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

2012 Technical Report

Submitted to
The Oklahoma State Department of Education
August 2012

ALWAYS LEARNING PEARSON

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 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 February 21-22, 2012. Five tests—grade 7 Geography, grades 7 and 8 Mathematics, and grades 7 and 8 Reading are primarily computer delivery (a paper form is also available), and were administered during the online test window from April 10, 2012 to May 4, 2012. The remaining tests were administered via paper between April 10, 2012 and April 24, 2012. This report provides technical details of work accomplished through the end of 2012 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 the validity of test scores and interpretation of results. This technical report covers all of these topics for the 2011-12 testing year.

Table of Contents

List of Tables

Section 1

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

The Oklahoma Core Curriculum Tests are state-mandated, criterion-referenced tests used to assess student proficiency. In the spring of 2012, 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 in grades 3 through 8, Science and Writing in grades 5 and 8, and Social Studies in grades 5, 7, and 8. The 2012 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 7th 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 5 are administered in two sessions. All tests in grades 6 through 8, as well as the grade 5 writing test, are administered in a single session. The grades 7 and 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 or for test test-takers with accommodations requiring a paper form). Tests for all other grades and subjects were administered exclusively in paper and pencil format.

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 operational set of items for the Reading, Mathematics, Science, and Social Studies assessments were composed of Multiple-Choice (MC) items only. For each subject and grade, there were between eight and twelve forms consisting of a common set of operational items and a unique set of 10 field-test items. Responses to the operational items were used to produce student scores. Responses to the field-test items were used to evaluate the psychometric properties of these newly developed items for possible inclusion on future forms. 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 2012 operational form was created as well. A student could receive an equivalent form for various reasons, including becoming ill during test administration or experiencing some kind of security breach. The State Department of Education Office of Accountability and Assessments determines eligibility for an equivalent form on a case-by-case basis. Responses for students who took an equivalent form were scored and reported using the scoring tables from the form’s previous administration.

1.1 Content 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://www.ok.gov/sde/test-support-teachers-and-administrators.

A list of the testable standards for each subject is provided in Table 1-1. For Math[1] and Reading[2], the same testable standards appear in each grade level.

The tables in Appendix A provide information drawn from the 2012 *PASS* blueprints. These tables show the *PASS* standards, objectives, skills, and processes, as well as the number of items allocated to each standard, objective/skill and/or process according to the blueprint and actual number of items appearing on the 2012 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
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

[1] 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.

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

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./4.	Causes and Results of the American Revolution
Standard 5.	Governing Documents/Early Federal Period
Standard 6./10.	Moving Toward the Civil War
Standard 7./8.	Early 19th Century America
Standard 9.	Westward Movement

1.2 Summary of Test Development and Content Validity

To ensure content validity of the OCCT tests, Pearson content specialists 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 were 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 2012 assessments.

1.2.a Aligning Tests to *PASS* Content Standards

In addition to the test blueprints provided by SDE (see Appendix A), Table 1-2 describes five criteria for test alignment with the *PASS* Standards and Objectives.

Table 1-2. Criteria for Aligning the Test with *PASS* Standards and Objectives.

1. Categorical Concurrence	The test is constructed so that there are at least six items measuring each *PASS* standard. 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 *PASS* objective.
3. Range of Knowledge Correspondence	The test is constructed so that at least 75% of the objectives for a *PASS* 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 *PASS* 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 *PASS* objective or concept being assessed, not from specialized knowledge or cultural background of the test-taker.

1.2.b Additional Considerations in Item Selection

The source of the operational items eligible for inclusion on the Spring 2012 form is a pool of previously field-tested or operationally-administered items ranging from the Spring 2005 through the Spring 2011 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/or student achievement 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 2012 operational assessments.

Table 1-3. Percentage of Items by Depth of Knowledge Levels

		DOK Level					
		1		2		3	
Subject	Grade	Target	Actual	Target	Actual	Target	Actual
Math	3	20-25	20	65-70	70	5-15	10
	4	20-25	24	65-70	64	5-15	12
	5	20-25	22	65-70	63	5-15	14
	6	10-15	14	65-70	70	15-25	16
	7	10-15	14	65-70	68	15-25	18
	8	10-15	14	65-70	68	15-25	18
Reading	3	20-25	16	65-70	68	5-15	16
	4	20-25	18	65-70	66	5-15	16
	5	20-25	14	65-70	72	5-15	14
	6	10-15	12	65-70	66	15-25	22
	7	10-15	10	65-70	72	15-25	18
	8	10-15	4	65-70	84	15-25	12
Science	5	20-25	18	65-70	67	5-15	16
	8	10-15	11	60-70	69	15-30	20
Social Studies	5	20-25	23	65-70	67	5-15	10
	7	10-15	11	65-70	69	15-25	20
	8	10-15	13	65-70	67	15-25	20

Note: All values are in percentages.

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

Table 1-4 provides an overview of the number of operational and field test items for the Spring 2012 OSTP-OCCT 3-8 assessments. The Spring 2012 test is comprised of a single core of operational items on each form. For each of the MC 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 item is detailed in Section 4). For the 17 MC tests, between eight and twelve 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. Additional Common Core-aligned and vertical linking field test items for Mathematics and Reading are not included in the counts in this table.

Table 1-4. Configuration of the OSTP-OCCT 3-8 Tests for Spring 2012

		Counts Across Forms					**Count of *PASS*-Aligned Items per FT Form**		
Subject	**Grade**	**Core Forms**	**FT Forms**	**OP Items**	**FT Items***	**Total Items**	**OP**	**FT***	**Total**
Math	3	1	10	50	40	90	50	5	55
	4	1	12	50	40	90	50	5	55
	5	1	12	49^	40	89	49^	5	54
	6	1	12	50	40	90	50	5	55
	7	1	12	50	40	90	50	5	55
	8	1	10	50	40	90	50	5	55
Reading	3	1	10	50	40	90	50	5	55
	4	1	12	50	40	90	50	5	55
	5	1	12	50	40	90	50	5	55
	6	1	12	50	40	90	50	5	55
	7	1	12	50	40	90	50	5	55
	8	1	10	50	40	90	50	5	55
Science	5	1	12	45	80	125	45	10	55
	8	1	8	45	80	125	45	10	55
Social Studies	5	1	12	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; *Common Core-aligned Field Test items and Vertical Linking items were administered in Math and Reading and do not contribute to these counts; ^One Math Grade 5 item initially defined as Operational was not scored.

Section 2

Administration of the OCCT in Grades 3 to 8

Valid and reliable assessment requires that tests 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 to receive 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 the PearsonAccess™ website, and optionally with data received directly from Oklahoma. Oklahoma educators also use the PearsonAccess™ 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 by 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 on materials with no discrepancies. This system allows Pearson to proceed in scoring clean batches while any discrepant material issues are being resolved. As discrepant materials are received, they are 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 of 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.

Pearson provides updates to the initial discrepancy reports on a daily basis 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 unaccounted for 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 to SDE. In addition, the Oklahoma Call Center is able to provide receipt status information as 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 these 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 2012 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 all statistical analyses for the OCCT assessments. The classical analyses described in this section (as well the calibration and equating of each test) were conducted using carefully selected samples of approximately 15,000 students for each grade and subject.

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 performed at this time. This 'cleaned' sample was used to create the subsample file to be used in subsequent 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 responses
- Scored item responses
- 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 eligible for inclusion in the data analyses.

Subsampling. Contractual requirements dictate that equated scale scores and performance levels be delivered to SDE within 48 hours of the close of the testing window for online tests and within 2 weeks of the close of paper-and-pencil tests. To meet this reporting schedule, student data were obtained prior to the close of the administration windows. To ensure that subsamples used for analyses and equating were representative of the population of Oklahoma students, Pearson research scientists pulled stratified subsamples of approximately 15,000 students for each grade and subject, conditioning on district representation, gender, and ethnicity. The sampling technique employed was approved by both the SDE and the Oklahoma Technical Advisory Committee (TAC), a panel of recognized experts in measurement and policy. The demographic breakdown of the students in Spring 2012 item analysis and calibration subsamples appears in Table 3-1 and for all students in Table 3-2. The subsamples used for analyses and equating were deemed to be appropriately representative of the test-taking population.

Table 3-1. Demographic Characteristics of the Student Subsample for Spring 2012

				African	Native			Pacific		
Subject/Grade		**Female**	**Male**	**American**	**American**	**Hispanic**	**Asian**	**Islander**	**White**	**Other**
Math	3	7587	7472	1466	2469	2109	295	31	7955	743
	4	7511	7579	1399	2523	2036	307	42	8077	726
	5	7474	7541	1367	2566	1952	293	31	8077	745
	6	7590	7584	1304	2647	1897	247	45	8363	684
	7	7567	7644	1249	2681	1876	266	33	8386	720
	8	7613	7591	1199	2783	1677	218	26	8543	760
	All	45342	45411	7984	15669	11547	1626	208	49401	4378
Reading	3	7644	7548	1478	2494	2130	298	38	8024	745
	4	7511	7566	1398	2514	2022	308	37	8081	733
	5	7459	7535	1364	2567	1931	294	32	8071	746
	6	7542	7452	1288	2617	1863	245	44	8286	676
	7	7670	7456	1366	2502	1854	316	50	8352	686
	8	7568	7573	1458	2429	1842	321	38	8275	784
	All	45394	45130	8352	15123	11642	1782	239	49089	4370
Science	5	7496	7624	1402	2587	1960	292	35	8114	742
	8	7515	7675	1358	2595	1777	270	49	8467	694
	All	15011	15299	2760	5182	3737	562	84	16581	1436
Social Studies	5	7443	7679	1492	2610	1962	279	32	8060	710
	7	7467	7765	1529	2686	1714	252	37	8269	745
	8	7384	7731	1350	2650	1746	271	51	8422	660
	All	22294	23175	4371	7946	5422	802	120	24751	2115

Table 3-2. Demographic Characteristics of the Student Population for Spring 2012

Subject/Grade		Female	Male	African American	Native American	Hispanic	Asian	Pacific Islander	White	Other
Math	3	22622	22727	4132	7026	6773	852	127	23938	2535
	4	21983	22049	4071	7096	6300	843	113	23173	2466
	5	21663	21881	3919	7193	6017	876	117	23087	2375
	6	21604	21642	4064	7211	5696	840	96	23079	2298
	7	20958	20845	3885	7071	5161	792	117	22773	2004
	8	20737	20782	3975	6937	5026	797	97	22660	2038
	All	129567	129926	24046	42534	34973	5000	667	138710	13716
Reading	3	22740	22556	4167	6975	6765	853	128	23912	2533
	4	22040	21819	4059	7030	6268	843	116	23102	2463
	5	21819	21762	3957	7183	5989	869	119	23117	2377
	6	21862	21828	4164	7271	5735	845	95	23282	2340
	7	21116	20907	3918	7081	5191	799	114	22913	2007
	8	20848	20873	3985	6959	5026	786	106	22807	2065
	All	130425	129745	24250	42499	34974	4995	678	139133	13785
Science	5	21892	22129	4012	7273	6055	878	117	23303	2411
	8	21319	21642	4179	7126	5277	825	109	23292	2195
	All	43211	43771	8191	14399	11332	1703	226	46595	4606
Social Studies	5	23068	24180	4423	7891	6508	898	119	24837	2621
	7	22193	22746	4330	7683	5593	811	118	24243	2161
	8	22353	23495	4565	7686	5618	830	111	24732	2379
	All	67614	70421	13318	23260	17719	2539	348	73812	7161

Statistical Key Check. Administering items that have only one 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 a 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 2012 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 response 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 distractor 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 2012 administration, all operational items preformed 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 3-3. The operational test results indicate that the omit rates were small for all assessments (grade 3, which is administered using a consumable booklet, is slightly higher). Across tests, the average p-value ranged from 0.60 to 0.77 and the average point biserial correlation ranged from 0.36 to 0.42. In tandem, these summary statistics indicate sets of operational items that are functioning appropriately.

Table 3-3. Test-Level Summaries of Classical Item Analyses for Spring 2012

Subject	Grade	Sample Size	Mean	Mean % of Max	Items Points	Mean P	Mean r_{pb}	Omit Min	Omit Max
Math	3	15068	38.07	0.76	50	0.76	0.41	0.09	2.38
	4	15110	37.99	0.76	50	0.76	0.41	0.01	0.89
	5	15031	35.57	0.73	49	0.73	0.39	0.00	0.37
	6	15187	33.08	0.66	50	0.66	0.41	0.03	0.39
	7	15211	31.61	0.63	50	0.63	0.40	0.00	0.05
	8	15206	32.08	0.64	50	0.64	0.41	0.00	0.07
Reading	3	15207	35.91	0.72	50	0.72	0.41	0.08	1.09
	4	15093	37.17	0.74	50	0.74	0.39	0.01	0.19
	5	15005	38.13	0.76	50	0.76	0.40	0.01	0.11
	6	15019	35.22	0.70	50	0.70	0.41	0.01	0.23
	7	15126	38.30	0.77	50	0.77	0.38	0.00	0.09
	8	15147	38.38	0.77	50	0.77	0.36	0.00	0.05
Science	5	15132	32.54	0.72	45	0.72	0.38	0.02	0.14
	8	15210	29.69	0.66	45	0.66	0.38	0.01	0.19
Social Studies	5	15145	35.97	0.60	60	0.60	0.37	0.01	0.24
	7	15232	30.83	0.69	45	0.69	0.37	0.00	0.05
	8	15150	28.66	0.64	45	0.64	0.42	0.02	0.12

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 achievement 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 matched-achievement students across groups. 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 achievement, 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. 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 students for race/ethnicity comparisons and male students for gender comparisons. The focal groups were members of minority racial groups and female students.

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 into the C category if MH D-DIF is significantly different from zero ($p < 0.05$), and its absolute value is greater than or equal to 1.5.
- The item is classified into the B category if MH D-DIF is significantly different from zero ($p < 0.05$), and its absolute value is between 1.0 and 1.5.
- The item is classified into the A category if 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 3-4 summarize the number of field test items in DIF categories for the 17 multiple choice tests for the OCCT Spring 2012 administrations. Items flagged for DIF were placed before content experts during the Spring 2012 field test data review (described in 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 3-4. DIF Flag Incidence Across All OSTP-OCCT 3-8 Field Test Items for Spring 2012

Subject/Grade		Total FT items	Female	African American	Native American	Hispanic	Asian
Math	3	40	4	3	0	2	4
	4	40	1	7	0	2	1
	5	40	6	1	0	3	1
	6	40	2	1	0	3	2
	7	40	2	4	0	1	0
	8	40	3	5	0	0	4
Reading	3	40	2	5	0	2	4
	4	40	0	3	1	2	0
	5	40	2	4	0	5	1
	6	40	1	2	0	3	2
	7	40	4	3	0	0	0
	8	40	2	6	0	6	3
Science	5	80	3	4	0	1	0
	8	80	2	4	0	1	7
Social Studies	5	80	4	4	0	0	4
	7	80	8	7	0	8	12
	8	80	0	5	0	0	5

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, field-tested multiple choice items across the 17 subjects and grades based on the Spring 2012 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 content 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 undertaken as 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 and a binder containing item images and statistics. Pearson discussed with SDE the analyses performed and the criteria for flagging items. Flagged items were then reviewed and decisions were made on an item-by-item basis as to whether to accept the item, accept the item with revisions (which would require re-field-testing prior to operational use), or reject the item. Review of the data included presentation of the item's *p*-value, point-biserial correlation, point-biserial correlation by response option, response distributions, mean overall score by response option, frequency distributions of response options by students in the lower, middle, and upper third of the score distribution, and indications of item DIF and IRT misfit. Items failing to meet the minimum performance requirements as set by the flagging criteria 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 Q_1 index (see Section 4.2)

Bias Review. One key goal of the data review meetings was to assess potential bias based on DIF results and item content. Although efforts were made to mitigate potential item bias through rigorous writer training and review processes, there remains potential for bias to be present in items, which may 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, which 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 subgroup for all items flagged for DIF. The data review panel was then asked if there was evidence of context (e.g., cultural barriers) or language in an item that might result in bias. The data review panel made the final determination regarding the presence of item bias.

3.4.b Results of Data Review

The number of items inspected during data review as a result of employing the previously-described flagging criteria for the classical item analyses, DIF, and IRT procedures is presented in Table 3-5.

Table 3-5. Spring 2012 Data Review Flagging and Outcomes Summary

Subject	Grade	FT Items	No. Flagged	Rejected	Accepted	Accepted with Edits
Math	3	40	19	5	31	4
	4	40	13	2	36	2
	5	40	13	1	34	5
	6	40	12	0	36	4
	7	40	14	1	32	7
	8	40	15	2	37	1
Reading	3	40	16	1	39	0
	4	40	9	4	36	0
	5	40	12	5	35	0
	6	40	13	6	34	0
	7	40	18	9	31	0
	8	40	20	6	34	0
Science	5	80	19	4	69	7
	8	80	29	9	68	3
Social Studies*	5	80	37	19	53	8
	7	80	39	22	52	6
	8	80	18	10	70	0

*Note. A large number of adequately-performing Social Studies items were rejected at data review due to a recent curriculum change, which resulted in these items no longer aligning to the content standards set to go into effect in 2012-13.

3.5 Test Reliability

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

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

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

Table 3-6 presents Cronbach's alpha for the operational tests by subject area for the Spring 2012 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 3-6 shows the reliability analysis results by the different reporting subgroups for the OSTP-OCCT assessments for Spring 2012 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 3-6. Test Reliability by Subgroup for Spring 2012

Subject	Grade	All	F	M	AA	NA	HI	AS	PI	WH	OT
Math	3	0.90	0.90	0.90	0.91	0.89	0.90	0.91	0.93	0.89	0.90
	4	0.90	0.90	0.90	0.91	0.89	0.90	0.91	0.93	0.89	0.90
	5	0.89	0.88	0.89	0.89	0.88	0.88	0.88	0.91	0.88	0.88
	6	0.90	0.90	0.90	0.89	0.89	0.89	0.91	0.91	0.89	0.89
	7	0.89	0.89	0.90	0.89	0.88	0.88	0.90	0.87	0.89	0.90
	8	0.90	0.89	0.90	0.90	0.89	0.89	0.91	-	0.90	0.89
Reading	3	0.90	0.89	0.90	0.90	0.88	0.90	0.91	0.90	0.89	0.90
	4	0.89	0.88	0.89	0.90	0.87	0.88	0.92	-	0.87	0.89
	5	0.89	0.89	0.90	0.90	0.89	0.89	0.89	0.87	0.88	0.89
	6	0.90	0.89	0.91	0.90	0.89	0.89	0.89	0.93	0.89	0.90
	7	0.88	0.87	0.89	0.89	0.87	0.89	0.90	0.89	0.87	0.87
	8	0.86	0.86	0.86	0.88	0.84	0.88	0.90	0.93	0.84	0.85
Science	5	0.87	0.86	0.88	0.86	0.87	0.86	0.86	0.86	0.85	0.86
	8	0.86	0.85	0.87	0.86	0.85	0.85	0.86	0.91	0.85	0.86
Social Studies	5	0.90	0.88	0.91	0.87	0.88	0.88	0.89	0.90	0.89	0.89
	7	0.86	0.84	0.87	0.85	0.84	0.85	0.87	0.89	0.85	0.85
	8	0.90	0.88	0.91	0.88	0.89	0.88	0.91	0.92	0.89	0.89

Note. Missing values in this table are reflective of subgroups with insufficient score variability for computation of reliability coefficients; F = Female, M = Male, AA = African American, NA = Native American, HI = Hispanic, AS = Asian, PI = Pacific Islander, WH = White, O = Other.

