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
2012–2013 Technical Report

FINAL

Submitted to
The Oklahoma State Department of Education
December 2013

Copyright

Revision History

Version 1.1

Table of Contents

List of Tables

List of Figures

List of Appendices

Acronyms and Abbreviations

2PPC Two Parameter Partial Credit model
3PL Three Parameter Logistic model
ACE Achieving Classroom Excellence
AERA American Educational Research Association
APA American Psychological Association
AYP Adequate Yearly Progress
BR Braille
BTC Building Test Coordinator
C^3 Oklahoma's Core curriculum, the College, Career and Citizen Ready
CCSSO Council of Chief State School Officers
CE Critical Element
CFA Confirmatory Factor Analysis
CR Constructed-Response
CSEM Conditional Standard Error of Measurement
DIF Differential Item Functioning
DOK Depth of Knowledge
DTC District Test Coordinator
EFA Exploratory Factor Analysis
EHS Electronic Handscoring System
ELL English Language Learners
EOI End-of-Instruction
EQ Equivalent
FN False Negative
FP False Positive
GRT General Research Tape
HOSS Highest Obtainable Scale Score
ICC Item Characteristic Curve
IEP Individualized Education Program
IRT Item Response Theory
LIU Language in Use
LOSS Lowest Obtainable Scale Score
MC Multiple-Choice
MH Mantel-Haenszel
NCES National Center for Education Statistics
NCLB No Child Left Behind
NCME National Council on Measurement in Education
NGA National Governors Association Center
OAAP Oklahoma Alternate Assessment Program
OAC Oklahoma Administrative Code
OCCT Oklahoma Core Curriculum Tests
OE Open-Ended
OMAAP Oklahoma Modified Alternate Assessment Program
OP Operational
OSTP Oklahoma School Testing Program
PASS Priority Academic Student Skills
RIBs Rater Item Blocks
RT Retest
SAS Statistical Analysis System
SD Standard Deviation
SDE Oklahoma State department of Education
SEM Standard Error of Measurement
SS Scale Score
TA Test Administrator
TAC Technical Advisory Committee
TCC Test Characteristic Curve
TP Test Proctor
TPM Test Preparation Manual
US DOE United States Department of Education
WP Writing Prompt

Introduction

This report summarizes the research data analyses conducted on the Oklahoma Core Curriculum Tests, Grades 3 through 8 (OCCT 3–8) test administrations and provides data evidences in supporting the validity and reliability of the tests.

For the OCCT 3–8, Reading and Mathematics tests are administered in Grades 3–8; Science, Social Studies, and Writing in Grade 5; Geography in Grade 7; and Science, U.S. History, and Writing in Grade 8. All students must take the OCCT for content areas in which a modified assessment is not available. The Department of Special Education oversees the implementation of the Oklahoma Alternate Assessment Program (OAAP), or portfolio assessment, which includes all of the Grades 3–8 contents.

Within the state of Oklahoma, the development of the *Priority Academic Student Skills (PASS)* and most recently Oklahoma's Core Curriculum, the College, Career and Citizen Ready (C^3) content standards, the development of the Oklahoma School Testing Program (OSTP) items and operational test forms, the review of the alignment of the content to the test, the administrations of the test, the machine scoring and hand scoring of student responses, the setting of cut scores, and the psychometric analyses are all important steps in the process of developing a valid assessment system (Barton, 2007). This document serves to capture a small portion of the enormous amount of time and effort devoted to one of the OSTP assessments, the OCCT for Grades 3 through 8, in relation to the importance, reliability, and validity of the assessment as part of the Oklahoma assessment system.

The American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education's (NCME) *Standards for Educational and Psychological Testing* (1999; *Standards for Testing*) are the guides by which this report provides various evidences of validity. The CTB/McGraw-Hill work process in the Oklahoma OCCT 3–8 for Spring 2013 has paid close attention to the *Standards for Testing,* and this report provides evidence that is appropriate to a statewide summative assessment. Because the results of this assessment are used as part of state and federal accountability systems, attention has also been paid to the federal guidance provided in the Standards and Assessments Peer Review Guidance (US DOE, 2004). Evidence within this report also relates to the Critical Elements (CE) as part of the guidance for Peer Review.

Section 1 – Overview

The Oklahoma Core Curriculum Tests (OCCT) assessments are state-mandated criterion-referenced tests that measure student proficiency in specific content areas. Each test has the purpose of measuring the student's knowledge relative to the Oklahoma Core Curriculum, the College, Career and Citizen Ready (C^3), Oklahoma's content standards. In Spring 2013, the OCCT assessments were administered to all eligible students in Grades 3 through 8. The OCCT covered: Mathematics and Reading for Grades 3 through 8; Science and Writing for Grades 5 and 8; and Social Studies for Grades 5, 7 (Geography), and 8 (U.S. History). Along with the Operational tests (OP), other form variations were administered for the OCCT: Equivalent forms (EQ), Braille forms (BR), and Large-Print forms. Field test forms were administered for Social Studies Grades 5, 7, and 8.

In the Fall of 2012, CTB/McGraw-Hill was contracted by the Oklahoma State Department of Education (SDE) to develop, administer, and maintain the Oklahoma School Testing Program (OSTP) OCCT and Oklahoma Modified Alternate Assessment Program (OMAAP) for Achieving Classroom Excellence (ACE) End-of-Instruction (EOI) and Grades 3–8. The purpose of this technical report is to provide objective information regarding technical aspects of the OCCT 3–8 assessments by specifying the technical details of the work accomplished from Summer 2012 (developed, administered, and processed by Pearson) through the end of Spring 2013 on these tests. 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 the OSTP-ACE G3–8 tests include the administration manual OSTP *2012–2013 Test Preparation Manual* found at: http://www.ok.gov/sde/documents/2013-10-17/2012-2013-test-prep-manual; interpretation manuals, implementation materials, and training materials for administrators, schools, and teachers, found at: http://www.ok.gov/sde/test-support-teachers-and-administrators; and teachers, students, and parent guides found at: http://ok.gov/sde/assessment-administrator-resources-administrators.

The Spring 2013 OCCT 3–8 field test items for the Mathematics and Reading Grades 3–8, Science Grades 5 and 8, and Social Studies Grades 5, 7, and 8 assessments were developed by Pearson in collaboration with the Oklahoma SDE. The assessments for these subjects were developed by CTB/McGraw-Hill in collaboration with the SDE, and were administered by the SDE. Note that there were other forms applied concurrently with the operational forms in each of the above administrations and for each of the contents: the Braille (BR) form, the Large-Print form, and the Equivalent (EQ) form.

Section 1.1. – Purpose

This report includes only data and analyses for the operational forms and content for the Spring 2013 administration. It begins with a description of the Oklahoma content standards, which are described in Section 1.2 – *Oklahoma* C^3 Content Standards. All operational and field test items

for OCCT 3–8 Spring 2013 were subjected to cycles of reviews by the SDE and Pearson. A description of the item development process, along with a description of the alignment process and test development, is presented in complete detail in **Section 2 – Item and Test Development**. A detailed description of the administration processes is found in **Section 3 – Administration,** and a discussion of the operational population and the research samples utilized in the analysis is found in **Section 5 – Sampling Plan and Field Test Design**.

The Spring 2013 OCCT 3–8 scores for Mathematics and Reading Grades 3 through 8 were based on a post-equating design. The Science Grades 5 and 8 scores were analyzed for new scaling. The Social Studies Grade 5, Grade 7, and Grade 8 items were field test items only. A complete description of the operational and field test item analyses and the calibration/scaling and equating analysis is found in **Section 6 – Methods** and **Section 7 – Results**.

Section 1.2. – *Oklahoma C³* Content Standards

CTB/McGraw-Hill developed the Spring 2013 Oklahoma OCCT 3–8 assessments to measure the *Oklahoma C³* Standards shown in Table 1.1. The objectives associated with the content and/or process standards tested are provided in Appendix A.

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 in parentheses)	
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	
Process/Inquiry Standards and Objectives	
Process 1.	Observe and Measure
Process 2.	Classify
Process 3.	Experiment
Process 4.	Interpret and Communicate
Grade 5 Content Standards	
Standard 1.	Properties of Matter and Energy
Standard 2.	Organisms and Environments
Standard 3.	Structures of the Earth and the Solar System

Table 1.1. Testable Standards for OCCT Grades 3 to 8 *(continued)*

Grade 8 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 1.	James Towne Settlement and Plimoth Plantation
Standard 2.	Colonial America
Standard 3.	American Revolution
Standard 4.	Early Federal Period
Social Studies Grade 7 (Geography)	
Standard 1.	Geographic Tools/Geography Skills
Standard 2.	Human and Physical Characteristics of Regions
Standard 3.	Patterns of the Earth
Standard 4.	Human Systems
Standard 5.	Human/Environment Interaction
Social Studies Grade 8 (U.S. History)	
Standard 1.	Causes and Events of the American Revolution
Standard 2.	Foundations and Founders of the American Nation
Standard 3.	Developing the American Government System
Standard 4.	The Transformation of the United States to the Mid-1800s
Standard 5.	Causes, Events, and Leadership in the Civil War

Section 2 – Item and Test Development

In the Spring 2013 administration, there was one Operational form with embedded sets of field test items for the tests administered for Mathematics and Reading Grades 3–8, and Science Grades 5 and 8. For the Grade 5 Social Studies, Grade 7 Geography, and Grade 8 U.S. History tests, field test forms were administered instead. There were six field test forms for Mathematics and Reading Grades 3–8, and Science Grades 5 and 8. There were two field test forms for Grade 5 Social Studies, Grade 7 Geography, and Grade 8 U.S. History. A Braille form and an Equivalent form were produced for Mathematics and Reading Grades 3–8 and Science Grades 5 and 8. Because it was a field test year, no Equivalent forms were produced for Grade 5 Social Studies, Grade 7 Geography, and Grade 8 U.S, History.

The Braille form is usually a mirror of the operational administration. The Equivalent is designated as a breach form. A student could receive an Equivalent form for various reasons, including becoming ill during the test administration or experiencing any 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.

Section 2.1 – Aligning Test to *Oklahoma C³* Content Standards

In general, alignment is a process that provides experts the opportunity to make item-level judgments about the grade level, standards, and indicators to which items should be aligned. There are multiple points in the alignment process at which assessment items are either created or evaluated for alignment to content. Most tests, particularly high-stakes, large-scale assessments, are built via rigorous and well-researched methodologies. They are guided by well-defined content and by the boundaries within the content that can be reasonably assessed in a testing environment. Such guidance is typically in the form of item specifications and test blueprints. The item specifications help define which content standards can be assessed by a test (and which content standards are better assessed in the classroom), the breadth and depth of the content that may be limited for the test, and the format and types of items, or test questions appropriate for the content being assessed (e.g., multiple-choice or open-ended).

A list of the assessable standards for each subject is provided in Table 2.1. For Mathematics and Reading, the same assessable standards appear in each grade level.

Table 2.1. 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 in parentheses)	
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	
OKC^3 Process/Inquiry Standards and Objectives	
Process 1.	Observe and Measure
Process 2.	Classify
Process 3.	Experiment
Process 4.	Interpret and Communicate
Grade 5 *OK C*^{3} Content Standards	
Standard 1.	Properties of Matter and Energy
Standard 2.	Organisms and Environments
Standard 3.	Structures of the Earth and the Solar System
Grade 8 *OK C*^{3} 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 1.	James Towne Settlement and Plimoth Plantation
Standard 2.	Colonial America
Standard 3.	American Revolution
Standard 4.	Early Federal Period
Social Studies Grade 7 (Geography)	
Standard 1.	Geographic Tools/Geography Skills
Standard 2.	Human and Physical Characteristics of Regions
Standard 3.	Physical Systems of the Earth
Standard 4.	Human Systems
Standard 5.	Human/Environment Interaction
Social Studies Grade 8 (U.S. History)	
Standard 1.	Causes and Events of the American Revolution
Standard 2	The Revolutionary Era
Standard 3.	Developing the American Government System
Standard 4.	The Transformation of the United States to the Mid-1800s
Standard 5.	Causes, Events, and Leadership in the Civil War

Blueprints

The test blueprint defines the proportion of the content to be covered on the test that best reflects the proportional importance and coverage of standards in the classroom.

In addition to the test Blueprints provided by the SDE (http://www.ok.gov/sde/test-support-teachers-and-administrators for blueprints), Table 2.2 describes four criteria for test alignment with the *Oklahoma C*^{3} Content Standards and objectives.

Table 2.2. Criteria for Aligning the Test with *Oklahoma C*^{3} Standards and Objectives

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

Data review represents a critical step in the test development cycle. At the data review meeting, the SDE and CTB/McGraw-Hill staff had the opportunity to review actual student performance on the newly-developed and field tested Multiple-Choice (MC) items across the subjects based on the Spring 2013 field test administrations. The data review focused on the content validity, curricular alignment, and statistical functioning of field tested items prior to selection for operational test forms. The field test results used in the data review provided evidence that the items were designed to yield valid results and were accessible for use by the widest possible range of students. The review of student performance should provide evidence regarding the fulfillment of requirement 200.2(b)(2) of NCLB. The purpose of the review meeting was to ensure that psychometrically-sound, fair, and aligned items are used in the construction of the OCCT 3–8 assessments and entered into the respective item banks. CTB/McGraw-Hill provided technical and psychometric expertise to provide a clear explanation about the items' content, the field test process, the scoring process, and the resulting field test data to ensure the success of these meetings and the defensibility of the program.

Data review meetings were a collaborative effort between the SDE and CTB/McGraw-Hill. The SDE administrators and content specialists attended the meetings facilitated by CTB/McGraw-Hill content specialists and research scientists who trained the SDE staff on how to interpret and review the field test data. Meeting materials included a document explaining the flagging criteria, a document containing flagged items, and the item images. CTB/McGraw-Hill discussed with the SDE the analyses performed and the criteria for flagging the items. Flagged items were then reviewed, and decisions were made as to whether to accept the item, accept the item for future re-field testing with revisions, or reject the item. Review of the data included presentation of *p*-value, point-biserial correlation, point-biserial correlation by response option, response distributions, mean overall score by response option, and indications of item Differential Item Functioning (DIF) and Item Response Theory (IRT) misfit. Items failing to meet the requirements of sound technical data were carefully considered for rejection by the review panel, thereby enhancing the reliability and improving the validity of the items left in the bank for future use. While the panel used the data as a tool to inform their judgments, the panel (and not the data alone) made the final determination as to the appropriateness or fairness of the assessment items. The flagging criteria for the OCCT 3–8 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 – 6.3 Calibration & Item Fit**)

Section 2.2 – Item Pool Development and Selection

To ensure content validity of the Oklahoma OCCT 3–8 tests, CTB/McGraw-Hill content experts closely studied the *Oklahoma C*^{3} Standards and worked with Oklahoma content area specialists, teachers, and assessment experts to gather a pool of existing items that measure Oklahoma's Assessment Frameworks (i.e., *Oklahoma C*^{3}) 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 C*^{3} in each subject was developed. These items were developed under universal design guidelines set by the SDE and carefully reviewed and discussed by Content and Bias/Sensitivity Review Committees to evaluate not only content validity, but also plain language and the quality and appropriateness of the items. These committees were comprised of Oklahoma teachers and SDE staff. The committees' recommendations were used to select and/or revise items from the item pool used to construct the field test portions of the Spring 2013 assessments.

The source of the operational items included a pool of previously field tested or operationally-administered items ranging from the Spring 2006 through the Spring 2012 administrations for Mathematics Grades 3–8, Reading Grades 3–8, and Science Grades 5 and 8. The items were calibrated live using data from the operational administrations to estimate parameters for these items.

Item selection and form development for Spring 2013 was completed as a collaborative effort between staff at the SDE and CTB/McGraw-Hill Content Development and psychometricians (Research). The primary criterion for the selection of items was to meet the content

specifications represented by test blueprints and statistical guidelines. Within the limits set by these requirements, such as classical and item response theory statistics, described in **Section 6 – Methods**, editors selected items with the best content-relevant and statistical characteristics.

The OCCT 3–8 Operational tests for the Spring 2013 cycle were built by including previously field tested and operational items. Content experts also targeted the percentage of items measuring various Depth of Knowledge (DOK) levels for assembling the tests. Table 2.3 provides the DOK level percentages for the Spring 2013 operational assessments.

Table 2.3 Percentage of Items by Depth of Knowledge Levels for OCCT 3–8 Assessments

		DOK Level					
		1		2		3	
Subject	Grade	Target	Actual	Target	Actual	Target	Actual
Mathematics	3	20–25	18	65-70	76	5–15	6
	4	20–25	28	65-70	66	5–15	6
	5	20–25	16	65-70	70	5–15	14
	6	10–15	12	65-70	76	15–25	12
	7	10–15	14	65-70	64	15–25	22
	8	10–15	10	65-70	78	15–25	12
Reading	3	20–25	16	65-70	68	5–15	16
	4	20–25	12	65-70	76	5–15	12
	5	20–25	10	65-70	78	5–15	12
	6	10–15	12	65-70	74	15–25	14
	7	10–15	10	65-70	76	15–25	14
	8	10–15	16	65-70	66	15–25	18
Science	5	20–25	18	65-70	62	5–15	20
	8	10–15	13	65-70	58	15–25	29

Bias and Sensitivity

One aspect of the data review meetings was to assess potential bias based on DIF results and item content. Although bias in the items had been deflected by writer training and review processes, there is always the potential for bias to be detected through statistical analysis. This step in the development cycle is essential because the SDE and CTB/McGraw-Hill seek to avoid inclusion of items biased in any manner against a group, because these items may lead to inequitable test results. As described earlier, all field test items were analyzed statistically for differential item functioning (DIF) using the field test data. At the data review meetings, a CTB/McGraw-Hill research scientist explained the significance of DIF, in terms of level and the direction of the DIF flags. The data review panel reviewed the item content, the percentage of students selecting each response option, and the point-biserial correlation for each response option by gender and ethnicity for all items flagged for DIF. The data review panel was then asked if there was context (for example, cultural barriers) or language in an item that might result in bias (i.e., an explanation for the existence of the statistical DIF flag).

Once items were written, they were reviewed to assure the items were appropriate for and aligned to the grade level, the *Oklahoma C*^{3} Standard and objective, and the DOK intended. The items were also reviewed to assure they were accurate, written at an appropriate reading level for the grade, written at an appropriate level of difficulty, and did not contain sensitive or potentially biasing issues.

Statistical bias analyses were performed as part of the development, review, and fairness efforts. Field test items were analyzed for statistical bias utilizing the Mantel-Haenszel method (Holland and Thayer, 1988; Michaelides, 2008). The results for Spring 2013 are found in **Section 7 – Results** of this report.

Section 2.3 – Configuration of the Spring 2013 Tests

For Spring 2013, CTB/McGraw-Hill Content Development selected items from the available item pools that had been previously field tested and approved by the SDE staff for usage on the operational assessments. The operational items on the Mathematics and Reading tests had appeared previously in the years 2008–2012. The operational items on the Science tests had appeared previously in the years 2007–2012. No operational items appeared on the Grade 5 Social Studies, Grade 7 Geography, and Grade 8 U.S. History tests. Field test items were selected from items that were approved by the SDE staff and Oklahoma teachers. Most of the field test items on the Mathematics, Reading, and Science tests had never been field tested; however, some items had been previously field tested, but required revisions and additional field testing. CTB/McGraw-Hill Research analyzed the selected items and provided feedback to Content Development regarding the best set of items to serve as the Spring 2013 operational form.

Section 2.4 – Operational and Field Test Items by Content Area

Table 2.4, provides an overview of the number of operational and field test items that composed the Spring 2013 OCCT 3–8 assessments. The Spring 2013 test was composed of one core operationally-scored form for each subject, except for Grade 5 Social Studies, Grade 7 Geography, and Grade 8 U.S. History. Field test items were embedded in the operational test forms for all content areas to build the item bank for future use. The forms in the Spring 2013 assessments were randomly assigned within classrooms to obtain randomly-equivalent samples of examinees for the field test items.

Table 2.4 Configuration of the OCCT 3–8 Tests for Spring 2013

			Item Counts (Per Form)			Maximum Possible Points on OP Test Items (Per Form)*			
						OP		FT	
Subject	Grade	Forms	OP	FT	Test	MC	OE	MC	OE
Mathematics	3	1	50	10	60	50		10	
	4	1	50	10	60	50		10	
	5	1	50	10	60	50		10	
	6**	1	50	10	60	50		10	
	7	1	50	10	60	50		10	
	8	1	50	10	60	50		10	
Reading	3	1	50	10	60	50		9	2
	4	1	50	10	60	50		9	2
	5	1	50	10	60	50		9	2
	6	1	50	10	60	50		9	2
	7	1	50	10	60	50		9	2
	8	1	50	10	60	50		9	2
Science	5	1	45	10	55	45		10	
	8	1	45	10	55	45		10	
Social Studies	5	1		60	60			60	
	7	1		60	60			60	
	8	1		60	60			60	

Note: OP = Operational; FT = Field Test; MC = Multiple Choice; OE = Open-ended; * = For the Mathematics and Reading Tests, five field test items were aligned to the Common Core standards on each test form. ** = one Grade 6 Operational item was suppressed on the test, reducing the total possible points to 49.

Section 3 – Administration

To ensure a valid and reliable assessment, the OCCT 3–8 assessments are first constructed in alignment with the *Oklahoma C*^{3} Standards (now called the Oklahoma Academic Standards) by the Oklahoma SDE in collaboration with CTB/McGraw-Hill. The tests are then administered and scored according to sound measurement principles for the purpose of evaluating validity. Additionally, best practices require that the test administrating and scoring entities perform their tasks in a consistent manner throughout the state so that all students have a fair and equitable opportunity for a score that reflects their achievement in each subject.

Schools play a key role in administering the OCCT 3–8 assessments in a manner that is consistent with established procedures, monitoring the fair administration of the assessment, and working with the SDE office to address deviations from established assessment administration best practice procedures. School faculty members play a vital role in the success of OCCT 3–8 assessments by ensuring fairness in administration of the test.

Section 3.1 – Packaging and Shipping

In order to provide secure and dependable services for the shipping of the OCCT 3–8 assessment materials, CTB/McGraw-Hill's Transportation Department maintains the quality and security of material distribution and return by hiring reputable carriers that possess the ability to trace shipments. CTB/McGraw-Hill 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 test coordinators. Each shipment to a district contains a shipping document set that includes a packing list for each school's materials.

Materials are packaged using information provided by the test coordinators through the Oklahoma WAVE system. Oklahoma educators also use this system to provide CTB/McGraw-Hill with the precode information needed to print student barcode labels, which are affixed on answer documents or consumable test books. The bar-coding of all secure materials at the time of production allows for accurate tracking of these materials through the entire packing, delivery, and return process. This allows CTB/McGraw-Hill to inventory all materials throughout the packaging and delivery process.

Section 3.2 – Materials Return

The Test Preparation Manual and Materials Return poster provide clear instructions on how to assemble, box, label, and return testing materials after test administration. CTB/McGraw-Hill utilizes double-column boxes to distribute and collect test materials, and makes additional cartons available for order to meet the various return needs of the districts.

Stack cards and paper bands are provided to group and secure used student response booklets for scoring. Color-coded return labels with pre-printed return information are also provided. These

labels facilitate the sorting of each carton and its contents upon receipt at CTB/McGraw-Hill's Data Processing Facility.

Section 3.3 – Materials Discrepancies Process

The scanning process allows CTB/McGraw-Hill to capture MC responses and student writing images. Test security form information is also captured electronically via a secure database. All scorable material discrepancies are captured, investigated by the CTB/McGraw-Hill Oklahoma Help Desk, and reported. The results are subsequently reported to the SDE.

A pre-determined date is set by the SDE and CTB/McGraw-Hill in order to account for any materials that arrive after the scheduled deadline. Late arriving material is processed up to the agreed upon date, at which point the Oklahoma SDE must be notified of any late arriving documents and render a processing decision. Following an initial call campaign to all districts with outstanding secure material, the CTB/McGraw-Hill Oklahoma Program Management team notifies the SDE regarding unresolved material discrepancies presented in a preliminary file. A subsequent call or email campaign may be conducted based on the results of the initial effort. Final missing inventory reports are then provided to the SDE. CTB/McGraw-Hill takes security seriously and makes every effort to recover missing material.

Section 4 – Scoring

The Oklahoma Spring 2013 OCCT Grades 3–8 test books included items that were machine scored (MC items), and extended-writing prompts (WPs) that were scored by trained human or "hand" scorers (raters). OCCT Grades 3–8 test books also included six field test constructed-response items (CR) for each of the six Reading test forms. These items were scored by trained "hand" scorers. The MC items were scanned and scored as correct or incorrect according to predefined answer keys. Items that had multiple marked answers or were blank were treated as incorrect.

The Writing test is one portion of the OCCT 3–8 test. Writing is assessed at Grades 5 and 8. Each writing response receives two types of scores. First, a series of analytic scores focus on specific writing traits. These traits receive scores of 1 to 4. Next, a composite score is derived by providing a differential weight or percentage to the score in each of the analytic traits and applying a formula to obtain the final Writing score. Condition codes are used if the student's writing response is unscorable. The results are reported with the MC results.

Scoring Rubrics

Analytic scoring rubrics were provided by the Oklahoma SDE. The rubrics focus on five specific writing skills: Ideas and Development; Organization, Unity, and Coherence; Word Choice; Sentences and Paragraphs; and Grammar, Usage, and Mechanics. Each is rated from 4 (the highest score) to 1 (the lowest score). In addition, the skill areas of Ideas and Development, and Organization, Unity, and Coherence require that the composition be written in a mode that is appropriate for both audience and purpose. Three modes are emphasized: Opinion/Argument, Informative, and Narrative.

