21.44
697	1142	2.77	9968	24.22
705	1388	3.37	11356	27.59
712	1515	3.68	12871	31.27
721	1632	3.97	14503	35.24
729	1822	4.43	16325	39.66
738	1981	4.81	18306	44.48
747	2204	5.35	20510	49.83
757	2347	5.7	22857	55.53
768	2541	6.17	25398	61.71
779	2550	6.2	27948	67.9
792	2390	5.81	30338	73.71
806	2502	6.08	32840	79.79

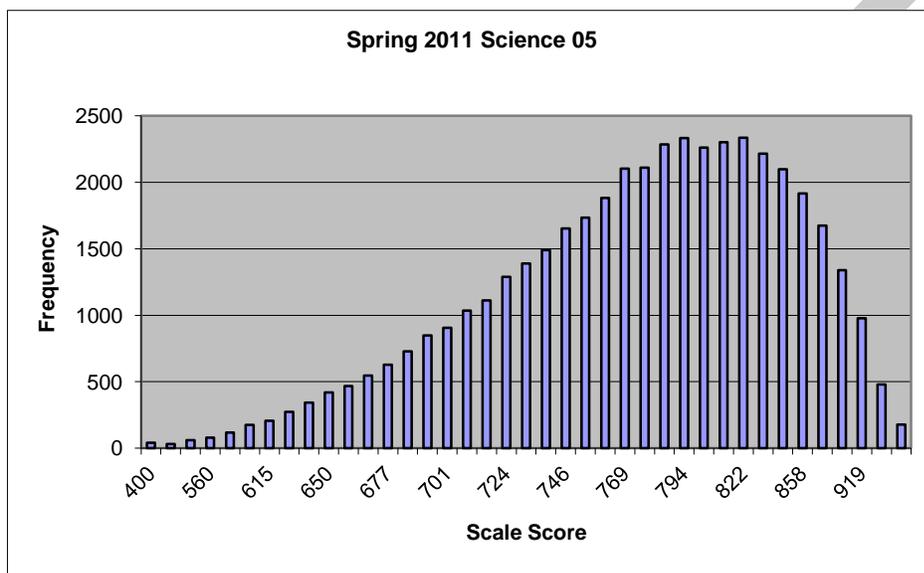
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
822	2361	5.74	35201	85.53
842	2052	4.99	37253	90.51
865	1700	4.13	38953	94.64
896	1185	2.88	40138	97.52
940	627	1.52	40765	99.05
990	393	0.95	41158	100



Science Grade 5 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	41	0.09	554	1.24
476	32	0.07	586	1.31
530	60	0.13	646	1.45
560	78	0.18	724	1.62
583	118	0.26	842	1.89
600	176	0.39	1018	2.28
615	206	0.46	1224	2.75
628	272	0.61	1496	3.36
639	342	0.77	1838	4.12
650	420	0.94	2258	5.07
660	467	1.05	2725	6.11
669	546	1.23	3271	7.34
677	628	1.41	3899	8.75
686	729	1.64	4628	10.38
694	847	1.9	5475	12.28
701	906	2.03	6381	14.32
709	1034	2.32	7415	16.64
717	1111	2.49	8526	19.13
724	1288	2.89	9814	22.02
731	1389	3.12	11203	25.14
739	1489	3.34	12692	28.48
746	1653	3.71	14345	32.19
754	1734	3.89	16079	36.08
761	1883	4.22	17962	40.3
769	2102	4.72	20064	45.02
777	2110	4.73	22174	49.75
785	2285	5.13	24459	54.88
794	2334	5.24	26793	60.11
803	2262	5.08	29055	65.19
812	2302	5.16	31357	70.35
822	2336	5.24	33693	75.6
833	2215	4.97	35908	80.57
845	2099	4.71	38007	85.27
858	1915	4.3	39922	89.57
874	1675	3.76	41597	93.33