3.6 Analysis of the Writing Tests

The administration of the Spring 2012 Writing assessment took place on February 21 and 22, 2012. Students in grades 5 and 8 responded to one operational writing prompt. The following sections describe the statistical analyses conducted to place the 2012 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 3-7. 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. Writing Analytic Traits and Scoring Weights

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:

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

3.6.b Adjustment for Rater-Year Effects

The baseline for each grade's operational writing scale was 2006. To place the 2012 operational prompt scores 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 writing prompts field-tested in 2007, ETS provided a set of unique transformation constants to adjust for prompt difficulty. Based on ETS' report, *OCCT Writing: Scaling the 2007 Field-Test Prompts* (ETS, 2007), the following equation was used to adjust the 2012 raw composite scores:

$$SCS_{12} = B_{07}\left(RCS_{12}\right) + A_{07} \quad (3)$$

Where SCS_{12} represents the scaled composite score after adjusting the 2012 prompt to the 2007 scale.

In 2012, Pearson also performed a rater drift study to adjust for the difference in raters between the 2007 administration to the current administration. Pearson's Performance Scoring Center (PSC) blindly rescored approximately 500 randomly-selected student responses from 2007 for each grade's prompt. Only prompts with valid scored responses (i.e., no condition codes such as off-topic) from 2007 were selected to be rescored in 2012 as part of the rater drift study. The rescored prompts were then linked to their original 2007 scores and formed the basis for computation of a second set of linear scaling constants.

The 2012 rater effect constants (C_{12} & D_{12}) were determined by using the means (M) and standard deviations (S) of the 2007 raw composite scores and the 2012 rescored raw composite scores as calculated below for each grade.

$$D_{12} = {}^{S_{07}}\!/_{S_{12}} \quad (4)$$

$$C_{12} = M_{07} - \left(M_{12} * D_{12}\right) \quad (5)$$

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 scaled composite scores where computed using the formula below:

$$SCS_{12} = B_{07}\left[\left(D_{12} * RCS_{12}\right) + C_{12}\right] + A_{07} \quad (6)$$

Table 3-8 provides the resulting score distribution statistics after performing the described compound adjustment. Final 2012 transformation constants are also provided within this table.

Table 3-8. Results of Grades 5 and 8 Writing Prompt Scoring and Scaling

Grade	Statistic	2012	2011	2010	2009
5	N	45,427	46057	44994	43665
	MIN	17	18	15	19
	MAX	60	60	60	60
	MEAN	42.39	46.21	43.67	44.57
	SD	8.84	7.99	8.25	8.54
	Constants				
	A	-0.7524			
	B	1.0284			
	C	1.7849			
	D	1.0395			
8	N	44,720	43051	40962	42271
	MIN	17	15	19	18
	MAX	60	60	60	60
	MEAN	47.35	45.76	45.73	45.5
	SD	8.09	7.28	7.42	7.04
	Constants				
	A	2.2187			
	B	0.9770			
	C	-1.7106			
	D	1.1107			

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 grades 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 traits in both grades. The final writing score for a student in a given trait is the average of the two scores. The inter-rater reliability coefficients for the operational prompt are presented in Table 3-9. The results show that exact and adjacent rater agreement on trait scores for both the grades 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 grades 5 and 8 Writing assessments' writing prompts fall within the moderate range.

Table 3-9. Inter-rater Reliability for Grades 5 and 8 Writing Prompts for Spring 2012

			Point Discrepancy Percentages							Agreement Percentages			
Trait	Max Points	Valid N	-3	-2	-1	0	1	2	3	Exact	Adjacent	+/- 2 or more	Kappa
							Grade 5						
1	4	45,427	0.00	0.56	18.02	62.82	18.06	0.53	0.01	62.82	36.08	1.10	0.44
2	4	45,427	0.01	0.70	18.43	61.58	18.70	0.59	0.00	61.58	37.13	1.30	0.42
3	4	45,427	0.00	0.66	18.42	61.81	18.53	0.56	0.00	61.81	36.95	1.22	0.43
4	4	45,427	0.00	0.76	18.99	60.24	19.30	0.71	0.00	60.24	38.29	1.47	0.43
5	4	45,427	0.01	0.78	18.91	60.34	19.14	0.83	0.00	60.34	38.05	1.62	0.44
							Grade 8						
1	4	44,720	0.01	0.42	16.76	65.56	16.82	0.42	0.00	65.56	33.58	0.85	0.40
2	4	44,720	0.01	0.62	17.44	63.71	17.62	0.59	0.00	63.71	35.06	1.22	0.43
3	4	44,720	0.00	0.54	17.44	63.82	17.70	0.49	0.00	63.82	35.14	1.03	0.40
4	4	44,720	0.00	0.61	18.00	62.65	18.11	0.61	0.00	62.65	36.11	1.22	0.43
5	4	44,720	0.01	0.68	19.02	60.72	18.89	0.68	0.00	60.72	37.91	1.37	0.40

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 achievement level of θ responds correctly to item *i* is

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

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 traditionally equal to 1.7. With multiple-choice items it is assumed that, due to guessing, examinees with very low achievement levels have a non-zero probability of responding correctly to an item. This probability is represented in the 3-PL model by the c_i parameter.

IRT models were fit to the 2012 assessment data using MULTILOG version 7.03 (Thissen, Chen, & Bock, 2003). MULTILOG estimates parameters simultaneously for dichotomous items via marginal maximum likelihood. All item and calibrations and scoring were independently conducted and verified by two Pearson research scientists.

4.2 Assessment of Item Fit to the IRT Model

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

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

where Q_{1i} is the fit of item *i*, 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 achievement- and item parameter estimates in Equation (7) and summing over examinees in cell *r*:

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

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

$$Z_j = \frac{Q_{1i} - df}{\sqrt{(2df)}}. \tag{10}$$

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 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 seven operational items (three in grade 3 Reading, and one each in grade 4 Mathematics, grade 8 Reading, grade 7 Geography, and grade 8 Science) 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 information on data review, see Section 3.4).

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 the base scale. Typically, for the CINEG design, common (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 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 2012 administration. For each test, the anchor set was comprised of at least 20% of all operational items. The anchor set was proportionally representative of the total test in terms of content assessed, and it mimicked the difficulty of the overall test as well.

Table 4-1. Number of Anchor Items per Grade and Subject for Spring 2012

			Initial Anchor Set		Final Anchor Set	
Subject	Grade	Operational Items	Items	%	Item	%
Math	3	50	19	38%	19	38%
	4	50	19	38%	17	34%
	5	49	18	37%	18	37%
	6	50	19	38%	18	36%
	7	50	18	36%	18	36%
	8	50	19	38%	18	36%
Reading	3	50	20	40%	17	34%
	4	50	20	40%	20	40%
	5	50	22	44%	22	44%
	6	50	23	46%	23	46%
	7	50	20	40%	16	32%
	8	50	19	38%	19	38%
Science	5	45	16	36%	13	29%
	8	45	15	33%	13	29%
Social Studies	5	60	20	33%	20	33%
	7	45	17	38%	15	33%
	8	45	17	38%	15	33%

Despite the careful selection and placement of anchor items, it is possible for these items to perform differentially across administrations. Dramatic changes in item parameter values can result in systematic errors in equating results (Kolen & Brennan, 2004). As a result, prior to finalizing the equating constants, Pearson evaluated changes in the item parameters from the item bank to the Spring 2012 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(θ_k), with 40 quadrature points.
2) Place the current anchor item parameters on the baseline scale by computing Stocking & Lord (SL) constants using STUIRT and all (*k*) anchor items.
3) Apply the SL anchor constants to the current item parameters, and compute the current raw score to scale score table. The results based on all *k* anchor items 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 computed from 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 on this theta distribution.

$$d_i^{\,2} = \sum^{k} \left[P_{ix}(\theta_k) - P_{iy}(\theta_k)\right]^2 \bullet g(\theta_k) \qquad (11)$$

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 4, dropping one item for each iteration, 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 raw 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 are 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, and only one item was removed at a time.

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

4.5 Scaling and Scoring Results

The lowest obtainable scale score (LOSS), highest obtainable scale score (HOSS), and final scaling constants for each of the subjects are shown in Table 4-2. The scaling constants, *M*1 (multiplicative) and *M*2 (additive), place the true scores associated with each raw score point onto the reporting (or operational) scale using a straightforward linear transformation:

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

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

The true-score equivalent corresponding to each raw score was estimated 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 expected to demonstrate. For the OCCT 3-8 assessments, there are three cut scores that divide scores into four performance levels: Unsatisfactory, Limited Knowledge, Proficient, and Advanced. The cut scores for each of the tests appear in Table 4-2. In addition, a conditional standard error of measurement (CSEM; see Section 6.3) was computed for each of the raw score points. The resulting raw score to scale score conversions, CSEMs, and performance levels for Spring 2012 are shown in Table 4-3 to Table 4-8. RSSS tables for grades 5 and 8 Writing are not included in these tables as the there no further transformation of the composite score beyond that described in Section 3.6.

Table 4-2. LOSS, HOSS, Scaling Constants, and Cut Scores by Subject

Subject	Grade	M1	M2	LOSS	HOSS	Limited Cut	Proficient Cut	Advanced Cut
Math	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	753.900	400	990	638	700	814
	8	70	745.500	400	990	647	700	829
Social Studies	5	70	713.810	400	990	645	700	786
	7	70	759.777	400	990	595	700	847
	8	70	709.940	400	990	622	700	821
Writing	5	NA	NA	15	60	26	36	54
	8	NA	NA	15	60	25	36	54

Table 4-3. Raw Score to Scale Score Conversion Tables for Mathematics (grades 3 to 5) Spring 2012

	Grade 3			Grade 4			Grade 5		
Raw Score	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	400	1	44	400	1	54
10	400	1	43	426	1	48	400	1	54
11	435	1	49	459	1	52	408	1	55
12	464	1	52	484	1	54	458	1	62
13	487	1	54	504	1	54	492	1	65
14	506	1	53	521	1	52	518	1	65
15	523	1	50	535	1	49	538	1	62
16	537	1	47	549	1	45	556	1	57
17	550	1	43	561	1	41	571	1	52
18	563	1	40	572	1	38	584	1	47
19	574	1	37	583	1	36	596	1	42
20	584	1	34	593	1	33	607	1	39
21	594	1	32	602	1	31	617	1	35
22	603	1	30	611	1	30	627	1	33
23	612	1	29	619	1	29	636	1	31
24	621	1	28	628	1	27	644	2	29
25	629	1	27	636	1	26	653	2	28
26	637	2	26	643	2	26	661	2	27
27	645	2	25	651	2	25	668	2	26
28	653	2	24	658	2	24	676	2	25
29	660	2	24	666	2	24	683	2	25
30	667	2	23	673	2	23	691	2	24
31	675	2	23	680	2	23	698	2	24
32	682	2	23	687	2	23	706	3	24
33	689	2	23	694	2	22	713	3	24

	Grade 3			Grade 4			Grade 5		
Raw Score	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
34	697	2	23	701	3	22	721	3	24
35	704	3	23	708	3	22	728	3	24
36	712	3	23	715	3	22	736	3	24
37	719	3	23	723	3	23	745	3	25
38	728	3	24	730	3	23	753	3	25
39	736	3	24	738	3	24	762	3	26
40	745	3	25	747	3	25	772	3	27
41	755	3	26	756	3	26	783	3	29
42	765	3	28	766	3	28	794	4	31
43	777	3	30	777	3	30	807	4	34
44	789	3	32	790	3	33	822	4	38
45	804	4	36	804	3	37	840	4	43
46	821	4	41	821	4	43	863	4	48
47	843	4	47	844	4	49	894	4	50
48	873	4	52	874	4	53	951	4	42
49	923	4	50	928	4	50	990	4	34
50	990	4	38	990	4	39			

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

Table 4-4. Raw Score to Scale Score Conversion Tables for Mathematics (grades 6 to 8) Spring 2012

	Grade 6			Grade 7			Grade 8		
Raw Score	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	63	400	1	67	400	1	76
1	400	1	63	400	1	67	400	1	76
2	400	1	63	400	1	67	400	1	76
3	400	1	63	400	1	67	400	1	76
4	400	1	63	400	1	67	400	1	76
5	400	1	63	400	1	67	400	1	76
6	400	1	63	400	1	67	400	1	76
7	400	1	63	400	1	67	400	1	76
8	400	1	63	400	1	67	400	1	76
9	400	1	63	400	1	67	400	1	76
10	424	1	65	400	1	67	400	1	76
11	483	1	71	471	1	74	400	1	76
12	518	1	72	513	1	77	479	1	80
13	543	1	70	542	1	76	529	1	82
14	563	1	66	565	1	71	559	1	80
15	579	1	60	584	1	65	580	1	74
16	594	1	53	600	1	58	597	1	66
17	607	1	47	614	1	51	610	1	57
18	618	1	42	627	1	45	622	1	49
19	629	1	37	638	1	40	633	1	42
20	639	1	34	648	1	36	642	2	37
21	648	1	31	658	1	33	651	2	32
22	656	1	29	667	1	30	659	2	29
23	664	2	27	676	2	29	667	2	27
24	672	2	26	684	2	27	674	2	25
25	680	2	25	691	2	26	681	2	24
26	687	2	24	699	2	25	687	2	23
27	694	2	23	706	3	24	694	2	22
28	700	3	22	713	3	23	700	3	21
29	707	3	22	720	3	22	707	3	21
30	714	3	21	727	3	22	713	3	20
31	720	3	21	734	3	21	719	3	20
32	726	3	21	740	3	21	725	3	20
33	733	3	20	747	3	21	731	3	20

Raw Score	Grade 6			Grade 7			Grade 8		
	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
34	739	3	20	754	3	20	737	3	20
35	746	3	20	760	3	20	744	3	20
36	752	3	20	767	3	20	750	3	20
37	759	3	21	774	3	20	757	3	20
38	766	3	21	781	3	20	764	3	20
39	774	3	21	788	3	21	771	3	21
40	781	3	22	796	3	21	778	4	21
41	789	3	23	804	4	22	786	4	22
42	798	4	24	812	4	23	794	4	24
43	807	4	26	821	4	24	804	4	25
44	818	4	28	831	4	27	814	4	28
45	830	4	32	843	4	30	826	4	31
46	844	4	36	856	4	34	840	4	36
47	861	4	42	874	4	39	858	4	41
48	887	4	46	898	4	42	883	4	46
49	931	4	44	940	4	40	925	4	46
50	990	4	34	990	4	31	990	4	34

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

Table 4-5. Raw Score to Scale Score Conversion Tables for Reading (grades 3 to 5) Spring 2012

	Grade 3			Grade 4			Grade 5		
Raw Score	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	54	400	1	47	400	1	51
1	400	1	54	400	1	47	400	1	51
2	400	1	54	400	1	47	400	1	51
3	400	1	54	400	1	47	400	1	51
4	400	1	54	400	1	47	400	1	51
5	400	1	54	400	1	47	400	1	51
6	400	1	54	400	1	47	400	1	51
7	400	1	54	400	1	47	400	1	51
8	400	1	54	400	1	47	400	1	51
9	409	1	55	406	1	48	400	1	51
10	466	1	61	455	1	54	400	1	51
11	498	1	64	484	1	56	442	1	55
12	522	1	63	505	1	56	479	1	58
13	540	1	59	522	1	53	503	1	59
14	556	1	54	536	1	49	521	1	57
15	569	1	49	548	1	44	537	1	53
16	581	1	44	559	1	40	550	1	48
17	592	1	39	569	1	36	561	1	44
18	602	1	36	579	1	33	572	1	39
19	611	1	33	587	1	30	582	1	36
20	619	1	30	596	1	28	591	1	33
21	627	1	29	603	1	27	600	1	30
22	635	1	27	611	1	26	608	1	28
23	643	1	26	618	1	25	616	1	27
24	650	2	25	625	1	24	623	1	26
25	657	2	25	632	1	23	630	1	25
26	664	2	24	639	1	23	637	1	24
27	671	2	24	646	1	22	644	2	23
28	678	2	23	652	1	22	651	2	23
29	684	2	23	659	2	22	658	2	22
30	691	2	23	665	2	21	664	2	22
31	698	2	23	672	2	21	671	2	22
32	705	3	23	679	2	21	678	2	22
33	712	3	23	685	2	21	684	2	22
34	719	3	23	692	2	21	691	2	22

Raw Score	Grade 3 OPI Score	Grade 3 Perf. Level	Grade 3 CSEM	Grade 4 OPI Score	Grade 4 Perf. Level	Grade 4 CSEM	Grade 5 OPI Score	Grade 5 Perf. Level	Grade 5 CSEM
35	727	3	24	699	2	22	698	2	22
36	735	3	24	706	3	22	706	3	23
37	743	3	24	714	3	22	713	3	23
38	751	3	25	721	3	23	721	3	24
39	760	3	26	729	3	23	729	3	24
40	770	3	27	738	3	24	738	3	25
41	780	3	28	747	3	26	747	3	26
42	790	3	30	756	3	27	757	3	28
43	802	3	32	767	3	30	769	3	30
44	815	3	36	779	3	33	781	3	33
45	830	3	40	794	3	37	795	3	37
46	849	3	46	811	3	43	813	3	42
47	873	3	51	832	3	49	834	4	49
48	908	4	52	863	4	56	864	4	55
49	975	4	39	915	4	55	914	4	54
50	990	4	36	990	4	41	990	4	40

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 Reading (grades 6 to 8) Spring 2012

	Grade 6			Grade 7			Grade 8		
Raw Score	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	54	400	1	58	400	1	46
1	400	1	54	400	1	58	400	1	46
2	400	1	54	400	1	58	400	1	46
3	400	1	54	400	1	58	400	1	46
4	400	1	54	400	1	58	400	1	46
5	400	1	54	400	1	58	400	1	46
6	400	1	54	400	1	58	400	1	46
7	400	1	54	400	1	58	400	1	46
8	400	1	54	400	1	58	400	1	46
9	446	1	58	400	1	58	400	1	46
10	489	1	61	428	1	60	443	1	52
11	514	1	61	483	1	65	475	1	56
12	533	1	58	514	1	66	498	1	57
13	548	1	53	536	1	64	517	1	55
14	560	1	47	553	1	59	533	1	52
15	571	1	42	567	1	53	547	1	48
16	581	1	37	579	1	47	559	1	44
17	590	1	33	589	1	41	570	1	40
18	598	1	31	598	1	36	580	1	37
19	607	1	29	606	1	32	590	1	34
20	615	1	27	614	1	28	598	1	32
21	622	1	26	621	1	26	607	1	30
22	630	1	25	628	1	24	615	1	29
23	637	1	25	634	1	23	623	1	28
24	644	1	24	640	1	21	630	1	27
25	652	2	24	646	1	21	638	1	26
26	659	2	23	652	1	20	645	1	26
27	666	2	23	657	1	19	652	1	25
28	673	2	23	663	1	19	659	2	25
29	679	2	23	668	2	19	666	2	25
30	686	2	22	674	2	19	674	2	25
31	693	2	22	679	2	19	681	2	25
32	700	3	22	685	2	19	688	2	25
33	707	3	23	690	2	19	696	2	26
34	715	3	23	696	2	19	704	3	26