Anchor Papers

The 2013 Operational Writing prompts for Grades 5 and 8 were new and required range-finding along with four new field test prompts for each of Grades 5 and 8. Prior to scoring, the prompts underwent extensive range-finding with the two SDE representatives, on May 6–7, 2013, in Oklahoma City, and in six subsequent conference calls. Rubrics and Writing prompts were reviewed. Anchor candidates were discussed and final anchors selected for use during the training and scoring activities. Three anchor papers per score point were selected for each writing skill area. The range-finding discussions were helpful in defining the parameters of the scoring requirements in the analytic Writing rubrics and in providing insights and summary statements for training of raters.

Section 4.1 – Hand Scoring

Hand scoring involves training and qualifying team leaders and raters, monitoring scoring accuracy and production, and ensuring the security of both the test materials and the scoring facilities. An explanation of the training and qualification procedures follows.

Training, Qualification, and Checkset Materials

All raters were trained and qualified in specific rater item blocks (RIBs), each of which consisted of a single writing prompt. The Operational prompts for Grades 5 and 8 were scored concurrently by two separate groups of raters. The Operational Writing prompts were scored prior to the field test prompts. Four field test prompts each for Grades 5 and 8 were scored concurrently by eight separate groups of raters. Raters and team leaders were trained using the following steps:

- Provide a general introduction to OCCT 3–8
- Introduce and review the writing prompts and scoring rubrics
- Review anchor papers and training papers, and answer questions arising from established scores
- Explain scoring strategies, followed by a question-and-answer period
- Administer Qualifying Round 1
- Review Qualifying Round 1 established scores and answer questions arising from the scores
- Administer Qualifying Round 2 (if necessary)
- Explain condition codes and sensitive paper procedures
- Explain unscannable image procedures

All raters were trained and qualified using the same procedures and criteria used for the team leaders, who had been trained prior to the training of the raters. The qualification process was conducted through the Online Training System and proctored by Handscoring Supervisors and team leaders. The Online Training System enabled supervisors to determine whether a rater had qualified upon completion of the set. The CTB/McGraw-Hill handscoring supervisors proctored the training of the team leaders.

Throughout the course of hand scoring, calibration sets of pre-scored papers (checksets/validity sets) were administered daily to each rater to monitor scoring accuracy and to maintain a consistent focus on the established rubrics and guidelines. Checksets were executed via imaging software that provided images in a manner so that the rater did not know when a checkset was being administered.

The CTB/McGraw-Hill Data Monitoring staff ran inter-rater reliability reports throughout live scoring to look for any raters who were struggling and in need of retraining. Retraining involved a one-on-one discussion between the team leader (or Handscoring Supervisor) and the rater, who discussed the scoring concerns as well as the scoring guides and, if necessary, training papers. If the rater's accuracy on checkset scores did not meet the quality standards after this retraining, they were dismissed from the project immediately.

In addition to the checkset process, CTB/McGraw-Hill's handscoring protocol included the use of read-behinds (spot-checks during live scoring). The read-behind was another valuable rater-reliability monitoring technique that allowed a team leader to review a rater's scored documents and provide feedback and counseling as appropriate.

Selection of Handscorers

CTB/McGraw-Hill and Kelly Services, Inc., strive to develop a highly qualified, experienced core of raters so that the integrity of all projects is appropriately maintained.

CTB/McGraw-Hill requires that all content experts, team leaders, and raters possess a bachelor's degree or higher. Kelly Services, Inc., carefully screened all new applicants and required them to produce either a transcript or a copy of the degree. Kelly Services, Inc., also required a one- to two-hour interview/screening process. Individuals who did not present proper documentation or had less than desirable work records were eliminated during this process. Kelly Services, Inc., verified that 100% of all potential raters met the degree requirement. All experienced raters and team leaders had already successfully completed the screening process.

All potential raters completed a pre-interview activity. For some parts of the pre-interview activity, applicants were shown examples of test responses and were supplied with a scoring guide. In a brief introduction, they became acquainted with the application of a rubric. After the introduction, applicants applied the scoring guide to score the sample responses.

Each applicant's scores were used for discussion during the interview process to determine the applicant's trainability as well as an ability to understand and implement the standards set forth in the sample scoring guide.

Kelly Services, Inc., interviewed each applicant and determined the applicant's suitability for a specific content area and grade level. Applicants with strong leadership skills were interviewed further to determine whether they were qualified to be team leaders.

When Kelly Services, Inc., determined that applicants were qualified, they were recommended for employment. All assignments were made according to availability and suitability. Before being hired, all employees were required to read, agree to, and sign a nondisclosure agreement outlining CTB/McGraw-Hill business ethics and security procedures.

Security guards were on-site whenever employees were present in the building. All employees were issued identification badges and required to wear them in plain view at all times. Visitors and employees who presented at the building entrance without their issued ID badges were issued temporary visitors' badges good for that one day only and were required to wear them in plain view. In addition, employees were advised to arrive the following day with their previously-issued ID badges worn in plain view. All employees and visitors were subject to inspection of their personal effects.

Handscoring Process

Writing prompts were evaluated on each of the five analytic traits and in accordance with Oklahoma's rubric. Using CTB/McGraw-Hill's Electronic Handscoring System (EHS), all writing responses were scored independently by two raters. The EHS employed an automated, random distribution of papers for first reads, second reads, and resolution reads across all readers

designated to score that item. No student biographical or identifiable information was available to raters; all imaged items were scored as blind reads.

Rater Reliability

Section 8 – Summary of Reliability and Validity describes the outcomes of inter-reliability. The inter-rater reliability coefficients for the operational writing prompt are presented in Table 4.1 for Grade 5 and Table 4.2 for Grade 8.

Table 4.1 Inter-Rater Reliability Coefficients for the Operational Writing Prompt, Grade 5

Item	Form	Trait	Score Points	% of Agreement			Checkset Average Agreement Percentages
				Perfect	Adjacent	Perfect + Adjacent	
Grade 5 Operational Writing	1	1	1-4	62.9	34.8	97.7	76.9
		2	1-4	65.3	33.4	98.7	79.5
		3	1-4	64.6	32.0	96.6	77.5
		4	1-4	64.4	33.4	97.8	77.2
		5	1-4	63.3	34.1	97.4	77.6
Grade 5 Field Test Writing	1	1	1-4	70.3	26.8	97.1	89.0
		2	1-4	66.7	30.4	97.1	90.9
		3	1-4	61.4	35.6	97.0	85.4
		4	1-4	63.0	34.6	97.6	83.1
		5	1-4	61.7	35.6	97.3	81.7
Grade 5 Field Test Writing	2	1	1-4	60.8	36.6	97.4	68.7
		2	1-4	60.8	37.0	97.8	60.1
		3	1-4	56.5	40.6	97.1	51.4
		4	1-4	55.6	42.4	98.0	62.1
		5	1-4	55.9	41.8	97.7	74.5
Grade 5 Field Test Writing	3	1	1-4	69.2	26.4	95.6	78.2
		2	1-4	72.6	22.0	94.6	78.2
		3	1-4	59.9	33.2	93.1	79.0
		4	1-4	62.2	31.4	93.6	77.4
		5	1-4	61.2	31.4	92.6	68.1
Grade 5 Field Test Writing	4	1	1-4	61.1	33.6	94.7	73.6
		2	1-4	67.6	31.0	98.6	64.8
		3	1-4	62.7	35.6	98.3	73.1
		4	1-4	67.3	32.0	99.3	63.4
		5	1-4	69.3	29.8	99.1	84.7

Table 4.2 Inter-Rater Reliability Coefficients for the Operational Writing Prompt, Grade 8

Item	Form	Trait	Score Points	% of Agreement			Checkset Average Agreement Percentages
				Perfect	Adjacent	Perfect + Adjacent	
Grade 8 Operational Writing	1	1	1-4	58.0	38.8	96.8	70.4
		2	1-4	57.6	39.3	96.9	71.7
		3	1-4	57.2	39.9	97.1	70.1
		4	1-4	57.8	39.5	97.3	70.2
		5	1-4	57.3	39.8	97.1	67.1
Grade 8 Field Test Writing	1	1	1-4	61.7	35.6	97.3	74.4
		2	1-4	61.4	36.6	98.0	76.2
		3	1-4	58.7	39.6	98.3	74.4
		4	1-4	57.7	40.0	97.7	71.3
		5	1-4	61.7	36.2	97.9	72.6
Grade 8 Field Test Writing	2	1	1-4	60.8	34.2	95.0	70.0
		2	1-4	58.7	37.0	95.7	72.8
		3	1-4	60.8	36.8	97.6	63.1
		4	1-4	58.7	38.8	97.5	58.5
		5	1-4	61.9	35.0	96.9	60.4
Grade 8 Field Test Writing	3	1	1-4	58.4	35.6	94.0	62.5
		2	1-4	60.7	34.6	95.3	69.2
		3	1-4	55.7	38.2	96.9	63.4
		4	1-4	58.7	38.2	96.9	69.2
		5	1-4	56.7	39.2	95.9	66.1
Grade 8 Field Test Writing	4	1	1-4	61.1	35.6	96.7	73.0
		2	1-4	61.1	35.0	96.1	72.6
		3	1-4	60.4	37.0	97.4	73.3
		4	1-4	60.4	36.6	97.0	68.4
		5	1-4	62.0	34.4	96.4	64.6

Section 5 – Sampling Plan and Field Test Design

Section 5.1 – Sampling Plan

A sample representative of the population of Oklahoma students was used for the Spring 2013 equating because final scale scores and performance levels should be reported within two weeks of the closed testing window. To meet this reporting schedule, some students' data were prioritized in the scanning and scoring process and used throughout item level analyses, calibration, and equating. Once the data was available, CTB Research conducted a data integrity check and compared the sample selection (expected) to the 2012 sample to assure that the sample was representative. Table 5.1 shows equating sample size and respective percentage of the population for each grade and content area of the Spring 2013 administration. Because population includes valid students only and a validation process happened after equating, samples can include more students than population, such as for Reading Grade 8. For Grades 3, 4, and 5, samples were used for equating; while for Grades 6, 7, and 8, almost whole population datasets were used. Grades 3, 4, and 5 students took paper-pencil tests and Grades 6, 7, and 8 students took online tests. CTB Research received paper-pencil tests results last due to necessary pickup and transit time.

Table 7, in the **Tables** section, provides the proportion of students in the sample and within the Spring 2013 population that came from each of the subgroups: gender, ethnicity, special population (ELL, IEP, Section 504, and accommodated), and socio-economic status (SES Low and SES High). SES Low flag is for students who received free lunch. It is clear from these tables that the sample is also representative of the state's population, even across most of the subgroups, with the exception of American Indian/Alaskan, which is overrepresented in Grades 3–5 Mathematics. The differences between the sample and the state tend to be less than +/- 6% with a median difference of 0.17 (absolute value).

No sampling decreases or increases were required since the sample received was well representative of the target or expected sample, and therefore, representative of the population of students in Oklahoma.

Table 5.1. Equating Sample Size for Spring 2013 and Respective Percentage of the Population for Each Grade and Content

Content	Grade	Samples	Population	Percent
Math	3	7064	46316	15
	4	6175	45383	14
	5	5707	44295	13
	6	42208	43222	98
	7	42422	43146	98
	8	41573	41377	100
Reading	3	6986	45683	15
	4	6094	44704	14
	5	5516	43798	13
	6	42173	42971	98
	7	42613	43368	98
	8	41836	42341	99
Science	5	36251	44805	81
	8	33941	44209	77

Section 5.2 – Field Test Design

New items are field tested to build up the item bank for future form selections. An embedded field test design was used 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 students' motivation for operational and field test items are the same. Ten multiple-choice field test items per form were placed in common positions across Mathematics, Reading, and Science forms.

Section 5.3 – Data Checking Activities

During the field test data analysis, CTB conducted detailed data checking and applied the following data cleaning exclusionary rules.

5.3.1 Suppressed/Omitted/Invalidated cases

Eliminate suppressed/omitted/invalidated cases flagged in the WinScore files. Eliminate cases that have five or fewer valid attempts.

5.3.2 Duplicate cases

Check and eliminate any duplicate cases by checking student ID (if available), first and last name, middle initial, GIS_CD (GIS code normally contain the district and school ID), teacher name, school, birthday, gender, and response vectors.

5.3.3 Non-public schools

The non-public schools were excluded. Those schools are:

• Oklahoma School for the Deaf
• Oklahoma School for the Blind
• Riverside Indian School
• Sequoia Indian School
• Jones Academy

5.3.4 Second-timers

Students who took the test for a second time were excluded as well.

Section 6 – Methods

The Spring 2013 OCCT 3–8 program was based on the application of a post-equating method using anchor items and equating samples. Verification of the equating samples was described in **Section 5 – Sampling Plan and Field Test Design**. A series of item-level analyses were conducted. These analyses were highly scrutinized to confirm that score keys were accurately and systematically applied and that the summary statistics, such as the item difficulties (*p*-values) and reliabilities (point biserial correlations), were comparable across administrations. CTB/McGraw-Hill Content Development completed a review of all items flagged for possible mis-keys and approved the score keys that were applied. The items were then scaled using the item response theory (IRT) models customary for OCCT 3–8: the three-parameter logistic (3PL) for MC items and the two-parameter partial credit (2PPC) models for Writing prompts. The following section describes the methods used in the analyses of the operational test items.

Section 6.1 – Classical Item Analyses

Item Level Analyses

Each operational test item was first reviewed in terms of classical raw score statistics. Each item was reviewed for frequency distribution (number of students responding for each answer choice or score level), overall *p*-value (proportion of students choosing the correct answer), and point biserial or item-test correlation (how correlated each individual item is with the test as a whole based on the correct response). Typically, *p*-values should range between 0.25 and 0.90. Items with a *p*-value less than 0.25 are considered more difficult because fewer than 25% of the students are achieving the correct answer. Values greater than 0.90 indicate a fairly easy item because more than 90% of students are achieving the correct answer. A small number of easy items are included to motivate low-performing students, and a small number of difficult items are included to motivate high-performing students. With newly-tested content, the *p*-values may dip lower than 0.25, at which point the item should be evaluated in light of the newness of content or students' opportunity to learn the content. Point biserials or item-test correlations are usually in the range of 0.30 and above, although some items can be acceptable when as low as 0.15. The point biserials of each item's distractors, or incorrect responses, were also analyzed, as well as any distractor with a positive point biserial, either of which was reviewed for the possibility of an additional correct response or no correct response.

It is also important to track the rate at which students do not respond to, or omit, items. Omitted items receive a zero score. The rate of omission often provides some information about test times, or speediness, particularly if there is a high rate of items omitted at the end of a test session. It also provides an indication of items that may simply be unclear or illogically presented. When more than 5% of students omit an item, the item is reviewed by both CTB Research and Publishing/Content Development and shared with the SDE.

A summary comparison of the classical statistics between the Spring 2011, Spring 2012, and Spring 2013 OCCT 3–8 results is presented in Table 8. Typically, differences less than about $\left|0.05\right|$ are desirable and, as can be seen, *p*-values and mean item-test correlation differences were within expectation.

A detailed summary of the item level classical raw score statistics and omission rates for Spring 2013 and a comparison to Spring 2011 is provided in **Section 7 – Results**.

Section 6.2 – Differential Item Functioning (DIF)

One of the goals of the 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 reference and focal subgroups (the latter being the group of interest). 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 the flagging of items for DIF might be related to actual differences in relevant knowledge or skills or statistical Type I errors. 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 OCCT 3–8, DIF analyses are conducted across gender (males/females) and ethnicity—focal subgroups African American (not Hispanic), American Indian/Alaskan Native, Asian, Hispanic, and Multiracial versus the reference group White (not Hispanic).

The Mantel-Haenszel DIF statistic (Holland and Thayer, 1988; Michaelides, 2008) was used for the OCCT 3–8 operational tests. It matches students across the reference and focal groups based on their overall test performance, and provides a chi-square to test whether the odds of answering an item correctly are similar for both the reference and focal groups. The items were classified into three categories on the basis of the MH DIF chi-square statistics and the MH delta (Δ) value of A, B, or C for either dichotomous or polytomous items (see Dorans & Holland, 1993; Zieky, 1993; and Michaelides, 2008), where items classified as A are interpreted as having no DIF and items classified as C are interpreted as having potentially severe DIF. The item flag classifications are made as follows:

- The item is classified into the C category if MH DIF is significantly different from zero ($p < 0.05$), and the absolute value of MH delta is greater than or equal to 1.5.
- The item is classified into the B category if MH DIF is significantly different from zero ($p < 0.05$), and the absolute value of MH delta is between 1.0 and 1.5.
- The item is classified into the A category if MH DIF is not significantly different from zero ($p \geq 0.05$), or if the absolute value of MH delta is less than 1.0.

Section 6.3 – Calibration & Item Fit

Item Response Theory (IRT) Models

Item response theory (IRT) allows comparisons between items and examinees, even those from different test forms, by using a common scale for all items and examinees (i.e., as if there were a hypothetical test that contained items from all forms). The three-parameter logistic (3PL) model (Lord & Novick, 1968; Lord, 1980) was used to analyze item responses on the MC items. For analysis of the CR items, the two-parameter partial credit model (2PPC) (Muraki, 1992; Yen, 1993) was used.

IRT is a statistical methodology that takes into account the fact that not all test items are alike and that all items do not provide the same amount of information in determining how much a student knows or can do. Computer programs that implement IRT models use actual student data to estimate the characteristics of the items on a test, called "parameters." The parameter estimation process is called "item calibration."

IRT models typically vary according to the number of parameters estimated. For the OCCT 3–8 tests, three parameters are estimated: the discrimination parameter, the difficulty parameter(s), and, for MC items, the guessing parameter. The discrimination parameter is an index of how well an item differentiates between high-performing and low-performing students. An item that cannot be answered correctly by low-performing students, but can be answered correctly by high-performing students, will have a high discrimination value. The difficulty parameter is an index of how easy or difficult an item is. An item will be more difficult if the difficulty parameter is higher. The guessing parameter is the probability that a student with very low ability will answer the item correctly.

Because the characteristics of MC and CR items are different, two IRT models were used in item calibration. The three-parameter logistic (3PL) model (Lord & Novick, 1968; Lord, 1980) was used in the analysis of MC items. In this model, the probability that a student with ability θ responds correctly to item *i* is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]}, \quad (1)$$

where a_i is the item discrimination, b_i is the item difficulty, and c_i is the probability of a correct response by a very low-scoring student.

For analysis of the CR items, the 2PPC model was used. The 2PPC model is a special case of Bock's (1972) nominal model. Bock's model states that the probability of an examinee with ability θ having a score $(k - 1)$ at the k^{th} level of the j^{th} item is

$$P_{jk}(\theta) = P(x_j = k - 1 \mid \theta) = \frac{\exp Z_{jk}}{\sum_{i=1}^{m_j} \exp Z_{ji}}, \; k = 1 \ldots m_j, \quad (2)$$

where

$$Z_{jk} = A_{jk}\theta + C_{jk} \tag{3}$$

and k is the item response category ($k = 1, 2, \ldots. m_j$). The m_j denotes the number of score levels for the j^{th} item, and typically the highest score level is assigned ($m_j - 1$) score points. For the special case of the 2PPC model used here, the following constraints were used:

$$A_{jk} = \alpha_j(k-1), \tag{4}$$

and

$$C_{jk} = -\sum_{i=0}^{k-1}\gamma_{ji}, \tag{5}$$

where

$$\gamma_{j0} = 0, \tag{6}$$

and α_j and γ_{ji} are the free parameters to be estimated from the data.

Each item has ($m_j - 1$) independent γ_{ji} parameters and one α_j parameter; a total of m_j parameters are estimated for each item.

The IRT model parameters were estimated using CTB/McGraw-Hill's PARDUX software (Burket, 2002). PARDUX estimates parameters simultaneously for MC and CR items using marginal maximum likelihood procedures implemented via the expectation-maximization algorithm (Bock & Aitkin, 1981; Thissen, 1982). Simulation studies have compared PARDUX with MULTILOG (Thissen, 1991), PARSCALE (Muraki & Bock, 1991), and BIGSTEPS (Wright & Linacre, 1992). PARSCALE, MULTILOG, and BIGSTEPS are among the most widely known and used IRT programs. PARDUX was found to perform at least as well as these other programs (Fitzpatrick, 1990; Fitzpatrick, 1994; Fitzpatrick & Julian, 1996).

Assessment of Item Fit to the IRT Model

Item-Model Fit

Item fit statistics discern the appropriateness of using an item in the 3PL or 2PPC model. A procedure described by Yen (1981) was used to measure fit to the 3PL model. Students are rank-ordered on the basis of $\hat{\theta}$ values and sorted into ten cells with 10% of the sample in each cell. For each item, the number of students in cell k who answered item i, N_{ik}, and the number of students in that cell who answered item i correctly, R_{ik}, were determined. The observed proportion in cell k passing item i, O_{ik}, is R_{ik}/N_{ik}. The fit index for item i is

$$Q_{1i} = \sum_{k=1}^{10}\frac{N_{ik}(O_{ik}-E_{ik})^2}{E_{ik}(1-E_{ik})}, \tag{7}$$

with

$$E_{ik} = \frac{1}{N_{ik}} \sum_{j \varepsilon \text{ cell } k}^{N_{ik}} P_i(\hat{\theta}_j). \quad (8)$$

A modification of this procedure was used to measure fit to the 2PPC model. For the 2PPC model, Q_{1j} was assumed to have approximately a chi-square distribution with the following degree of freedom:

$$df = I(m_j - 1) - m_j, \quad (9)$$

where I is the total number of cells (usually 10) and m_j is the possible number of score levels for item j.

To adjust for differences in degrees of freedom among items, Q_1 was transformed to Z_{Q_1}

where

$$Z_{Q_1} = (Q_1 - df)/(2df)^{1/2}. \quad (10)$$

The value of Z will increase with sample size, all else being equal. To use this standardized statistic to flag items for potential misfit, it has been CTB/McGraw-Hill's practice to vary the critical value for Z as a function of sample size. For the OP tests, which have large calibration sample sizes, the criterion $Z_{Q_1}Crit$ used to flag items was calculated using the expression

$$Z_{Q_1}Crit = \left(\frac{N}{1500}\right) * 4, \quad (11)$$

where N is the calibration sample size.

Items were considered to have poor fit if the value of the obtained Z_{Q1} was greater than the value of Z_{Q1} critical. If the obtained Z_{Q1} was less than Z_{Q1} critical, the items were rated as having acceptable fit.

Section 6.4 – Equating

Test Scaling and Equating

Once all item-level analyses were conducted, each Spring 2013 OCCT 3–8 form was calibrated and equated using the Stocking and Lord procedure (Stocking & Lord, 1983), a standard method of equating a new test form onto an existing scale. The Stocking and Lord procedure is based on the test characteristic curve (TCC) from the anchor items, which were selected to be representative of reference forms and Spring 2013 operational forms by statistics and content. CTB PARDUX software was applied to equating (2011). TCC plots for Mathematics, Reading, and Science are found in Figures 15–28.

Stability of Anchor Item

Stability of anchor items for equating procedure is important. The following method was applied to drop anchor items before equating:

1.) Items flagged using the TCC method are considered for exclusion when the correlation between the input and estimated item parameters is below 0.80 for the *a*-parameter and below 0.90 for the *b*-parameter. If the exclusion of an outlying anchor item increases the *a*-parameter correlation to above 0.80 or increases the *b*-parameter correlation to above 0.90, then the anchor is a candidate for removal.
2.) An anchor is a candidate for removal when the item is flagged on four of the seven statistics considered when examining the severe differences between the IRT regression curves: Item characteristic curves (ICCs) for anchor items before and after equating.
3.) An outlier for *a*-parameter or *b*-parameter can be a candidate based on anchor item plot, which shows the relationships of anchor item parameters before and after equating (Kolen and Brennan, 2004).
4.) Removal of the item may not significantly alter the content distribution of the anchor set. The distribution of items across the content standards must remain within 10% of the test blueprint for Reading and Mathematics.
5.) The mean difference and standard deviation ratio are also referenced.
6.) It is important to recognize that differential item performance in two test administrations does not necessarily indicate item flaws and may be affected by population differences, differences in teaching strategies, curriculum changes, etc. Therefore, CTB/McGraw-Hill recommended that Oklahoma SDE consider item content-related factors in addition to statistical evidence of differential item performance in two test administrations.

Items removed from the anchor set based on the flags from the evaluation procedure were still scored as part of the whole test. After an anchor item was removed from the anchor set based on the above criteria, the anchor file needed to be adjusted and a second version of the calibration and equating were produced. All outputs in the second version were to be evaluated following the same guidelines as the original calibration runs.

Section 6.5 – Writing Scoring

Writing prompts were administered in Spring 2013. Students in Grades 5 and 8 responded to one operational writing prompt. 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 6.1. 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 6.1. 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 composite score (CS) is calculated as a weighted composite of the average of two independent ratings for each of the five analytic traits:

$$CS = 15(0.30ID + 0.25OUC + 0.15WC + 0.15SP + 0.15GUM)$$

No IRT model was applied to the Writing prompts.

Section 7 – Results

This section provides the data analysis results for the Spring 2013 OCCT 3–8. Item level analyses for operational and field test items are presented below. Standard, test, and proficiency level student performances are summarized and presented as well. Note that there was one operational MC item suppressed in Mathematics Grade 6 as a result of the SDE review of the operational items. That item did not contribute to students' scores and is not included in the analyses results in this section.

Section 7.1 – Item Level Performance

A summary comparison of the classical item statistics between the Spring 2011, Spring 2012, and Spring 2013 OCCT 3–8 results is presented in Table 8. Typically, differences of less than about $\left|0.05\right|$ are desirable. As can be seen in Table 8, between Spring 2012 and Spring 2013, *p*-values had a slight increase across grades and content areas, with the largest difference seen in Reading Grade 6 (0.03). The mean item-test correlation showed the largest change in Mathematics Grade 3 and Reading Grade 8 (0.03), but few *p*-values changed. The two-year difference was of 0.00% for most grades and content areas.