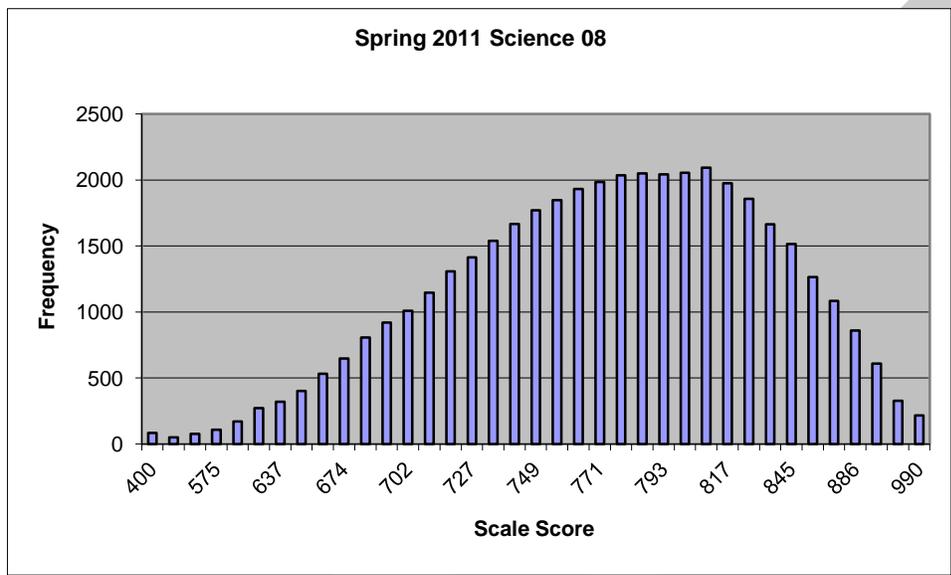
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
893	1339	3	42936	96.33
919	978	2.19	43914	98.53
962	479	1.07	44393	99.6
990	177	0.4	44570	100



Science Grade 8 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	83	0.2	823	1.94
430	50	0.12	873	2.06
533	77	0.18	950	2.24
575	109	0.26	1059	2.5
601	171	0.4	1230	2.9
621	272	0.64	1502	3.54
637	319	0.75	1821	4.3
651	403	0.95	2224	5.25
663	531	1.25	2755	6.5
674	648	1.53	3403	8.03
684	807	1.9	4210	9.93
694	921	2.17	5131	12.1
702	1009	2.38	6140	14.48
711	1146	2.7	7286	17.18
719	1308	3.09	8594	20.27
727	1414	3.34	10008	23.6
734	1538	3.63	11546	27.23
742	1666	3.93	13212	31.16
749	1771	4.18	14983	35.34
757	1848	4.36	16831	39.7
764	1932	4.56	18763	44.25
771	1984	4.68	20747	48.93
778	2036	4.8	22783	53.74
786	2050	4.84	24833	58.57
793	2042	4.82	26875	63.39
801	2055	4.85	28930	68.23
809	2093	4.94	31023	73.17
817	1975	4.66	32998	77.83
826	1858	4.38	34856	82.21
835	1665	3.93	36521	86.14
845	1516	3.58	38037	89.71
857	1265	2.98	39302	92.7
870	1084	2.56	40386	95.25
886	859	2.03	41245	97.28
907	608	1.43	41853	98.71