	Grade 6			Grade 7			Grade 8		
Raw Score	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
35	722	3	23	702	3	20	712	3	27
36	730	3	23	708	3	21	721	3	27
37	737	3	24	715	3	21	729	3	28
38	746	3	24	722	3	23	739	3	29
39	754	3	25	730	3	24	749	3	31
40	763	3	26	738	3	26	760	3	32
41	773	3	27	748	3	28	771	3	35
42	783	3	29	758	3	32	784	3	38
43	795	3	31	770	3	36	799	3	41
44	807	3	34	785	3	41	816	3	46
45	822	3	38	802	4	49	836	4	50
46	840	4	43	824	4	57	861	4	54
47	862	4	49	856	4	63	894	4	55
48	894	4	51	908	4	61	942	4	49
49	952	4	42	990	4	48	990	4	39
50	990	4	35	990	4	48	990	4	39

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 Science Spring 2012

	Grade 5			Grade 8		
Raw Score	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	76	400	1	80
1	400	1	76	400	1	80
2	400	1	76	400	1	80
3	400	1	76	400	1	80
4	400	1	76	400	1	80
5	400	1	76	400	1	80
6	400	1	76	400	1	80
7	400	1	76	400	1	80
8	400	1	76	457	1	83
9	494	1	83	533	1	88
10	541	1	85	570	1	86
11	570	1	81	595	1	79
12	591	1	73	614	1	70
13	608	1	64	629	1	60
14	622	1	55	642	1	51
15	635	1	47	654	2	44
16	646	2	41	664	2	38
17	656	2	36	673	2	33
18	665	2	32	682	2	30
19	674	2	29	691	2	28
20	682	2	28	699	2	26
21	690	2	26	706	3	25
22	698	2	25	714	3	24
23	705	3	24	721	3	23
24	713	3	24	728	3	22
25	720	3	23	735	3	22
26	727	3	23	742	3	21
27	734	3	22	748	3	21
28	741	3	22	755	3	21
29	748	3	22	761	3	20
30	755	3	22	768	3	20
31	763	3	22	775	3	20
32	770	3	23	781	3	20
33	778	3	23	788	3	21
34	786	3	24	796	3	21

Raw Score	Grade 5			Grade 8		
	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
35	794	3	24	803	3	22
36	803	3	25	812	3	23
37	813	3	27	820	3	24
38	823	4	29	830	4	26
39	835	4	32	840	4	28
40	849	4	35	853	4	32
41	865	4	39	867	4	36
42	886	4	43	886	4	40
43	915	4	44	912	4	41
44	966	4	34	958	4	35
45	990	4	29	990	4	28

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 Social Studies Spring 2012

	Grade 5			Grade 7			Grade 8		
Raw Score	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
0	400	1	75	400	1	58	400	1	73
1	400	1	75	400	1	58	400	1	73
2	400	1	75	400	1	58	400	1	73
3	400	1	75	400	1	58	400	1	73
4	400	1	75	400	1	58	400	1	73
5	400	1	75	400	1	58	400	1	73
6	400	1	75	400	1	58	400	1	73
7	400	1	75	400	1	58	400	1	73
8	400	1	75	400	1	58	400	1	73
9	400	1	75	442	1	64	431	1	75
10	400	1	75	487	1	69	507	1	81
11	400	1	75	519	1	71	546	1	81
12	400	1	75	545	1	69	571	1	77
13	446	1	78	566	1	65	590	1	70
14	510	1	82	584	1	60	606	1	62
15	544	1	83	600	2	55	620	1	53
16	568	1	80	615	2	50	632	2	46
17	586	1	74	629	2	46	643	2	40
18	601	1	67	642	2	43	653	2	35
19	614	1	59	654	2	40	662	2	32
20	626	1	52	666	2	38	671	2	29
21	636	1	45	678	2	36	679	2	27
22	645	2	40	689	2	35	687	2	26
23	654	2	36	699	2	34	695	2	25
24	662	2	32	709	3	33	702	3	24
25	670	2	30	719	3	32	710	3	23
26	677	2	28	729	3	32	717	3	22
27	684	2	26	739	3	31	723	3	22
28	690	2	25	749	3	31	730	3	21
29	696	2	24	759	3	31	737	3	21
30	702	3	23	769	3	31	744	3	21
31	708	3	22	779	3	31	751	3	21
32	714	3	21	789	3	31	758	3	22
33	720	3	21	800	3	31	765	3	22
34	725	3	20	811	3	32	773	3	22

Raw Score	Grade 5			Grade 7			Grade 8		
	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM	OPI Score	Perf. Level	CSEM
35	730	3	20	823	3	33	781	3	23
36	736	3	20	836	3	34	790	3	24
37	741	3	19	849	4	35	799	3	26
38	746	3	19	863	4	36	810	3	28
39	752	3	19	879	4	38	821	4	31
40	757	3	19	896	4	40	834	4	35
41	763	3	19	916	4	40	850	4	41
42	768	3	19	942	4	37	871	4	46
43	774	3	19	977	4	29	901	4	48
44	780	3	19	990	4	26	958	4	40
45	786	4	20	990	4	26	990	4	33
46	792	4	20						
47	798	4	20						
48	805	4	21						
49	812	4	22						
50	820	4	22						
51	828	4	24						
52	837	4	25						
53	847	4	27						
54	857	4	29						
55	870	4	33						
56	885	4	37						
57	904	4	40						
58	931	4	40						
59	982	4	29						
60	990	4	27						

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 would 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 the proportion of correct classifications across all levels. The overall consistency index is computed as the sum of the diagonal cells in a consistency table. Another way to express overall consistency is to use the kappa coefficient, as used in the inter-rater reliability studies in Section 3. 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 agreement due to 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 presents the overall accuracy and consistency indices for the Spring 2012 OCCT 3-8 tests.

Table 5-1. Estimates of Accuracy and Consistency in Performance Classifications

Subject	Grade	Accuracy	Consistency	Kappa (K)	False Positive	False Negative
Math	3	0.75	0.69	0.54	0.20	0.06
	4	0.77	0.71	0.54	0.08	0.14
	5	0.74	0.68	0.54	0.06	0.20
	6	0.74	0.71	0.57	0.15	0.11
	7	0.78	0.71	0.55	0.12	0.10
	8	0.77	0.70	0.57	0.10	0.13
Reading	3	0.87	0.81	0.61	0.06	0.08
	4	0.82	0.76	0.58	0.09	0.09
	5	0.76	0.71	0.52	0.11	0.13
	6	0.79	0.74	0.55	0.07	0.14
	7	0.72	0.66	0.44	0.23	0.05
	8	0.73	0.67	0.41	0.22	0.05
Science	5	0.80	0.75	0.54	0.10	0.11
	8	0.82	0.77	0.50	0.06	0.12
Social Studies	5	0.77	0.71	0.57	0.06	0.17
	7	0.80	0.73	0.53	0.09	0.11
	8	0.78	0.73	0.57	0.10	0.12

As shown in Table 5-1, the overall accuracy indices range between 72 and 87 percent, and overall consistency ranges between 66 and 81 for the Spring 2012 OCCT administration. Kappa coefficients range from 0.41 and 0.61. The rate of estimated false positives ranges from 6 to 23 and estimated false negative rates range from 5 to 20 percent.

Table 5-2 provides the accuracy, consistency, false positive, and false negative rates by cut score for Spring 2012. 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 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 being proficient); 2) the true achievement level is below the standard, but the observed test score is above the standard (i.e., a false positive); and 3) the true achievement 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, we estimate that 90 percent of grade 3 Mathematics students were correctly classified as proficient or not proficient based on their performance (scenario 1), 8 percent were considered proficient but their true performance is below the standard (scenario 2), and 2 percent were not considered proficient although their true performance is above the standard (scenario 3). Overall, the estimated rates for accurate classification are above 85% for the administration of all subjects and grades – 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	U / (L+P+A)	(U+L) / (P+A)	(U+L+P) / A	U / (L+P+A)	(U+L) / (P+A)	(U+L+P) / A	U / (L+P+A)	(U+L) / (P+A)	(U+L+P) / A
MATH	3	0.95	0.90	0.89	0.94	0.89	0.86	0.04	0.08	0.08	0.01	0.02	0.03
	4	0.95	0.91	0.91	0.94	0.89	0.87	0.03	0.01	0.04	0.01	0.08	0.05
	5	0.95	0.91	0.88	0.93	0.88	0.87	0.01	0.03	0.02	0.04	0.06	0.10
	6	0.92	0.90	0.91	0.91	0.88	0.90	0.06	0.08	0.01	0.02	0.02	0.07
	7	0.93	0.91	0.93	0.90	0.87	0.91	0.03	0.05	0.04	0.04	0.04	0.03
	8	0.94	0.90	0.92	0.93	0.88	0.89	0.04	0.02	0.05	0.02	0.09	0.03
READING	3	0.96	0.92	0.99	0.94	0.89	0.97	0.01	0.03	0.01	0.03	0.04	0.00
	4	0.94	0.90	0.98	0.92	0.87	0.97	0.01	0.06	0.02	0.05	0.04	0.00
	5	0.96	0.91	0.89	0.94	0.88	0.89	0.03	0.06	0.03	0.02	0.03	0.08
	6	0.95	0.91	0.94	0.93	0.88	0.92	0.03	0.02	0.03	0.03	0.08	0.04
	7	0.96	0.91	0.85	0.94	0.89	0.82	0.01	0.07	0.15	0.03	0.02	0.00
	8	0.96	0.91	0.86	0.94	0.90	0.82	0.01	0.08	0.14	0.03	0.02	0.00
SCIENCE	5	0.98	0.94	0.88	0.98	0.93	0.84	0.01	0.05	0.04	0.01	0.01	0.08
	8	0.97	0.94	0.91	0.97	0.91	0.89	0.01	0.02	0.03	0.01	0.05	0.06
SOCIAL STUDIES	5	0.95	0.89	0.92	0.92	0.88	0.90	0.02	0.01	0.03	0.03	0.09	0.05
	7	0.98	0.93	0.89	0.97	0.90	0.85	0.01	0.03	0.05	0.01	0.04	0.05
	8	0.94	0.91	0.92	0.93	0.88	0.91	0.01	0.02	0.07	0.05	0.06	0.01

Note: U =Unsatisfactory; L = Limited Knowledge; P = Proficient; and A = Advanced.

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

Section 6

Summary Statistics

6.1 Descriptive Statistics

The summary descriptive statistics of the scale scores for the Spring 2012 test-taking population appears in Table 6-1 through Table 6-4. The scales scores presented exclude invalid student cases.

Table 6-1. Descriptive Statistics of Scale Scores for Spring 2012 - Overall

Subject/Grade		Scale Score			
		N	Mean	SD	Median
Math	3	45237	741	88	745
	4	43951	746	88	747
	5	43478	742	86	745
	6	43228	734	80	739
	7	41329	736	79	740
	8	41015	727	82	731
Reading	3	44542	743	82	743
	4	43183	725	73	721
	5	42925	734	79	738
	6	43009	731	79	737
	7	41541	740	69	738
	8	41226	759	81	760
Science	5	43989	783	71	786
	8	42935	769	64	775
Social Studies	5	47169	730	77	736
	7	44890	783	90	789
	8	45794	736	85	737

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

Table 6-2. Descriptive Statistics of Scale Scores for Spring 2012 by Gender

		Female				Male			
Subject	**Grade**	**N**	**Mean**	**SD**	**Med.**	**N**	**Mean**	**SD**	**Med.**
Math	3	22550	739	88	736	22653	743	88	745
	4	21938	744	86	738	21986	748	89	747
	5	21608	740	83	736	21831	745	88	745
	6	21576	732	77	733	21614	737	82	739
	7	20758	735	77	740	20571	738	81	740
	8	20509	727	80	731	20495	728	84	731
Reading	3	22372	750	81	751	22136	735	82	743
	4	21738	728	71	729	21425	721	75	721
	5	21510	738	79	738	21388	731	79	729
	6	21541	735	75	737	21432	727	82	730
	7	20914	747	69	748	20627	733	70	730
	8	20618	767	82	760	20595	751	79	749
Science	5	21874	780	69	778	22088	786	73	786
	8	21290	767	61	768	21603	770	67	775
Social Studies	5	23017	727	72	730	24103	732	81	736
	7	22169	774	85	779	22721	792	93	800
	8	22305	730	79	730	23416	742	90	751

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

Table 6-3. Descriptive Statistics of Scale Scores for Spring 2012 by Race/Ethnicity

		African American				Native American				Hispanic				Asian			
Subject	**Grade**	**N**	**Mean**	**SD**	**Med.**	**N**	**Mean**	**SD**	**Med.**	**N**	**Mean**	**SD**	**Med.**	**N**	**Mean**	**SD**	**Med.**
Math	3	4087	698	91	704	7012	739	84	736	6746	715	86	719	852	780	94	777
	4	4027	703	88	708	7082	738	83	738	6287	725	86	723	843	798	98	790
	5	3883	701	86	706	7171	731	82	736	6007	722	86	721	876	796	91	794
	6	4063	695	85	700	7188	728	75	733	5693	715	77	720	840	786	88	789
	7	3836	701	84	706	6944	729	74	734	5108	711	79	720	790	790	83	788
	8	3928	693	90	707	6817	720	77	725	4973	703	85	707	794	786	90	778
Reading	3	4060	710	85	712	6888	742	77	743	6632	713	83	719	823	763	87	760
	4	3973	692	73	692	6957	718	69	721	6142	698	72	699	819	753	81	756
	5	3859	698	79	698	7083	726	76	729	5871	706	79	706	850	759	89	757
	6	4070	697	77	700	7140	725	76	730	5621	706	76	707	821	757	78	763
	7	3871	712	70	715	6958	735	65	730	5132	715	69	715	792	760	79	758
	8	3943	724	83	729	6845	756	76	760	4966	727	85	729	783	782	95	784
Science	5	4001	739	71	741	7257	777	69	778	6054	758	68	755	877	801	73	803
	8	4173	733	68	735	7077	765	59	768	5272	746	63	748	825	795	68	803
Social Studies	5	4409	687	81	696	7867	724	72	730	6483	706	77	714	897	759	83	763
	7	4330	733	95	739	7637	775	85	779	5590	756	89	759	811	826	89	836
	8	4547	698	85	702	7617	732	78	737	5605	712	83	710	830	780	92	781

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

Table 6-3. Descriptive Statistics of Scale Scores for Spring 2012 by Race/Ethnicity (cont.)

		Pacific Islander				White				Other			
Subject	**Grade**	**N**	**Mean**	**SD**	**Med.**	**N**	**Mean**	**SD**	**Med.**	**N**	**Mean**	**SD**	**Med.**
Math	3	127	723	95	719	23882	755	85	755	2531	741	87	745
	4	113	717	90	723	23143	760	85	756	2456	746	86	747
	5	117	727	88	728	23053	757	82	753	2371	740	84	736
	6	96	727	79	723	23051	747	77	746	2297	733	77	733
	7	112	722	82	734	22555	749	76	754	1984	737	79	740
	8	94	719	91	731	22393	739	77	737	2016	725	79	725
Reading	3	122	722	84	719	23534	756	79	760	2483	745	80	751
	4	110	707	75	714	22768	738	71	738	2414	725	72	729
	5	112	715	74	717	22805	750	76	747	2345	736	78	738
	6	95	720	86	730	22963	744	77	746	2299	730	79	737
	7	110	721	76	722	22690	751	67	748	1988	744	68	738
	8	104	713	106	721	22540	772	76	771	2045	761	79	760
Science	5	117	754	67	755	23279	798	67	794	2404	784	71	786
	8	109	737	76	748	23284	780	61	781	2195	769	63	775
Social Studies	5	119	720	68	725	24787	744	72	746	2607	728	75	730
	7	118	768	97	769	24243	800	85	800	2161	786	89	789
	8	111	717	99	730	24708	749	83	751	2376	735	86	737

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

Table 6-4. Descriptive Statistics of Scale Scores for Spring 2012 by Free/Reduced Lunch Status

		Free/Reduced Lunch = No				Free/Reduced Lunch = Yes			
Subject	**Grade**	**N**	**Mean**	**SD**	**Median**	**N**	**Mean**	**SD**	**Median**
Math	3	18383	767	86	765	26854	724	85	728
	4	17945	772	87	766	26006	728	83	730
	5	18434	768	83	772	25044	724	83	728
	6	18632	757	77	759	24596	717	77	720
	7	18855	760	75	760	22474	717	77	720
	8	19271	749	78	750	21744	708	81	713
Reading	3	18212	768	78	770	26330	726	80	727
	4	17723	748	73	747	25460	708	69	706
	5	18283	760	77	757	24642	715	76	713
	6	18613	753	77	754	24396	714	76	715
	7	18979	759	68	758	22562	724	67	722
	8	19373	781	77	771	21853	739	79	739
Science	5	18578	805	68	803	25411	766	69	770
	8	20097	787	61	788	22838	753	62	755
Social Studies	5	19163	757	71	763	28006	711	75	714
	7	19676	814	82	811	25214	759	88	769
	8	20801	762	82	765	24993	715	81	717

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 the test-taking population's performance in the Spring 2012 administration are presented in Table 6-5 (also, see Appendix B). The percentage distributions for each of the content areas are comparable to previous administrations (e.g., Spring 2011).

Table 6-5. Percentage of Students by Performance Level for Spring 2012

Subject/Grade		N	Unsatisfactory	Limited Knowledge	Proficient	Advanced
Math	3	45237	10.1%	20.0%	45.3%	24.7%
	4	43951	9.8%	17.2%	53.2%	19.8%
	5	43478	10.4%	19.4%	42.5%	27.8%
	6	43228	15.5%	13.7%	50.3%	20.5%
	7	41329	17.1%	12.6%	51.6%	18.7%
	8	41015	10.8%	20.6%	42.7%	26.0%
	All	258238	12.2%	17.3%	47.6%	22.9%
Reading	3	44542	11.3%	16.7%	68.9%	3.1%
	4	43183	15.1%	21.5%	59.0%	4.5%
	5	42925	10.5%	21.8%	56.2%	11.5%
	6	43009	14.1%	17.0%	60.2%	8.7%
	7	41541	11.7%	13.5%	55.9%	18.9%
	8	41226	8.5%	12.1%	61.4%	18.0%
	All	256426	11.9%	17.1%	60.3%	10.6%
Science	5	43989	2.6%	9.2%	58.5%	29.8%
	8	42935	3.5%	9.9%	70.0%	16.7%
	All	86924	3.0%	9.5%	64.2%	23.3%
Social Studies	5	47169	10.5%	20.2%	46.4%	22.9%
	7	44890	2.7%	14.1%	58.2%	25.0%
	8	45794	8.4%	21.6%	54.3%	15.7%
	All	137853	7.3%	18.7%	52.9%	21.2%

6.3 Conditional Standard Error of Measurement

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

$$CSEM\ (O_X \mid \theta) = \sqrt{\left[\sum_{X=0}^{MaxX} O_X^2\, p(X \mid \theta)\right] - \left[\sum_{X=0}^{MaxX} O_X \cdot p(X \mid \theta)\right]^2} \qquad (13)$$

where O_X is the observed scaled score for a particular number-correct score *X*, *θ* is the IRT achievement scale value conditioned on, and $p(\bullet)$ is the probability function. Pearson has implemented a computational approach for estimating *CSEM*(O_x | *θ*) in which *p*(*X* | *θ*) 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 in Table 4-3 to Table 4-8 for the Spring 2012 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:

$$SEM = \sqrt{\frac{\sum_{j}(CSEM_{j}^{2} \cdot N_{j})}{N_{T}}} \tag{14}$$

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 OCCT assessments. Table 6-6 presents the overall estimates of *SEM* for each of the content areas for the Spring 2012 administration.