A summary of the range of *p*-values and item-test correlations for all operational and field test items for Spring 2013 is presented in Table 9. (Note that item-test correlations were calculated by correlating the correct response of the focal item to the remainder of the items in the test, focal item excluded.) For analysis, the Writing trait was treated as an item. As shown in Table 9, the average *p*-values for the operational test items are from the low 0.60s to mid 0.70s in Mathematics; from the low to mid 0.70s in Reading; from the mid 0.50s to upper 0.60s in Science; and in the mid 0.50s in Writing. The range of the *p*-values dips below 0.25 in Mathematics Grade 7and Science Grade 8. Item-test correlations across most grades and content areas are within typical and acceptable ranges; except for Science Grade 5 where one item shows an item-test correlation lower than 0.15. For the field test items, the average *p*-values are in the mid 0.40s to low 0.60s for Mathematics, mid 0.50s to low 0.60s for Reading, mid 0.40s and mid 0.50s for Science, and low to mid 50s for Writing. The item-test correlations for field test items are in the upper 0.20s to low 0.30s for Mathematics and Reading, and in the low to mid 0.20s for Science.

The item omission rates for operational and field test items are presented in Table 10. The operational items for Mathematics, Reading, and Science had omission rate across grades less than 1.4% (well below the 5% criteria), indicating acceptable administration times for the number of items in each test session. The MC field test items for Mathematics, Reading, and Science show omission rates well below the 5% criteria for the MC items; for CR items in the lower Reading grades, omission rates were well above the 5% criteria. For Writing Grade 8, the omission rate reached 4.17%.

DIF results are reported for Mathematics, Reading, and Science in Table 11 for gender and Tables 12–12.3 for ethnicity. The results indicate that the majority of operational test items did not exhibit bias. For operational items on gender DIF, there were a total of 22 items (3.14%) flagged for moderate "B" DIF and 6 items (0.86%) flagged for severe "C" DIF. For operational

items in the four ethnicity groups DIF analyses, there were between 0.00% to 8.86% of items flagged for moderate "B" DIF and between 0.14% and 2.71% of items flagged for severe "C" DIF.

DIF results for the field test items in Mathematics, Reading, and Science show that for gender DIF, there were a total of 23 items (2.64%) flagged for moderate "B" DIF and 2 items (0.23%) flagged for severe "C" DIF. For field test items in the four ethnicity groups included in the DIF analyses, there were between 0.34% to 10.22% of items flagged for moderate "B" DIF, and between 0.00% and 4.94% of items flagged for severe "C" DIF.

All of the items flagged were reviewed by CTB content experts who cross-referenced all teacher judgments and comments from across the content reviews, bias and sensitivity reviews, as well as alignment workshops to make decisions with the SDE about suppressions from operational scoring and use of flagged operational and/or field test items in future test forms.

Problematic Items

Piloting or field testing items is the best way to find potentially problematic items in the item pool. However, even during an operational administration, there are times that items become unstable or do not exhibit the highest expected qualities. Therefore, the evaluation of items across administrations from the content reviews, bias and sensitivity reviews, alignment workshops, and the various statistical analyses can be exhaustive and must be sensitive to the test blueprints, which can sometimes result in the suppression of some operational items from student scores and of some field test items from the item pool. Sometimes, OE items that do not show enough/adequate case counts at a given score level, resulting in score collapses, and items that do not converge during scaling or that exhibit extreme misfit are also suppressed.

During the Spring 2013 OCCT 3–8 operational and field test administration, items were reviewed for their classical statistics, and when those statistics were outside the range of difficulty (*p*-values less than 0.25 or greater than 0.90) or showed low item-test-correlations (less than 0.15) for a specific item, the item was used or kept as a "good item" in the pool only when the content of the item justified its use (e.g., it was for a new standard or new approach that was expected to be difficult). For the Spring 2013 OCCT 3–8 operational test, there was one item in Mathematics Grade 6 suppressed before scoring began. The item was suppressed because the wrong art was in the item. Out of the 896 newly field tested items, the ones with less than desirable *p*-values and item-test correlations were reviewed by CTB/McGraw-Hill content experts and Research, and 25 were considered to have less than desirable statistics and were suppressed from the item pool.

Section 7.2 – Standards Level Performance

A review of the item difficulty across standards within each grade and content area is provided to illustrate at which standards items were more or less difficult for students. The summaries are presented in Tables 13–15. The tables provide the number of operational items, the reliability (coefficient alpha), and standard error of measurement (SEM) (formulas for which are found in **Section 8 – Summary of Reliability and Validity**), and the average difficulty or IRT location (*b*

parameter) value. The tables also provide the average *p*-values for the state and for each proficiency level for each standard.

The reliability at each standard, which is influenced by the number of items contributing to each standard, ranges from 0.51 to 0.82 in Mathematics, from 0.42 to 0.84 in Reading, and from 0.48 to 0.68 in Science. Across all content areas, the standard errors are no greater than 1.98, and the maximum amount of IRT information is 0.15. IRT locations should be reviewed within each grade by standard, as should the *p*-values.

IRT locations and *p*-values can also be reviewed within each grade by standard in Tables 13–15 as well. The IRT scale locations provide an indication as to the average *b*-parameters or location values of a set of items contributing to each of the standards. The distinction from average *p*-values is that the IRT locations provide information about where the items are found along the scale score continuum, such that higher values indicate a lower probability of students with low estimated ability of answering those items correctly. The *p*-values provide only the proportion of students in each group answering the items correctly, averaged across items within each of the standards.

Section 7.3 – Test Level Performance

Total Group Scale Scores

The OCCT 3–8 applies a number-correct scoring method based on the 3PL IRT model, which is used to estimate scale scores corresponding to each raw score. In this method, all students who have the same raw score get the same scale score regardless of which items are correct.

Tables 16 and 17 provide the state-level distribution of the scale scores across grades and content areas for Spring 2012 and Spring 2013, respectively. Table 17 provides the state-level distribution of the scale scores across grades and content areas as well as the distribution across the 25th, 50th, and 75th percentiles for Spring 2013. (Spring 2013 scale score reliability as coefficient alpha and standard error of measurement are also provided and further explained in **Section 8 – Summary of Reliability and Validity**.) As a reference only, Table 16 also shows those results for Spring 2012. Histograms and associated skewness and kurtosis of the data for Spring 2013 Mathematics, Reading, and Science are provided in Figures 1–14. The data are close to normally distributed with a very minimal positive skew in most content areas and grades.

Subgroup Scale Scores and Mean Differences

Subgroup-level scale score performance data (scale score means and standard deviations, minimum and maximum scale scores, reliability and standard error of measurement) are provided along with state-level data in Tables 18–20. Mean differences were subjected to independent sample *t*-tests for accommodated students, ELL, gender, IEP, Section 504, and Low SES subgroups, in each content area. One way Analysis of Variance (ANOVA) tests were conducted across ethnicities, for which equal variances were not assumed and the level of significance was set at 0.05. Results of the *t*-tests and ANOVAs are found in Tables 21–38 and Tables 39–41, respectively.

As presented in Tables 21–38, results of the *t*-tests show that males outperform females in most grades in Mathematics, with the exception of Grade 8, and in Science Grades 5 and 8, while females outperform males in all grades in Reading. Mean differences are not statistically significant at Mathematics Grades 5 and 8 only. At the subgroup level, the results show that accommodated students tend to have the lowest performance of the "special population" subgroups. Overall results of the *t*-tests within each category indicate that accommodated, ELL, IEP, and Low SES students all score significantly lower than the rest of the population in all grades and content areas (mean differences ranging from 31 to 98 fewer scale score points), as expected. For the Section 504 group, the same is true in all grades and content areas, but the average differences compared with the rest of the state are less than 20 scale score points and not significant at Science Grade 8.

Statistically significant differences exist between the ethnicity groups in all content areas and grades as presented in the ANOVA results in Tables 39–41.

A *post-hoc Dunnett's C pair-wise* comparison analysis was conducted to identify potential pairs of significant differences ($p = 0.05$), the results of which are found in Tables 42–44. In comparing ethnicities across all grades and content areas, students identified as Asian and White (not Hispanic) tended to outperform the other ethnicities. As shown in Tables 42–44, Asian outperformed all other ethnicities in all content areas and grades. White (not Hispanic) outperformed all other ethnicities, except Asian, in all content areas and grades. Most pairs were significantly different, with the following exceptions which were **not** significant: American Indian/Alaskan Native is not significantly different from Multiracial in Mathematics Grades 3, 4, 6, 7 and 8 and Reading Grades 7 and 8.

Section 7.4 – Proficiency Level Performance

Table 45 shows the scale score means and standard deviations for the state and for students in each proficiency level. Tables 46 and 47 provide the statewide distribution (or "impact data") of students within each proficiency level (*Unsatisfactory, Limited Knowledge, Proficient, and Advanced*) and the overall pass rates defined as the total percentage of students in both the *Proficient and Advanced* proficiency levels for Spring 2012 and 2013, respectively. Table 46 provides Spring 2012 data as a reference. Please note that Tables 45 and 47 do not include the number of students considered *Undetermined* (invalid) in the denominator of calculation.

Impact data across proficiency levels are also provided for each gender, ethnicity, and special population subgroups in Tables 48–50, where comparative performance across subgroups mimics what was provided for the scale score descriptions.

Section 8 – Summary of Reliability and Validity

This section summarizes some of the evidence in the earlier sections and provides additional evidence to support the degree to which the OCCT 3–8 tests are reliable and valid. For the OCCT 3–8, several measures of reliability are available. First, the tests are administered in standard fashion to all students. When students needed accommodations, such accommodations were provided with specific guidance from the *OSTP 2012–2013 Test Preparation Manual* (http://ok.gov/sde/sites/ok.gov.sde/files/documents/files/2705543-W_tpm_w13OK.pdf) under General Guidance, which describes details about the tests as well as specific administration policies, procedures, and accommodation guidelines.

Section 8.1 – Item Level Reliability

Item-specific reliability statistics include inter-rater reliability, item-test correlations, and differential item functioning (DIF) or item bias. The inter-rater reliabilities of OE items rely heavily on the solid and consistent training of the hand scorers, as was described in **Section 4 – Scoring**. Tables 4.1 and 4.2, for Grade 5 and Grade 8, respectively, provide the relevant inter-rater reliability statistics, which are presented in terms of the percentage of perfect and adjacent agreement and checkset average agreement.

The point biserial, or item-test correlation, a type of internal consistency measure, is one measure of the correlation between each item and the overall test as described in **Section 6 – Methods**, results of which were described in **Section 7 – Results**. The item-test correlations for each content area, grade, and item type are shown in Table 56. The operational item-test average correlations range from 0.39 to 0.44 (Mathematics); from 0.39 to 0.42 (Reading); and 0.37 and 0.36 (Science). One operational item in Spring 2013 OCCT 3–8 Science Grade 5 presented an item-test correlation less than 0.15. That item was investigated by Content Development and found to be correctly scored. Any operational items with extremely low point biserial that may remain in the OCCT 3–8 item pool will be avoided in future operational forms.

DIF statistics (described in **Section 6 – Methods** and **Section 7 – Results**) provide a measure of the systematic errors by subgroups that are specifically attributed to some bias or systematic over- or under-representation of subgroup performance when compared to the total group performance. As discussed in **Section 7 – Results** and is apparent in Tables 11–12.3 (last rows), the percentage of operational and field test items that exhibited DIF at the moderate and severe levels was 3.37% for gender and between 0.25% and 13.56% for the four ethnicity groups.

Section 8.2 – Test Level Reliability

Total test reliability statistics (alpha and CSEMs) measure the level of consistency (reliability) of performance over all test questions in a given form, the results of which imply how well the questions measure the content domain and could continue to do so over repeated administrations. Total test reliability coefficients (in this case measured by Cronbach's alpha [α; 1951]) may range from 0.00 to 1.00, where 1.00 refers to a perfectly reliable test. The OCCT 3–8 reliability data are based on Oklahoma-specific representative samples from each grade (the scaling sample), and the results for 2013 are typical of the results obtained for all previous OCCT 3–8

operational tests. The total test reliabilities of the operational forms were evaluated first by Cronbach's α (1951) index of internal consistency. The specific calculation for Cronbach's α is

$$\hat{\alpha} = \frac{k}{k-1}\left(1 - \frac{\sum \hat{\sigma}_i^2}{\hat{\sigma}_X^2}\right), \tag{12}$$

where k is the number of items on the test form, $\hat{\sigma}_i^2$ is the variance of item i, and $\hat{\sigma}_X^2$ is the total test variance and the summation is over all the items ($i = 1, \ldots, k$) on the test. Achievement tests are typically considered of sound reliability when their reliability coefficients are in the range of 0.80 and above.

Table 57 shows the reliability coefficients for each scored operational test form for each content area and grade for both Spring 2012 and Spring 2013. Alpha reliability coefficients for Spring 2012 and Spring 2013 are quite similar. Reliability for Spring 2013 ranged between 0.85 (Science Grade 8) and 0.91 (Mathematics Grade 3). Such a range is indicative of the high reliability of Spring 2013 OCCT 3–8 operational tests. As is evident in Tables 18–20, for Spring 2013 state and subgroup data, the coefficients are quite high and similar to the state, even at the subgroup levels. The mean (and range) of the state-level reliability coefficients for each content area are as follows: Mathematics 0.90 (range 0.89–0.91), Reading 0.89 (range 0.88–0.90), Science 0.86 (range 0.85–0.86). At the subgroup level, the lowest reliability (0.76) was found for the ELL students in Science Grade 8.

The SEM is another measure of reliability and is a direct estimate of the degree of measurement error in students' total scores (per the alpha reliability coefficient). The SEM represents the number of score points about which a given score can vary, similar to the standard deviation of a score; the smaller the SEM, the smaller the variability of the estimate, and the higher the reliability. The total SEMs are computed with the following formula:

$$SEM = SD\_TT(\sqrt{1-\hat{\alpha}}) , \tag{13}$$

where *SD_TT* is the standard deviation for the total test, and $\hat{\alpha}$ is the result of the calculation of Cronbach's α in Equating 12.

The CSEMs conditional on each scale score are computed with the following formula:

$$CSEM = SD\_SS(\sqrt{1-\hat{\alpha}}) , \tag{14}$$

SD_SS is the standard deviation of the scale score. The total test SEMs for each test form are provided for each content area and grade at the state and subgroup levels in Tables 18–20. Scale score specific SEMs are given in Tables 58–60, which also provide the raw scores associated with each scale score.

Section 8.3 – Test Level Validity

Exploratory Factor Analysis (EFA) and Confirmatory Factor Analyses (CFA) were conducted to evaluate the unidimensionality assumption of the OCCT 3–8 test scores for the total population and then various subgroups such as accommodated, ELL, Section 504, Low SES, and IEP. In factor analyses, the "construct" is referred to as a factor. If the data are essentially unidimensional, a single factor should account for most of the variation in the data.

Accordingly, a unidimensional factor model was tested using polychoric correlation coefficients against the obtained covariance matrix using maximum likelihood estimation (Bentler & Bonett, 1980, Jöreskog & Sorbom, 1989) for each content area and grade for the total population and each subgroup using SAS version 9.1. The polychoric correlation is most appropriate when variables are dichotomous or ordinal and together are assumed to reflect a single, underlying construct (Byrne, 1998).

First, the factorability of the correlation matrix was examined before conducting the CFA (Is the data adequately correlated and thus analyzable or "factorable" to move forward?). The Kaiser-Meyer Olkin (KMO; Kaiser, 1970, 1974) measure of sampling adequacy was used through an EFA procedure to evaluate the strength of the linear relationship among the items within each correlation matrix. KMO values in the 0.90 and greater range are considered "marvelous" according to Kaiser's (1974) criteria. As shown in Tables 61–64, KMO values for the total group ranged from 0.96 to 0.97, and for each subgroup from 0.87 to 0.96 (Accommodated), from 0.85 to 0.96 (ELL), from 0.95 to 0.97 (Free Lunch), and from 0.92 to 0.97 (IEP). That most of the KMO values are in the "marvelous" range suggests that the matrix is appropriate for CFA for each analysis.

As a rough estimate of the number of factors (dimensions or constructs) that might be present in the data, the Kaiser criterion of computing the eigenvalues for the correlation matrix was examined next. Eigenvalues represent how much variability is accounted for by each factor not in sum but out of the total amount of variance, which means there will be times the percentages can be greater than 100%. Tables 61–64 also show the total amount of variance that exists in each form, as well as the percent of variance accounted for by the initial eigenvalue. For the total group analyses, the first eigenvalue's measure of the amount of variance in relation to the total variance is 86–97% (Mathematics), 91–103% (Reading), and 104–106% (Science). The range of variance by the first eigenvalue in each content area and subgroup is as follows:

- Accommodated: 81–88% (Mathematics), 88–93% (Reading), 98% and 91% (Science).
- ELL: 79–87% (Mathematics), 84–93% (Reading), 91% and 85% (Science).
- Free Lunch: 85–96% (Mathematics), 91–102% (Reading), 104% and 106% (Science).
- IEP: 85–92% (Mathematics), 92–95% (Reading), 100% and 97% (Science).

Such values indicate one major factor is present in each of the content assessments. It is interesting to note that the range of variance for Science is mostly higher than the other two content areas for the total population and each subgroup.

As a rule, "essential unidimensionality" is assumed when the ratio of the first eigenvalue to the second eigenvalue is at least three. The final column of Tables 61–64 provides the ratio of the first and second eigenvalues. All grades and content areas for the total population and each subgroup have no ratios less than six; therefore, the OCCT 3–8 tests are demonstrating essential unidimensionality per the eigenvalue ratio criterion.

An additional available criterion used in EFA to judge the number of factors present is the scree test (Cattell, 1966) of eigenvalues plotted against factors. Examinations of the scree plots (Figures 29–98) for all grades and content areas for the total population and each subgroup indicate a single factor model is present and similar patterns between the total population and subgroups.

Summary inspection across all the criteria—variance, ratio of eigenvalues, and scree plots—seems to indicate that the tests for each content area and grade, and for each subgroup, are essentially unidimensional. It is important to review the relationships of factors in conjunction with all other data, particularly where items may be dependent (for example, where all open-ended items are scored twice).

Section 8.4 – Performance Level Reliability

Proficiency Level Reliability

One of the cornerstones of the federal Elementary and Secondary Education Act (ESEA) emphasizes the need for all students to score in the "Proficient" category on English Language Arts, Mathematics and Science. Because of a heavy emphasis on moving all students to or above the "Proficient" category, the consistency and accuracy of the classification of students into these proficiency categories is of particular interest. The statistical quality of cut scores that define the proficiency levels in which students are classified based on their performance serves as additional validity evidence. Details about the Science standard setting workshop and the Bookmark Standard Setting Procedure used to set the cut scores are given in the Oklahoma School Testing Program Standard Setting Technical Report for OCCT Grades 5 and 8 Science and Writing (CTB/McGraw-Hill, 2013). It may be useful to note that the Bookmark Procedure (Mitzel, Lewis, Patz, & Green, 2001) is a well-documented and highly regarded procedure that has been demonstrated by independent research to produce reasonable cut scores on tests across the country.

It is also important to review the specific scale score SEM for each cut score. Table 65 shows the Spring 2013 SEMs estimated for each of the cut scores for each content area and grade. Comparison of these SEMs to the SEMs associated with other OCCT 3–8 scale scores for each test (shown in Tables 58–60) reveals that these values are almost always among the lowest, meaning that the OCCT 3–8 tests tend to measure most accurately near the cut score. This is a desirable quality when cut scores are used to classify examinees. (Not every scale score possible, sometimes including the cut score, is shown in Tables 58–60; there are more scale scores possible at each raw score than can be shown in these tables.)

Not only is it important that the amount of measurement error around the cut score is minimal, but also important is the expected consistency with which students would be classified into performance levels if given the test over repeated occasions.

Classification consistency is defined as the extent to which two classifications of a single student agree from two independent administrations of the same test (or two parallel forms of the test). Classification consistency and accuracy are additional measures of test reliability as well as validity. Reliability coefficients, such as Cronbach's alpha, are used to check for the internal consistency within a single test. Test-retest reliability requires two administrations of the same test, which requires another test as an external reference. Consistency in the classification sense represents how well two forms of an assessment with equal difficulty agree (Livingston & Lewis, 1995). It is estimated using actual response data and total test reliability from an administered form of an assessment, from which two parallel forms of the assessment are statistically modeled and classifications compared.

Classification accuracy is defined as the agreement between the actual classifications using observed cut scores and true classifications based on known true cut scores (Livingston & Lewis, 1995). It is common to estimate classification accuracy by utilizing a psychometric model to find true scores corresponding to observed scores.

In other words, classification consistency refers to the agreement between two observed classification results, while classification accuracy refers to the agreement between the observed classification outcome and the true classification result. A straightforward approach to classification consistency estimation can be expressed in terms of a contingency table representing the probability of a particular classification outcome under specific scenarios. For example, the following is a contingency table of (H+1) × (H+1), where H is the number of cut scores such that two cut scores yield a 3x3 contingency table as follows.

	Level 1	**Level 2**	**Level 3**	**Sum**
Level 1	P_{11}	P_{21}	P_{31}	$P_{.1}$
Level 2	P_{12}	P_{22}	P_{32}	$P_{.2}$
Level 3	P_{13}	P_{23}	P_{33}	$P_{.3}$
Sum	$P_{1.}$	$P_{2.}$	$P_{3.}$	1.0

To report classification consistency, Swaminathan, Hambleton, and Algina (1974) suggest using Cohen's kappa (1960):

$$\text{kappa} = \frac{P - P_c}{1 - P_c}, \tag{15}$$

where P is defined as the sum of diagonal values of the contingency table (shaded above) and P_c is the chance probability of a consistent classification under two completely random assignments. This probability, P_c, is the sum of the probabilities obtained by multiplying the

marginal probability of the first administration and the corresponding marginal probability of the second administration:

$$P_c = (P_{1\cdot} \times P_{\cdot 1}) + (P_{2\cdot} \times P_{\cdot 2}) + (P_{3\cdot} \times P_{\cdot 3}). \quad (16)$$

The Livingston and Lewis (1995) method based on the binomial error model and the four-parameter beta true score distribution was applied to OCCT 3–8. Tables 66 and 67 show the classification consistency and classification accuracy indices. Note that the values of all indices depend on several factors, such as the reliability of the actual test form, the distribution of scores, the number of cut scores, and the location of each cut score. The probability of a correct classification (Consistency) is the probability that the classification the student received is consistent with the classification that the student would have received on a parallel form; in other words, that the classification is correct. This is akin to the exact agreement rate in inter-rater reliability, and the expectation is that this probability would be high.

Table 66 shows the average consistency is 0.69 across all grades and content areas, and ranges from 0.6 (Science Grades 5 & 8) to 0.78 (Reading Grade 3). The average accuracy is 0.77 across all grades and content areas, and ranges from 0.70 (Science Grades 5 and 8) to 0.84 (Reading Grade 3). Cohen's kappa (Kappa) provides the same type of reliability or agreement statistic as in the inter-rater reliabilities. In this context, it represents the agreement of the classifications between the two parallel forms with consideration of the probability of a correct classification by chance (Consistency–Chance[1])/(1–Chance). In general, the value of Kappa is lower than the value of Consistency because the probability of a correct classification by chance is greater than 0. This is true of the OCCT 3–8 data in Table 66. The average Kappa is 0.54 over all grades and content areas and ranges from 0.46 (Science Grade 8) to 0.59 (Mathematics Grade 3).

Consistency and accuracy are important to consider together. The probability of accuracy (Accuracy) represents the agreement between the observed classification, based on the actual test form, and the true classification given the modeled form. Table 67 shows consistency and accuracy at the cut score level. The average consistency across grades and cut score level is 0.89, ranging from 0.82 (Science Grade 8, at the Proficient and Advanced proficiency levels) to 0.97 (Reading Grade 3, at the Advanced proficiency level). The average accuracy across grades and cut score level is 0.92, ranging from 0.87 (Science Grade 8, at the Proficient and Advanced proficiency levels) to 0.98 (Reading Grade 3, at the Advanced proficiency level). Finally, Table 68 provides the probability of false positives (FP) and false negatives (FN) as measures of error in the data table, and these are low (no greater than 0.06), as expected.

[1] The probability of a correct classification by chance (Chance) is the probability that the classification is correct and is due to chance alone. The probability of Chance is estimated under a complete random assignment procedure using the marginal distribution of each form. The Chance probabilities are expected to be low.

Section 9 – Online Disruption in Spring 2013 Administration

Background

Online testing was administered to all Oklahoma OCCT Grades 6–8 Reading and Mathematics tests in the Spring 2013 administration. While the online testing window started in mid-April and was progressing well, on Monday, April 29th and Tuesday, April 30th, students taking online assessments in the state experienced system interruptions, which led some local districts to temporarily suspend testing on those days. CTB's diagnostic efforts revealed that two separate issues with CTB/McGraw-Hill's Online Assessment System (OAS) caused the interruptions.

CTB/McGraw-Hill utilizes industry standard tools and current state software engineering practices for building and evaluating our systems. In preparation for the 2013 testing season, additional capacity was added and extensive load testing was performed. However, the factors that triggered the system interruptions on April 29th and 30th had not been detected in testing or prior year production use.

The key difference between what was seen in actual use versus CTB's performance evaluation was high utilization sustained over a longer period of time, which consumed more physical memory in the first outage and subsequently more virtual memory in the second. This memory utilization was more than the load tests predicted.

CTB engineers worked quickly to address both issues. In the first event, CTB restarted systems and added physical memory. In the second event, CTB identified necessary memory configuration changes and implemented them as rapidly as possible to allow the testing to continue. Changes included increased physical memory on nodes hosting applications, increased virtual memory Heap size on each node, and finally after analyzing utilization, adding more nodes on the following weekend. Once all these upgrades were completed, the OAS application was able to manage the volume of requests without degraded performance.