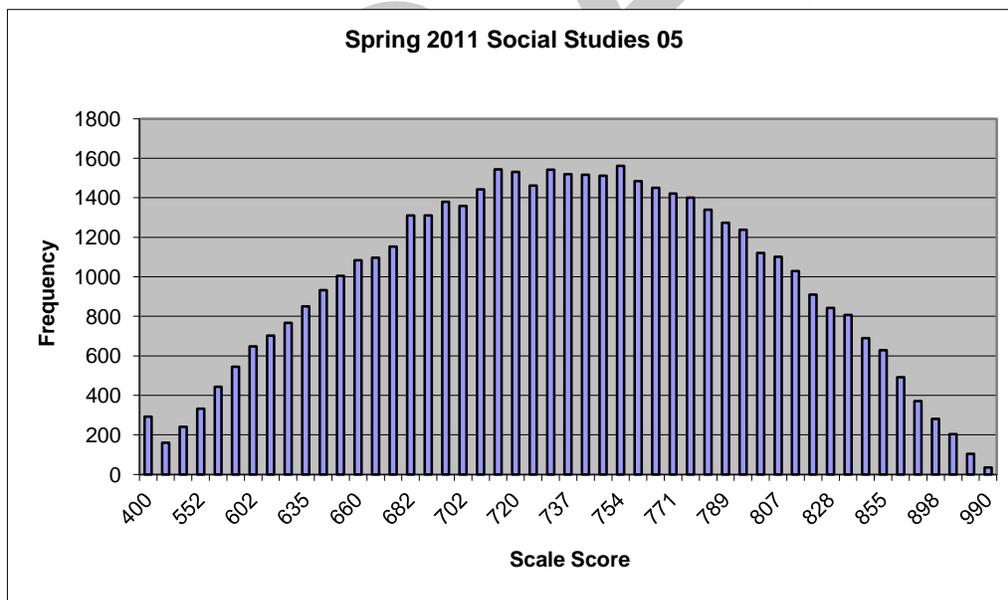
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
936	328	0.77	42181	99.49
990	217	0.51	42398	100



Social Studies Grade 5 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	292	0.61	805	1.68
482	160	0.33	965	2.01
525	241	0.5	1206	2.51
552	333	0.69	1539	3.21
572	444	0.93	1983	4.13
588	545	1.14	2528	5.27
602	648	1.35	3176	6.62
614	702	1.46	3878	8.08
625	766	1.6	4644	9.68
635	850	1.77	5494	11.45
644	932	1.94	6426	13.39
653	1004	2.09	7430	15.49
660	1083	2.26	8513	17.74
668	1097	2.29	9610	20.03
675	1153	2.4	10763	22.43
682	1311	2.73	12074	25.17
689	1310	2.73	13384	27.9
695	1379	2.87	14763	30.77
702	1359	2.83	16122	33.6
708	1443	3.01	17565	36.61
714	1544	3.22	19109	39.83
720	1531	3.19	20640	43.02
726	1462	3.05	22102	46.07
731	1542	3.21	23644	49.28
737	1520	3.17	25164	52.45
743	1517	3.16	26681	55.61
748	1512	3.15	28193	58.76
754	1562	3.26	29755	62.02
760	1484	3.09	31239	65.11
765	1451	3.02	32690	68.14
771	1422	2.96	34112	71.1
777	1400	2.92	35512	74.02
783	1340	2.79	36852	76.81
789	1273	2.65	38125	79.47

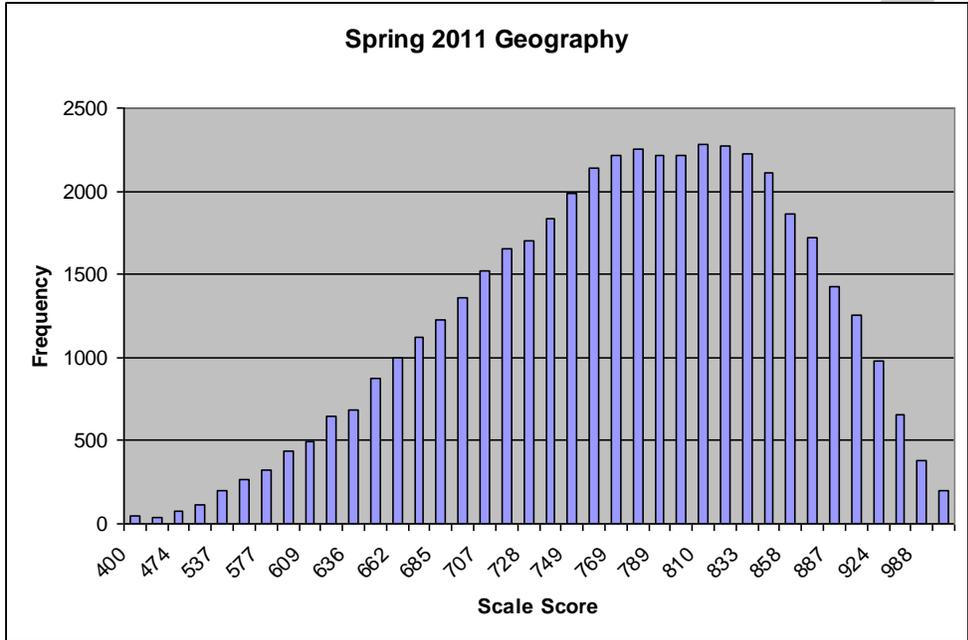
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
795	1238	2.58	39363	82.05
801	1121	2.34	40484	84.38
807	1102	2.3	41586	86.68
814	1029	2.14	42615	88.82
821	910	1.9	43525	90.72
828	842	1.76	44367	92.48
836	807	1.68	45174	94.16
845	689	1.44	45863	95.59
855	629	1.31	46492	96.9
866	491	1.02	46983	97.93
880	371	0.77	47354	98.7
898	280	0.58	47634	99.29
924	204	0.43	47838	99.71
969	104	0.22	47942	99.93
990	35	0.07	47977	100