Table 6-6. Overall Estimates of *SEM* by Test

Subject	Grade	SEM in OPI Units
Math	3	32
	4	33
	5	32
	6	29
	7	30
	8	32
Reading	3	32
	4	30
	5	32
	6	30
	7	35
	8	37
Science	5	30
	8	28
Social Studies	5	31
	7	36
	8	34

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

*Note: In 2012, field test sets in Mathematics and Reading included Common Core-aligned items as well as vertical linking items; these items are not included in the counts presented in this appendix.

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 2012 Test	Number of Items Field-Tested in 2012
Algebraic Reasoning: Patterns and Relationships	**7**	**7**	**6**
Algebra Patterns (1.1)	2	2	1
Equations (1.2)	2	2	3
Number Properties (1.3)	3	3	2
Number Sense and Operation	**20**	**20**	**17**
Number Sense (2.1)	10	10	8
Number Operations (2.2)	10	10	9
Geometry	**7**	**7**	**4**
Properties of shapes (3.1)	3	3	1
Spatial Reasoning (3.2)	2	2	2
Coordinate Geometry (3.3)	2	2	1
Measurement	**9**	**9**	**9**
Measurement (4.1)	4	4	3
Time and Temperature (4.2)	2	2	3
Money (4.3)	3	3	3
Data Analysis	**7**	**7**	**4**
Data Analysis (5.1)	4	4	2
Probability (5.2)	3	3	2
Total Test	**50**	**50**	**40**

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 2012 Test**	**Number of Items Field-Tested in 2012**
Algebraic Reasoning: Patterns and Relationships	**7**	**7**	**6**
Algebra Patterns (1.1)	3	3	2
Equations (1.2)	2	2	0
Number Properties (1.3)	2	2	4
Number Sense and Operation	**18**	**18**	**13**
Number Sense (2.1)	8	8	5
Number Operations (2.2)	10	10	8
Geometry	**9**	**9**	**7**
Lines (3.1)	2	2	1
Angles (3.2)	2	2	1
Polygons (3.3)	3	3	5
Transformations (3.4)	2	2	0
Measurement	**9**	**9**	**7**
Measurement (4.1)	5	5	4
Time and Temperature (4.2)	2	2	1
Money (4.3)	2	2	2
Data Analysis	**7**	**7**	**7**
Data Analysis (5.1)	2	2	1
Probability (5.2)	2	2	2
Central Tendency (5.3)	3	3	4
Total Test	**50**	**50**	**40**

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 2012 Test	Number of Items Field-Tested in 2012
Algebraic Reasoning: Patterns and Relationships	**13**	**13**	**5**
Algebra Patterns (1.1)	5	5	1
Equations (1.2)	4	4	1
Number Properties (1.3)	4	4	3
Number Sense and Operation	**16**	**16**	**15**
Number Sense (2.1)	8	8	8
Number Operations (2.2)	8	8	7
Geometry	**7**	**7**	**6**
Circles and Polygons (3.1)	4	4	3
Angles (3.2)	3	3	3
Measurement	**7**	**7**	**8**
Measurement (4.1)	5	5	4
Money (4.2)	2	2	4
Data Analysis	**7**	**6**	**6**
Data Analysis (5.1)	3	3	1
Probability (5.2)	2	2	3
Central Tendency (5.3)	2	1	2
Total Test	**50**	**49**	**40**

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 2012 Test	Number of Items Field-Tested in 2012
Algebraic Reasoning: Patterns and Relationships	**13**	**13**	**10**
Algebra Patterns (1.1)	4	4	2
Expressions and Equations (1.2)	4	4	3
Number Properties (1.3)	3	3	2
Solving Equations (1.4)	2	2	3
Number Sense and Operation	**15**	**15**	**12**
Number Sense (2.1)	5	5	0
Number Operations (2.2)	10	10	12
Geometry	**8**	**8**	**7**
Three Dimensional Figures (3.1)	2	2	2
Congruent and Similar Figures (3.2)	2	2	1
Coordinate Geometry (3.3)	4	4	4
Measurement	**7**	**7**	**5**
Circles (4.1)	4	4	5
Conversions (4.2)	3	3	0
Data Analysis	**7**	**7**	**6**
Data Analysis (5.1)	3	3	1
Probability (5.2)	2	2	3
Central Tendency (5.3)	2	2	2
Total Test	**50**	**50**	**40**

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 2012 Test**	**Number of Items Field-Tested in 2012**
Algebraic Reasoning: Patterns and Relationships	**15**	**15**	**14**
Linear Relationships (1.1)	5	5	6
Solving Equations (1.2)	5	5	4
Solving and Graphing Inequalities (1.3)	5	5	4
Number Sense and Operation	**11**	**11**	**8**
Number Sense (2.1)	5	5	5
Number Operations (2.2)	6	6	3
Geometry	**8**	**8**	**4**
Classifying Figures (3.1)	1-3	2	1
Lines and Angles (3.2)	1-3	2	2
Transformations (3.3)	4	4	1
Measurement	**9**	**9**	**7**
Perimeter and Area (4.1)	5	5	2
Circles (4.2)	2	2	2
Composite Figures (4.3)	2	2	3
Data Analysis	**7**	**7**	**7**
Data Analysis (5.1)	2	2	3
Probability (5.2)	2	2	1
Central Tendency (5.3)	3	3	3
Total Test	**50**	**50**	**40**

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 2012 Test	Number of Items Field-Tested in 2012
Algebraic Reasoning: Patterns and Relationships	**16**	**16**	**14**
Equations (1.1)	10-12	11	11
Inequalities (1.2)	4-6	5	3
Number Sense and Operation	**11**	**11**	**10**
Number Sense (2.1)	3-4	4	3
Number Operations (2.2)	7-8	7	7
Geometry	**9**	**9**	**6**
Three Dimensional Figures (3.1)	5	5	4
Pythagorean Theorem (3.2)	4	4	2
Measurement	**7**	**7**	**4**
Surface Area and Volume (4.1)	3	3	1
Ratio and Proportions (4.2)	2	2	1
Composite Figures (4.3)	2	2	2
Data Analysis	**7**	**7**	**6**
Data Analysis (5.1)	3	3	2
Central Tendency (5.3)	4	4	4
Total Test	**50**	**50**	**40**

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 2012 Test**	**Number of Items Field-Tested in 2012**
Vocabulary	**12**	**12**	**10**
Words in Context (2.1)	2-4	2	3
Affixes, Roots, and Stems (2.2)	2-4	3	1
Synonyms, Antonyms, and Homonyms (2.3)	2-4	3	3
Using Resource Materials (2.4)	2-4	4	3
Comprehension/Critical Literacy	**24**	**24**	**19**
Literal Understanding (4.1)	5	5	4
Inferences and Interpretation (4.2)	7	7	6
Summary and Generalization (4.3)	6	6	5
Analysis and Evaluation (4.4)	6	6	4
Literature	**8**	**8**	**5**
Literary Elements (5.2)	3-4	3	3
Figurative Language/Sound Devices (5.3)	4-5	5	2
Research and Information	**6**	**6**	**6**
Accessing Information (6.1)	6	6	6
Total Test	**50**	**50**	**40**

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 2012 Test	Number of Items Field-Tested in 2012
Vocabulary	**12**	**12**	**9**
Words in Context (1.1)	4	4	3
Affixes, Roots, and Stems (1.2)	4	4	3
Synonyms, Antonyms, and Homonyms (1.3)	4	4	3
Comprehension/Critical Literacy	**23**	**23**	**17**
Literal Understanding (3.1)	4	4	4
Inferences and Interpretation (3.2)	6	6	5
Summary and Generalization (3.3)	7	7	4
Analysis and Evaluation (3.4)	6	6	4
Literature	**9**	**9**	**7**
Literary Elements (4.2)	5	5	5
Figurative Language/Sound Devices (4.3)	4	4	2
Research and Information	**6**	**6**	**7**
Accessing Information (5.1)	6	6	7
Total Test	**50**	**50**	**40**

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 2012 Test	Number of Items Field-Tested in 2012
Vocabulary	**12**	**12**	**9**
Words in Context (1.1)	4	4	3
Affixes, Roots, and Stems (1.2)	4	4	3
Synonyms, Antonyms, and Homonyms (1.3)	4	4	3
Comprehension/Critical Literacy	**20**	**19**	**16**
Literal Understanding (3.1)	4	4	1
Inferences and Interpretation (3.2)	4-6	5	4
Summary and Generalization (3.3)	4-6	5	6
Analysis and Evaluation (3.4)	4-6	5	5
Literature	**12**	**12**	**8**
Literary Genre (4.1)	4	4	2
Literary Elements (4.2)	4	4	3
Figurative Language/Sound Devices (4.3)	4	4	3
Research and Information	**6**	**7**	**7**
Accessing Information (5.1)	2-4	4	3
Interpreting Information (5.2)	2-4	3	4
Total Test	**50**	**50**	**40**

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 2012 Test	Number of Items Field-Tested in 2012
Vocabulary	**8**	**8**	**6**
Words in Context (1.1)	4	4	4
Word Origins (1.2)	4	4	2
Comprehension/Critical Literacy	**20**	**19**	**15**
Literal Understanding (3.1)	4	4	2
Inferences and Interpretation (3.2)	4-6	5	5
Summary and Generalization (3.3)	4-6	5	4
Analysis and Evaluation (3.4)	4-6	5	4
Literature	**14**	**15**	**12**
Literary Genres (4.1)	4	4	4
Literary Elements (4.2)	4-6	5	3
Figurative Language/Sound Devices (4.3)	4-6	6	5
Research and Information	**8**	**8**	**7**
Accessing Information (5.1)	4	4	5
Interpreting Information (5.2)	4	4	2
Total Test	**50**	**50**	**40**

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 2012 Test	Number of Items Field-Tested in 2012
Vocabulary	**10**	**10**	**8**
Words in Context (1.1)	3-4	3	2
Word Origins (1.2)	3-4	3	3
Idioms and Comparisons (1.3)	3-4	4	3
Comprehension/Critical Literacy	**20**	**20**	**14**
Literal Understanding (3.1)	4-5	5	3
Inferences and Interpretation (3.2)	4-6	5	4
Summary and Generalization (3.3)	4-6	5	3
Analysis and Evaluation (3.4)	4-6	5	4
Literature	**12**	**12**	**12**
Literary Genres (4.1)	4	4	4
Literary Elements (4.2)	4	4	4
Figurative Language/Sound Devices (4.3)	4	4	4
Research and Information	**8**	**8**	**6**
Accessing Information (5.1)	4	4	3
Interpreting Information (5.2)	4	4	4
Total Test	**50**	**50**	**40**

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 2012 Test	Number of Items Field-Tested in 2012
Vocabulary	**6**	**6**	**5**
Words in Context (1.1)	2	2	3
Word Origins (1.2)	2	2	0
Idioms and Comparisons (1.3)	2	2	2
Comprehension/Critical Literacy	**21**	**21**	**17**
Literal Understanding (3.1)	4-5	5	4
Inferences and Interpretation (3.2)	4-6	5	5
Summary and Generalization (3.3)	5-7	5	4
Analysis and Evaluation (3.4)	6-8	6	4
Literature	**15**	**15**	**10**
Literary Genre (4.1)	4-5	5	3
Literary Elements (4.2)	5-7	5	3
Figurative Language/Sound Devices (4.3)	4-6	5	4
Research and Information	**8**	**8**	**8**
Accessing Information (5.1)	4	4	4
Interpreting Information (5.2)	4	4	4
Total Test	**50**	**50**	**40**

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 2012 Test	Number of Items Field-Tested in 2012
Process Standards			
Observe and Measure	**10**	**10**	**16**
SI Metric (P1.1)	5	5	7
Similar/different characteristics (P1.2)	5	5	9
Classify	**10**	**10**	**20**
Observable properties (P2.1)	5	5	10
Serial order (P2.2)	5	5	10
Experiment	**11**	**11**	**15**
Experimental design (P3.2)	7	7	10
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	5	10
Prediction based on data (P4.3)	4	5	9
Explanations based on data (P4.4)	4	4	10
Total Test	**45**	**45**	**80**
Content Standards			
Properties of Matter and Energy	**18**	**18**	**30**
Matter has physical properties (1.1)	6	6	9
Physical properties can be measured (1.2)	6	6	8
Energy can be transferred (1.3)	6	6	6
Potential/Kinetic Energy (1.4)	0	0	7
Organisms and Environments	**12**	**12**	**20**
Dependence upon community (2.1)	6	6	11
Individual organism and species survival (2.2)	6	6	9
Structures of the Earth and the Solar System	**11**	**11**	**25**
Properties of Soils (3.1)	0	0	6
Weather patterns (3.2)	6	7	9
Earth as a planet (3.3)	5	4	10
Total Test	**41**	**41**	**75**

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

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 2012 Test**	**Number of Items Field-Tested in 2012**
Process Standards			
Observe and Measure	**8**	**6**	**15**
Qualitative/quantitative observations/changes (P1.1)	4	3	9
SI (metrics) units/appropriate tools (P1.2 and P1.3)	4	3	6
Classify	**8**	**10**	**17**
Classification system (P2.1)	4	6	9
Properties ordered (P2.2)	4	4	8
Experiment	**16**	**16**	**26**
Experimental design (P3.2)	6	6	10
Identify variables (P3.3)	6	6	11
Hazards/practice safety (P3.6)	4	4	5
Interpret and Communicate	**13**	**13**	**22**
Data tables/line/bar/trend and circle graphs (P4.2)	7	7	12
Explanations/prediction (P4.3)	6	6	10
Total Test	**45**	**45**	**80**
Content Standards			
Properties and Chemical Changes in Matter	**7-8**	**8**	**15**
Chemical reactions (1.1)	3-4	4	8
Conservation of matter (1.2)	3-4	4	7
Motion and Forces	**8**	**8**	**14**
Motion of an object (2.1)	4	4	7
Object subjected to a force (2.2)	4	4	7
Diversity and Adaptations of Organisms	**9**	**9**	**13**
Classification (3.1)	5	5	7
Internal and external structures (3.2)	4	4	6
Structures/Forces of the Earth/Solar System	**8**	**7**	**19**
Landforms result from constructive and destructive forces (4.1)	4	4	7
Rock cycle (4.2)	4	3	6
Global Weather Patterns (4.3)	0	0	6
Earth's History	**7-8**	**9**	**14**
Catastrophic events (5.1)	3-4	5	6
Fossil evidence (5.2)	3-4	4	8
Total Test	**41**	**41**	**75**

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

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 2012 Test	Number of Items Field-Tested in 2012
Early Exploration	**8**	**8**	**12**
Expeditions (2.1)	4	4	4
Native American Reaction (2.2)	4	4	8
Colonial America	**12**	**12**	**8**
Settlements and Migration (3.1)	4	4	1
Colonial Life (3.2)	4	4	5
Individuals and Groups (3.3)	4	4	2
American Revolution	**12**	**12**	**10**
Causes and Results (4.1)	4	4	5
Declaration of Independence (4.3)	4	4	1
Individuals (4.4)	4	4	4
Early Federal Period	**8**	**8**	**7**
Constitutional Provisions (5.2)	4	4	4
Ratification and Rights (5.3)	4	4	3
Geographic Skills	**20**	**20**	**43**
Maps/Charts/Graphs Usage (7.1)	7	7	14
Human/Environment Interaction (7.2)	5	5	11
Historical Places (7.3)	4	4	9
Westward Movement (7.4)	4	4	9
Total Test	**60**	**60**	**80**

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 2012 Test	Number of Items Field-Tested in 2012
Geographic Tools/Geography Skills	**9**	**9**	**12**
Map Concepts (1.2)	4	4	9
Maps/Charts/Graphs (6.1)	5	5	3
Regions	**12**	**12**	**20**
Regional Characteristics (2.1)	4	4	6
Conflict/Cooperation (2.2)	4	4	8
Locations (2.4)	4	4	6
Physical Systems	**8**	**8**	**13**
Climate/Weather (3.2)	4	4	8
Natural Disasters (3.3)	4	4	5
Human Systems	**8**	**8**	**17**
World Cultures (4.1)	4	4	8
Population Issues (4.5)	4	4	9
Human/Environment Interaction	**8**	**8**	**18**
Natural Resources (5.1)	4	4	10
Human Modification (5.2)	4	4	8
Total Test	**45**	**45**	**80**

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 2012 Test**	**Number of Items Field-Tested in 2012**
Social Studies Process Skills (1.0)	**6**	**6**	**13**
Causes and Results of the American Revolution (3.0/4.0)	**10**	**10**	**12**
Causes of the American Revolution (3.0)	5	5	8
Results of the American Revolution (4.0)	5	5	4
Governing Documents/Early Federal Period (5.0)	**6**	**6**	**10**
Moving Toward the Civil War (6.0/10.0)	**9**	**9**	**20**
Northern/Southern Economic Growth (6.0)	4	4	7
Eve of War (10.0)	5	5	13
Early 19th Century America (7.0/8.0)	**8**	**8**	**14**
Jacksonian Era (7.0)	4	4	10
Cultural Growth and Reform (8.0)	4	4	4
Westward Movement (9.0)	**6**	**6**	**11**
Total Test	**45**	**45**	**80**