On April 30th, while CTB engineers were trying to restore the system health back to normal, one activity was to "cut over" to the disaster recovery site. While the system remained accessible, this "cut over" caused interruptions for almost all students who were active in the system. Also, as the system was moved from the working servers to the disaster recovery servers, some student responses were not immediately accessible to students when they logged back into that test session. All of the students' responses had been saved, but they were not immediately available due to the time it takes for the disaster recovery systems to be synchronized with live testing data from the primary servers. Based on the severity of the interruption and a recommendation, the State Superintendent requested that students complete their current test and then schools suspend online testing. For OCCT Grades 6–8 testing, some students, who were testing operational forms switched to an equivalent form. On May 1st, online testing returned to normal, and students did not experience further interruptions.

Section 9.1 – Interruption Study

The SDE requested an independent third-party study to investigate the impact of the disruptions on the test scores, and HumRRO was selected by the SDE to perform the study. Technical Advisory Committee (TAC) consists of researchers and assessment experts selected by the SDE. TAC members participated in a conference call with HumRRO researchers to recommend a design for the disruption study.

HumRRO applied several different methods using propensity matching between disrupted students, non-disrupted students and linear regression. Their study indicated that in general, students appeared to do as well on the test regardless of disruptions in testing (For detailed description, see "A Statistical Investigation of Oklahoma Computer Disruptions" in Appendix B).

Section 9.2 – Actions for Disrupted Students

There were two types of disruption issues: 1) There were students whose performance was considered to be impacted and 2) There were students whose immediate scores (displayed on the online test delivery client when the test is finished) and two-week scores were different.

For the first issue, the SDE decided on retaining all scores of impacted students who scored Proficient or Advanced. Though students as a whole did not experience depressions in scores, the SDE wanted no lasting effect on student performance. Also, the SDE decided to not report scores for impacted students who scored Limited Knowledge or Unsatisfactory. Though the HumRRO study suggested no systematic impact of the disruption on test scores, in the event a particular student was impacted, not reporting the scores would ensure that the individual student was not wrongly scored. The SDE selected this approach for the benefit of the students and schools.

For the second issue, about 1,400 students received an immediate raw score that differed from the score they received on the two-week preliminary reports. Several factors contributed to this issue, one of which was caused by local sites downloading the wrong test system, and the other where it appears that during a short time-span on the second day of the interruptions (4/30), some of the student responses were not saved for inclusion in the two-week reports. Students received the higher of the two scores, the immediate and the two-week report scores. Students with higher immediate raw scores received a letter from CTB indicating their proficiency level. They were included in the report card, but not in the school or district summary report. Students with a higher two-week score received a traditional score report that included performance level and performance by content standard.

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.

Barton, K. (2007). Validity and accommodations: The journey toward accessible assessments. In C. Cahalan Laitusis & L. Cook (Eds). *Large-scale assessment and accommodations: What works?* Council for Exceptional Children.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* 37:29–51.

Bock, R.D. and M. Aitkin. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika* 46:443–459.

Burket, G.R. (2002). *PARDUX* [Computer program].

Byrne, B. M. (1998). *Structural equation modeling with LISRELA, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334.

CTB/McGraw-Hill. (2011). *PARDUX Software version 1.66.* Monterey, CA.

CTB/McGraw-Hill. (2013). *Oklahoma School Testing Program: 2013–2014 Test Preparation Manual.* Monterey, CA: CTB/McGraw-Hill LLC.

CTB/McGraw-Hill. (2013). *Oklahoma School Testing Program: Standard setting technical report for OCCT Grades 5 and 8 Science and Writing*. Monterey, CA: Author.

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.

Fitzpatrick, A.R. (1990). *Status report on the results of preliminary analysis of dichotomous and multi-level items using the PARMATE program*.

Fitzpatrick, A.R. (1994). *Two studies comparing parameter estimates produced by PARDUX and BIGSTEPS*.

Fitzpatrick, A.R. & M.W. Julian (1996). *Two studies comparing the parameter estimates produced by PARDUX and PARSCLE*. Monterey, CA: CTB/McGraw-Hill.

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

Jöreskog, K., & Sorbom, D. (1989). *LISREL 7 User's Reference Guide*, Chicago: Scientific Software International.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika, 35*(4), 401–415.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36.

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

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

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.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22, 719–748.

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.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249–281). Mahwah, NJ: Lawrence Erlbaum.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement* 16:159–176.

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.

Muraki, E. & R.D. Bock (1991). *PARSCALE*: *Parameter Scaling of Rating Data* [Computer program]. Chicago, IL: Scientific Software, Inc.

Oklahoma State Department of Education. (2013). *Test Support (Teachers and Administrators).* Available online: http://www.ok.gov/sde/test-support-teachers-and-administrators.

Pearson. (2011). *Oklahoma School Testing Program: Oklahoma Core Curriculum Tests Grades 3 to 8 Assessments 2011 Technical Report.* Pearson.

Pearson. (2012). *Oklahoma School Testing Program: Oklahoma Core Curriculum Tests Grades 3 to 8 Assessments 2012 Technical Report.* Pearson.

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

Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion-referenced tests: A decision-theoretic formulation. *Journal of Educational Measurement*, *11*(4), 263–267.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* 47:175–186.

Thissen, D. (1991). MULTILOG: Multiple category item analysis and test scoring using item response theory [Computer software]. Chicago, IL: Scientific Software International, Inc.

Thissen, D., Chen, W-H., & Bock, R. D. (2003). *MUTILOG; Multiple-category item analysis and test scoring using item response theory (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.

United States Department of Education. (2002). *No Child Left Behind Act of 2001.* (U.S. Department of Education, Office of Elementary and Secondary Education). Washington, DC.

United States Department of Education. (2004). *Standards and assessment peer review guidance: Information and examples for meeting requirements for the No Child Left Behind Act of 2001.* (U.S. Department of Education, Office of Elementary and Secondary Education). Washington, DC.

Wright, B.D. & J. M. Linacre. (1992). *BIGSTEPS Rasch Analysis* [Computer program]. Chicago, IL: MESA Press.

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.

Yen, W.M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement* 30:187–213.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–347). Hillsdale, NJ: Lawrence Erlbaum.

Tables

Table 7. Subgroup Representativeness of Scaling Sample Compared to Total Population, Spring 2013

		Female			Male		
Content	Grade	Sample	State	Diff.	Sample	State	Diff.
Mathematics	3	50.26	49.44	0.81	49.74	50.56	-0.81
	4	49.53	49.79	-0.26	50.47	50.21	0.26
	5	49.73	49.94	-0.21	50.27	50.06	0.21
	6	50.05	49.58	0.47	49.95	50.42	-0.47
	7	50.28	49.83	0.45	49.72	50.17	-0.45
	8	50.47	50.17	0.30	49.53	49.83	-0.30
Reading	3	50.70	49.70	1.00	49.30	50.30	-1.00
	4	50.09	50.15	-0.06	49.91	49.85	0.06
	5	50.27	50.22	0.05	49.73	49.78	-0.05
	6	50.27	49.89	0.38	49.73	50.11	-0.38
	7	50.38	49.99	0.39	49.62	50.01	-0.39
	8	50.48	50.17	0.31	49.52	49.83	-0.31
Science	5	50.04	49.95	0.09	49.96	50.05	-0.09
	8	50.04	49.80	0.24	49.96	50.20	-0.24

Table 7. Subgroup Representativeness of Scaling Sample Compared to Total Population, Spring 2013 *(continued)*

		African American (Not Hispanic)			American Indian/Alaskan		
Content	Grade	Sample	State	Diff.	Sample	State	Diff.
Mathematics	3	7.82	8.91	-1.09	18.73	14.97	3.76
	4	8.44	8.94	-0.49	18.35	15.21	3.14
	5	8.17	9.10	-0.93	20.90	15.97	4.93
	6	8.63	8.69	-0.06	16.28	16.11	0.17
	7	8.99	8.96	0.03	16.55	16.45	0.10
	8	8.85	8.88	-0.03	16.42	16.31	0.11
Reading	3	7.92	8.89	-0.97	18.94	14.99	3.95
	4	8.71	8.94	-0.22	18.14	15.18	2.96
	5	7.78	9.09	-1.31	21.07	15.97	5.10
	6	8.68	8.74	-0.06	16.42	16.26	0.15
	7	9.13	9.12	0.00	16.54	16.46	0.08
	8	9.04	9.05	-0.01	16.46	16.40	0.07
Science	5	9.90	9.22	0.68	16.01	15.97	0.04
	8	10.06	9.33	0.73	16.67	16.37	0.30

Table 7. Subgroup Representativeness of Scaling Sample Compared to Total Population, Spring 2013 *(continued)*

		Hispanic			Asian		
Content	Grade	Sample	State	Diff.	Sample	State	Diff.
Mathematics	3	13.55	15.81	-2.26	1.73	2.05	-0.32
	4	14.52	15.16	-0.65	1.49	1.87	-0.38
	5	13.29	14.61	-1.32	1.32	1.93	-0.61
	6	12.23	13.88	-1.65	2.1	2.02	0.08
	7	12.03	13.2	-1.17	1.94	1.88	0.06
	8	11.5	12.68	-1.18	2.02	1.98	0.04
Reading	3	13.49	15.71	-2.22	1.63	2.02	-0.4
	4	14.12	15.09	-0.96	1.44	1.86	-0.42
	5	12.88	14.49	-1.61	1.36	1.92	-0.55
	6	12.22	13.69	-1.47	2.05	1.96	0.08
	7	12.02	13.07	-1.05	1.94	1.88	0.06
	8	11.42	12.58	-1.16	1.95	1.91	0.04
Science	5	14.13	14.61	-0.48	1.93	1.92	0.01
	8	12.84	12.92	-0.08	1.98	1.89	0.09

Table 7. Subgroup Representativeness of Scaling Sample Compared to Total Population, Spring 2013 (*continued*)

		Hawaiian/Pacific Islander			White (Not Hispanic)			Multiracial		
Content	Grade	Sample	State	Diff.	Sample	State	Diff.	Sample	State	Diff.
Mathematics	3	0.35	0.26	0.09	50.59	51.37	-0.78	7.23	6.63	0.60
	4	0.45	0.30	0.15	50.22	52.13	-1.91	6.53	6.39	0.14
	5	0.27	0.26	0.01	49.57	51.81	-2.25	6.48	6.31	0.17
	6	0.27	0.26	0.00	53.97	52.50	1.47	6.51	6.54	-0.02
	7	0.22	0.22	0.00	53.76	52.76	1.00	6.50	6.52	-0.02
	8	0.30	0.29	0.01	54.72	53.61	1.10	6.20	6.25	-0.05
Reading	3	0.34	0.26	0.08	50.49	51.49	-1.00	7.19	6.64	0.55
	4	0.45	0.30	0.15	50.58	52.29	-1.71	6.56	6.35	0.21
	5	0.24	0.25	-0.01	49.97	51.93	-1.96	6.69	6.36	0.33
	6	0.27	0.27	0.01	53.83	52.49	1.35	6.53	6.60	-0.06
	7	0.21	0.22	0.00	53.70	52.74	0.95	6.46	6.50	-0.04
	8	0.25	0.25	0.00	54.67	53.54	1.12	6.21	6.27	-0.06
Science	5	0.24	0.25	0.00	51.68	51.72	-0.03	6.10	6.32	-0.21
	8	0.31	0.29	0.02	52.61	53.37	-0.76	5.54	5.82	-0.28

Table 7. Subgroup Representativeness of Scaling Sample Compared to Total Population, Spring 2013 *(continued)*

		ELL			IEP			Section 504		
Content	Grade	Sample	State	Diff.	Sample	State	Diff.	Sample	State	Diff.
Mathematics	3	8.96	10.48	-1.53	11.20	11.71	-0.52	0.35	0.63	-0.28
	4	6.58	6.93	-0.35	10.76	11.63	-0.88	0.54	0.79	-0.25
	5	5.32	5.80	-0.48	10.67	11.07	-0.40	0.53	0.84	-0.31
	6	3.11	4.39	-1.27	7.31	9.89	-2.58	0.89	0.92	-0.03
	7	3.56	4.38	-0.82	7.17	9.16	-1.99	0.92	0.93	-0.01
	8	3.43	4.31	-0.89	5.98	7.87	-1.89	0.88	0.92	-0.04
Reading	3	8.55	10.21	-1.66	10.30	10.70	-0.41	0.38	0.64	-0.26
	4	5.97	6.60	-0.62	9.53	10.50	-0.97	0.60	0.81	-0.21
	5	4.84	5.46	-0.62	9.22	10.19	-0.97	0.52	0.85	-0.32
	6	3.04	4.06	-1.02	7.22	9.52	-2.30	0.91	0.93	-0.02
	7	3.43	4.13	-0.70	7.31	9.14	-1.83	0.93	0.95	-0.02
	8	3.22	4.03	-0.81	6.46	8.17	-1.71	0.88	0.92	-0.04
Science	5	5.55	5.79	-0.24	11.91	12.14	-0.23	0.76	0.83	-0.07
	8	4.35	4.59	-0.24	10.34	10.43	-0.09	0.91	0.90	0.01

Table 7. Subgroup Representativeness of Scaling Sample Compared to Total Population, Spring 2013 *(continued)*

		SES - Low			SES - High			Accommodated		
Content	Grade	Sample	State	Diff.	Sample	State	Diff.	Sample	State	Diff.
Mathematics	3	36.95	38.92	-1.96	63.05	61.08	1.96	8.86	10.15	-1.29
	4	36.51	39.60	-3.09	63.49	60.40	3.09	8.47	9.51	-1.04
	5	34.59	40.18	-5.59	65.41	59.82	5.59	7.67	9.16	-1.48
	6	44.09	42.66	1.44	55.91	57.34	-1.44	3.60	6.71	-3.11
	7	45.09	44.16	0.93	54.91	55.84	-0.93	3.85	6.13	-2.28
	8	47.60	46.65	0.95	52.40	53.35	-0.95	2.77	5.00	-2.23
Reading	3	36.76	39.04	-2.29	63.24	60.96	2.29	7.72	7.92	-0.21
	4	37.08	39.84	-2.77	62.92	60.16	2.77	6.97	7.56	-0.59
	5	35.46	40.43	-4.96	64.54	59.57	4.96	6.24	7.65	-1.41
	6	43.87	42.59	1.27	56.13	57.41	-1.27	3.04	5.57	-2.53
	7	44.93	44.02	0.91	55.07	55.98	-0.91	3.21	5.14	-1.93
	8	47.41	46.43	0.97	52.59	53.57	-0.97	2.27	4.13	-1.86
Science	5	39.52	39.95	-0.43	60.48	60.05	0.43	9.31	9.92	-0.61
	8	44.65	45.69	-1.04	55.35	54.31	1.04	6.22	6.35	-0.13

Table 8. Summary of P-Values and Item-Test Correlations Statistics for Operational Test Forms, Spring 2011 to Spring 2013

		Operational Mean *P*-Values*				Operational Mean Item-Test Correlations*			
Content	Grade	2011	2012	2013	Diff. S13-S12	2011	2012	2013	Diff. S13-S12
Mathematics	3	0.73	0.76	0.76	0.00	0.37	0.41	0.44	0.03
	4	0.73	0.76	0.76	0.00	0.36	0.41	0.42	0.01
	5	0.67	0.73	0.73	0.00	0.36	0.39	0.40	0.01
	6	0.59	0.66	0.66	0.00	0.36	0.41	0.41	0.00
	7	0.57	0.63	0.62	-0.01	0.35	0.40	0.39	-0.01
	8	0.59	0.64	0.65	0.01	0.36	0.41	0.42	0.01
Reading	3	0.66	0.72	0.72	0.00	0.37	0.41	0.42	0.01
	4	0.67	0.74	0.72	-0.02	0.33	0.39	0.40	0.01
	5	0.71	0.76	0.77	0.01	0.35	0.40	0.42	0.02
	6	0.67	0.70	0.73	0.03	0.35	0.41	0.40	-0.01
	7	0.69	0.77	0.77	0.00	0.36	0.38	0.40	0.02
	8	0.72	0.77	0.76	-0.01	0.32	0.36	0.39	0.03
Science	5	.	.	0.68	.	.	.	0.37	.
	8	.	.	0.57	.	.	.	0.36	.

Note: *Census Data; Suppressed items are not included in the data; Science cannot be compared across years because Science was rescaled in Spring 2013.

Table 9. Summary of Range of P-Values and Item-Test Correlations Statistics for Operational and Field Test, Spring 2013

		Mean *P*-Values*						Mean Item-Test Correlations*					
		Operational Items			Field-Test Items			Operational Items			Field-Test Items		
Content	Grade	Low	Mean	High	Low	Mean	High	Low	Mean	High	Low	Mean	High
Mathematics	3	0.38	0.76	0.93	0.09	0.62	0.96	0.25	0.44	0.60	0.03	0.31	0.57
	4	0.51	0.76	0.93	0.22	0.59	0.96	0.26	0.42	0.57	0.06	0.32	0.51
	5	0.41	0.73	0.91	0.17	0.55	0.89	0.23	0.40	0.57	0.02	0.32	0.50
	6	0.36	0.66	0.89	0.09	0.51	0.91	0.20	0.41	0.58	0.03	0.30	0.48
	7	0.24	0.62	0.88	0.08	0.46	0.85	0.22	0.39	0.55	0.00	0.28	0.50
	8	0.42	0.65	0.88	0.19	0.45	0.76	0.28	0.42	0.55	0.02	0.29	0.57
Reading	3	0.41	0.72	0.92	0.13	0.56	0.91	0.18	0.42	0.60	0.10	0.31	0.49
	4	0.43	0.72	0.92	0.13	0.58	0.94	0.24	0.40	0.54	0.09	0.31	0.49
	5	0.54	0.77	0.96	0.18	0.62	0.92	0.25	0.42	0.55	0.09	0.33	0.48
	6	0.43	0.73	0.96	0.10	0.61	0.96	0.27	0.40	0.52	0.08	0.31	0.49
	7	0.51	0.77	0.97	0.10	0.57	0.95	0.17	0.40	0.55	0.01	0.28	0.47
	8	0.47	0.76	0.97	0.22	0.63	0.94	0.24	0.39	0.50	0.01	0.27	0.44
Science	5	0.49	0.68	0.97	0.20	0.55	0.95	0.11	0.37	0.50	0.03	0.26	0.44
	8	0.23	0.57	0.87	0.18	0.47	0.94	0.20	0.36	0.51	0.01	0.22	0.45
Writing	5	0.51	0.55	0.59	0.39	0.53	0.62	.	.	.	.	.	.
	8	0.55	0.56	0.56	0.51	0.57	0.60	.	.	.	.	.	.

Note: *Census Data; Suppressed items are not included in data.

Table 10. Summary of Range of Omission Rates for Operational and Field Test by Item Type, Spring 2013

Content	Grade	Item Type	Omission Rates*					
			Operational Items			Field-Test Items		
			Low	Mean	High	Low	Mean	High
Mathematics	3	MC	0.06%	0.34%	1.36%	0.01%	0.43%	2.03%
	4	MC	0.03%	0.12%	0.32%	0.01%	0.12%	0.27%
	5	MC	0.03%	0.11%	0.28%	0.03%	0.14%	0.36%
	6	MC	0.01%	0.06%	0.14%	0.03%	0.57%	1.03%
	7	MC	0.01%	0.06%	0.25%	0.06%	0.60%	0.97%
	8	MC	0.00%	0.04%	0.07%	0.08%	0.78%	1.49%
Reading	3	CR	.	.	.	3.04%	3.86%	5.05%
		MC	0.05%	0.42%	1.16%	0.03%	0.36%	0.99%
	4	CR	.	.	.	8.25%	9.45%	11.29%
		MC	0.06%	0.17%	0.30%	0.06%	0.28%	0.96%
	5	CR	.	.	.	4.90%	6.24%	8.69%
		MC	0.05%	0.15%	0.25%	0.03%	0.20%	0.88%
	6	CR	.	.	.	0.91%	1.62%	2.69%
		MC	0.02%	0.05%	0.08%	0.07%	0.50%	1.18%
	7	CR	.	.	.	1.25%	2.32%	5.32%
		MC	0.01%	0.05%	0.09%	0.10%	0.38%	0.95%
	8	CR	.	.	.	0.66%	1.76%	3.43%
		MC	0.00%	0.05%	0.08%	0.01%	0.31%	0.90%
Science	5	MC	0.01%	0.10%	0.21%	0.00%	0.10%	0.29%
	8	MC	0.03%	0.09%	0.14%	0.01%	0.10%	0.22%
Writing	5	CR	3.86%	3.86%	3.86%	3.49%	3.54%	3.61%
	8	CR	3.33%	3.33%	3.33%	2.63%	3.64%	4.17%

Note: *Census Data; Suppressed items are not included in the data.

Table 11. Spring 2013 Summary of Operational and Field Test Items Flagged for Mantel-Haenszel Differential Item Functioning, by Item Type: Gender Male/Female

Content	Grade	Item Type	Operational Items		Field Test Items		Total DIF Flags B+C
			B	C	B	C	
Mathematics	3	MC	4	.	.	.	4
	4	MC	1	.	2	.	3
	5	MC	1	1	4	.	6
	6	MC	.	1	2	.	4
	7	MC	2	.	3	.	5
	8	MC	5	.	2	.	7
Reading	3	CR	.	.	.	.	.
		MC	.	.	.	.	.
	4	CR	.	.	.	.	.
		MC	.	.	.	.	.
	5	CR	.	.	1	.	1
		MC	.	.	.	2	2
	6	CR	.	.	1	.	1
		MC	1	1	.	.	2
	7	CR	.	.	2	.	2
		MC	3	1	2	.	6
	8	CR	.	.	3	.	3
		MC	2	1	1	.	4
Science	5	MC	3	.	.	.	3
	8	MC	.	.	.	.	.
Total Items Flagged			22	6	23	2	53
Total Items Tested			700		871		1571
Percentage of Items Flagged			3.14%	0.86%	2.64%	0.23%	3.37%

Note: Census Data; Suppressed items are not included in the data.

Table 12. Spring 2013 Summary of Operational and Field Test Items Flagged for Mantel-Haenszel Differential Item Functioning, by Item Type: Ethnicity White/Asian

Content	Grade	Item Type	Operational Items		Field Test Items		Total DIF Flags B+C
			B	C	B	C	
Mathematics	3	MC	3	.	8	.	11
	4	MC	6	4	8	1	19
	5	MC	5	.	7	6	18
	6	MC	4	.	10	6	21
	7	MC	3	2	9	2	16
	8	MC	4	.	9	3	16
Reading	3	CR	.	.	.	.	.
		MC	2	.	4	3	9
	4	CR	.	.	.	.	.
		MC	5	.	5	1	11
	5	CR	.	.	.	.	.
		MC	1	2	7	3	13
	6	CR	.	.	.	1	1
		MC	8	3	2	3	16
	7	CR	.	.	.	1	1
		MC	5	3	6	4	18
	8	CR	.	.	1	2	3
		MC	8	3	3	4	18
Science	5	MC	5	1	7	1	14
	8	MC	3	.	3	2	8
Total Items Flagged			62	19	89	43	213
Total Items Tested			700		871		1571
Percentage of Items Flagged			8.86%	2.71%	10.22%	4.94%	13.56%

Note: Census Data; Suppressed items are not included in the data.

Table 12.1 Spring 2013 Summary of Operational and Field Test Items Flagged for Mantel-Haenszel Differential Item Functioning, by Item Type: Ethnicity White/African American

Content	Grade	Item Type	Operational Items		Field Test Items		Total DIF Flags B+C
			B	C	B	C	
Mathematics	3	MC	.	.	7	1	8
	4	MC	3	1	9	1	14
	5	MC	.	1	7	3	11
	6	MC	2	.	3	1	7
	7	MC	.	.	3	.	3
	8	MC	2	.	3	.	5
Reading	3	CR	.	.	.	.	.
		MC	1	.	4	.	5
	4	CR	.	.	.	.	.
		MC	.	1	6	.	7
	5	CR	.	.	.	.	.
		MC	.	.	3	2	5
	6	CR	.	.	.	.	.
		MC	1	.	2	.	3
	7	CR	.	.	.	.	.
		MC	1	.	6	1	8
	8	CR	.	.	.	.	.
		MC	2	1	6	1	10
Science	5	MC	.	.	5	.	5
	8	MC	3	.	4	.	7
Total Items Flagged			15	5	68	10	98
Total Items Tested			700		871		1571
Percentage of Items Flagged			2.14%	0.71%	7.81%	1.15%	6.24%

Note: Census Data; Suppressed items are not included in the data.

Table 12.2 Spring 2013 Summary of Operational and Field Test Items Flagged for Mantel-Haenszel Differential Item Functioning, by Item Type: Ethnicity White/Hispanic

Content	Grade	Item Type	Operational Items		Field Test Items		Total DIF Flags B+C
			B	C	B	C	
Mathematics	3	MC	2	.	5	.	7
	4	MC	4	.	4	.	8
	5	MC	1	.	4	.	5
	6	MC	.	.	.	.	1
	7	MC	.	.	1	.	1
	8	MC	1	.	1	.	2
Reading	3	CR	.	.	.	.	.
		MC	2	.	5	1	8
	4	CR	.	.	.	.	.
		MC	.	.	1	1	2
	5	CR	.	.	.	.	.
		MC	.	1	1	3	5
	6	CR	.	.	.	.	.
		MC	1	3	1	1	6
	7	CR	.	.	1	.	1
		MC	4	1	2	.	7
	8	CR	.	.	.	.	.
		MC	2	3	4	2	11
Science	5	MC	.	.	2	.	2
	8	MC	.	.	.	.	.
Total Items Flagged			17	9	32	8	66
Total Items Tested			700		871		1571
Percentage of Items Flagged			2.43%	1.29%	3.67%	0.92%	4.20%

Note: Census Data; Suppressed items are not included in the data.