Social Studies Grade 7 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	45	0.1	564	1.21
413	35	0.08	599	1.29
474	78	0.17	677	1.45
510	114	0.24	791	1.7
537	196	0.42	987	2.12
559	265	0.57	1252	2.69
577	319	0.69	1571	3.38
594	433	0.93	2004	4.31
609	493	1.06	2497	5.37
623	648	1.39	3145	6.76
636	688	1.48	3833	8.24
649	873	1.88	4706	10.11
662	1002	2.15	5708	12.27
673	1125	2.42	6833	14.68
685	1230	2.64	8063	17.33
696	1363	2.93	9426	20.25
707	1523	3.27	10949	23.53
718	1654	3.55	12603	27.08
728	1700	3.65	14303	30.73
738	1834	3.94	16137	34.68
749	1987	4.27	18124	38.95
759	2139	4.6	20263	43.54
769	2216	4.76	22479	48.3
779	2255	4.85	24734	53.15
789	2214	4.76	26948	57.91
800	2215	4.76	29163	62.67
810	2279	4.9	31442	67.56
822	2270	4.88	33712	72.44
833	2227	4.79	35939	77.23
845	2114	4.54	38053	81.77
858	1866	4.01	39919	85.78
872	1718	3.69	41637	89.47
887	1428	3.07	43065	92.54
904	1259	2.71	44324	95.24
924	979	2.1	45303	97.35

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
950	653	1.4	45956	98.75
988	377	0.81	46333	99.56
990	204	0.44	46537	100



Social Studies Grade 8 Scale Score Distribution for Spring 2011

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	373	0.82	1070	2.36
467	292	0.64	1362	3
535	389	0.86	1751	3.86
569	502	1.11	2253	4.97
592	629	1.39	2882	6.36
610	812	1.79	3694	8.15
625	897	1.98	4591	10.13
638	1009	2.23	5600	12.35
650	1132	2.5	6732	14.85
661	1200	2.65	7932	17.5
671	1302	2.87	9234	20.37
680	1403	3.09	10637	23.46
688	1448	3.19	12085	26.66
697	1541	3.4	13626	30.06
704	1513	3.34	15139	33.39
712	1594	3.52	16733	36.91
719	1575	3.47	18308	40.38
726	1572	3.47	19880	43.85
732	1541	3.4	21421	47.25
739	1706	3.76	23127	51.01
746	1623	3.58	24750	54.59
752	1645	3.63	26395	58.22
759	1596	3.52	27991	61.74
766	1639	3.62	29630	65.36
772	1694	3.74	31324	69.09
779	1677	3.7	33001	72.79
787	1585	3.5	34586	76.29
795	1580	3.49	36166	79.77
804	1527	3.37	37693	83.14
813	1491	3.29	39184	86.43
824	1419	3.13	40603	89.56
836	1309	2.89	41912	92.45
851	1181	2.6	43093	95.05
870	921	2.03	44014	97.08
897	727	1.6	44741	98.69

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
944	421	0.93	45162	99.62
990	174	0.38	45336	100