Appendix B

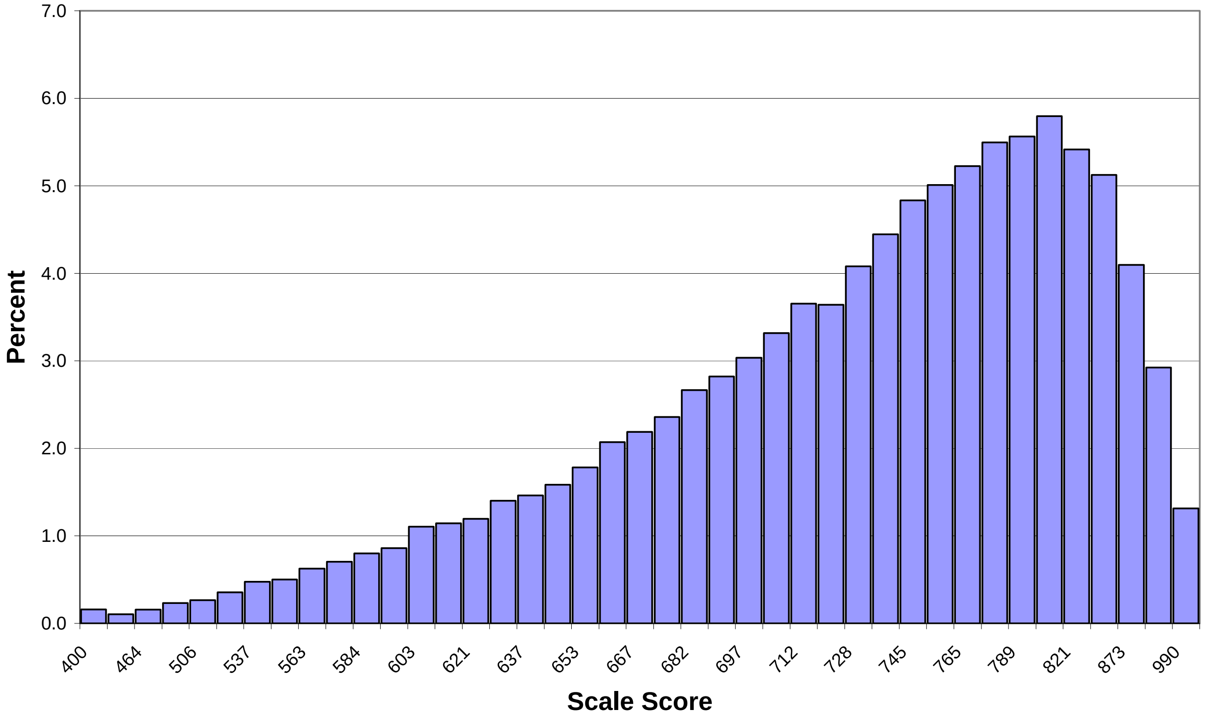
Scale Score Distributions for Spring 2012

Mathematics Grade 3 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	73	0.2	73	0.2
435	48	0.1	121	0.3
464	71	0.2	192	0.4
487	106	0.2	298	0.7
506	120	0.3	418	0.9
523	161	0.4	579	1.3
537	215	0.5	794	1.8
550	227	0.5	1,021	2.3
563	284	0.6	1,305	2.9
574	319	0.7	1,624	3.6
584	362	0.8	1,986	4.4
594	389	0.9	2,375	5.3
603	500	1.1	2,875	6.4
612	517	1.1	3,392	7.5
621	541	1.2	3,933	8.7
629	634	1.4	4,567	10.1
637	661	1.5	5,228	11.6
645	717	1.6	5,945	13.1
653	806	1.8	6,751	14.9
660	937	2.1	7,688	17.0
667	990	2.2	8,678	19.2
675	1067	2.4	9,745	21.5
682	1206	2.7	10,951	24.2
689	1276	2.8	12,227	27.0
697	1374	3.0	13,601	30.1
704	1500	3.3	15,101	33.4
712	1653	3.7	16,754	37.0
719	1647	3.6	18,401	40.7
728	1845	4.1	20,246	44.8
736	2011	4.4	22,257	49.2
745	2187	4.8	24,444	54.0
755	2266	5.0	26,710	59.0
765	2364	5.2	29,074	64.3
777	2486	5.5	31,560	69.8

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
789	2517	5.6	34,077	75.3
804	2622	5.8	36,699	81.1
821	2450	5.4	39,149	86.5
843	2318	5.1	41,467	91.7
873	1853	4.1	43,320	95.8
923	1323	2.9	44,643	98.7
990	594	1.3	45,237	100.0

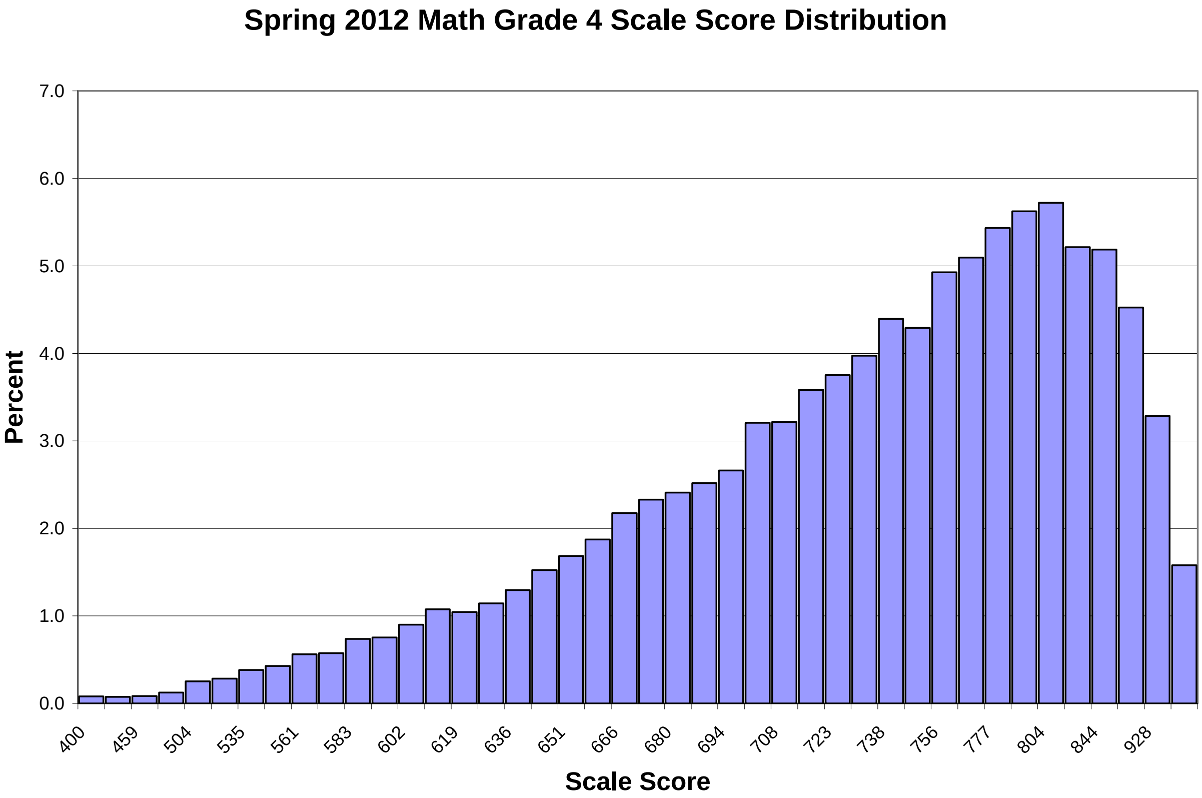
Spring 2012 Math Grade 3 Scale Score Distribution



Mathematics Grade 4 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	35	0.1	35	0.1
426	33	0.1	68	0.2
459	38	0.1	106	0.2
484	55	0.1	161	0.4
504	111	0.3	272	0.6
521	124	0.3	396	0.9
535	169	0.4	565	1.3
549	188	0.4	753	1.7
561	247	0.6	1,000	2.3
572	252	0.6	1,252	2.8
583	324	0.7	1,576	3.6
593	332	0.8	1,908	4.3
602	396	0.9	2,304	5.2
611	473	1.1	2,777	6.3
619	460	1.0	3,237	7.4
628	503	1.1	3,740	8.5
636	570	1.3	4,310	9.8
643	670	1.5	4,980	11.3
651	741	1.7	5,721	13.0
658	823	1.9	6,544	14.9
666	956	2.2	7,500	17.1
673	1024	2.3	8,524	19.4
680	1059	2.4	9,583	21.8
687	1106	2.5	10,689	24.3
694	1170	2.7	11,859	27.0
701	1409	3.2	13,268	30.2
708	1414	3.2	14,682	33.4
715	1575	3.6	16,257	37.0
723	1650	3.8	17,907	40.7
730	1747	4.0	19,654	44.7
738	1932	4.4	21,586	49.1
747	1886	4.3	23,472	53.4
756	2165	4.9	25,637	58.3
766	2239	5.1	27,876	63.4
777	2388	5.4	30,264	68.9

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
790	2472	5.6	32,736	74.5
804	2515	5.7	35,251	80.2
821	2292	5.2	37,543	85.4
844	2280	5.2	39,823	90.6
874	1989	4.5	41,812	95.1
928	1445	3.3	43,257	98.4
990	694	1.6	43,951	100.0

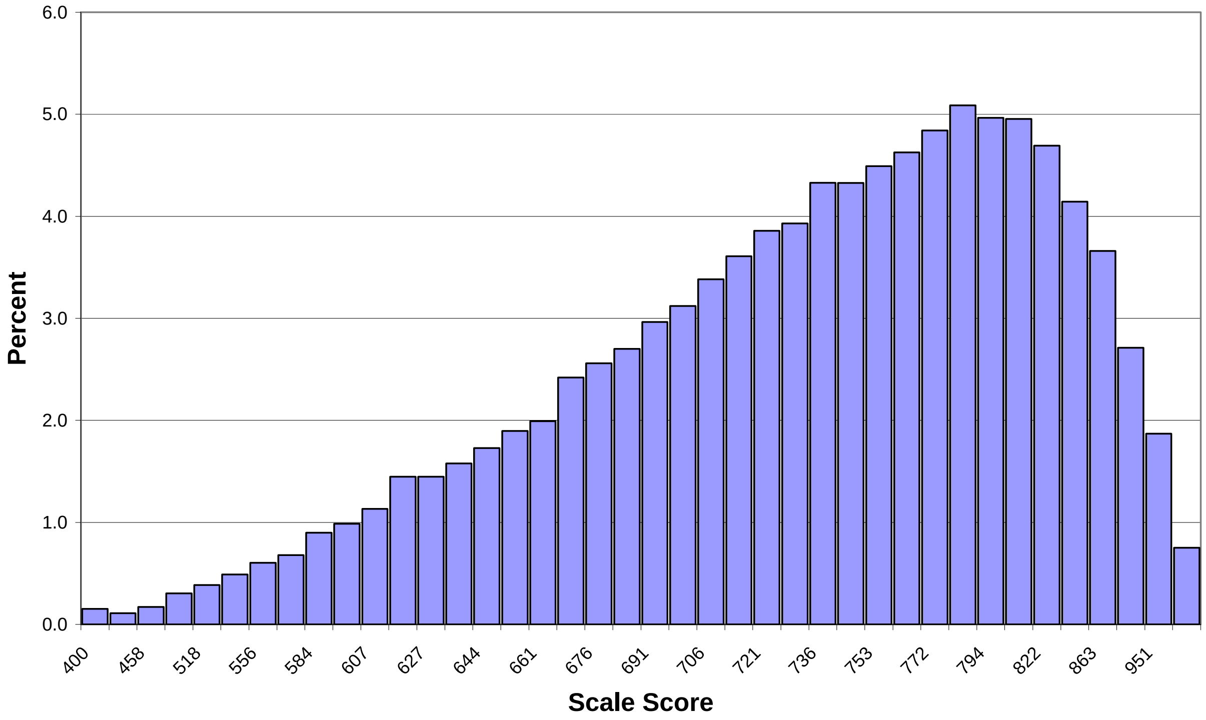


Mathematics Grade 5 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	66	0.2	66	0.2
408	48	0.1	114	0.3
458	75	0.2	189	0.4
492	133	0.3	322	0.7
518	168	0.4	490	1.1
538	213	0.5	703	1.6
556	263	0.6	966	2.2
571	295	0.7	1,261	2.9
584	391	0.9	1,652	3.8
596	429	1.0	2,081	4.8
607	492	1.1	2,573	5.9
617	630	1.4	3,203	7.4
627	630	1.4	3,833	8.8
636	686	1.6	4,519	10.4
644	751	1.7	5,270	12.1
653	824	1.9	6,094	14.0
661	866	2.0	6,960	16.0
668	1052	2.4	8,012	18.4
676	1113	2.6	9,125	21.0
683	1174	2.7	10,299	23.7
691	1288	3.0	11,587	26.7
698	1357	3.1	12,944	29.8
706	1471	3.4	14,415	33.2
713	1569	3.6	15,984	36.8
721	1677	3.9	17,661	40.6
728	1709	3.9	19,370	44.6
736	1882	4.3	21,252	48.9
745	1881	4.3	23,133	53.2
753	1953	4.5	25,086	57.7
762	2012	4.6	27,098	62.3
772	2105	4.8	29,203	67.2
783	2212	5.1	31,415	72.3
794	2158	5.0	33,573	77.2
807	2154	5.0	35,727	82.2
822	2040	4.7	37,767	86.9

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
840	1801	4.1	39,568	91.0
863	1591	3.7	41,159	94.7
894	1179	2.7	42,338	97.4
951	813	1.9	43,151	99.2
990	327	0.8	43,478	100.0

Spring 2012 Math Grade 5 Scale Score Distribution

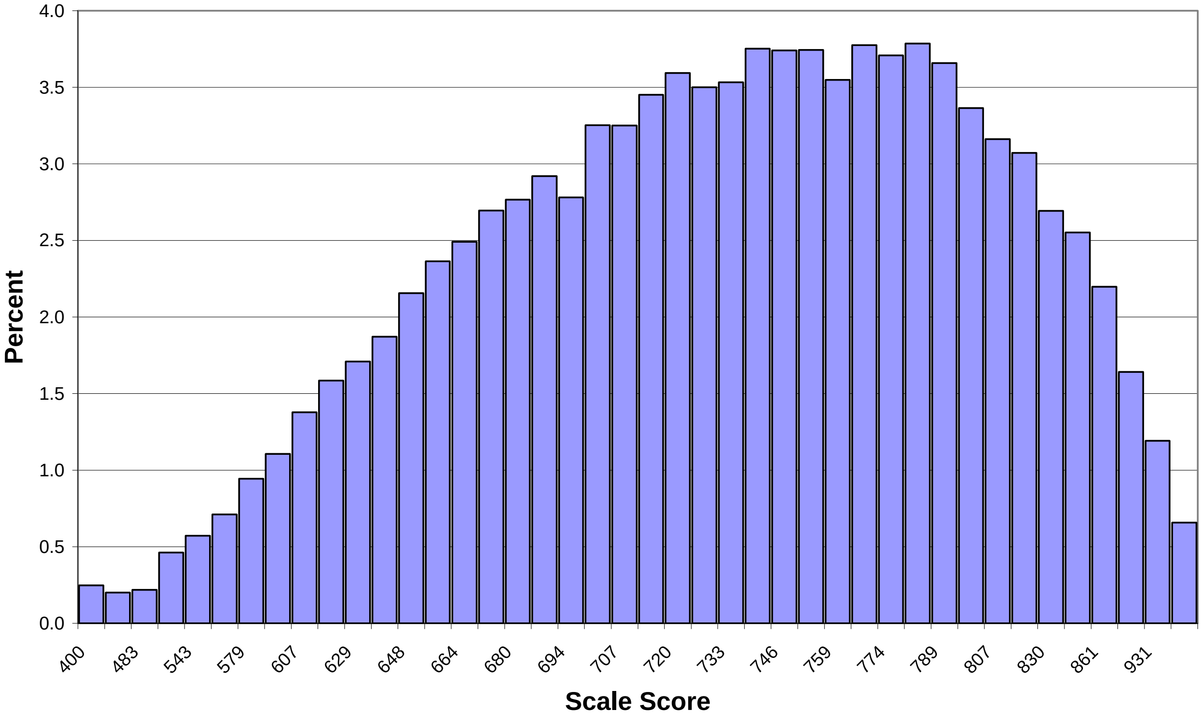


Mathematics Grade 6 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	107	0.2	107	0.2
424	87	0.2	194	0.4
483	95	0.2	289	0.7
518	200	0.5	489	1.1
543	247	0.6	736	1.7
563	307	0.7	1,043	2.4
579	408	0.9	1,451	3.4
594	478	1.1	1,929	4.5
607	596	1.4	2,525	5.8
618	685	1.6	3,210	7.4
629	739	1.7	3,949	9.1
639	809	1.9	4,758	11.0
648	932	2.2	5,690	13.2
656	1022	2.4	6,712	15.5
664	1077	2.5	7,789	18.0
672	1165	2.7	8,954	20.7
680	1196	2.8	10,150	23.5
687	1262	2.9	11,412	26.4
694	1202	2.8	12,614	29.2
700	1406	3.3	14,020	32.4
707	1405	3.3	15,425	35.7
714	1492	3.5	16,917	39.1
720	1553	3.6	18,470	42.7
726	1513	3.5	19,983	46.2
733	1527	3.5	21,510	49.8
739	1622	3.8	23,132	53.5
746	1617	3.7	24,749	57.3
752	1618	3.7	26,367	61.0
759	1534	3.5	27,901	64.5
766	1632	3.8	29,533	68.3
774	1603	3.7	31,136	72.0
781	1636	3.8	32,772	75.8
789	1581	3.7	34,353	79.5
798	1454	3.4	35,807	82.8
807	1367	3.2	37,174	86.0

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
818	1328	3.1	38,502	89.1
830	1164	2.7	39,666	91.8
844	1103	2.6	40,769	94.3
861	950	2.2	41,719	96.5
887	710	1.6	42,429	98.2
931	515	1.2	42,944	99.3
990	284	0.7	43,228	100.0