Table 12.3 Spring 2013 Summary of Operational and Field Test Items Flagged for Mantel-Haenszel Differential Item Functioning, by Item Type: Ethnicity White/American Indian

Content	Grade	Item Type	Operational Items B	Operational Items C	Field Test Items B	Field Test Items C	Total DIF Flags B+C
Mathematics	3	MC	.	.	.	.	.
	4	MC	.	.	.	.	.
	5	MC	.	.	.	.	.
	6	MC	.	.	.	.	1
	7	MC	.	.	.	.	.
	8	MC	.	.	.	.	.
Reading	3	CR	.	.	.	.	.
		MC	.	.	1	.	1
	4	CR	.	.	.	.	.
		MC	.	.	.	.	.
	5	CR	.	.	.	.	.
		MC	.	.	2	.	2
	6	CR	.	.	.	.	.
		MC	.	.	.	.	.
	7	CR	.	.	.	.	.
		MC	.	.	.	.	.
	8	CR	.	.	.	.	.
		MC	.	.	.	.	.
Science	5	MC	.	.	.	.	.
	8	MC	.	.	.	.	.
Total Items Flagged			.	1	3	.	4
Total Items Tested			700		871		1571
Percentage of Items Flagged			.	0.14%	0.34%	.	0.25%

Note: Census Data; Suppressed items are not included in the data.

Table 13. Mathematics Grades 3–5 Standards Level Summary Data, Spring 2013

Grade/ Obj.	Standard Reference	No. of Items	Average Difficulty (IRT Loc)	Average IRT Information	Objective % Correct State Mean	Average P-value					
						State	P.L. 1	P.L. 2	P.L. 3	P.L. 4	Pass
3.1	Algebraic Reasoning: Patterns and Relationships	7	675.00	0.05	77.39	0.78	0.40	0.63	0.83	0.96	0.87
3.2	Number Sense & Operation	20	678.70	0.08	75.58	0.76	0.39	0.60	0.81	0.96	0.86
3.3	Geometry	7	714.57	0.05	67.29	0.67	0.38	0.54	0.70	0.89	0.76
3.4	Measurement	9	669.56	0.14	77.57	0.78	0.37	0.61	0.84	0.97	0.88
3.5	Data Analysis	7	661.57	0.08	80.77	0.81	0.38	0.66	0.88	0.98	0.91
4.1	Algebraic Reasoning: Patterns and Relationships	7	650.86	0.05	82.69	0.83	0.48	0.71	0.88	0.97	0.90
4.2	Number Sense & Operation	18	692.56	0.08	74.97	0.75	0.37	0.56	0.80	0.96	0.85
4.3	Geometry	9	670.89	0.04	75.86	0.76	0.46	0.61	0.78	0.95	0.83
4.4	Measurement	9	687.22	0.05	75.11	0.75	0.39	0.60	0.80	0.93	0.84
4.5	Data Analysis	7	698.00	0.06	74.54	0.75	0.37	0.56	0.79	0.96	0.84
5.1	Algebraic Reasoning: Patterns and Relationships	13	701.92	0.06	71.55	0.72	0.38	0.56	0.76	0.91	0.81
5.2	Number Sense & Operation	16	698.63	0.07	74.92	0.75	0.41	0.59	0.79	0.94	0.84
5.3	Geometry	7	724.00	0.04	67.99	0.68	0.36	0.52	0.71	0.89	0.77
5.4	Measurement	7	705.00	0.06	71.04	0.71	0.39	0.55	0.74	0.93	0.80
5.5	Data Analysis	7	661.29	0.05	78.92	0.79	0.46	0.67	0.83	0.95	0.87

Note: Obj. = Objective; P.L. = Performance Level.

Table 13. Mathematics Grades 3–5 Standards Level Summary Data, Spring 2013 *(continued)*

Grade/ Obj.	Standard Reference	No. of Items	Alpha	SEM
3.1	Algebraic Reasoning: Patterns and Relationships	7	0.60	0.98
3.2	Number Sense & Operation	20	0.82	1.64
3.3	Geometry	7	0.51	1.07
3.4	Measurement	9	0.73	1.07
3.5	Data Analysis	7	0.69	0.90
4.1	Algebraic Reasoning: Patterns and Relationships	7	0.57	0.91
4.2	Number Sense & Operation	18	0.80	1.62
4.3	Geometry	9	0.63	1.11
4.4	Measurement	9	0.61	1.16
4.5	Data Analysis	7	0.63	1.02
5.1	Algebraic Reasoning: Patterns and Relationships	13	0.70	1.44
5.2	Number Sense & Operation	16	0.75	1.54
5.3	Geometry	7	0.55	1.10
5.4	Measurement	7	0.55	1.08
5.5	Data Analysis	7	0.53	0.98

Table 13.1 Mathematics Grades 6–8 Standards Level Summary Data, Spring 2013

Grade/ Obj.	Standard Reference	No. of Items	Average Difficulty (IRT Loc)	Average IRT Information	Objective % Correct State Mean	Average P-value					
						State	P.L. 1	P.L. 2	P.L. 3	P.L. 4	Pass
6.1	Algebraic Reasoning: Patterns and Relationships	13	718.46	0.07	68.57	0.69	0.36	0.53	0.72	0.91	0.78
6.2	Number Sense & Operation	15	731.80	0.09	63.84	0.64	0.32	0.45	0.67	0.89	0.74
6.3	Geometry	7	698.86	0.05	69.04	0.69	0.45	0.57	0.71	0.89	0.76
6.4	Measurement	7	736.14	0.08	59.69	0.60	0.26	0.40	0.63	0.87	0.71
6.5	Data Analysis	7	692.00	0.07	68.25	0.68	0.37	0.52	0.72	0.91	0.78
7.1	Algebraic Reasoning: Patterns and Relationships	15	745.67	0.07	62.89	0.63	0.38	0.51	0.67	0.88	0.72
7.2	Number Sense & Operation	11	707.55	0.07	68.03	0.68	0.42	0.56	0.73	0.93	0.78
7.3	Geometry	8	736.63	0.05	60.43	0.60	0.34	0.48	0.65	0.86	0.70
7.4	Measurement	9	754.44	0.13	52.99	0.53	0.31	0.39	0.54	0.84	0.62
7.5	Data Analysis	7	725.14	0.07	67.03	0.67	0.39	0.56	0.74	0.89	0.77
8.1	Algebraic Reasoning: Patterns and Relationships	16	704.94	0.06	67.97	0.68	0.37	0.53	0.71	0.88	0.78
8.2	Number Sense & Operation	11	745.00	0.05	59.82	0.60	0.29	0.44	0.61	0.83	0.71
8.3	Geometry	9	740.67	0.06	61.93	0.62	0.32	0.46	0.63	0.86	0.73
8.4	Measurement	7	723.57	0.11	62.45	0.62	0.22	0.39	0.67	0.91	0.77
8.5	Data Analysis	7	672.29	0.06	76.68	0.77	0.41	0.66	0.82	0.93	0.87

Note: Obj. = Objective; P.L. = Performance Level.

Table 13.1 Mathematics Grades 6–8 Standards Level Summary Data, Spring 2013 *(continued)*

Grade/ Obj.	Standard Reference	No. of Items	Alpha	SEM
6.1	Algebraic Reasoning: Patterns and Relatic	13	0.72	1.49
6.2	Number Sense & Operation	15	0.76	1.62
6.3	Geometry	8	0.51	1.07
6.4	Measurement	7	0.65	1.13
6.5	Data Analysis	7	0.58	1.10
7.1	Algebraic Reasoning: Patterns and Relatic	15	0.70	1.67
7.2	Number Sense & Operation	11	0.66	1.38
7.3	Geometry	8	0.60	1.20
7.4	Measurement	9	0.62	1.25
7.5	Data Analysis	7	0.55	1.12
8.1	Algebraic Reasoning: Patterns and Relatic	16	0.73	1.68
8.2	Number Sense & Operation	11	0.67	1.48
8.3	Geometry	9	0.67	1.27
8.4	Measurement	7	0.72	1.09
8.5	Data Analysis	7	0.62	0.99

Table 14. Reading Grades 3–5 Standards Level Summary Data, Spring 2013

Grade/ Obj.	Standard Reference	No. of Items	Average Difficulty (IRT Loc)	Average IRT Information	Objective % Correct State Mean	Average P-value					
						State	P.L. 1	P.L. 2	P.L. 3	P.L. 4	Pass
3.2	Vocabulary	12	709.67	0.05	73.14	0.73	0.39	0.59	0.81	0.98	0.82
3.4	Comprehension/ Critical Literacy	26	708.04	0.07	71.66	0.72	0.36	0.56	0.80	0.97	0.81
3.5	Literature	6	710.67	0.06	71.37	0.72	0.36	0.56	0.80	0.97	0.81
3.6	Research and Information	6	723.67	0.05	66.60	0.67	0.32	0.51	0.75	0.96	0.77
4.1	Vocabulary	12	689.33	0.06	72.51	0.73	0.41	0.61	0.81	0.96	0.82
4.3	Comprehension/ Critical Literacy	23	686.91	0.06	73.30	0.73	0.41	0.63	0.81	0.95	0.83
4.4	Literature	9	700.33	0.07	71.05	0.71	0.36	0.58	0.80	0.96	0.82
4.5	Research and Information	6	704.33	0.06	69.06	0.69	0.42	0.57	0.76	0.95	0.78
5.1	Vocabulary	11	666.82	0.08	80.22	0.80	0.49	0.70	0.87	0.98	0.88
5.3	Comprehension/ Critical Literacy	22	675.73	0.10	78.92	0.79	0.42	0.67	0.87	0.98	0.89
5.4	Literature	11	706.82	0.06	71.49	0.71	0.37	0.57	0.79	0.96	0.81
5.5	Research and Information	6	676.33	0.04	76.43	0.77	0.44	0.66	0.84	0.97	0.85

Note: Obj. = Objective; P.L. = Performance Level.

Table 14. Reading Grades 3–5 Standards Level Summary Data, Spring 2013 *(continued)*

Grade/ Obj.	Standard Reference	No. of Items	Alpha	SEM
3.2	Vocabulary	12	0.70	1.38
3.4	Comprehension/Critical Literacy	26	0.84	1.98
3.5	Literature	6	0.56	0.97
3.6	Research and Information	6	0.57	1.01
4.1	Vocabulary	12	0.66	1.39
4.3	Comprehension/Critical Literacy	23	0.78	1.91
4.4	Literature	9	0.65	1.20
4.5	Research and Information	6	0.42	1.01
5.1	Vocabulary	11	0.63	1.17
5.3	Comprehension/Critical Literacy	22	0.83	1.67
5.4	Literature	11	0.65	1.37
5.5	Research and Information	6	0.46	0.95

Table 14.1 Reading Grades 6–8 Standards Level Summary Data, Spring 2013

Grade/ Obj.	Standard Reference	No. of Items	Average Difficulty (IRT Loc)	Average IRT Information	Objective % Correct State Mean	Average P-value					
						State	P.L. 1	P.L. 2	P.L. 3	P.L. 4	Pass
6.1	Vocabulary	8	664.50	0.12	76.99	0.77	0.43	0.66	0.85	0.97	0.87
6.3	Comprehension/ Critical Literacy	22	699.64	0.06	70.15	0.70	0.38	0.58	0.78	0.93	0.80
6.4	Literature	12	670.58	0.06	76.75	0.77	0.46	0.68	0.84	0.95	0.86
6.5	Research and Information	8	686.88	0.06	71.40	0.71	0.41	0.61	0.78	0.93	0.80
7.1	Vocabulary	10	695.60	0.09	74.56	0.75	0.52	0.65	0.79	0.93	0.81
7.3	Comprehension/ Critical Literacy	20	675.45	0.11	78.59	0.79	0.49	0.68	0.85	0.97	0.87
7.4	Literature	12	680.67	0.15	78.38	0.78	0.49	0.70	0.85	0.96	0.86
7.5	Research and Information	8	700.63	0.10	72.84	0.73	0.40	0.60	0.80	0.95	0.82
8.1	Vocabulary	7	662.00	0.07	81.31	0.81	0.51	0.70	0.85	0.96	0.87
8.3	Comprehension/ Critical Literacy	21	684.43	0.05	76.03	0.76	0.41	0.60	0.81	0.95	0.84
8.4	Literature	14	671.29	0.05	80.22	0.80	0.46	0.67	0.85	0.96	0.87
8.5	Research and Information	8	734.63	0.03	64.25	0.64	0.34	0.49	0.67	0.87	0.71

Note: Obj. = Objective; P.L. = Performance Level.

Table 14.1 Reading Grades 6 – 8 Standards Level Summary Data, Spring 2013 *(continued)*

Grade/ Obj.	Standard Reference	No. of Items	Alpha	SEM
6.1	Vocabulary	8	0.64	1.00
6.3	Comprehension/Critical Literacy	22	0.77	1.93
6.4	Literature	12	0.65	1.32
6.5	Research and Information	8	0.53	1.13
7.1	Vocabulary	10	0.48	1.23
7.3	Comprehension/Critical Literacy	20	0.78	1.62
7.4	Literature	12	0.66	1.27
7.5	Research and Information	8	0.56	1.15
8.1	Vocabulary	7	0.49	0.93
8.3	Comprehension/Critical Literacy	21	0.78	1.75
8.4	Literature	14	0.69	1.36
8.5	Research and Information	8	0.45	1.25

Table 15. Science Grades 5 & 8 Standards Level Summary Data, Spring 2013

Grade/ Obj.	Standard Reference	No. of Items	Average Difficulty (IRT Loc)	Average IRT Information	Objective % Correct State Mean	Average P-value					
						State	P.L. 1	P.L. 2	P.L. 3	P.L. 4	Pass
5.P1	Observe and Measure	11	724.27	0.09	68.61	0.69	0.42	0.62	0.79	0.92	0.83
5.P2	Classify	9	735.67	0.07	65.75	0.66	0.40	0.60	0.75	0.89	0.80
5.P3	Experiment	13	720.00	0.09	70.02	0.70	0.47	0.65	0.78	0.90	0.82
5.P4	Interpret and Communicate	12	713.00	0.08	68.71	0.69	0.44	0.63	0.78	0.91	0.82
5.S1	Properties of Matter and Energy	17	724.94	0.09	68.44	0.69	0.45	0.63	0.77	0.90	0.81
5.S2	Organisms and Environments	12	719.50	0.06	5.97	0.68	0.42	0.64	0.78	0.90	0.82
5.S3	Structures of the Earth and the Solar System	12	740.92	0.09	5.55	0.63	0.36	0.55	0.74	0.90	0.79
8.P1	Observe and Measure	8	738.38	0.12	56.65	0.56	0.30	0.48	0.66	0.84	0.71
8.P2	Classify	9	718.78	0.08	59.02	0.59	0.35	0.53	0.67	0.82	0.71
8.P3	Experiment	16	732.75	0.10	55.74	0.56	0.31	0.47	0.65	0.83	0.70
8.P4	Interpret and Communicate	12	730.42	0.11	58.67	0.59	0.34	0.51	0.68	0.83	0.72
8.S1	Properties and Chemical Changes in Matter	8	735.63	0.15	53.34	0.53	0.27	0.44	0.62	0.83	0.68
8.S2	Motion and Forces	8	730.75	0.11	4.69	0.55	0.29	0.45	0.66	0.85	0.71
8.S3	Diversity and Adaptations of Organisms	7	699.14	0.10	5.15	0.66	0.40	0.59	0.76	0.90	0.79
8.S4	Structures/Forces of the Earth/Solar System	11	756.64	0.07	4.77	0.51	0.30	0.44	0.59	0.75	0.63
8.S5	Earth's History	8	730.13	0.10	5.10	0.60	0.35	0.52	0.69	0.85	0.74

Note: Obj. = Objective; P.L. = Performance Level.

Table 15. Science Grades 5 & 8 Standards Level Summary Data, Spring 2013 *(continued)*

Grade/ Obj.	Standard Reference	No. of Items	Alpha	SEM
5.P1	Observe and Measure	11	0.62	1.40
5.P2	Classify	9	0.55	1.31
5.P3	Experiment	13	0.61	1.48
5.P4	Interpret and Communicate	12	0.60	1.49
5.S1	Properties of Matter and Energy	17	0.67	1.77
5.S2	Organisms and Environments	12	0.62	1.48
5.S3	Structures of the Earth and the Solar System	12	0.67	1.51
8.P1	Observe and Measure	8	0.53	1.27
8.P2	Classify	9	0.49	1.32
8.P3	Experiment	16	0.68	1.79
8.P4	Interpret and Communicate	12	0.59	1.52
8.S1	Properties and Chemical Changes in Matter	8	0.55	1.25
8.S2	Motion and Forces	8	0.56	1.27
8.S3	Diversity and Adaptations of Organisms	7	0.50	1.11
8.S4	Structures/Forces of the Earth/Solar System	11	0.48	1.52
8.S5	Earth's History	8	0.51	1.27

Table 16. Spring 2012 Scale Score Statistics

Content	Grade	N Count	Mean	SD	LOSS	N Min.	Scale Score Percentile 25th	50th	75th	N Max.	HOSS	Alpha	SEM
Mathematics	3	45419	741.04	88.37	400	89	689	745	789	597	990	0.90	2.61
	4	44146	745.79	87.78	400	42	694	747	804	695	990	0.90	2.66
	5	43641	742.14	85.78	400	75	691	745	794	327	990	0.89	2.77
	6	43415	734.25	79.51	400	105	687	739	781	284	990	0.90	2.95
	7	42540	735.15	80.15	400	239	691	740	788	182	990	0.89	3.01
	8	42076	726.02	83.07	400	476	681	731	778	234	990	0.90	3.03
Reading	3	44742	742.68	82.26	400	64	691	743	790	122	990	0.90	2.82
	4	43366	724.38	73.38	400	32	679	721	767	208	990	0.88	2.75
	5	43073	734.25	79.60	400	78	684	738	781	349	990	0.89	2.69
	6	43195	730.75	78.68	400	57	679	730	783	102	990	0.89	2.85
	7	42500	738.85	70.35	400	46	696	738	785	365	990	0.87	2.70
	8	42135	757.77	81.81	400	49	712	760	799	499	990	0.86	2.73
Science	5	44116	782.53	71.36	400	47	741	786	823	193	990	0.87	2.71
	8	43131	768.51	63.84	400	38	735	775	812	69	990	0.86	2.85

Note: Statistics re-calculated by CTB based on 2012 GRT.

Table 17. Spring 2013 Scale Score Statistics

Content	Grade	N Count	Mean	SD	LOSS	N Min.	Scale Score Percentile 25th	50th	75th	N Max.	HOSS	Alpha	SEM
Mathematics	3	46316	739.00	88.33	400	151	691	740	792	864	990	0.91	2.62
	4	45383	745.43	90.14	400	133	693	744	798	858	990	0.90	2.69
	5	44295	740.71	86.98	400	137	690	741	788	430	990	0.89	2.81
	6	43221	737.10	78.84	400	129	687	737	787	241	990	0.90	2.94
	7	43146	732.30	80.70	400	239	688	730	782	126	990	0.89	3.02
	8	41377	732.09	83.25	400	143	685	735	786	309	990	0.90	3.00
Reading	3	45683	741.22	86.35	400	221	690	748	799	167	990	0.90	2.81
	4	44704	729.59	77.54	400	130	683	733	771	106	990	0.89	2.85
	5	43798	735.55	84.47	400	187	683	739	788	468	990	0.90	2.64
	6	42971	731.18	77.53	400	49	685	730	782	115	990	0.89	2.78
	7	43368	729.88	67.56	400	91	688	728	766	239	990	0.88	2.66
	8	42341	750.16	82.06	400	74	701	753	803	178	990	0.88	2.72
Science	5	44805	695.10	72.00	400	221	656	700	742	96	990	0.86	2.84
	8	44209	694.21	57.11	400	252	665	700	733	5	990	0.85	2.98

Note: Census Data; Suppressed items are not included in data.

Table 18. Spring 2013, Mathematics Grades 3 & 4 State and Subgroup Scale Score Descriptive Data

Grade	No. of Items	Subgroup	Sample Size	Scale Score Mean	Scale Score SD	Min Scale Score Obtained	Max Scale Score Obtained	Coefficient Alpha	SEM
3	50	Whole State	46316	739.00	88.33	400	990	0.91	2.62
	50	Female	22881	735.33	87.00	400	990	0.91	2.64
	50	Male	23395	742.62	89.46	400	990	0.91	2.60
	50	Asian	950	778.70	92.68	479	990	0.91	2.35
	50	African American	4126	690.63	93.23	400	990	0.92	2.88
	50	Hispanic	7322	711.96	85.05	400	990	0.91	2.79
	50	Indian American	6934	738.06	83.19	400	990	0.90	2.63
	50	White	23794	754.52	84.70	400	990	0.90	2.52
	50	Multiracial	3190	737.98	85.75	400	990	0.91	2.63
	50	ELL	4856	695.73	84.62	400	990	0.91	2.88
	50	IEP	5425	700.13	93.72	400	990	0.92	2.85
	50	Section 504	291	721.85	92.04	400	990	0.92	2.72
	50	Low SES	28291	719.48	85.55	400	990	0.91	2.75
	50	Accommodated	4702	676.81	85.45	400	990	0.90	2.98
4	50	Whole State	45383	745.43	90.14	400	990	0.90	2.69
	50	Female	22577	743.67	88.37	400	990	0.90	2.70
	50	Male	22768	747.24	91.79	400	990	0.91	2.67
	50	Asian	850	792.93	95.07	400	990	0.90	2.34
	50	African American	4056	700.87	92.34	400	990	0.91	2.94
	50	Hispanic	6882	720.35	89.24	400	990	0.90	2.84
	50	Indian American	6901	740.12	84.73	400	990	0.89	2.73
	50	White	23658	760.45	87.01	400	990	0.89	2.58
	50	Multiracial	3036	743.62	87.52	400	990	0.90	2.70
	50	ELL	3145	690.44	90.27	400	990	0.90	2.99
	50	IEP	5280	700.62	92.27	400	990	0.91	2.95
	50	Section 504	358	728.45	84.18	433	990	0.89	2.83
	50	Low SES	27413	725.22	86.26	400	990	0.90	2.82
	50	Accommodated	4314	678.26	85.73	400	990	0.89	3.07

Note: SEM = Standard Error of Measurement; ELL = English Language Learner; IEP = Individualized Education Program.

Table 18.1 Spring 2013, Mathematics Grades 5 & 6 State and Subgroup Scale Score Descriptive Data

Grade	No. of Items	Subgroup	Sample Size	Scale Score Mean	Scale Score SD	Min Scale Score Obtained	Max Scale Score Obtained	Coefficient Alpha	SEM
5	50	Whole State	44295	740.71	86.98	400	990	0.89	2.81
	50	Female	22105	740.57	84.76	400	990	0.89	2.81
	50	Male	22162	740.89	89.13	400	990	0.90	2.80
	50	Asian	857	793.53	91.96	400	990	0.90	2.43
	50	African American	4030	700.98	88.34	400	990	0.89	3.01
	50	Hispanic	6472	721.58	86.66	400	990	0.89	2.92
	50	Indian American	7075	729.19	82.77	400	990	0.88	2.88
	50	White	22950	754.83	84.01	400	990	0.88	2.72
	50	Multiracial	2911	739.28	84.39	400	990	0.89	2.82
	50	ELL	2570	684.53	88.79	400	990	0.89	3.09
	50	IEP	4904	690.29	89.55	400	990	0.89	3.06
	50	Section 504	372	723.38	80.87	505	990	0.88	2.93
	50	Low SES	26497	721.57	83.23	400	990	0.89	2.92
	50	Accommodated	4057	673.80	84.86	400	990	0.87	3.14
6	49	Whole State	43222	737.09	78.86	400	990	0.90	2.94
	49	Female	21425	734.23	76.48	400	990	0.90	2.95
	49	Male	21791	739.91	81.02	400	990	0.90	2.92
	49	Asian	875	785.87	85.70	400	990	0.91	2.62
	49	African American	3754	697.31	78.99	400	990	0.89	3.09
	49	Hispanic	5999	715.37	78.64	400	990	0.89	3.04
	49	Indian American	6964	729.55	74.65	400	990	0.89	2.99
	49	White	22691	750.32	75.55	400	990	0.89	2.88
	49	Multiracial	2939	733.37	78.78	400	990	0.90	2.95
	49	ELL	1896	676.20	82.90	400	990	0.88	3.14
	49	IEP	4275	679.37	83.75	400	990	0.89	3.13
	49	Section 504	397	726.55	72.03	400	939	0.89	3.01
	49	Low SES	24785	717.79	75.41	400	990	0.89	3.04
	49	Accommodated	2900	666.63	79.15	400	990	0.87	3.16

Note: SEM = Standard Error of Measurement; ELL = English Language Learner; IEP = Individualized Education Program.

Table 18.2 Spring 2013, Mathematics Grades 7 & 8 State and Subgroup Scale Score Descriptive Data

Grade	No. of Items	Subgroup	Sample Size	Scale Score Mean	Scale Score SD	Min Scale Score Obtained	Max Scale Score Obtained	Coefficient Alpha	SEM
7	50	Whole State	43146	732.30	80.70	400	990	0.89	3.02
	50	Female	21497	730.67	76.80	400	990	0.88	3.03
	50	Male	21641	733.92	84.37	400	990	0.90	3.00
	50	Asian	813	787.77	84.70	400	990	0.90	2.69
	50	African American	3868	691.93	80.87	400	968	0.87	3.14
	50	Hispanic	5695	709.57	77.29	400	990	0.87	3.11
	50	Indian American	7099	726.05	75.10	400	990	0.87	3.06
	50	White	22762	745.90	78.50	400	990	0.89	2.96
	50	Multiracial	2909	723.75	81.03	400	990	0.88	3.05
	50	ELL	1889	673.07	83.51	400	990	0.85	3.19
	50	IEP	3953	669.02	87.38	400	990	0.86	3.19
	50	Section 504	401	713.35	85.27	400	968	0.89	3.08
	50	Low SES	24091	712.31	77.49	400	990	0.87	3.10
	50	Accommodated	2644	659.24	85.68	400	916	0.85	3.21
8	50	Whole State	41377	732.09	83.25	400	990	0.90	3.00
	50	Female	20755	732.31	81.87	400	990	0.90	3.00
	50	Male	20615	731.87	84.61	400	990	0.91	2.99
	50	Asian	820	793.94	93.46	400	990	0.91	2.59
	50	African American	3675	697.81	84.60	400	990	0.90	3.13
	50	Hispanic	5245	709.20	82.49	400	990	0.90	3.10
	50	Indian American	6747	724.12	79.06	400	990	0.90	3.05
	50	White	22184	743.91	80.14	400	990	0.90	2.94
	50	Multiracial	2706	727.21	84.23	400	990	0.90	3.02
	50	ELL	1785	666.86	80.55	400	990	0.88	3.22
	50	IEP	3258	666.94	82.85	400	990	0.88	3.22
	50	Section 504	382	716.76	89.62	400	990	0.91	3.07
	50	Low SES	22074	711.44	80.31	400	990	0.89	3.10
	50	Accommodated	2070	655.55	78.46	400	990	0.86	3.25

Note: SEM = Standard Error of Measurement; ELL = English Language Learner; IEP = Individualized Education Program.

Table 19. Spring 2013, Reading Grades 3 & 4 State and Subgroup Scale Score Descriptive Data

Grade	No. of Items	Subgroup	Sample Size	Scale Score Mean	Scale Score SD	Min Scale Score Obtained	Max Scale Score Obtained	Coefficient Alpha	SEM
3	50	Whole State	45683	741.22	86.35	400	990	0.90	2.81
	50	Female	22691	747.94	83.61	400	990	0.90	2.76
	50	Male	22967	734.66	88.45	400	990	0.91	2.85
	50	Asian	923	768.66	92.10	400	990	0.91	2.63
	50	African American	4061	705.13	89.78	400	990	0.91	2.99
	50	Hispanic	7177	710.84	85.97	400	990	0.90	2.98
	50	Indian American	6846	736.95	82.58	400	990	0.90	2.84
	50	White	23524	756.68	82.16	400	990	0.89	2.72
	50	Multiracial	3152	742.78	85.75	400	990	0.90	2.80
	50	ELL	4665	687.62	83.57	400	990	0.89	3.09
	50	IEP	4889	687.55	101.37	400	990	0.92	3.03
	50	Section 504	294	721.37	84.21	400	961	0.90	2.93
	50	Low SES	27847	721.91	84.78	400	990	0.90	2.92
	50	Accommodated	3620	654.85	93.22	400	990	0.90	3.16
4	50	Whole State	44704	729.59	77.54	400	990	0.89	2.85
	50	Female	22405	735.77	75.52	400	990	0.88	2.81
	50	Male	22269	723.42	79.01	400	990	0.89	2.88
	50	Asian	832	754.15	82.30	400	990	0.89	2.66
	50	African American	3995	695.10	78.64	400	990	0.89	3.03
	50	Hispanic	6744	704.38	77.12	400	990	0.89	2.99
	50	Indian American	6786	724.79	75.10	400	990	0.88	2.88
	50	White	23377	743.20	74.64	400	990	0.88	2.76
	50	Multiracial	2970	730.24	73.00	400	990	0.88	2.85
	50	ELL	2950	668.09	77.07	400	890	0.88	3.15
	50	IEP	4695	675.59	89.64	400	990	0.91	3.09
	50	Section 504	361	718.86	71.78	513	951	0.87	2.95
	50	Low SES	26892	711.31	75.03	400	990	0.88	2.96
	50	Accommodated	3381	649.10	81.98	400	951	0.89	3.20

Note: SEM = Standard Error of Measurement; ELL = English Language Learner; IEP = Individualized Education Program.

Table 19.1 Spring 2013, Reading Grades 5 & 6 State and Subgroup Scale Score Descriptive Data

Grade	No. of Items	Subgroup	Sample Size	Scale Score		Min Scale Score Obtained	Max Scale Score Obtained	Coefficient Alpha	SEM
				Mean	SD				
5	50	Whole State	43798	735.55	84.47	400	990	0.90	2.64
	50	Female	21983	742.93	82.41	400	990	0.89	2.59
	50	Male	21790	728.16	85.85	400	990	0.90	2.69
	50	Asian	839	763.51	94.40	487	990	0.91	2.43
	50	African American	3980	698.29	83.91	400	990	0.90	2.90
	50	Hispanic	6345	706.17	82.52	400	990	0.90	2.85
	50	Indian American	6995	727.64	81.70	400	990	0.89	2.70
	50	White	22743	751.63	80.96	400	990	0.88	2.52
	50	Multiracial	2896	735.83	83.34	400	990	0.90	2.64
	50	ELL	2391	655.52	84.15	400	990	0.89	3.11
	50	IEP	4463	670.80	95.97	400	990	0.92	3.00
	50	Section 504	371	721.48	82.14	400	990	0.90	2.76
	50	Low SES	26092	715.32	82.14	400	990	0.90	2.79
	50	Accommodated	3350	645.29	89.46	400	990	0.90	3.13
6	50	Whole State	42971	731.18	77.53	400	990	0.89	2.78
	50	Female	21436	736.54	75.81	400	990	0.88	2.75
	50	Male	21530	725.86	78.84	400	990	0.89	2.81
	50	Asian	844	759.34	84.15	400	990	0.89	2.58
	50	African American	3754	697.73	74.66	400	990	0.89	2.98
	50	Hispanic	5881	704.73	75.47	400	990	0.89	2.93
	50	Indian American	6989	724.31	74.02	400	990	0.88	2.83
	50	White	22554	745.01	75.34	400	990	0.87	2.70
	50	Multiracial	2949	728.96	78.10	400	990	0.88	2.80
	50	ELL	1744	652.86	75.21	400	966	0.88	3.14
	50	IEP	4091	666.62	83.42	400	990	0.90	3.08
	50	Section 504	398	719.01	77.09	400	899	0.89	2.84
	50	Low SES	24669	712.05	74.40	400	990	0.88	2.90
	50	Accommodated	2392	649.18	78.53	400	990	0.89	3.16

Note: SEM = Standard Error of Measurement; ELL = English Language Learner; IEP = Individualized Education Program.

Table 19.2 Spring 2013, Reading Grades 7 & 8 State and Subgroup Scale Score Descriptive Data

Grade	No. of Items	Subgroup	Sample Size	Scale Score Mean	Scale Score SD	Min Scale Score Obtained	Max Scale Score Obtained	Coefficient Alpha	SEM
7	50	Whole State	43368	729.88	67.56	400	990	0.88	2.66
	50	Female	21678	736.47	66.40	400	990	0.87	2.60
	50	Male	21687	723.31	68.06	400	990	0.89	2.72
	50	Asian	814	753.21	75.92	400	990	0.89	2.44
	50	African American	3957	697.58	67.04	400	990	0.89	2.90
	50	Hispanic	5668	706.73	66.07	400	990	0.89	2.84
	50	Indian American	7140	728.02	63.74	400	990	0.87	2.68
	50	White	22874	741.55	65.33	400	990	0.87	2.56
	50	Multiracial	2915	725.28	67.59	400	990	0.88	2.70
	50	ELL	1792	663.91	65.72	400	990	0.88	3.11
	50	IEP	3965	675.21	73.89	400	990	0.90	3.03
	50	Section 504	412	717.01	64.65	515	990	0.88	2.79
	50	Low SES	24278	714.68	65.30	400	990	0.88	2.79
	50	Accommodated	2229	660.59	72.11	400	990	0.89	3.10
8	50	Whole State	42341	750.16	82.06	400	990	0.88	2.72
	50	Female	21238	759.25	81.30	400	990	0.87	2.66
	50	Male	21098	741.03	81.79	400	990	0.88	2.77
	50	Asian	810	781.96	90.99	400	990	0.89	2.51
	50	African American	3830	715.05	82.88	400	990	0.89	2.93
	50	Hispanic	5326	723.37	85.36	400	990	0.89	2.88
	50	Indian American	6943	744.76	79.20	400	990	0.87	2.76
	50	White	22671	763.06	78.23	400	990	0.86	2.64
	50	Multiracial	2761	748.81	80.02	400	990	0.87	2.73
	50	ELL	1705	665.33	82.15	400	990	0.88	3.15
	50	IEP	3459	676.62	86.02	400	990	0.90	3.09
	50	Section 504	391	740.33	80.77	400	990	0.87	2.78
	50	Low SES	22681	728.96	80.12	400	990	0.88	2.86
	50	Accommodated	1747	660.33	83.62	400	977	0.89	3.15

Note: SEM = Standard Error of Measurement; ELL = English Language Learner; IEP = Individualized Education Program.

Table 20. Spring 2013, Science Grades 5 & 8 State and Subgroup Scale Score Descriptive Data

Grade	No. of Items	Subgroup	Sample Size	Scale Score Mean	Scale Score SD	Min Scale Score Obtained	Max Scale Score Obtained	Coefficient Alpha	SEM
5	45	Whole State	44805	695.10	72.00	400	990	0.86	2.84
	45	Female	22365	692.06	68.78	400	990	0.85	2.86
	45	Male	22409	698.15	74.99	400	990	0.87	2.81
	45	Asian	859	717.47	76.01	400	990	0.86	2.70
	45	African American	4132	653.82	72.74	400	990	0.85	3.01
	45	Hispanic	6546	671.43	69.80	400	990	0.84	2.96
	45	Indian American	7154	691.19	68.73	400	990	0.84	2.87
	45	White	23172	709.50	68.38	400	990	0.84	2.77
	45	Multiracial	2942	695.23	72.31	400	990	0.86	2.84
	45	ELL	2596	638.30	72.44	400	919	0.82	3.07
	45	IEP	5438	653.01	78.48	400	990	0.86	3.01
	45	Section 504	374	683.93	73.99	400	990	0.85	2.90
	45	Low SES	26904	678.71	70.25	400	990	0.85	2.93
	45	Accommodated	4444	637.41	73.88	400	990	0.83	3.07
8	45	Whole State	44209	694.21	57.11	400	990	0.85	2.98
	45	Female	21997	693.66	53.86	400	990	0.84	2.99
	45	Male	22171	694.79	60.14	400	990	0.86	2.97
	45	Asian	836	721.10	55.71	400	990	0.86	2.82
	45	African American	4124	664.80	61.66	400	849	0.82	3.05
	45	Hispanic	5712	674.71	59.09	400	894	0.83	3.03
	45	Indian American	7239	690.60	53.78	400	990	0.83	3.01
	45	White	23596	704.20	53.48	400	990	0.84	2.95
	45	Multiracial	2702	694.39	57.07	400	894	0.85	2.98
	45	ELL	2028	642.62	62.65	400	824	0.76	3.08
	45	IEP	4613	649.11	67.02	400	990	0.81	3.07
	45	Section 504	397	692.84	56.48	400	894	0.85	2.99
	45	Low SES	24011	679.97	58.17	400	990	0.83	3.03
	45	Accommodated	2807	639.76	65.72	400	824	0.77	3.08

Note: SEM = Standard Error of Measurement; ELL = English Language Learner; IEP = Individualized Education Program.

Table 21. Spring 2013, Mathematics Subgroup Scale Score Mean Differences, t-test: Accommodated/Non Accommodated

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-52.62	5826.14	<.0001	-69.21	85.82
4	-53.98	5305.59	<.0001	-74.23	87.48
5	-52.72	4898.13	<.0001	-73.65	84.34
6	-49.75	3299.11	<.0001	-75.53	76.56
7	-45.49	2936.30	<.0001	-77.83	78.51
8	-45.45	2310.57	<.0001	-80.57	81.37

Note: SEM at or closest above the cut scores.

Table 22. Spring 2013, Reading Subgroup Scale Score Mean Differences, t-test: Accommodated/Non Accommodated

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-58.64	4110.01	<.0001	-93.81	82.55
4	-59.83	3836.03	<.0001	-87.07	74.04
5	-61.25	3800.95	<.0001	-97.73	80.38
6	-52.69	2652.50	<.0001	-86.84	74.93
7	-46.80	2429.73	<.0001	-73.05	65.61
8	-45.94	1885.24	<.0001	-93.70	79.91

Note: SEM at or closest above the cut scores.

Table 23. Spring 2013, Science Subgroup Scale Score Mean Differences, t-test: Accommodated/Non Accommodated

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
5	-55.20	5329.43	<.0001	-64.04	69.41
8	-45.81	3073.80	<.0001	-58.14	55.33

Note: SEM at or closest above the cut scores.

Table 24. Spring 2013, Mathematics Subgroup Scale Score Mean Differences, t-test: ELL/Non ELL

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-37.53	6131.80	<.0001	-48.33	87.08
4	-35.46	3611.86	<.0001	-59.09	88.89
5	-33.12	2871.50	<.0001	-59.64	85.85
6	-32.80	2049.96	<.0001	-63.68	77.77
7	-31.58	2047.87	<.0001	-61.93	79.70
8	-34.95	1955.07	<.0001	-68.17	82.09

Note: SEM at or closest above the cut scores.

Table 25. Spring 2013, Reading Subgroup Scale Score Mean Differences, t-test: ELL/Non ELL

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-46.18	5803.59	<.0001	-59.7	84.44
4	-44.9	3363.69	<.0001	-65.84	75.8
5	-47.89	2659.77	<.0001	-84.65	82.25
6	-44.39	1896.15	<.0001	-81.64	75.84
7	-43.39	1950.8	<.0001	-68.82	66.16
8	-43.56	1842.58	<.0001	-88.38	80.2

Note: SEM at or closest above the cut scores.

Table 26. Spring 2013, Science Subgroup Scale Score Mean Differences, t-test: ELL/Non ELL

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
5	-41.22	2905.52	<.0001	-60.29	70.61
8	-38.14	2183.50	<.0001	-54.06	55.98

Note: SEM at or closest above the cut scores.

Table 27. Spring 2013, Mathematics Subgroup Scale Score Mean Differences, t-test: Female/Male

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-8.89	46272.55	<.0001	-7.30	88.25
4	-4.22	45303.53	<.0001	-3.57	90.10
5	-0.38	44164.41	0.7006	-0.32	86.98
6	-7.51	43142.92	<.0001	-5.69	78.80
7	-4.18	42811.18	<.0001	-3.25	80.69
8	0.54	41302.88	0.5867	0.45	83.25

Note: SEM at or closest above the cut scores.

Table 28. Spring 2013, Reading Subgroup Scale Score Mean Differences, t-test: Female/Male

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	16.49	45566.97	<.0001	13.28	86.08
4	16.89	44555.12	<.0001	12.35	77.28
5	18.36	43663.49	<.0001	14.77	84.14
6	14.30	42912.16	<.0001	10.67	77.34
7	20.38	43337.62	<.0001	13.16	67.24
8	22.99	42327.29	<.0001	18.22	81.54

Note: SEM at or closest above the cut scores.

Table 29. Spring 2013, Science Subgroup Scale Score Mean Differences, t-test: Female/Male

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
5	-8.95	44456.64	<.0001	-6.09	71.95
8	-2.08	43711.20	0.0378	-1.13	57.10

Note: SEM at or closest above the cut scores.

Table 30. Spring 2013, Mathematics Subgroup Scale Score Mean Differences, t-test: IEP/Non IEP

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-32.80	6701.13	<.0001	-44.03	87.19
4	-37.74	6612.51	<.0001	-50.72	88.67
5	-42.07	6043.16	<.0001	-56.70	85.14
6	-47.91	5069.74	<.0001	-64.05	76.50
7	-48.26	4595.32	<.0001	-69.66	78.16
8	-46.85	3807.15	<.0001	-70.71	81.04

Note: SEM at or closest above the cut scores.

Table 31. Spring 2013, Reading Subgroup Scale Score Mean Differences, t-test: IEP/Non IEP

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-39.92	5681.47	<.0001	-60.10	84.33
4	-44.41	5458.57	<.0001	-60.34	75.30
5	-48.33	5186.22	<.0001	-72.09	81.61
6	-52.60	4785.31	<.0001	-71.36	74.65
7	-49.42	4589.85	<.0001	-60.17	65.30
8	-52.83	3986.36	<.0001	-80.08	79.07

Note: SEM at or closest above the cut scores.

Table 32. Spring 2013, Science Subgroup Scale Score Mean Differences, t-test: IEP/Non IEP

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
5	-42.78	6652.51	<.0001	-47.90	70.28
8	-49.24	5316.69	<.0001	-50.36	55.00

Note: SEM at or closest above the cut scores.

Table 33. Spring 2013, Mathematics Subgroup Scale Score Mean Differences, t-test: Low SES/High SES

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-62.28	38957.87	<.0001	-50.15	84.88
4	-61.27	38143.00	<.0001	-51.05	86.62
5	-58.48	37742.18	<.0001	-47.64	83.78
6	-61.44	39581.00	<.0001	-45.23	75.62
7	-60.23	40875.23	<.0001	-45.26	77.51
8	-55.96	40653.87	<.0001	-44.26	80.27

Note: SEM at or closest above the cut scores.

Table 34. Spring 2013, Reading Subgroup Scale Score Mean Differences, t-test: Low SES/High SES

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-63.00	39626.97	<.0001	-49.45	82.91
4	-64.36	38869.40	<.0001	-45.89	74.21
5	-64.10	39028.10	<.0001	-50.05	80.82
6	-62.01	39510.23	<.0001	-44.92	74.28
7	-54.61	40956.43	<.0001	-34.53	65.35
8	-59.58	41853.13	<.0001	-45.66	78.84

Note: SEM at or closest above the cut scores.

Table 35. Spring 2013, Science Subgroup Scale Score Mean Differences, t-test: Low SES/High SES

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
5	-62.02	39410.37	<.0001	-41.03	69.14
8	-60.08	44138.96	<.0001	-31.17	54.96

Note: SEM at or closest above the cut scores.

Table 36. Spring 2013, Mathematics Subgroup Scale Score Mean Differences, t-test: Section 504/Non Section 504

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-3.32	46314.00	0.0009	-17.25	88.32
4	-3.58	45381.00	0.0003	-17.12	90.13
5	-3.86	44293.00	0.0001	-17.47	86.96
6	-2.68	43220.00	0.0075	-10.64	78.85
7	-4.72	43144.00	<.0001	-19.12	80.68
8	-3.61	41375.00	0.0003	-15.47	83.24

Note: SEM at or closest above the cut scores.

Table 37. Spring 2013, Reading Subgroup Scale Score Mean Differences, t-test: Section 504/Non Section 504

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
3	-3.96	45681.00	<.0001	-19.98	86.33
4	-2.64	44702.00	0.0083	-10.81	77.53
5	-3.22	43796.00	0.0013	-14.19	84.46
6	-3.15	42969.00	0.0016	-12.29	77.52
7	-3.89	43366.00	0.0001	-12.99	67.55
8	-2.38	42339.00	0.0173	-9.92	82.05

Note: SEM at or closest above the cut scores.

Table 38. Spring 2013, Science Subgroup Scale Score Mean Differences, t-test: Section 504/Non Section 504

Grade	*t*	DF	Sig. (2-tailed)	Mean Difference	Standard Error Difference
5	-3.01	44803.00	0.0026	-11.27	72.00
8	-0.48	44207.00	0.6322	-1.38	57.12

Note: SEM at or closest above the cut scores.

Table 39. Spring 2013, Mathematics Subgroup Mean Differences, ANOVA: Ethnicity

Grade	Categories	Sum of Squares	df	Mean Square	*F*	Sig.
3	Between Groups	22245144.80	5	4449029.00	607.58	<.0001
3	Within Groups	339104510.20	46310	7322.50	.	.
3	Total	361349655.00	46315	.	.	.
4	Between Groups	19842641.70	5	3968528.30	516.09	<.0001
4	Within Groups	348931756.70	45377	7689.60	.	.
4	Total	368774398.40	45382	.	.	.
5	Between Groups	16643559.70	5	3328711.90	462.97	<.0001
5	Within Groups	318431688.10	44289	7189.90	.	.
5	Total	335075247.80	44294	.	.	.
6	Between Groups	15261649.40	5	3052329.90	520.36	<.0001
6	Within Groups	253495412.10	43216	5865.80	.	.
6	Total	268757061.40	43221	.	.	.
7	Between Groups	16447623.50	5	3289524.70	536.44	<.0001
7	Within Groups	264538408.40	43140	6132.10	.	.
7	Total	280986031.90	43145	.	.	.
8	Between Groups	13792216.50	5	2758443.30	418.09	<.0001
8	Within Groups	272952949.10	41371	6597.70	.	.
8	Total	286745165.60	41376	.	.	.

Note: df = Degrees of Freedom

Table 40. Spring 2013, Reading Subgroup Mean Differences, ANOVA: Ethnicity

Grade	Categories	Sum of Squares	df	Mean Square	*F*	Sig.
3	Between Groups	18364022.20	5	3672804.40	520.60	<.0001
3	Within Groups	322245961.10	45677	7054.90	.	.
3	Total	340609983.30	45682	.	.	.
4	Between Groups	14026352.50	5	2805270.50	492.22	<.0001
4	Within Groups	254745023.10	44698	5699.20	.	.
4	Total	268771375.60	44703	.	.	.
5	Between Groups	17976784.00	5	3595356.80	534.59	<.0001
5	Within Groups	294518082.00	43792	6725.40	.	.
5	Total	312494866.00	43797	.	.	.
6	Between Groups	13643341.20	5	2728668.20	479.22	<.0001
6	Within Groups	244644053.60	42965	5694.00	.	.
6	Total	258287394.80	42970	.	.	.
7	Between Groups	10806463.60	5	2161292.70	500.81	<.0001
7	Within Groups	187131950.80	43362	4315.60	.	.
7	Total	197938414.40	43367	.	.	.
8	Between Groups	13343716.30	5	2668743.30	415.75	<.0001
8	Within Groups	271751838.10	42335	6419.10	.	.
8	Total	285095554.40	42340	.	.	.

Note: df = Degrees of Freedom

Table 41. Spring 2013, Science Subgroup Mean Differences, ANOVA: Ethnicity

Grade	Categories	Sum of Squares	df	Mean Square	*F*	Sig.
5	Between Groups	16053502.70	5	3210700.50	665.21	<.0001
5	Within Groups	216225933.40	44799	4826.60	.	.
5	Total	232279436.10	44804	.	.	.
8	Between Groups	8792562.80	5	1758512.60	574.02	<.0001
8	Within Groups	135417058.00	44203	3063.50	.	.
8	Total	144209620.80	44208	.	.	.