Spring 2012 Math Grade 6 Scale Score Distribution

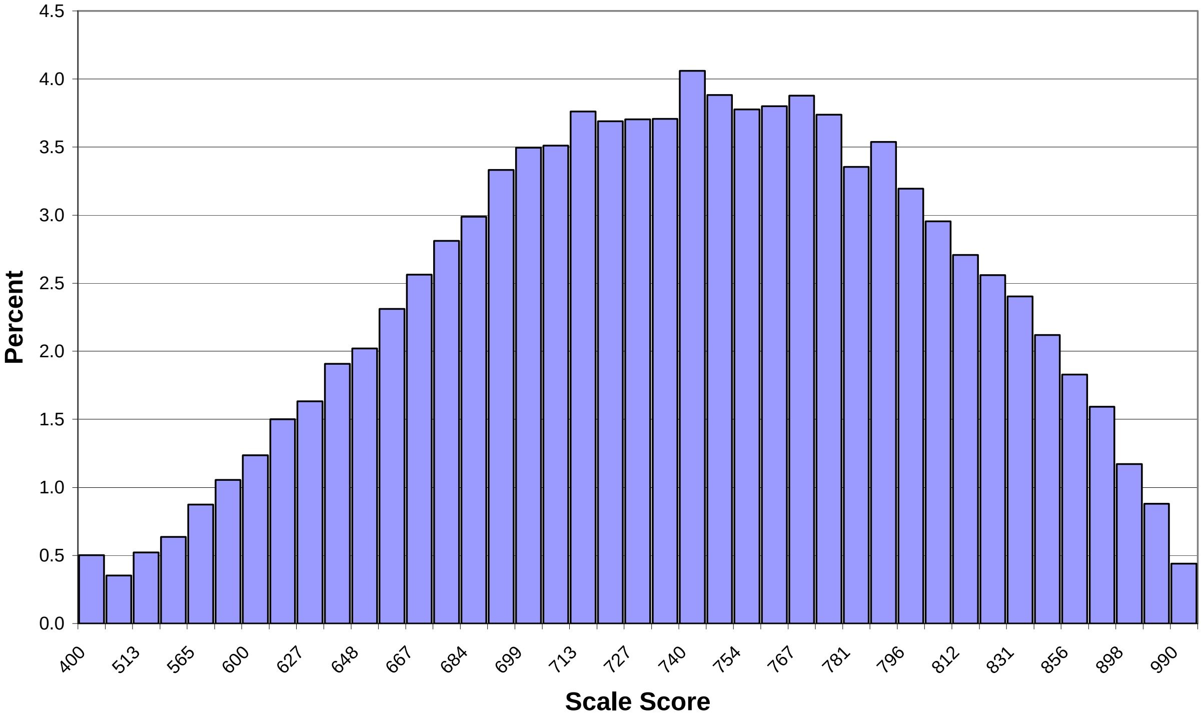


Mathematics Grade 7 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	208	0.5	208	0.5
471	146	0.4	354	0.9
513	216	0.5	570	1.4
542	263	0.6	833	2.0
565	361	0.9	1,194	2.9
584	436	1.1	1,630	3.9
600	511	1.2	2,141	5.2
614	620	1.5	2,761	6.7
627	675	1.6	3,436	8.3
638	789	1.9	4,225	10.2
648	835	2.0	5,060	12.2
658	955	2.3	6,015	14.6
667	1059	2.6	7,074	17.1
676	1162	2.8	8,236	19.9
684	1235	3.0	9,471	22.9
691	1377	3.3	10,848	26.2
699	1445	3.5	12,293	29.7
706	1451	3.5	13,744	33.3
713	1555	3.8	15,299	37.0
720	1525	3.7	16,824	40.7
727	1531	3.7	18,355	44.4
734	1532	3.7	19,887	48.1
740	1678	4.1	21,565	52.2
747	1605	3.9	23,170	56.1
754	1561	3.8	24,731	59.8
760	1571	3.8	26,302	63.6
767	1603	3.9	27,905	67.5
774	1545	3.7	29,450	71.3
781	1386	3.4	30,836	74.6
788	1462	3.5	32,298	78.1
796	1320	3.2	33,618	81.3
804	1221	3.0	34,839	84.3
812	1119	2.7	35,958	87.0
821	1058	2.6	37,016	89.6
831	993	2.4	38,009	92.0

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
843	876	2.1	38,885	94.1
856	756	1.8	39,641	95.9
874	658	1.6	40,299	97.5
898	484	1.2	40,783	98.7
940	364	0.9	41,147	99.6
990	182	0.4	41,329	100.0

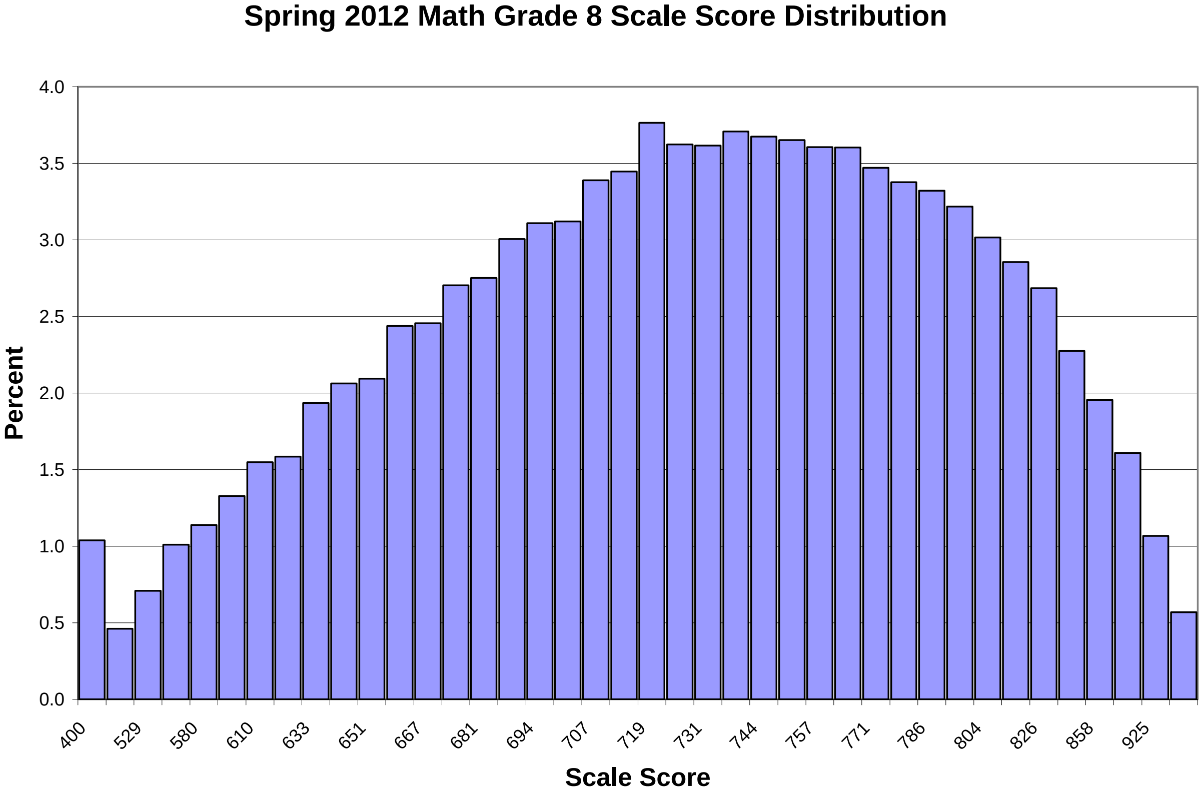
Spring 2012 Math Grade 7 Scale Score Distribution



Mathematics Grade 8 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	426	1.0	426	1.0
479	189	0.5	615	1.5
529	291	0.7	906	2.2
559	414	1.0	1,320	3.2
580	467	1.1	1,787	4.4
597	545	1.3	2,332	5.7
610	635	1.5	2,967	7.2
622	650	1.6	3,617	8.8
633	794	1.9	4,411	10.8
642	846	2.1	5,257	12.8
651	859	2.1	6,116	14.9
659	1000	2.4	7,116	17.3
667	1007	2.5	8,123	19.8
674	1109	2.7	9,232	22.5
681	1129	2.8	10,361	25.3
687	1233	3.0	11,594	28.3
694	1275	3.1	12,869	31.4
700	1280	3.1	14,149	34.5
707	1390	3.4	15,539	37.9
713	1414	3.4	16,953	41.3
719	1544	3.8	18,497	45.1
725	1486	3.6	19,983	48.7
731	1483	3.6	21,466	52.3
737	1521	3.7	22,987	56.0
744	1507	3.7	24,494	59.7
750	1498	3.7	25,992	63.4
757	1479	3.6	27,471	67.0
764	1478	3.6	28,949	70.6
771	1424	3.5	30,373	74.1
778	1385	3.4	31,758	77.4
786	1362	3.3	33,120	80.8
794	1320	3.2	34,440	84.0
804	1237	3.0	35,677	87.0
814	1171	2.9	36,848	89.8
826	1101	2.7	37,949	92.5

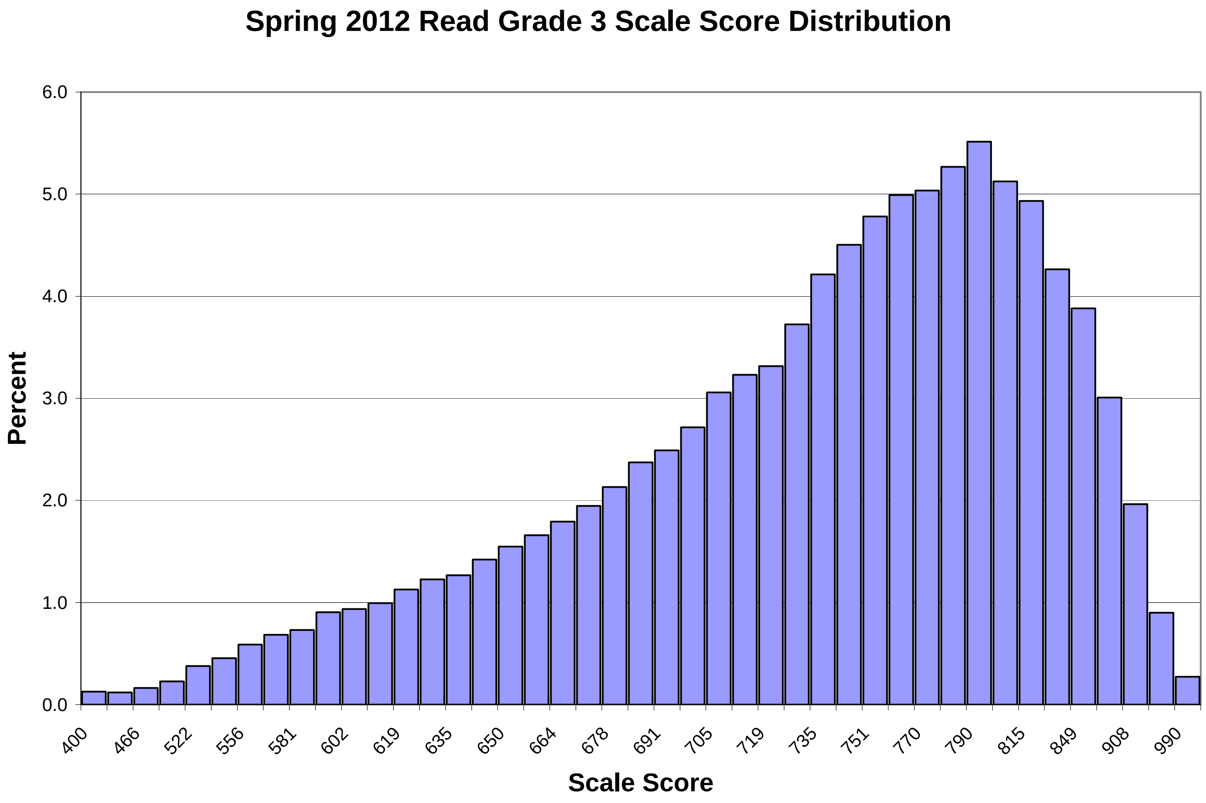
Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
840	933	2.3	38,882	94.8
858	802	2.0	39,684	96.8
883	660	1.6	40,344	98.4
925	438	1.1	40,782	99.4
990	233	0.6	41,015	100.0



Reading Grade 3 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	56	0.1	56	0.1
409	54	0.1	110	0.2
466	73	0.2	183	0.4
498	101	0.2	284	0.6
522	168	0.4	452	1.0
540	203	0.5	655	1.5
556	262	0.6	917	2.1
569	304	0.7	1,221	2.7
581	325	0.7	1,546	3.5
592	403	0.9	1,949	4.4
602	417	0.9	2,366	5.3
611	443	1.0	2,809	6.3
619	503	1.1	3,312	7.4
627	547	1.2	3,859	8.7
635	565	1.3	4,424	9.9
643	633	1.4	5,057	11.4
650	689	1.5	5,746	12.9
657	739	1.7	6,485	14.6
664	798	1.8	7,283	16.4
671	867	1.9	8,150	18.3
678	950	2.1	9,100	20.4
684	1057	2.4	10,157	22.8
691	1109	2.5	11,266	25.3
698	1210	2.7	12,476	28.0
705	1362	3.1	13,838	31.1
712	1439	3.2	15,277	34.3
719	1477	3.3	16,754	37.6
727	1659	3.7	18,413	41.3
735	1877	4.2	20,290	45.6
743	2006	4.5	22,296	50.1
751	2130	4.8	24,426	54.8
760	2224	5.0	26,650	59.8
770	2243	5.0	28,893	64.9
780	2346	5.3	31,239	70.1

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
790	2456	5.5	33,695	75.6
802	2283	5.1	35,978	80.8
815	2198	4.9	38,176	85.7
830	1899	4.3	40,075	90.0
849	1729	3.9	41,804	93.9
873	1340	3.0	43,144	96.9
908	875	2.0	44,019	98.8
975	401	0.9	44,420	99.7
990	122	0.3	44,542	100.0

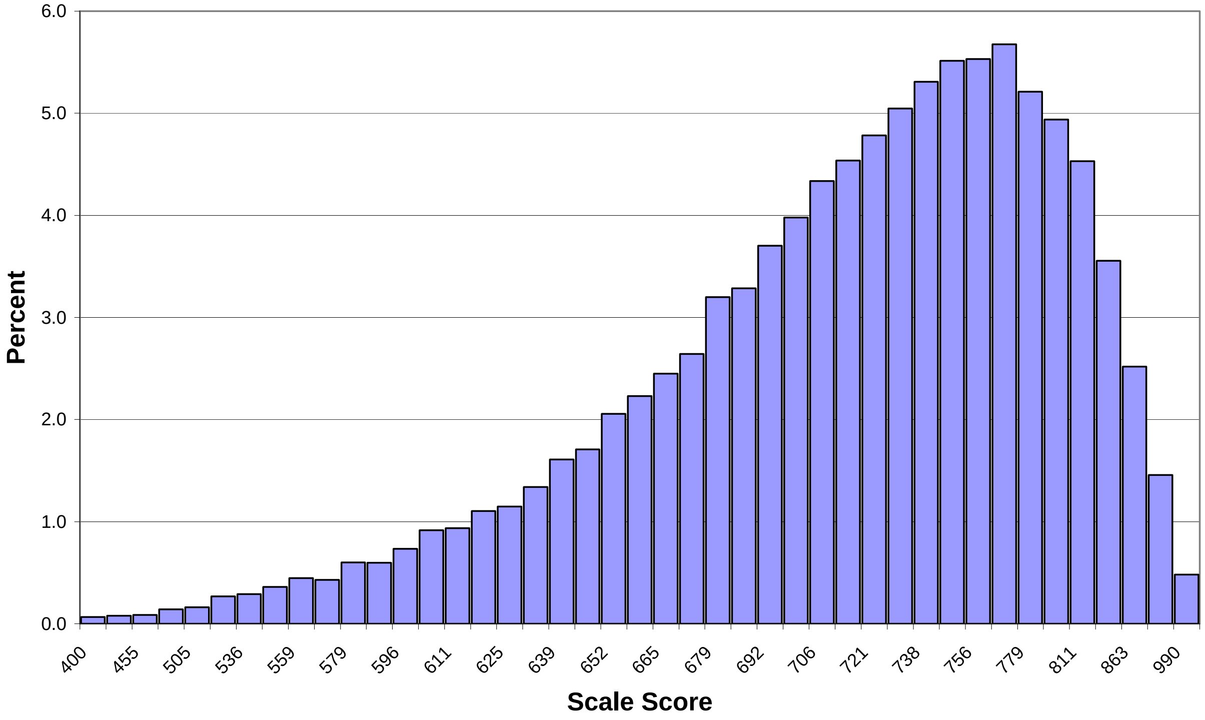


Reading Grade 4 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	29	0.1	29	0.1
406	34	0.1	63	0.1
455	37	0.1	100	0.2
484	61	0.1	161	0.4
505	70	0.2	231	0.5
522	116	0.3	347	0.8
536	125	0.3	472	1.1
548	156	0.4	628	1.5
559	193	0.4	821	1.9
569	186	0.4	1,007	2.3
579	260	0.6	1,267	2.9
587	258	0.6	1,525	3.5
596	317	0.7	1,842	4.3
603	395	0.9	2,237	5.2
611	405	0.9	2,642	6.1
618	477	1.1	3,119	7.2
625	496	1.1	3,615	8.4
632	578	1.3	4,193	9.7
639	695	1.6	4,888	11.3
646	738	1.7	5,626	13.0
652	888	2.1	6,514	15.1
659	963	2.2	7,477	17.3
665	1058	2.5	8,535	19.8
672	1141	2.6	9,676	22.4
679	1382	3.2	11,058	25.6
685	1419	3.3	12,477	28.9
692	1599	3.7	14,076	32.6
699	1718	4.0	15,794	36.6
706	1873	4.3	17,667	40.9
714	1959	4.5	19,626	45.4
721	2066	4.8	21,692	50.2
729	2179	5.0	23,871	55.3
738	2292	5.3	26,163	60.6
747	2381	5.5	28,544	66.1
756	2389	5.5	30,933	71.6

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
767	2451	5.7	33,384	77.3
779	2250	5.2	35,634	82.5
794	2133	4.9	37,767	87.5
811	1956	4.5	39,723	92.0
832	1535	3.6	41,258	95.5
863	1088	2.5	42,346	98.1
915	629	1.5	42,975	99.5
990	208	0.5	43,183	100.0

Spring 2012 Read Grade 4 Scale Score Distribution

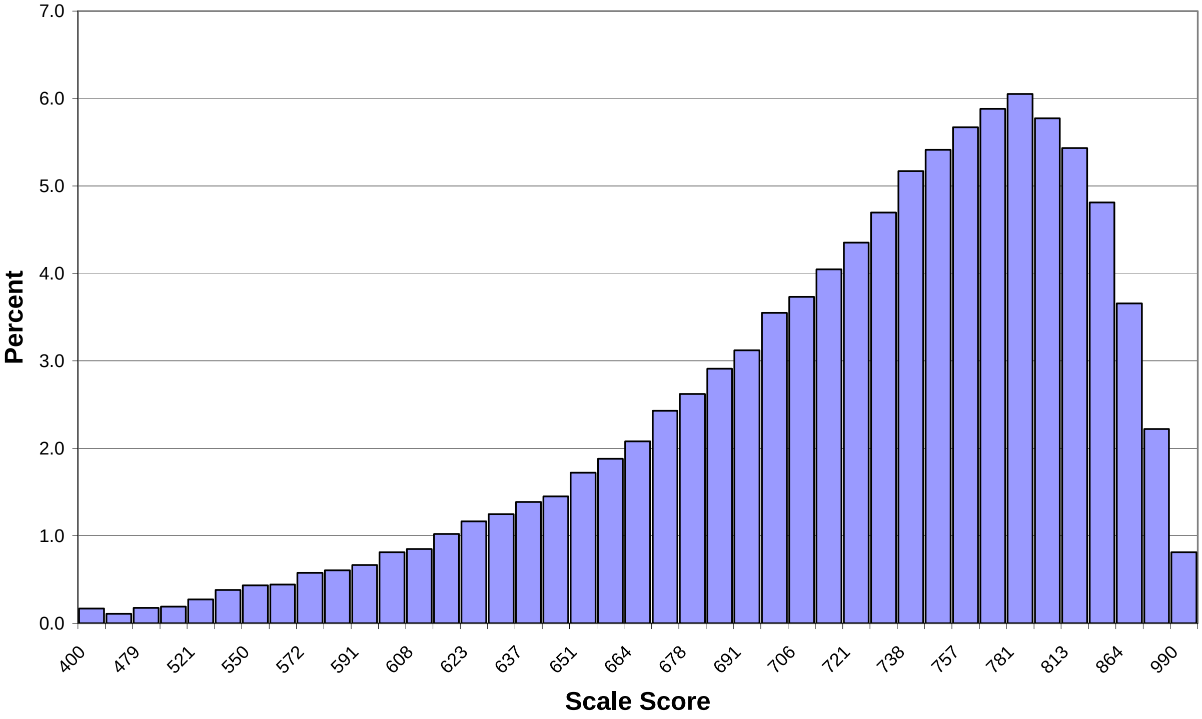


Reading Grade 5 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	72	0.2	72	0.2
442	46	0.1	118	0.3
479	75	0.2	193	0.4
503	82	0.2	275	0.6
521	117	0.3	392	0.9
537	164	0.4	556	1.3
550	186	0.4	742	1.7
561	191	0.4	933	2.2
572	247	0.6	1,180	2.7
582	260	0.6	1,440	3.4
591	286	0.7	1,726	4.0
600	349	0.8	2,075	4.8
608	365	0.9	2,440	5.7
616	438	1.0	2,878	6.7
623	500	1.2	3,378	7.9
630	536	1.2	3,914	9.1
637	595	1.4	4,509	10.5
644	623	1.5	5,132	12.0
651	739	1.7	5,871	13.7
658	807	1.9	6,678	15.6
664	893	2.1	7,571	17.6
671	1043	2.4	8,614	20.1
678	1126	2.6	9,740	22.7
684	1250	2.9	10,990	25.6
691	1340	3.1	12,330	28.7
698	1524	3.6	13,854	32.3
706	1602	3.7	15,456	36.0
713	1737	4.0	17,193	40.1
721	1868	4.4	19,061	44.4
729	2016	4.7	21,077	49.1
738	2219	5.2	23,296	54.3
747	2324	5.4	25,620	59.7
757	2434	5.7	28,054	65.4
769	2525	5.9	30,579	71.2
781	2598	6.1	33,177	77.3

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
795	2479	5.8	35,656	83.1
813	2332	5.4	37,988	88.5
834	2065	4.8	40,053	93.3
864	1570	3.7	41,623	97.0
914	953	2.2	42,576	99.2
990	349	0.8	42,925	100.0

Spring 2012 Read Grade 5 Scale Score Distribution

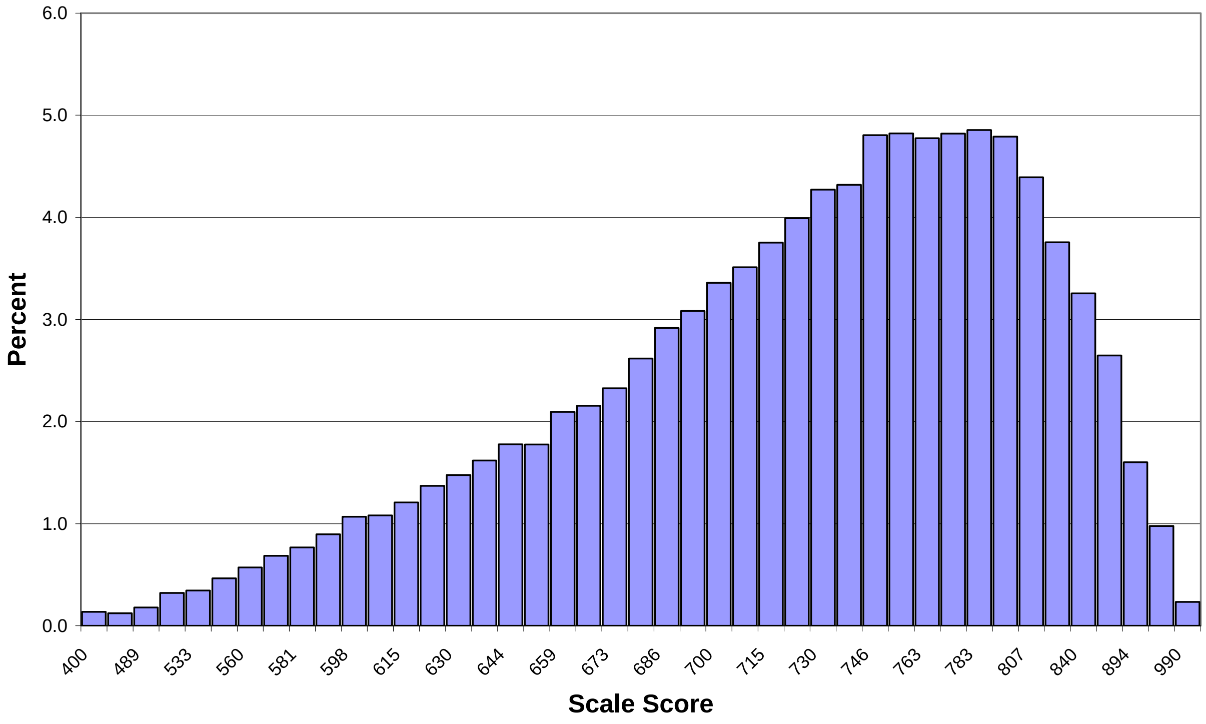


Reading Grade 6 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	59	0.1	59	0.1
446	53	0.1	112	0.3
489	77	0.2	189	0.4
514	139	0.3	328	0.8
533	149	0.3	477	1.1
548	200	0.5	677	1.6
560	246	0.6	923	2.1
571	295	0.7	1,218	2.8
581	330	0.8	1,548	3.6
590	385	0.9	1,933	4.5
598	459	1.1	2,392	5.6
607	465	1.1	2,857	6.6
615	519	1.2	3,376	7.8
622	590	1.4	3,966	9.2
630	635	1.5	4,601	10.7
637	696	1.6	5,297	12.3
644	764	1.8	6,061	14.1
652	763	1.8	6,824	15.9
659	901	2.1	7,725	18.0
666	927	2.2	8,652	20.1
673	1000	2.3	9,652	22.4
679	1126	2.6	10,778	25.1
686	1254	2.9	12,032	28.0
693	1326	3.1	13,358	31.1
700	1445	3.4	14,803	34.4
707	1510	3.5	16,313	37.9
715	1614	3.8	17,927	41.7
722	1716	4.0	19,643	45.7
730	1837	4.3	21,480	49.9
737	1858	4.3	23,338	54.3
746	2067	4.8	25,405	59.1
754	2074	4.8	27,479	63.9
763	2054	4.8	29,533	68.7
773	2073	4.8	31,606	73.5
783	2088	4.9	33,694	78.3

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
795	2060	4.8	35,754	83.1
807	1889	4.4	37,643	87.5
822	1616	3.8	39,259	91.3
840	1400	3.3	40,659	94.5
862	1139	2.6	41,798	97.2
894	689	1.6	42,487	98.8
952	421	1.0	42,908	99.8
990	101	0.2	43,009	100.0

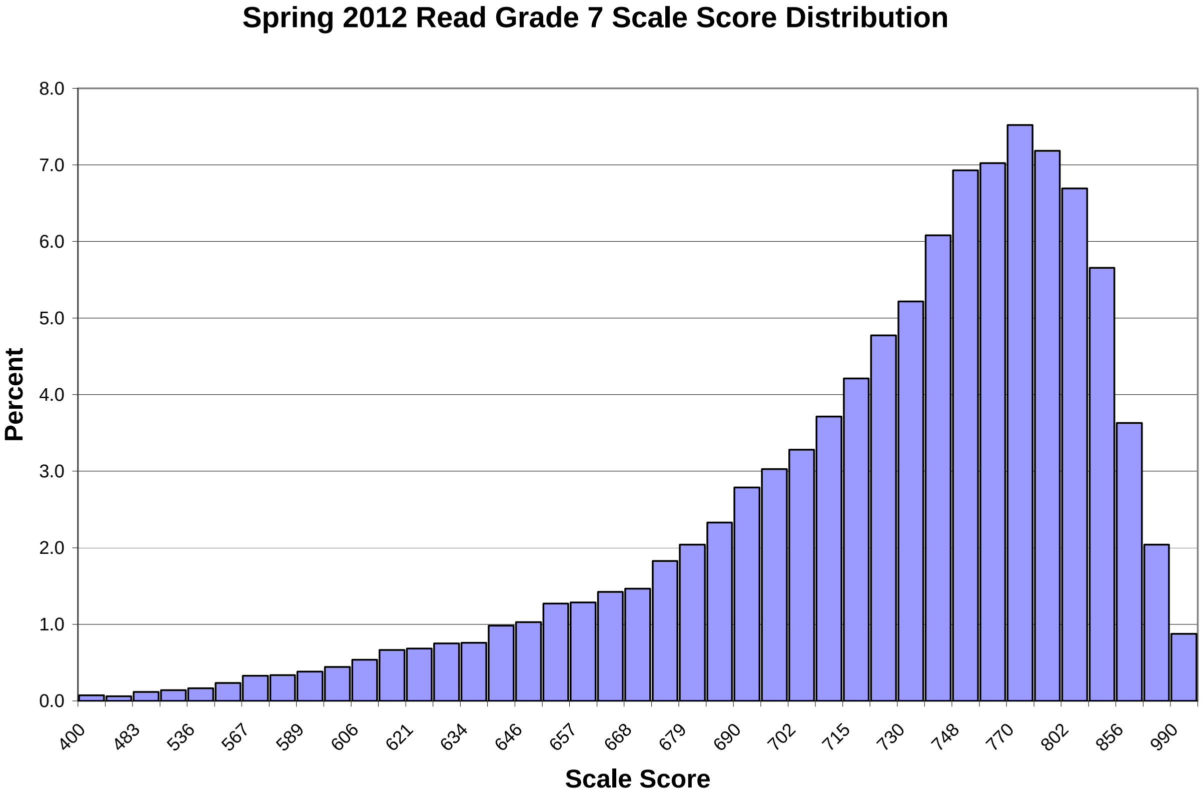
Spring 2012 Read Grade 6 Scale Score Distribution



Reading Grade 7 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	31	0.1	31	0.1
428	25	0.1	56	0.1
483	49	0.1	105	0.3
514	58	0.1	163	0.4
536	69	0.2	232	0.6
553	98	0.2	330	0.8
567	137	0.3	467	1.1
579	141	0.3	608	1.5
589	159	0.4	767	1.8
598	185	0.4	952	2.3
606	224	0.5	1,176	2.8
614	276	0.7	1,452	3.5
621	284	0.7	1,736	4.2
628	312	0.8	2,048	4.9
634	315	0.8	2,363	5.7
640	408	1.0	2,771	6.7
646	427	1.0	3,198	7.7
652	529	1.3	3,727	9.0
657	535	1.3	4,262	10.3
663	591	1.4	4,853	11.7
668	610	1.5	5,463	13.2
674	759	1.8	6,222	15.0
679	848	2.0	7,070	17.0
685	968	2.3	8,038	19.3
690	1158	2.8	9,196	22.1
696	1258	3.0	10,454	25.2
702	1363	3.3	11,817	28.4
708	1543	3.7	13,360	32.2
715	1750	4.2	15,110	36.4
722	1983	4.8	17,093	41.1
730	2168	5.2	19,261	46.4
738	2526	6.1	21,787	52.4
748	2878	6.9	24,665	59.4
758	2917	7.0	27,582	66.4

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
770	3125	7.5	30,707	73.9
785	2984	7.2	33,691	81.1
802	2780	6.7	36,471	87.8
824	2350	5.7	38,821	93.5
856	1508	3.6	40,329	97.1
908	848	2.0	41,177	99.1
990	364	0.9	41,541	100.0

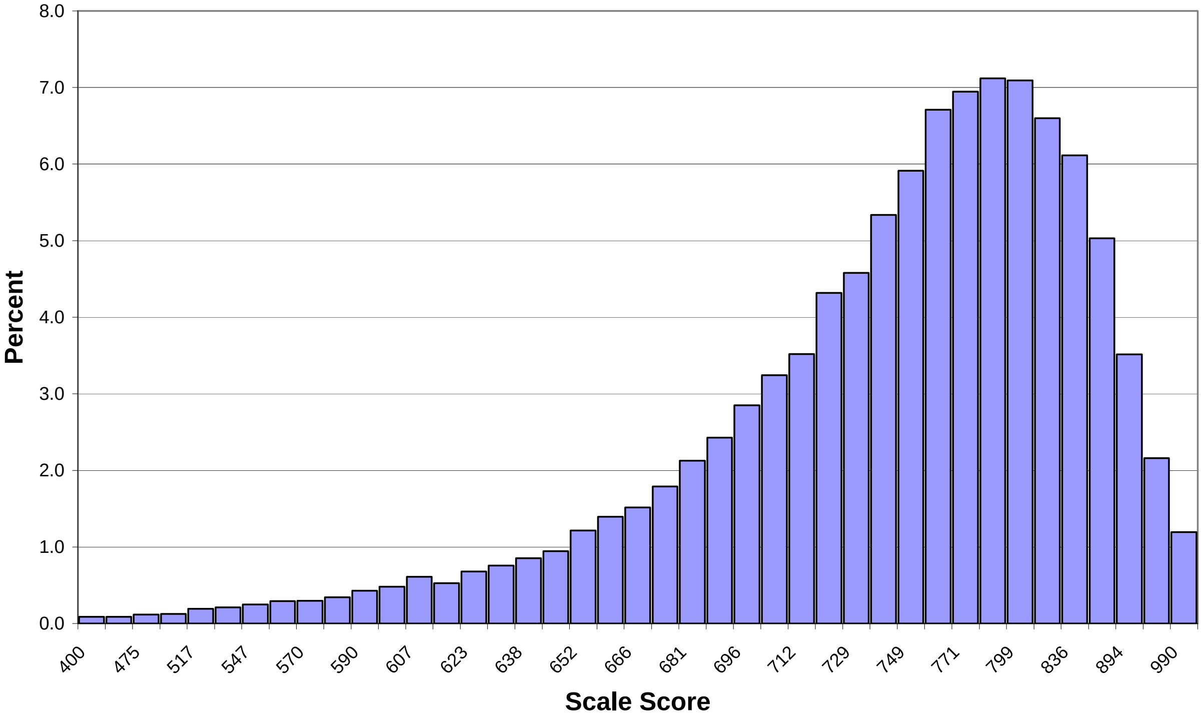


Reading Grade 8 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	37	0.1	37	0.1
443	37	0.1	74	0.2
475	48	0.1	122	0.3
498	52	0.1	174	0.4
517	79	0.2	253	0.6
533	87	0.2	340	0.8
547	103	0.2	443	1.1
559	121	0.3	564	1.4
570	123	0.3	687	1.7
580	142	0.3	829	2.0
590	177	0.4	1,006	2.4
598	199	0.5	1,205	2.9
607	252	0.6	1,457	3.5
615	217	0.5	1,674	4.1
623	280	0.7	1,954	4.7
630	312	0.8	2,266	5.5
638	352	0.9	2,618	6.4
645	390	0.9	3,008	7.3
652	501	1.2	3,509	8.5
659	575	1.4	4,084	9.9
666	625	1.5	4,709	11.4
674	738	1.8	5,447	13.2
681	877	2.1	6,324	15.3
688	1000	2.4	7,324	17.8
696	1175	2.9	8,499	20.6
704	1337	3.2	9,836	23.9
712	1450	3.5	11,286	27.4
721	1780	4.3	13,066	31.7
729	1888	4.6	14,954	36.3
739	2199	5.3	17,153	41.6
749	2438	5.9	19,591	47.5
760	2766	6.7	22,357	54.2
771	2863	6.9	25,220	61.2
784	2935	7.1	28,155	68.3
799	2924	7.1	31,079	75.4

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
816	2720	6.6	33,799	82.0
836	2521	6.1	36,320	88.1
861	2074	5.0	38,394	93.1
894	1449	3.5	39,843	96.6
942	891	2.2	40,734	98.8
990	492	1.2	41,226	100.0

Spring 2012 Read Grade 8 Scale Score Distribution

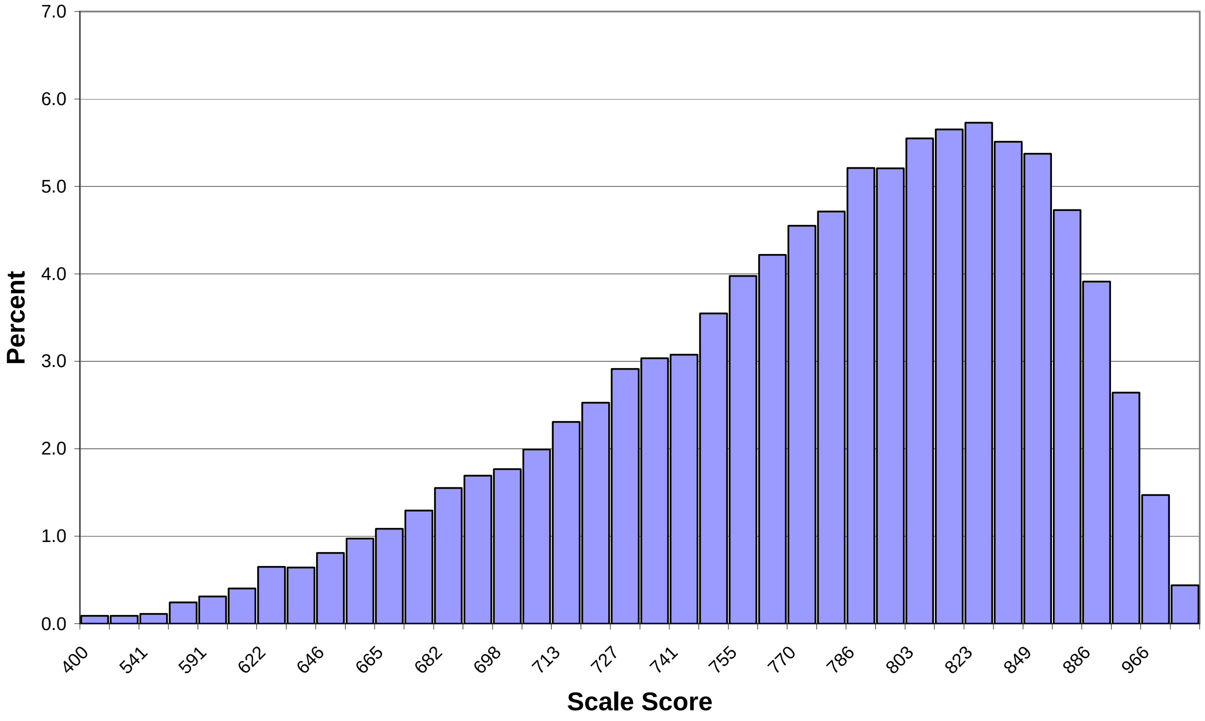


Science Grade 5 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	40	0.1	40	0.1
494	40	0.1	80	0.2
541	50	0.1	130	0.3
570	107	0.2	237	0.5
591	137	0.3	374	0.9
608	178	0.4	552	1.3
622	286	0.7	838	1.9
635	282	0.6	1,120	2.5
646	356	0.8	1,476	3.4
656	428	1.0	1,904	4.3
665	477	1.1	2,381	5.4
674	569	1.3	2,950	6.7
682	683	1.6	3,633	8.3
690	744	1.7	4,377	10.0
698	777	1.8	5,154	11.7
705	877	2.0	6,031	13.7
713	1015	2.3	7,046	16.0
720	1112	2.5	8,158	18.5
727	1281	2.9	9,439	21.5
734	1336	3.0	10,775	24.5
741	1353	3.1	12,128	27.6
748	1561	3.5	13,689	31.1
755	1749	4.0	15,438	35.1
763	1855	4.2	17,293	39.3
770	2002	4.6	19,295	43.9
778	2073	4.7	21,368	48.6
786	2292	5.2	23,660	53.8
794	2291	5.2	25,951	59.0
803	2441	5.5	28,392	64.5
813	2486	5.7	30,878	70.2
823	2520	5.7	33,398	75.9
835	2424	5.5	35,822	81.4
849	2364	5.4	38,186	86.8
865	2080	4.7	40,266	91.5
886	1721	3.9	41,987	95.4