Note: df = Degrees of Freedom

Table 42. Mathematics, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
3	SS	American Indian/Alaskan	Asian	40.64(*)	2.53	33.14	48.13
			African Amer. (Not Hispanic)	-47.43(*)	2.53	-51.69	-43.17
			Hispanic	-26.1(*)	2.53	-29.73	-22.47
			Multiracial	-0.09	2.53	-4.72	4.55
			White (Not Hispanic)	16.46(*)	2.53	13.50	19.41
		Asian	Amer. Indian/Alaskan	-40.64(*)	2.32	-47.49	-33.78
			African Amer. (Not Hispanic)	-88.07(*)	2.32	-95.20	-80.94
			Hispanic	-66.74(*)	2.32	-73.57	-59.90
			Multiracial	-40.72(*)	2.32	-48.05	-33.40
			White (Not Hispanic)	-24.18(*)	2.32	-30.73	-17.62
		African American (Not Hispanic)	Amer. Indian/Alaskan	47.43(*)	2.49	43.23	51.63
			Asian	88.07(*)	2.49	80.39	95.75
			Hispanic	21.33(*)	2.49	17.17	25.49
			Multiracial	47.34(*)	2.49	42.31	52.38
			White (Not Hispanic)	63.89(*)	2.49	60.29	67.49
		Hispanic	Amer. Indian/Alaskan	26.1(*)	2.53	22.47	29.74
			Asian	66.74(*)	2.53	59.26	74.21
			African Amer. (Not Hispanic)	-21.33(*)	2.53	-25.55	-17.11
			Multiracial	26.01(*)	2.53	21.42	30.61
			White (Not Hispanic)	42.56(*)	2.53	39.66	45.46
		Multiracial	Amer. Indian/Alaskan	0.09	2.47	-4.44	4.61
			Asian	40.72(*)	2.47	32.91	48.54
			African Amer. (Not Hispanic)	-47.34(*)	2.47	-52.33	-42.36
			Hispanic	-26.01(*)	2.47	-30.50	-21.53
			White (Not Hispanic)	16.55(*)	2.47	12.56	20.53
		White (Not Hispanic)	Amer. Indian/Alaskan	-16.46(*)	2.56	-19.45	-13.46
			Asian	24.18(*)	2.56	16.92	31.44
			African Amer. (Not Hispanic)	-63.89(*)	2.56	-67.59	-60.19
			Hispanic	-42.56(*)	2.56	-45.49	-39.63
			Multiracial	-16.55(*)	2.56	-20.68	-12.41

(*)Significant differences

Table 42. Mathematics, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
4	SS	American Indian/Alaskan	Asian	52.81(*)	2.53	44.74	60.88
			African Amer. (Not Hispanic)	-39.25(*)	2.53	-43.65	-34.86
			Hispanic	-19.77(*)	2.53	-23.56	-15.99
			Multiracial	3.50	2.53	-1.33	8.34
			White (Not Hispanic)	20.33(*)	2.53	17.29	23.37
		Asian	Amer. Indian/Alaskan	-52.81(*)	2.31	-60.16	-45.45
			African Amer. (Not Hispanic)	-92.06(*)	2.31	-99.69	-84.43
			Hispanic	-72.58(*)	2.31	-79.93	-65.23
			Multiracial	-49.3(*)	2.31	-57.15	-41.46
			White (Not Hispanic)	-32.48(*)	2.31	-39.54	-25.42
		African American (Not Hispanic)	Amer. Indian/Alaskan	39.25(*)	2.50	34.92	43.58
			Asian	92.06(*)	2.50	83.80	100.32
			Hispanic	19.48(*)	2.50	15.15	23.81
			Multiracial	42.76(*)	2.50	37.50	48.01
			White (Not Hispanic)	59.58(*)	2.50	55.86	63.30
		Hispanic	Amer. Indian/Alaskan	19.77(*)	2.53	15.99	23.56
			Asian	72.58(*)	2.53	64.51	80.65
			African Amer. (Not Hispanic)	-19.48(*)	2.53	-23.87	-15.08
			Multiracial	23.28(*)	2.53	18.44	28.11
			White (Not Hispanic)	40.1(*)	2.53	37.06	43.14
		Multiracial	Amer. Indian/Alaskan	-3.50	2.47	-8.22	1.22
			Asian	49.3(*)	2.47	40.90	57.71
			African Amer. (Not Hispanic)	-42.76(*)	2.47	-47.95	-37.56
			Hispanic	-23.28(*)	2.47	-28.00	-18.56
			White (Not Hispanic)	16.83(*)	2.47	12.65	21.00
		White (Not Hispanic)	Amer. Indian/Alaskan	-20.33(*)	2.56	-23.41	-17.25
			Asian	32.48(*)	2.56	24.63	40.33
			African Amer. (Not Hispanic)	-59.58(*)	2.56	-63.40	-55.76
			Hispanic	-40.1(*)	2.56	-43.18	-37.02
			Multiracial	-16.83(*)	2.56	-21.16	-12.49

(*)Significant differences

Table 42. Mathematics, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
5	SS	American Indian/Alaskan	Asian	64.34(*)	2.53	56.57	72.11
			African Amer. (Not Hispanic)	-28.21(*)	2.53	-32.46	-23.97
			Hispanic	-7.61(*)	2.53	-11.31	-3.92
			Multiracial	10.09(*)	2.53	5.36	14.82
			White (Not Hispanic)	25.64(*)	2.53	22.72	28.56
		Asian	Amer. Indian/Alaskan	-64.34(*)	2.31	-71.42	-57.25
			African Amer. (Not Hispanic)	-92.55(*)	2.31	-99.92	-85.18
			Hispanic	-71.95(*)	2.31	-79.07	-64.83
			Multiracial	-54.25(*)	2.31	-61.86	-46.64
			White (Not Hispanic)	-38.7(*)	2.31	-45.51	-31.88
		African American (Not Hispanic)	Amer. Indian/Alaskan	28.21(*)	2.50	24.04	32.39
			Asian	92.55(*)	2.50	84.59	100.52
			Hispanic	20.6(*)	2.50	16.35	24.85
			Multiracial	38.3(*)	2.50	33.15	43.45
			White (Not Hispanic)	53.85(*)	2.50	50.24	57.47
		Hispanic	Amer. Indian/Alaskan	7.61(*)	2.53	3.92	11.30
			Asian	71.95(*)	2.53	64.16	79.75
			African Amer. (Not Hispanic)	-20.6(*)	2.53	-24.90	-16.30
			Multiracial	17.7(*)	2.53	12.92	22.49
			White (Not Hispanic)	33.25(*)	2.53	30.24	36.27
		Multiracial	Amer. Indian/Alaskan	-10.09(*)	2.47	-14.70	-5.48
			Asian	54.25(*)	2.47	46.12	62.38
			African Amer. (Not Hispanic)	-38.3(*)	2.47	-43.39	-33.21
			Hispanic	-17.7(*)	2.47	-22.37	-13.03
			White (Not Hispanic)	15.55(*)	2.47	11.43	19.67
		White (Not Hispanic)	Amer. Indian/Alaskan	-25.64(*)	2.56	-28.60	-22.68
			Asian	38.7(*)	2.56	31.13	46.26
			African Amer. (Not Hispanic)	-53.85(*)	2.56	-57.57	-50.14
			Hispanic	-33.25(*)	2.56	-36.31	-30.19
			Multiracial	-15.55(*)	2.56	-19.83	-11.27

(*)Significant differences

Table 42. Mathematics, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
6	SS	American Indian/Alaskan	Asian	56.33(*)	2.54	49.36	63.29
			African Amer. (Not Hispanic)	-32.23(*)	2.54	-36.17	-28.30
			Hispanic	-14.17(*)	2.54	-17.59	-10.75
			Multiracial	3.83	2.54	-0.45	8.10
			White (Not Hispanic)	20.78(*)	2.54	18.12	23.44
		Asian	Amer. Indian/Alaskan	-56.33(*)	2.32	-62.69	-49.96
			African Amer. (Not Hispanic)	-88.56(*)	2.32	-95.22	-81.90
			Hispanic	-70.5(*)	2.32	-76.92	-64.08
			Multiracial	-52.5(*)	2.32	-59.33	-45.67
			White (Not Hispanic)	-35.55(*)	2.32	-41.66	-29.44
		African American (Not Hispanic)	Amer. Indian/Alaskan	32.23(*)	2.49	28.37	36.10
			Asian	88.56(*)	2.49	81.39	95.73
			Hispanic	18.06(*)	2.49	14.09	22.04
			Multiracial	36.06(*)	2.49	31.35	40.76
			White (Not Hispanic)	53.01(*)	2.49	49.65	56.38
		Hispanic	Amer. Indian/Alaskan	14.17(*)	2.53	10.76	17.58
			Asian	70.5(*)	2.53	63.49	77.50
			African Amer. (Not Hispanic)	-18.06(*)	2.53	-22.09	-14.03
			Multiracial	18(*)	2.53	13.64	22.35
			White (Not Hispanic)	34.95(*)	2.53	32.14	37.76
		Multiracial	Amer. Indian/Alaskan	-3.83	2.47	-7.99	0.34
			Asian	52.5(*)	2.47	45.21	59.79
			African Amer. (Not Hispanic)	-36.06(*)	2.47	-40.72	-31.40
			Hispanic	-18(*)	2.47	-22.26	-13.73
			White (Not Hispanic)	16.95(*)	2.47	13.24	20.66
		White (Not Hispanic)	Amer. Indian/Alaskan	-20.78(*)	2.56	-23.47	-18.09
			Asian	35.55(*)	2.56	28.78	42.32
			African Amer. (Not Hispanic)	-53.01(*)	2.56	-56.47	-49.55
			Hispanic	-34.95(*)	2.56	-37.80	-32.10
			Multiracial	-16.95(*)	2.56	-20.80	-13.10

(*)Significant differences

Table 42. Mathematics, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
7	SS	American Indian/Alaskan	Asian	61.72(*)	2.54	54.37	69.08
			African Amer. (Not Hispanic)	-34.12(*)	2.54	-38.09	-30.15
			Hispanic	-16.48(*)	2.54	-20.02	-12.95
			Multiracial	-2.31	2.54	-6.68	2.07
			White (Not Hispanic)	19.85(*)	2.54	17.15	22.55
		Asian	Amer. Indian/Alaskan	-61.72(*)	2.31	-68.42	-55.03
			African Amer. (Not Hispanic)	-95.84(*)	2.31	-102.81	-88.87
			Hispanic	-78.21(*)	2.31	-84.98	-71.43
			Multiracial	-64.03(*)	2.31	-71.20	-56.86
			White (Not Hispanic)	-41.87(*)	2.31	-48.33	-35.42
		African American (Not Hispanic)	Amer. Indian/Alaskan	34.12(*)	2.50	30.21	38.03
			Asian	95.84(*)	2.50	88.29	103.39
			Hispanic	17.63(*)	2.50	13.56	21.71
			Multiracial	31.81(*)	2.50	27.01	36.61
			White (Not Hispanic)	53.97(*)	2.50	50.57	57.37
		Hispanic	Amer. Indian/Alaskan	16.48(*)	2.52	12.97	20.00
			Asian	78.21(*)	2.52	70.79	85.62
			African Amer. (Not Hispanic)	-17.63(*)	2.52	-21.75	-13.52
			Multiracial	14.18(*)	2.52	9.67	18.68
			White (Not Hispanic)	36.33(*)	2.52	33.40	39.26
		Multiracial	Amer. Indian/Alaskan	2.31	2.47	-1.95	6.57
			Asian	64.03(*)	2.47	56.35	71.71
			African Amer. (Not Hispanic)	-31.81(*)	2.47	-36.56	-27.06
			Hispanic	-14.18(*)	2.47	-18.59	-9.76
			White (Not Hispanic)	22.16(*)	2.47	18.34	25.97
		White (Not Hispanic)	Amer. Indian/Alaskan	-19.85(*)	2.56	-22.58	-17.12
			Asian	41.87(*)	2.56	34.71	49.04
			African Amer. (Not Hispanic)	-53.97(*)	2.56	-57.46	-50.47
			Hispanic	-36.33(*)	2.56	-39.31	-33.36
			Multiracial	-22.16(*)	2.56	-26.11	-18.20

(*)Significant differences

Table 42. Mathematics, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
8	SS	American Indian/Alaskan	Asian	69.82(*)	2.54	62.20	77.44
			African Amer. (Not Hispanic)	-26.31(*)	2.54	-30.53	-22.08
			Hispanic	-14.92(*)	2.54	-18.71	-11.12
			Multiracial	3.09	2.54	-1.60	7.77
			White (Not Hispanic)	19.79(*)	2.54	16.92	22.65
		Asian	Amer. Indian/Alaskan	-69.82(*)	2.32	-76.78	-62.86
			African Amer. (Not Hispanic)	-96.13(*)	2.32	-103.39	-88.86
			Hispanic	-84.74(*)	2.32	-91.80	-77.67
			Multiracial	-66.73(*)	2.32	-74.23	-59.23
			White (Not Hispanic)	-50.03(*)	2.32	-56.73	-43.34
		African American (Not Hispanic)	Amer. Indian/Alaskan	26.31(*)	2.50	22.15	30.47
			Asian	96.13(*)	2.50	88.29	103.96
			Hispanic	11.39(*)	2.50	7.03	15.75
			Multiracial	29.39(*)	2.50	24.25	34.53
			White (Not Hispanic)	46.09(*)	2.50	42.48	49.71
		Hispanic	Amer. Indian/Alaskan	14.92(*)	2.52	11.14	18.69
			Asian	84.74(*)	2.52	77.04	92.43
			African Amer. (Not Hispanic)	-11.39(*)	2.52	-15.80	-6.98
			Multiracial	18(*)	2.52	13.15	22.85
			White (Not Hispanic)	34.7(*)	2.52	31.56	37.85
		Multiracial	Amer. Indian/Alaskan	-3.09	2.47	-7.65	1.48
			Asian	66.73(*)	2.47	58.74	74.73
			African Amer. (Not Hispanic)	-29.39(*)	2.47	-34.47	-24.31
			Hispanic	-18(*)	2.47	-22.75	-13.25
			White (Not Hispanic)	16.7(*)	2.47	12.62	20.79
		White (Not Hispanic)	Amer. Indian/Alaskan	-19.79(*)	2.57	-22.68	-16.89
			Asian	50.03(*)	2.57	42.62	57.44
			African Amer. (Not Hispanic)	-46.09(*)	2.57	-49.80	-42.38
			Hispanic	-34.7(*)	2.57	-37.90	-31.50
			Multiracial	-16.7(*)	2.57	-20.94	-12.46

(*)Significant differences

Table 43. Reading, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
3	SS	American Indian/Alaskan	Asian	31.71(*)	2.53	24.26	39.16
			African Amer. (Not Hispanic)	-31.82(*)	2.53	-36.03	-27.61
			Hispanic	-26.11(*)	2.53	-29.70	-22.52
			Multiracial	5.83(*)	2.53	1.26	10.41
			White (Not Hispanic)	19.73(*)	2.53	16.81	22.65
		Asian	Amer. Indian/Alaskan	-31.71(*)	2.31	-38.52	-24.89
			African Amer. (Not Hispanic)	-63.53(*)	2.31	-70.62	-56.44
			Hispanic	-57.82(*)	2.31	-64.61	-51.02
			Multiracial	-25.88(*)	2.31	-33.15	-18.60
			White (Not Hispanic)	-11.98(*)	2.31	-18.50	-5.45
		African American (Not Hispanic)	Amer. Indian/Alaskan	31.82(*)	2.49	27.67	35.97
			Asian	63.53(*)	2.49	55.89	71.17
			Hispanic	5.71(*)	2.49	1.60	9.83
			Multiracial	37.65(*)	2.49	32.68	42.63
			White (Not Hispanic)	51.55(*)	2.49	47.99	55.11
		Hispanic	Amer. Indian/Alaskan	26.11(*)	2.53	22.51	29.70
			Asian	57.82(*)	2.53	50.38	65.26
			African Amer. (Not Hispanic)	-5.71(*)	2.53	-9.89	-1.53
			Multiracial	31.94(*)	2.53	27.40	36.49
			White (Not Hispanic)	45.84(*)	2.53	42.97	48.71
		Multiracial	Amer. Indian/Alaskan	-5.83(*)	2.47	-10.30	-1.36
			Asian	25.88(*)	2.47	18.11	33.65
			African Amer. (Not Hispanic)	-37.65(*)	2.47	-42.58	-32.73
			Hispanic	-31.94(*)	2.47	-36.38	-27.51
			White (Not Hispanic)	13.90(*)	2.47	9.96	17.84
		White (Not Hispanic)	Amer. Indian/Alaskan	-19.73(*)	2.56	-22.69	-16.77
			Asian	11.98(*)	2.56	4.75	19.21
			African Amer. (Not Hispanic)	-51.55(*)	2.56	-55.21	-47.89
			Hispanic	-45.84(*)	2.56	-48.75	-42.94
			Multiracial	-13.90(*)	2.56	-17.98	-9.81

(*)Significant differences

Table 43. Reading, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
4	SS	American Indian/Alaskan	Asian	29.36(*)	2.53	22.34	36.38
			African Amer. (Not Hispanic)	-29.69(*)	2.53	-33.50	-25.88
			Hispanic	-20.41(*)	2.53	-23.69	-17.12
			Multiracial	5.45(*)	2.53	1.25	9.66
			White (Not Hispanic)	18.41(*)	2.53	15.77	21.05
		Asian	Amer. Indian/Alaskan	-29.36(*)	2.31	-35.76	-22.97
			African Amer. (Not Hispanic)	-59.05(*)	2.31	-65.68	-52.42
			Hispanic	-49.77(*)	2.31	-56.17	-43.37
			Multiracial	-23.91(*)	2.31	-30.74	-17.08
			White (Not Hispanic)	-10.95(*)	2.31	-17.09	-4.81
		African American (Not Hispanic)	Amer. Indian/Alaskan	29.69(*)	2.50	25.93	33.45
			Asian	59.05(*)	2.50	51.87	66.23
			Hispanic	9.28(*)	2.50	5.52	13.05
			Multiracial	35.14(*)	2.50	30.58	39.71
			White (Not Hispanic)	48.10(*)	2.50	44.87	51.33
		Hispanic	Amer. Indian/Alaskan	20.41(*)	2.53	17.12	23.69
			Asian	49.77(*)	2.53	42.74	56.79
			African Amer. (Not Hispanic)	-9.28(*)	2.53	-13.10	-5.47
			Multiracial	25.86(*)	2.53	21.65	30.07
			White (Not Hispanic)	38.82(*)	2.53	36.17	41.46
		Multiracial	Amer. Indian/Alaskan	-5.45(*)	2.47	-9.56	-1.35
			Asian	23.91(*)	2.47	16.59	31.22
			African Amer. (Not Hispanic)	-35.14(*)	2.47	-39.66	-30.62
			Hispanic	-25.86(*)	2.47	-29.97	-21.75
			White (Not Hispanic)	12.96(*)	2.47	9.32	16.59
		White (Not Hispanic)	Amer. Indian/Alaskan	-18.41(*)	2.56	-21.08	-15.74
			Asian	10.95(*)	2.56	4.12	17.78
			African Amer. (Not Hispanic)	-48.10(*)	2.56	-51.41	-44.78
			Hispanic	-38.82(*)	2.56	-41.49	-36.14
			Multiracial	-12.96(*)	2.56	-16.73	-9.18

(*)Significant differences

Table 43. Reading, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
5	SS	American Indian/Alaskan	Asian	35.87(*)	2.53	28.28	43.47
			African Amer. (Not Hispanic)	-29.35(*)	2.53	-33.48	-25.22
			Hispanic	-21.46(*)	2.53	-25.07	-17.86
			Multiracial	8.20(*)	2.53	3.60	12.79
			White (Not Hispanic)	23.99(*)	2.53	21.15	26.84
		Asian	Amer. Indian/Alaskan	-35.87(*)	2.31	-42.79	-28.95
			African Amer. (Not Hispanic)	-65.22(*)	2.31	-72.41	-58.03
			Hispanic	-57.34(*)	2.31	-64.29	-50.38
			Multiracial	-27.68(*)	2.31	-35.10	-20.25
			White (Not Hispanic)	-11.88(*)	2.31	-18.54	-5.22
		African American (Not Hispanic)	Amer. Indian/Alaskan	29.35(*)	2.50	25.28	33.41
			Asian	65.22(*)	2.50	57.44	73.00
			Hispanic	7.88(*)	2.50	3.74	12.03
			Multiracial	37.54(*)	2.50	32.54	42.55
			White (Not Hispanic)	53.34(*)	2.50	49.82	56.86
		Hispanic	Amer. Indian/Alaskan	21.46(*)	2.53	17.87	25.06
			Asian	57.34(*)	2.53	49.72	64.96
			African Amer. (Not Hispanic)	-7.88(*)	2.53	-12.08	-3.69
			Multiracial	29.66(*)	2.53	25.01	34.31
			White (Not Hispanic)	45.46(*)	2.53	42.51	48.40
		Multiracial	Amer. Indian/Alaskan	-8.20(*)	2.47	-12.67	-3.72
			Asian	27.68(*)	2.47	19.74	35.61
			African Amer. (Not Hispanic)	-37.54(*)	2.47	-42.49	-32.60
			Hispanic	-29.66(*)	2.47	-34.20	-25.12
			White (Not Hispanic)	15.80(*)	2.47	11.80	19.79
		White (Not Hispanic)	Amer. Indian/Alaskan	-23.99(*)	2.56	-26.87	-21.12
			Asian	11.88(*)	2.56	4.48	19.27
			African Amer. (Not Hispanic)	-53.34(*)	2.56	-56.96	-49.73
			Hispanic	-45.46(*)	2.56	-48.44	-42.47
			Multiracial	-15.80(*)	2.56	-19.95	-11.65

(*)Significant differences

Table 43. Reading, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
6	SS	American Indian/Alaskan	Asian	35.03(*)	2.54	28.06	42.01
			African Amer. (Not Hispanic)	-26.58(*)	2.54	-30.45	-22.71
			Hispanic	-19.58(*)	2.54	-22.96	-16.19
			Multiracial	4.65(*)	2.54	0.44	8.85
			White (Not Hispanic)	20.70(*)	2.54	18.08	23.32
		Asian	Amer. Indian/Alaskan	-35.03(*)	2.31	-41.39	-28.67
			African Amer. (Not Hispanic)	-61.61(*)	2.31	-68.26	-54.97
			Hispanic	-54.61(*)	2.31	-61.04	-48.19
			Multiracial	-30.39(*)	2.31	-37.20	-23.58
			White (Not Hispanic)	-14.33(*)	2.31	-20.45	-8.21
		African American (Not Hispanic)	Amer. Indian/Alaskan	26.58(*)	2.49	22.77	30.39
			Asian	61.61(*)	2.49	54.44	68.78
			Hispanic	7.00(*)	2.49	3.07	10.93
			Multiracial	31.23(*)	2.49	26.59	35.86
			White (Not Hispanic)	47.28(*)	2.49	43.96	50.60
		Hispanic	Amer. Indian/Alaskan	19.58(*)	2.53	16.21	22.95
			Asian	54.61(*)	2.53	47.60	61.63
			African Amer. (Not Hispanic)	-7.00(*)	2.53	-10.98	-3.02
			Multiracial	24.22(*)	2.53	19.92	28.53
			White (Not Hispanic)	40.28(*)	2.53	37.49	43.07
		Multiracial	Amer. Indian/Alaskan	-4.65(*)	2.47	-8.74	-0.55
			Asian	30.39(*)	2.47	23.10	37.67
			African Amer. (Not Hispanic)	-31.23(*)	2.47	-35.82	-26.63
			Hispanic	-24.22(*)	2.47	-28.43	-20.01
			White (Not Hispanic)	16.06(*)	2.47	12.40	19.71
		White (Not Hispanic)	Amer. Indian/Alaskan	-20.7(*)	2.56	-23.35	-18.05
			Asian	14.33(*)	2.56	7.55	21.12
			African Amer. (Not Hispanic)	-47.28(*)	2.56	-50.69	-43.87
			Hispanic	-40.28(*)	2.56	-43.11	-37.45
			Multiracial	-16.06(*)	2.56	-19.85	-12.27

(*)Significant differences

Table 43. Reading, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
7	SS	American Indian/Alaskan	Asian	25.19(*)	2.54	19.03	31.36
			African Amer. (Not Hispanic)	-30.43(*)	2.54	-33.73	-27.13
			Hispanic	-21.28(*)	2.54	-24.25	-18.32
			Multiracial	-2.74	2.54	-6.40	0.93
			White (Not Hispanic)	13.53(*)	2.54	11.27	15.79
		Asian	Amer. Indian/Alaskan	-25.19(*)	2.31	-30.80	-19.58
			African Amer. (Not Hispanic)	-55.62(*)	2.31	-61.46	-49.79
			Hispanic	-46.47(*)	2.31	-52.16	-40.79
			Multiracial	-27.93(*)	2.31	-33.94	-21.92
			White (Not Hispanic)	-11.66(*)	2.31	-17.07	-6.25
		African American (Not Hispanic)	Amer. Indian/Alaskan	30.43(*)	2.50	27.18	33.68
			Asian	55.62(*)	2.50	49.31	61.94
			Hispanic	9.15(*)	2.50	5.75	12.55
			Multiracial	27.7(*)	2.50	23.69	31.70
			White (Not Hispanic)	43.96(*)	2.50	41.14	46.79
		Hispanic	Amer. Indian/Alaskan	21.28(*)	2.52	18.33	24.23
			Asian	46.47(*)	2.52	40.26	52.69
			African Amer. (Not Hispanic)	-9.15(*)	2.52	-12.59	-5.72
			Multiracial	18.55(*)	2.52	14.77	22.32
			White (Not Hispanic)	34.81(*)	2.52	32.35	37.27
		Multiracial	Amer. Indian/Alaskan	2.74	2.47	-0.83	6.30
			Asian	27.93(*)	2.47	21.49	34.36
			African Amer. (Not Hispanic)	-27.7(*)	2.47	-31.66	-23.73
			Hispanic	-18.55(*)	2.47	-22.25	-14.84
			White (Not Hispanic)	16.27(*)	2.47	13.07	19.46
		White (Not Hispanic)	Amer. Indian/Alaskan	-13.53(*)	2.56	-15.82	-11.25
			Asian	11.66(*)	2.56	5.65	17.67
			African Amer. (Not Hispanic)	-43.96(*)	2.56	-46.86	-41.06
			Hispanic	-34.81(*)	2.56	-37.31	-32.31
			Multiracial	-16.27(*)	2.56	-19.58	-12.95

(*)Significant differences

Table 43. Reading, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
8	SS	American Indian/Alaskan	Asian	37.19(*)	2.54	29.65	44.74
			African Amer. (Not Hispanic)	-29.71(*)	2.54	-33.80	-25.62
			Hispanic	-21.39(*)	2.54	-25.09	-17.69
			Multiracial	4.05	2.54	-0.53	8.62
			White (Not Hispanic)	18.30(*)	2.54	15.51	21.09
		Asian	Amer. Indian/Alaskan	-37.19(*)	2.31	-44.07	-30.32
			African Amer. (Not Hispanic)	-66.91(*)	2.31	-74.07	-59.74
			Hispanic	-58.58(*)	2.31	-65.57	-51.60
			Multiracial	-33.15(*)	2.31	-40.55	-25.75
			White (Not Hispanic)	-18.89(*)	2.31	-25.52	-12.27
		African American (Not Hispanic)	Amer. Indian/Alaskan	29.71(*)	2.50	25.68	33.74
			Asian	66.91(*)	2.50	59.16	74.65
			Hispanic	8.32(*)	2.50	4.08	12.57
			Multiracial	33.76(*)	2.50	28.76	38.76
			White (Not Hispanic)	48.01(*)	2.50	44.52	51.51
		Hispanic	Amer. Indian/Alaskan	21.39(*)	2.52	17.71	25.07
			Asian	58.58(*)	2.52	50.96	66.20
			African Amer. (Not Hispanic)	-8.32(*)	2.52	-12.60	-4.04
			Multiracial	25.43(*)	2.52	20.70	30.17
			White (Not Hispanic)	39.69(*)	2.52	36.61	42.77
		Multiracial	Amer. Indian/Alaskan	-4.05	2.47	-8.50	0.41
			Asian	33.15(*)	2.47	25.24	41.05
			African Amer. (Not Hispanic)	-33.76(*)	2.47	-38.70	-28.82
			Hispanic	-25.43(*)	2.47	-30.08	-20.79
			White (Not Hispanic)	14.26(*)	2.47	10.27	18.24
		White (Not Hispanic)	Amer. Indian/Alaskan	-18.3(*)	2.57	-21.12	-15.48
			Asian	18.89(*)	2.57	11.54	26.24
			African Amer. (Not Hispanic)	-48.01(*)	2.57	-51.60	-44.42
			Hispanic	-39.69(*)	2.57	-42.82	-36.56
			Multiracial	-14.26(*)	2.57	-18.40	-10.11