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
915	1162	2.6	43,149	98.1
966	647	1.5	43,796	99.6
990	193	0.4	43,989	100.0

Spring 2012 Scie Grade 5 Scale Score Distribution

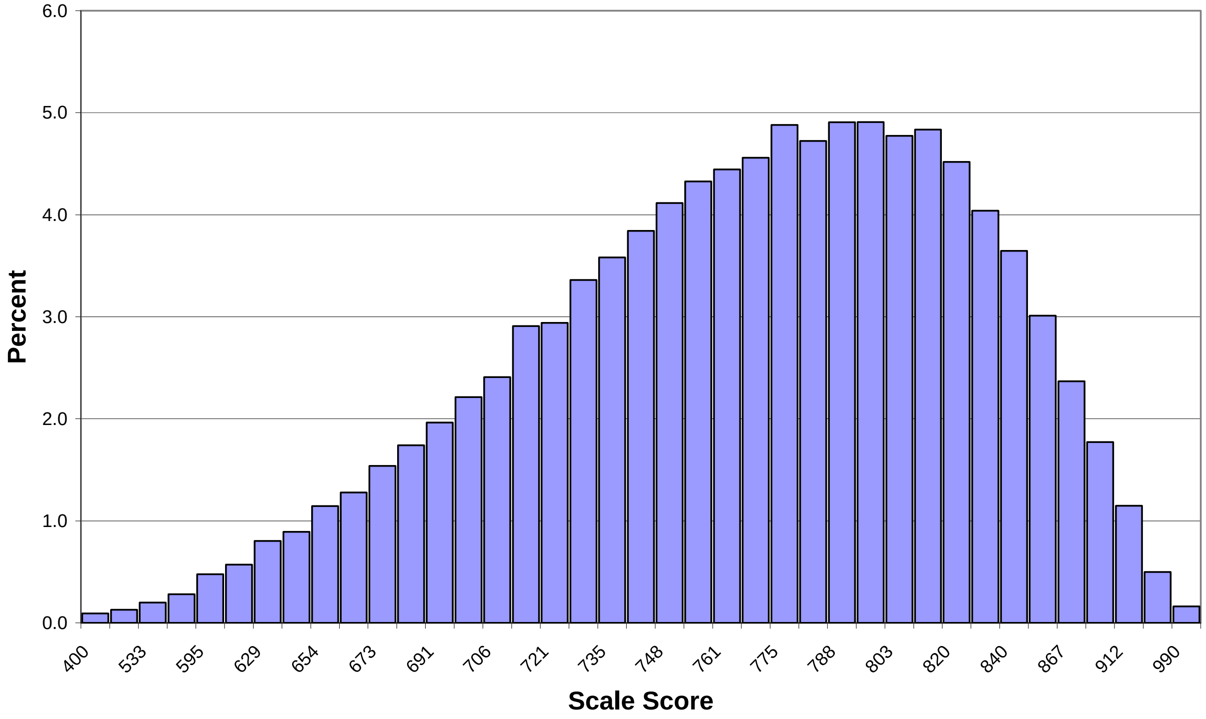


Science Grade 8 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	40	0.1	40	0.1
457	55	0.1	95	0.2
533	86	0.2	181	0.4
570	121	0.3	302	0.7
595	205	0.5	507	1.2
614	245	0.6	752	1.8
629	345	0.8	1,097	2.6
642	383	0.9	1,480	3.4
654	492	1.1	1,972	4.6
664	549	1.3	2,521	5.9
673	661	1.5	3,182	7.4
682	747	1.7	3,929	9.2
691	843	2.0	4,772	11.1
699	950	2.2	5,722	13.3
706	1034	2.4	6,756	15.7
714	1249	2.9	8,005	18.6
721	1262	2.9	9,267	21.6
728	1443	3.4	10,710	24.9
735	1538	3.6	12,248	28.5
742	1650	3.8	13,898	32.4
748	1767	4.1	15,665	36.5
755	1857	4.3	17,522	40.8
761	1908	4.4	19,430	45.3
768	1957	4.6	21,387	49.8
775	2095	4.9	23,482	54.7
781	2028	4.7	25,510	59.4
788	2106	4.9	27,616	64.3
796	2107	4.9	29,723	69.2
803	2049	4.8	31,772	74.0
812	2076	4.8	33,848	78.8
820	1939	4.5	35,787	83.4
830	1735	4.0	37,522	87.4
840	1566	3.6	39,088	91.0
853	1293	3.0	40,381	94.1
867	1017	2.4	41,398	96.4

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
886	761	1.8	42,159	98.2
912	493	1.1	42,652	99.3
958	214	0.5	42,866	99.8
990	69	0.2	42,935	100.0

Spring 2012 Scie Grade 8 Scale Score Distribution

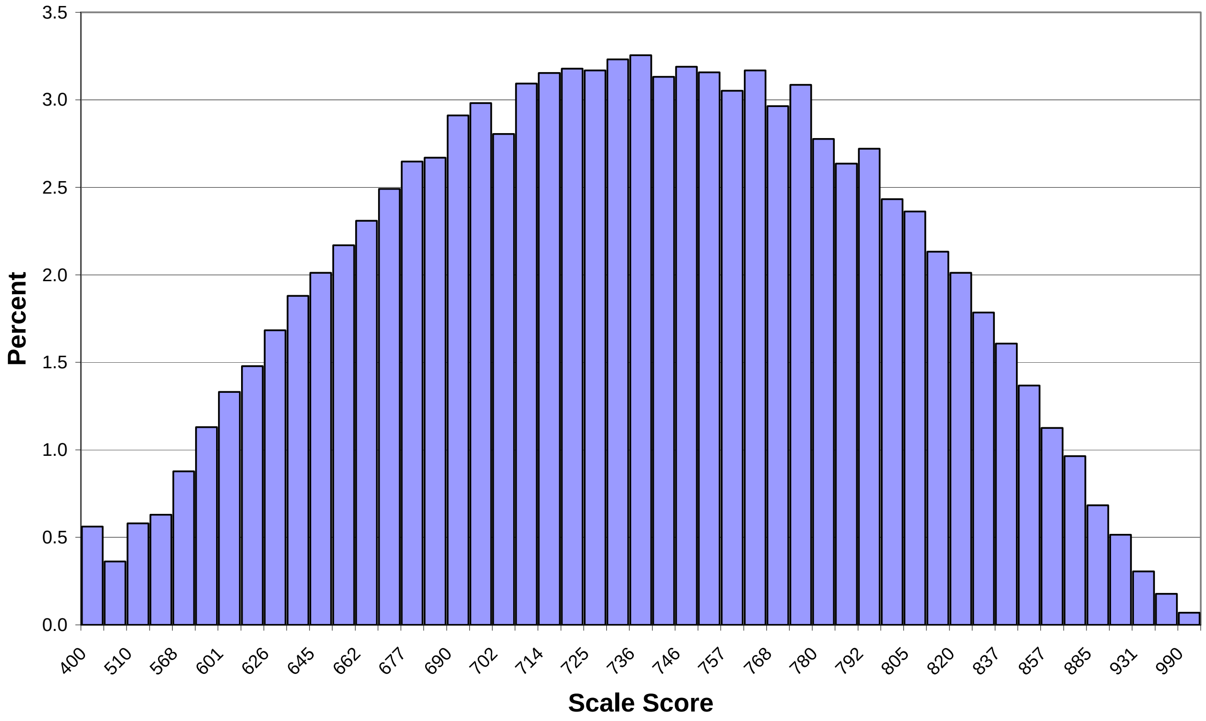


Social Studies Grade 5 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	265	0.6	265	0.6
446	171	0.4	436	0.9
510	274	0.6	710	1.5
544	297	0.6	1,007	2.1
568	414	0.9	1,421	3.0
586	533	1.1	1,954	4.1
601	628	1.3	2,582	5.5
614	697	1.5	3,279	7.0
626	794	1.7	4,073	8.6
636	887	1.9	4,960	10.5
645	949	2.0	5,909	12.5
654	1023	2.2	6,932	14.7
662	1089	2.3	8,021	17.0
670	1175	2.5	9,196	19.5
677	1249	2.6	10,445	22.1
684	1259	2.7	11,704	24.8
690	1373	2.9	13,077	27.7
696	1406	3.0	14,483	30.7
702	1323	2.8	15,806	33.5
708	1459	3.1	17,265	36.6
714	1487	3.2	18,752	39.8
720	1499	3.2	20,251	42.9
725	1494	3.2	21,745	46.1
730	1524	3.2	23,269	49.3
736	1535	3.3	24,804	52.6
741	1477	3.1	26,281	55.7
746	1504	3.2	27,785	58.9
752	1489	3.2	29,274	62.1
757	1439	3.1	30,713	65.1
763	1494	3.2	32,207	68.3
768	1398	3.0	33,605	71.2
774	1455	3.1	35,060	74.3
780	1310	2.8	36,370	77.1
786	1243	2.6	37,613	79.7

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
792	1283	2.7	38,896	82.5
798	1147	2.4	40,043	84.9
805	1114	2.4	41,157	87.3
812	1006	2.1	42,163	89.4
820	949	2.0	43,112	91.4
828	842	1.8	43,954	93.2
837	758	1.6	44,712	94.8
847	645	1.4	45,357	96.2
857	531	1.1	45,888	97.3
870	455	1.0	46,343	98.2
885	322	0.7	46,665	98.9
904	243	0.5	46,908	99.4
931	144	0.3	47,052	99.8
982	84	0.2	47,136	99.9
990	33	0.1	47,169	100.0

Spring 2012 Soci Grade 5 Scale Score Distribution

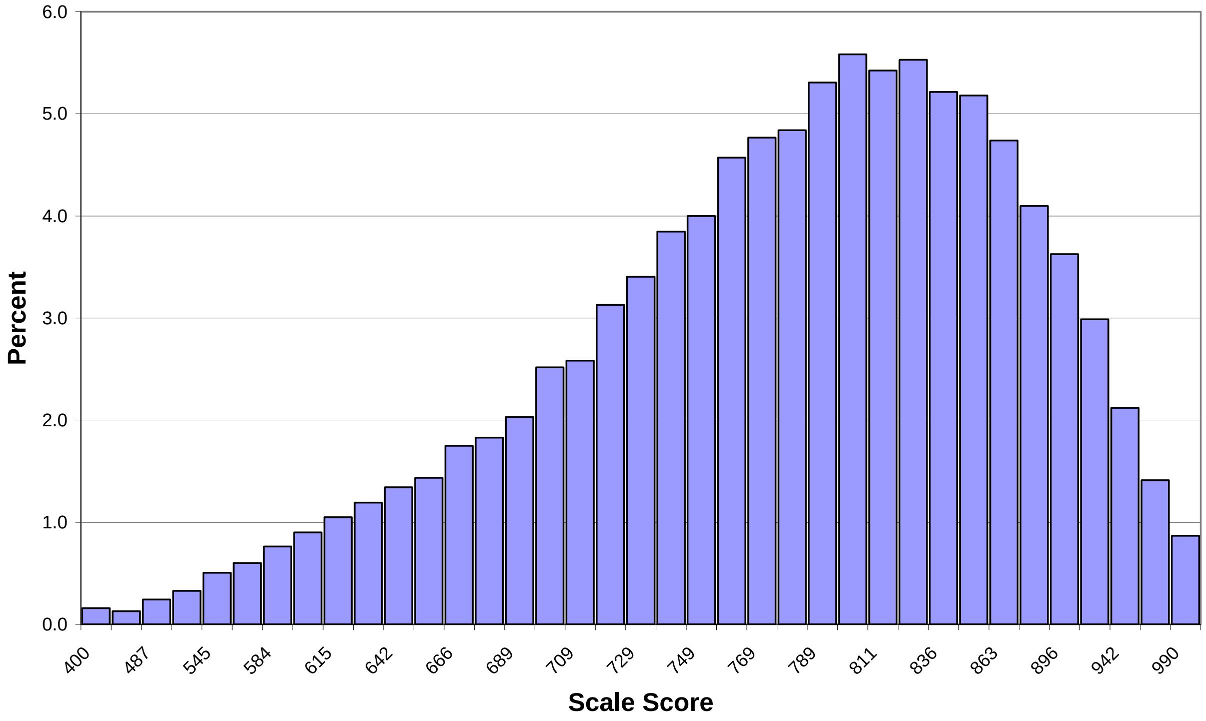


Social Studies Grade 7 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	71	0.2	71	0.2
442	58	0.1	129	0.3
487	109	0.2	238	0.5
519	148	0.3	386	0.9
545	227	0.5	613	1.4
566	270	0.6	883	2.0
584	342	0.8	1,225	2.7
600	404	0.9	1,629	3.6
615	471	1.0	2,100	4.7
629	535	1.2	2,635	5.9
642	603	1.3	3,238	7.2
654	644	1.4	3,882	8.6
666	785	1.7	4,667	10.4
678	821	1.8	5,488	12.2
689	912	2.0	6,400	14.3
699	1130	2.5	7,530	16.8
709	1160	2.6	8,690	19.4
719	1405	3.1	10,095	22.5
729	1528	3.4	11,623	25.9
739	1727	3.8	13,350	29.7
749	1795	4.0	15,145	33.7
759	2052	4.6	17,197	38.3
769	2140	4.8	19,337	43.1
779	2172	4.8	21,509	47.9
789	2382	5.3	23,891	53.2
800	2506	5.6	26,397	58.8
811	2435	5.4	28,832	64.2
823	2482	5.5	31,314	69.8
836	2340	5.2	33,654	75.0
849	2325	5.2	35,979	80.1
863	2127	4.7	38,106	84.9
879	1840	4.1	39,946	89.0
896	1627	3.6	41,573	92.6
916	1341	3.0	42,914	95.6
942	952	2.1	43,866	97.7

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
977	634	1.4	44,500	99.1
990	390	0.9	44,890	100.0

Spring 2012 Soci Grade 7 Scale Score Distribution

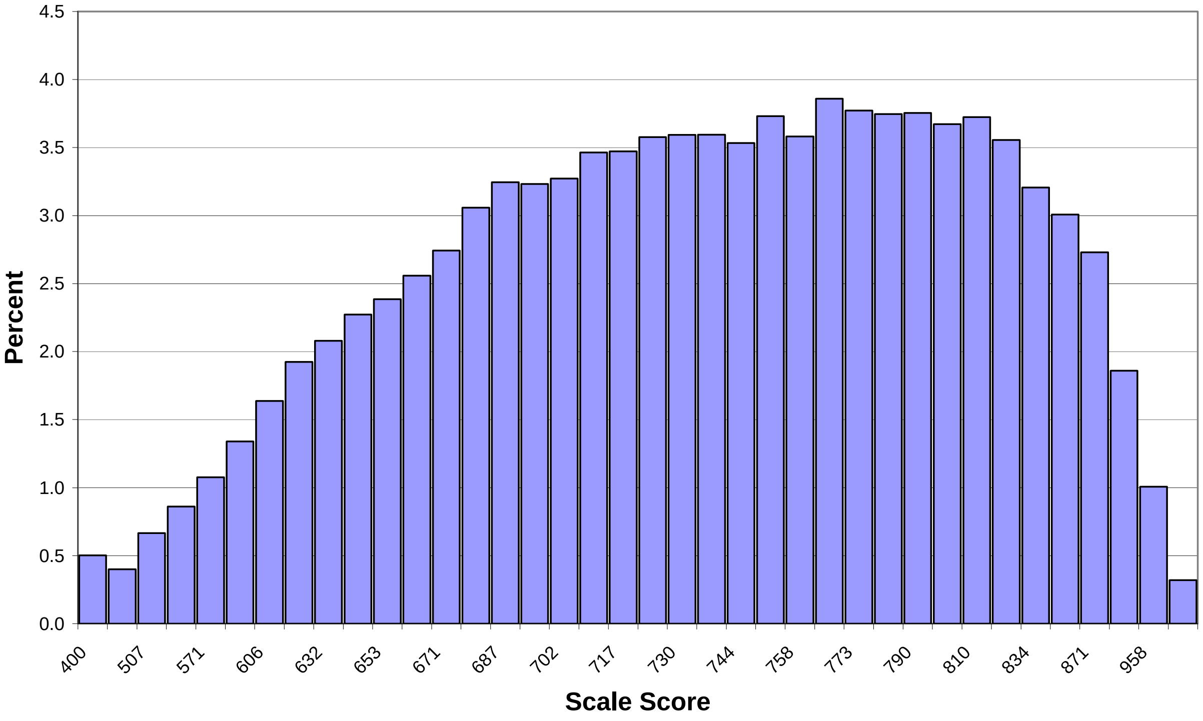


Social Studies Grade 8 Scale Score Distribution for Spring 2012

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
400	230	0.5	230	0.5
431	183	0.4	413	0.9
507	305	0.7	718	1.6
546	394	0.9	1,112	2.4
571	493	1.1	1,605	3.5
590	613	1.3	2,218	4.8
606	750	1.6	2,968	6.5
620	881	1.9	3,849	8.4
632	952	2.1	4,801	10.5
643	1041	2.3	5,842	12.8
653	1092	2.4	6,934	15.1
662	1171	2.6	8,105	17.7
671	1256	2.7	9,361	20.4
679	1400	3.1	10,761	23.5
687	1486	3.2	12,247	26.7
695	1480	3.2	13,727	30.0
702	1498	3.3	15,225	33.2
710	1586	3.5	16,811	36.7
717	1590	3.5	18,401	40.2
723	1638	3.6	20,039	43.8
730	1645	3.6	21,684	47.4
737	1646	3.6	23,330	50.9
744	1618	3.5	24,948	54.5
751	1709	3.7	26,657	58.2
758	1640	3.6	28,297	61.8
765	1767	3.9	30,064	65.7
773	1727	3.8	31,791	69.4
781	1715	3.7	33,506	73.2
790	1719	3.8	35,225	76.9
799	1681	3.7	36,906	80.6
810	1705	3.7	38,611	84.3
821	1628	3.6	40,239	87.9
834	1468	3.2	41,707	91.1
850	1377	3.0	43,084	94.1
871	1250	2.7	44,334	96.8

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
901	852	1.9	45,186	98.7
958	461	1.0	45,647	99.7
990	147	0.3	45,794	100.0

Spring 2012 Soci Grade 8 Scale Score Distribution


Percent
Scale Score