(*)Significant differences

Table 44. Science, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
5	SS	American Indian/Alaskan	Asian	26.28(*)	2.53	19.92	32.64
			African Amer. (Not Hispanic)	-37.37(*)	2.53	-40.81	-33.93
			Hispanic	-19.77(*)	2.53	-22.78	-16.76
			Multiracial	4.04(*)	2.53	0.18	7.89
			White (Not Hispanic)	18.31(*)	2.53	15.93	20.69
		Asian	Amer. Indian/Alaskan	-26.28(*)	2.31	-32.07	-20.49
			African Amer. (Not Hispanic)	-63.65(*)	2.31	-69.66	-57.63
			Hispanic	-46.05(*)	2.31	-51.87	-40.22
			Multiracial	-22.24(*)	2.31	-28.46	-16.02
			White (Not Hispanic)	-7.97(*)	2.31	-13.54	-2.40
		African American (Not Hispanic)	Amer. Indian/Alaskan	37.37(*)	2.50	33.98	40.76
			Asian	63.65(*)	2.50	57.14	70.16
			Hispanic	17.60(*)	2.50	14.15	21.05
			Multiracial	41.40(*)	2.50	37.22	45.59
			White (Not Hispanic)	55.68(*)	2.50	52.75	58.61
		Hispanic	Amer. Indian/Alaskan	19.77(*)	2.53	16.76	22.77
			Asian	46.05(*)	2.53	39.67	52.42
			African Amer. (Not Hispanic)	-17.60(*)	2.53	-21.09	-14.11
			Multiracial	23.80(*)	2.53	19.90	27.70
			White (Not Hispanic)	38.08(*)	2.53	35.62	40.54
		Multiracial	Amer. Indian/Alaskan	-4.04(*)	2.47	-7.79	-0.28
			Asian	22.24(*)	2.47	15.59	28.89
			African Amer. (Not Hispanic)	-41.40(*)	2.47	-45.54	-37.27
			Hispanic	-23.80(*)	2.47	-27.61	-20.00
			White (Not Hispanic)	14.27(*)	2.47	10.92	17.63
		White (Not Hispanic)	Amer. Indian/Alaskan	-18.31(*)	2.56	-20.72	-15.90
			Asian	7.97(*)	2.56	1.78	14.16
			African Amer. (Not Hispanic)	-55.68(*)	2.56	-58.69	-52.67
			Hispanic	-38.08(*)	2.56	-40.57	-35.58
			Multiracial	-14.27(*)	2.56	-17.76	-10.79

(*)Significant differences

Table 44. Science, Pair-Wise Dunnett's C Post-Hoc Comparison of Spring 2013 Scale Score Mean Differences by Ethnicity

Grade	Dependent Variable	(I) Ethnicity	(J) Ethnicity	Mean Difference (J-I)	Dunnett's C	95% Confidence Interval	
						Lower Bound	Upper Bound
8	SS	American Indian/Alaskan	Asian	30.50(*)	2.54	25.37	35.62
			African Amer. (Not Hispanic)	-25.8(*)	2.54	-28.54	-23.06
			Hispanic	-15.89(*)	2.54	-18.38	-13.41
			Multiracial	3.79(*)	2.54	0.62	6.95
			White (Not Hispanic)	13.60(*)	2.54	11.71	15.49
		Asian	Amer. Indian/Alaskan	-30.50(*)	2.31	-35.17	-25.82
			African Amer. (Not Hispanic)	-56.30(*)	2.31	-61.15	-51.44
			Hispanic	-46.39(*)	2.31	-51.12	-41.65
			Multiracial	-26.71(*)	2.31	-31.77	-21.65
			White (Not Hispanic)	-16.90(*)	2.31	-21.40	-12.39
		African American (Not Hispanic)	Amer. Indian/Alaskan	25.80(*)	2.50	23.10	28.50
			Asian	56.30(*)	2.50	51.04	61.55
			Hispanic	9.91(*)	2.50	7.08	12.74
			Multiracial	29.59(*)	2.50	26.16	33.01
			White (Not Hispanic)	39.40(*)	2.50	37.06	41.74
		Hispanic	Amer. Indian/Alaskan	15.89(*)	2.52	13.42	18.36
			Asian	46.39(*)	2.52	41.21	51.56
			African Amer. (Not Hispanic)	-9.91(*)	2.52	-12.76	-7.05
			Multiracial	19.68(*)	2.52	16.42	22.94
			White (Not Hispanic)	29.49(*)	2.52	27.43	31.55
		Multiracial	Amer. Indian/Alaskan	-3.79(*)	2.46	-6.86	-0.71
			Asian	26.71(*)	2.46	21.31	32.11
			African Amer. (Not Hispanic)	-29.59(*)	2.46	-32.96	-26.21
			Hispanic	-19.68(*)	2.46	-22.86	-16.49
			White (Not Hispanic)	9.81(*)	2.46	7.04	12.58
		White (Not Hispanic)	Amer. Indian/Alaskan	-13.60(*)	2.57	-15.51	-11.69
			Asian	16.90(*)	2.57	11.90	21.89
			African Amer. (Not Hispanic)	-39.40(*)	2.57	-41.80	-37.00
			Hispanic	-29.49(*)	2.57	-31.58	-27.40
			Multiracial	-9.81(*)	2.57	-12.70	-6.93

(*)Significant differences

Table 45. Spring 2013, Mean Scale Score and Standard Deviations for State and Each Proficiency Level

Content	Grade	N Count	Total		Pass	
			Mean	SD	Mean	SD
Mathematics	3	46316	739.00	88.33	781.69	62.03
	4	45383	745.43	90.14	784.75	65.46
	5	44295	740.71	86.98	782.33	61.00
	6	43222	737.09	78.86	773.36	54.68
	7	43146	732.30	80.70	772.10	54.63
	8	41377	732.09	83.25	775.66	57.78
Reading	3	45683	741.22	86.35	781.55	57.37
	4	44704	729.59	77.54	768.70	51.34
	5	43798	735.55	84.47	777.59	57.88
	6	42971	731.18	77.53	772.72	53.76
	7	43368	729.88	67.56	760.02	49.14
	8	42341	750.16	82.06	781.85	59.93
Science	5	44805	695.10	72.00	748.12	44.80
	8	44209	694.21	57.11	734.31	28.40

Note: Undetermined (invalid) students not included; Pass = Proficient + Advanced.

Table 45. Spring 2013, Mean Scale Score and Standard Deviations for State and Each Proficiency Level *(continued)*

Content	Grade	Unsatisfactory		Limited		Proficient		Advanced	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mathematics	3	575.73	52.99	671.25	19.30	748.05	27.17	855.64	52.24
	4	586.55	53.98	671.70	15.74	749.80	30.07	865.67	52.32
	5	587.72	53.88	673.82	17.10	746.54	25.97	848.78	50.83
	6	613.82	50.37	681.10	9.82	744.04	26.22	838.67	44.02
	7	623.96	56.12	687.81	6.08	746.96	28.88	848.01	42.50
	8	595.15	51.03	675.94	15.73	736.65	20.47	827.45	50.11
Reading	3	587.84	58.50	676.18	15.02	772.73	45.21	929.27	32.60
	4	601.73	52.68	679.69	12.93	757.82	36.61	886.29	39.02
	5	588.43	56.47	674.93	16.42	761.28	36.84	895.44	44.07
	6	603.73	43.21	678.81	15.06	754.85	31.89	869.32	44.29
	7	627.21	44.42	684.01	7.71	744.44	28.71	851.49	44.23
	8	604.03	48.85	677.95	11.79	759.87	35.51	881.22	45.52
Science	5	596.87	48.61	672.17	14.49	724.68	17.34	800.10	43.25
	8	611.15	50.98	677.73	11.95	720.62	14.29	771.77	23.19

Note: Undetermined (invalid) students not included.

Table 46. Spring 2012 Proficiency Level Impact Data, (% rounded)

Content	Grade	N Count	Pass	Unsatisfactory	Limited Knowledge	Proficient	Advanced
Mathematics	3	45419	69.86	10.17	19.97	45.22	24.64
	4	44146	72.94	9.89	17.17	53.21	19.73
	5	43641	70.16	10.45	19.39	42.48	27.68
	6	43415	70.76	15.58	13.67	50.28	20.48
	7	42540	69.74	17.69	12.57	51.36	18.38
	8	42076	67.99	11.22	20.79	42.34	25.65
Reading	3	44742	71.93	11.42	16.65	68.80	3.13
	4	43366	63.34	15.15	21.51	58.90	4.44
	5	43073	67.67	10.56	21.77	56.19	11.48
	6	43195	68.91	14.11	16.98	60.21	8.70
	7	42500	74.16	12.28	13.56	55.51	18.65
	8	42135	78.86	8.95	12.19	61.01	17.85
Science	5	44116	88.25	2.58	9.17	58.46	29.79
	8	43131	86.68	3.43	9.88	70.07	16.61

Note: Statistics re-calculated by CTB based on the 2012 GRT.

Table 47. Spring 2013 Proficiency Level Impact Data, (% rounded)

Content	Grade	N Count	Pass	Unsatisfactory	Limited Knowledge	Proficient	Advanced
Mathematics	3	46316	69.63	9.58	20.79	47.86	21.77
	4	45383	73.05	10.39	16.56	51.01	22.03
	5	44295	70.17	10.75	19.08	45.60	24.56
	6	43222	71.83	15.28	12.89	49.57	22.25
	7	43146	68.89	21.27	9.84	51.75	17.14
	8	41377	66.71	12.84	20.46	38.05	28.66
Reading	3	45683	71.67	11.86	16.47	67.63	4.04
	4	44704	68.42	14.11	17.47	62.63	5.80
	5	43798	68.67	11.42	19.91	60.33	8.35
	6	42971	66.08	12.90	21.01	55.77	10.31
	7	43368	71.68	15.16	13.16	61.24	10.44
	8	42341	77.10	10.68	12.22	63.13	13.97
Science	5	44805	51.00	20.99	28.01	35.15	15.85
	8	44209	52.43	19.81	27.77	38.39	14.03

Note: 2013 Science cuts are new and different from 2012 cuts; Undetermined (invalid) students not included.

Table 48. Spring 2013, Mathematics State and Subgroup Proficiency Level Impact Data

Grade	Subgroup	N Count	Pass	Unsatisfactory	Limited Knowledge	Proficient	Advanced
3	Whole State	46316	69.63	9.58	20.79	47.86	21.77
	Female	22881	68.10	9.89	22.00	47.73	20.37
	Male	23395	71.14	9.27	19.59	48.00	23.14
	Asian	950	81.58	5.16	13.26	43.47	38.11
	Black	4126	48.18	22.47	29.35	38.46	9.72
	Hispanic	7322	57.68	14.64	27.68	44.97	12.70
	Indian American	6934	69.50	8.19	22.31	48.79	20.71
	White	23794	76.48	6.34	17.18	50.05	26.44
	Multiracial	3190	70.41	9.75	19.84	49.66	20.75
	ELL	4856	49.36	19.32	31.32	40.94	8.42
	IEP	5425	51.85	20.52	27.63	39.61	12.24
	Section 504	291	60.48	13.06	26.46	44.33	16.15
	Low SES	28291	61.60	13.06	25.33	46.73	14.88
	Accommodated	4702	40.88	26.20	32.92	35.43	5.44
4	Whole State	45383	73.05	10.39	16.56	51.01	22.03
	Female	22577	72.60	10.40	17.01	51.65	20.95
	Male	22768	73.50	10.37	16.14	50.37	23.13
	Asian	850	87.29	4.24	8.47	46.71	40.59
	Black	4056	54.09	22.58	23.32	43.66	10.43
	Hispanic	6882	63.09	16.27	20.63	48.63	14.46
	Indian American	6901	71.79	10.09	18.13	53.27	18.52
	White	23658	79.09	6.85	14.06	52.43	26.65
	Multiracial	3036	72.76	10.74	16.50	51.25	21.51
	ELL	3145	49.98	25.53	24.48	42.67	7.31
	IEP	5280	53.33	21.93	24.73	43.07	10.27
	Section 504	358	65.92	14.53	19.55	49.72	16.20
	Low SES	27413	65.65	14.03	20.32	50.86	14.79
	Accommodated	4314	43.35	28.37	28.28	38.57	4.78

Note: ELL = English Language Learner; IEP = Individualized Education Program; SES = Socio-economic Status.

Table 48. Spring 2013, Mathematics State and Subgroup Proficiency Level Impact Data *(continued)*

Grade	Subgroup	N Count	Pass	Unsatisfactory	Limited Knowledge	Proficient	Advanced
5	Whole State	44295	70.17	10.75	19.08	45.60	24.56
	Female	22105	70.25	10.42	19.33	45.89	24.35
	Male	22162	70.11	11.05	18.83	45.33	24.79
	Asian	857	83.90	4.20	11.90	32.91	50.99
	African American	4030	52.41	22.16	25.43	39.50	12.90
	Hispanic	6472	61.57	15.22	23.21	43.63	17.94
	Indian American	7075	65.98	12.34	21.68	46.50	19.48
	White	22950	76.52	7.23	16.24	47.38	29.14
	Multiracial	2911	69.91	10.82	19.27	45.96	23.94
	ELL	2570	42.37	27.55	30.08	33.15	9.22
	IEP	4904	46.39	25.16	28.45	36.34	10.05
	Section 504	372	60.75	15.86	23.39	43.55	17.20
	Low SES	26497	62.32	14.47	23.20	45.21	17.11
	Accommodated	4057	37.66	29.97	32.36	31.75	5.92
6	Whole State	43222	71.83	15.28	12.89	49.57	22.25
	Female	21425	70.80	15.59	13.61	50.10	20.70
	Male	21791	72.85	14.96	12.18	49.07	23.78
	Asian	875	86.74	6.74	6.51	41.14	45.60
	African American	3754	52.08	29.75	18.17	42.59	9.48
	Hispanic	5999	62.23	22.00	15.77	47.86	14.37
	Indian American	6964	69.49	16.76	13.76	51.77	17.72
	White	22691	78.04	10.81	11.14	50.97	27.08
	Multiracial	2939	69.79	16.57	13.64	48.55	21.23
	ELL	1896	41.46	41.30	17.25	35.44	6.01
	IEP	4275	41.92	39.30	18.78	35.04	6.88
	Section 504	397	67.25	16.62	16.12	49.62	17.63
	Low SES	24785	63.37	20.74	15.89	48.94	14.42
	Accommodated	2900	35.66	45.93	18.41	31.66	4.00

Note: ELL = English Language Learner; IEP = Individualized Education Program; SES = Socio-economic Status.

Table 48. Spring 2013, Mathematics State and Subgroup Proficiency Level Impact Data *(continued)*

Grade	Subgroup	N Count	Pass	Unsatisfactory	Limited Knowledge	Proficient	Advanced
7	Whole State	43146	68.89	21.27	9.84	51.75	17.14
	Female	21497	69.18	20.47	10.35	53.90	15.28
	Male	21641	68.62	22.07	9.32	49.62	18.99
	Asian	813	88.07	7.13	4.80	48.46	39.61
	African American	3868	49.12	38.24	12.64	42.71	6.41
	Hispanic	5695	58.12	29.41	12.47	48.69	9.43
	Indian American	7099	67.32	22.21	10.47	53.80	13.52
	White	22762	75.27	16.13	8.60	53.69	21.59
	Multiracial	2909	64.76	24.68	10.55	50.50	14.27
	ELL	1889	37.32	49.81	12.86	32.77	4.55
	IEP	3953	36.73	51.45	11.81	31.87	4.86
	Section 504	401	57.11	29.43	13.47	43.64	13.47
	Low SES	24091	60.03	28.30	11.68	49.97	10.05
	Accommodated	2644	31.32	56.05	12.63	28.21	3.10
8	Whole State	41377	66.71	12.84	20.46	38.05	28.66
	Female	20755	67.21	12.46	20.32	39.04	28.17
	Male	20615	66.20	13.21	20.59	37.06	29.15
	Asian	820	87.20	4.39	8.41	30.61	56.59
	African American	3675	50.67	24.00	25.33	34.29	16.38
	Hispanic	5245	55.42	19.47	25.11	35.71	19.71
	Indian American	6747	62.93	13.84	23.23	38.40	24.53
	White	22184	72.71	9.30	17.99	39.48	33.23
	Multiracial	2706	64.34	13.90	21.77	37.32	27.01
	ELL	1785	32.77	37.59	29.64	24.82	7.96
	IEP	3258	33.12	37.54	29.34	24.31	8.81
	Section 504	382	59.16	18.59	22.25	36.91	22.25
	Low SES	22074	57.04	18.13	24.83	37.23	19.81
	Accommodated	2070	27.05	42.75	30.19	21.30	5.75

Note: ELL = English Language Learner; IEP = Individualized Education Program; SES = Socio-economic Status.

Table 49. Spring 2013, Reading State and Subgroup Proficiency Level Impact Data

Grade	Subgroup	N Count	Pass	Unsatisfactory	Limited Knowledge	Proficient	Advanced
3	Whole State	45683	71.67	11.86	16.47	67.63	4.04
	Female	22691	74.72	9.84	15.45	70.24	4.48
	Male	22967	68.69	13.85	17.46	65.08	3.61
	Asian	923	80.17	8.56	11.27	70.21	9.97
	African American	4061	55.45	23.02	21.52	53.80	1.65
	Hispanic	7177	57.61	19.88	22.50	56.03	1.59
	Indian American	6846	70.83	11.71	17.46	67.84	2.99
	White	23524	78.47	7.69	13.84	73.21	5.26
	Multiracial	3152	73.16	11.64	15.20	69.07	4.09
	ELL	4665	45.89	27.46	26.65	45.17	0.73
	IEP	4889	47.72	31.13	21.15	45.94	1.78
	Section 504	294	62.93	15.65	21.43	61.56	1.36
	Low SES	27847	63.76	16.13	20.11	61.58	2.18
	Accommodated	3620	32.57	42.85	24.59	32.18	0.39
4	Whole State	44704	68.42	14.11	17.47	62.63	5.80
	Female	22405	71.52	11.88	16.60	65.01	6.50
	Male	22269	65.32	16.32	18.36	60.23	5.09
	Asian	832	77.16	9.25	13.58	65.38	11.78
	African American	3995	50.89	25.63	23.48	49.24	1.65
	Hispanic	6744	55.71	22.03	22.26	53.11	2.59
	Indian American	6786	66.09	14.40	19.51	61.23	4.86
	White	23377	75.34	10.00	14.66	67.78	7.56
	Multiracial	2970	69.29	13.64	17.07	64.07	5.22
	ELL	2950	36.03	37.97	26.00	35.46	0.58
	IEP	4695	40.49	37.06	22.45	38.23	2.26
	Section 504	361	62.33	15.79	21.88	57.89	4.43
	Low SES	26892	59.80	19.01	21.20	56.91	2.89
	Accommodated	3381	26.50	48.92	24.58	26.09	0.41

Note: ELL = English Language Learner; IEP = Individualized Education Program; SES = Socio-economic Status.

Table 49. Spring 2013, Reading State and Subgroup Proficiency Level Impact Data *(continued)*

Grade	Subgroup	N Count	Pass	Unsatisfactory	Limited Knowledge	Proficient	Advanced
5	Whole State	43798	68.67	11.42	19.91	60.33	8.35
	Female	21983	71.87	9.50	18.63	62.43	9.45
	Male	21790	65.47	13.34	21.19	58.23	7.24
	Asian	839	77.00	8.94	14.06	59.12	17.88
	African American	3980	50.00	22.91	27.09	46.68	3.32
	Hispanic	6345	53.70	18.19	28.12	49.82	3.88
	Indian American	6995	65.52	12.59	21.89	59.31	6.20
	White	22743	76.70	7.28	16.02	65.95	10.75
	Multiracial	2896	69.30	11.22	19.48	60.77	8.53
	ELL	2391	26.98	38.85	34.17	25.89	1.09
	IEP	4463	38.49	34.62	26.89	35.76	2.73
	Section 504	371	59.57	16.17	24.26	53.10	6.47
	Low SES	26092	59.50	15.85	24.65	54.86	4.64
	Accommodated	3350	25.34	44.63	30.03	24.27	1.07
6	Whole State	42971	66.08	12.90	21.01	55.77	10.31
	Female	21436	68.60	11.09	20.31	57.19	11.41
	Male	21530	63.59	14.70	21.71	54.36	9.22
	Asian	844	77.37	8.06	14.57	56.40	20.97
	African American	3754	48.69	24.27	27.04	44.59	4.10
	Hispanic	5881	51.93	20.29	27.78	46.59	5.34
	Indian American	6989	63.69	14.25	22.06	55.60	8.08
	White	22554	73.19	8.84	17.97	60.22	12.96
	Multiracial	2949	64.56	12.95	22.48	54.46	10.11
	ELL	1744	23.05	46.10	30.85	21.56	1.49
	IEP	4091	32.78	40.82	26.40	29.70	3.08
	Section 504	398	62.06	15.33	22.61	53.27	8.79
	Low SES	24669	56.33	17.86	25.81	50.41	5.92
	Accommodated	2392	23.41	49.41	27.17	21.74	1.67

Note: ELL = English Language Learner; IEP = Individualized Education Program; SES = Socio-economic Status.

Table 49. Spring 2013, Reading State and Subgroup Proficiency Level Impact Data *(continued)*

Grade	Subgroup	N Count	Pass	Unsatisfactory	Limited Knowledge	Proficient	Advanced
7	Whole State	43368	71.68	15.16	13.16	61.24	10.44
	Female	21678	75.05	12.70	12.24	62.84	12.22
	Male	21687	68.31	17.62	14.07	59.65	8.66
	Asian	814	80.22	10.07	9.71	58.97	21.25
	African American	3957	53.70	29.24	17.06	50.09	3.61
	Hispanic	5668	58.36	24.93	16.71	53.18	5.19
	Indian American	7140	70.87	14.85	14.29	61.83	9.03
	White	22874	78.31	10.35	11.34	65.13	13.18
	Multiracial	2915	69.57	17.02	13.41	60.79	8.78
	ELL	1792	28.57	50.89	20.54	27.46	1.12
	IEP	3965	39.52	43.08	17.40	37.10	2.42
	Section 504	412	59.71	21.36	18.93	51.46	8.25
	Low SES	24278	63.50	20.56	15.94	57.23	6.26
	Accommodated	2229	30.51	51.59	17.90	29.34	1.17
8	Whole State	42341	77.10	10.68	12.22	63.13	13.97
	Female	21238	80.07	8.79	11.14	63.44	16.64
	Male	21098	74.11	12.57	13.32	62.83	11.28
	Asian	810	84.69	8.52	6.79	58.02	26.67
	African American	3830	60.05	21.64	18.30	52.92	7.13
	Hispanic	5326	64.96	19.10	15.94	55.90	9.07
	Indian American	6943	75.69	10.92	13.39	64.05	11.64
	White	22671	82.93	6.83	10.24	66.27	16.66
	Multiracial	2761	77.51	10.90	11.59	64.61	12.89
	ELL	1705	35.54	42.82	21.64	33.78	1.76
	IEP	3459	42.79	37.87	19.34	40.30	2.49
	Section 504	391	74.17	12.79	13.04	64.45	9.72
	Low SES	22681	68.62	15.60	15.78	60.15	8.47
	Accommodated	1747	34.46	45.11	20.44	33.31	1.14

Note: ELL = English Language Learner; IEP = Individualized Education Program; SES = Socio-economic Status.

Table 50. Spring 2013, Science State and Subgroup Proficiency Level Impact Data

Grade	Subgroup	N Count	Pass	Unsatisfactory	Limited Knowledge	Proficient	Advanced
5	Whole State	44805	51.00	20.99	28.01	35.15	15.85
	Female	22365	48.79	21.71	29.50	35.01	13.78
	Male	22409	53.22	20.27	26.51	35.29	17.93
	Asian	859	63.91	14.32	21.77	38.53	25.38
	African American	4132	27.13	40.92	31.95	21.81	5.32
	Hispanic	6546	35.96	31.12	32.92	27.45	8.51
	Indian American	7154	48.04	21.90	30.05	34.76	13.28
	White	23172	59.90	14.58	25.52	39.56	20.34
	Multiracial	2942	51.33	20.67	28.01	36.30	15.02
	ELL	2596	17.72	51.35	30.93	14.06	3.66
	IEP	5438	28.30	43.05	28.65	21.79	6.51
	Section 504	374	43.05	25.94	31.02	30.75	12.30
	Low SES	26904	40.93	27.94	31.13	30.82	10.11
	Accommodated	4444	19.33	52.25	28.42	15.80	3.53
8	Whole State	44209	52.43	19.81	27.77	38.39	14.03
	Female	21997	51.00	19.08	29.92	38.46	12.54
	Male	22171	53.86	20.51	25.62	38.33	15.53
	Asian	836	71.29	9.09	19.62	43.42	27.87
	African American	4124	30.29	37.49	32.23	25.00	5.29
	Hispanic	5712	36.99	30.81	32.20	30.02	6.97
	Indian American	7239	48.67	20.06	31.28	38.32	10.35
	White	23596	60.50	14.37	25.13	42.60	17.90
	Multiracial	2702	52.59	19.65	27.76	38.42	14.17
	ELL	2028	15.43	53.94	30.62	13.81	1.63
	IEP	4613	21.31	48.69	30.00	17.78	3.53
	Section 504	397	49.62	22.17	28.21	35.01	14.61
	Low SES	24011	40.98	26.89	32.13	32.82	8.15
	Accommodated	2807	15.71	55.75	28.54	13.82	1.89

Note: ELL = English Language Learner; IEP = Individualized Education Program; SES = Socio-economic Status.

Table 51. Spring 2013 Operational Test Parameters Correlations

		a-parameter					*b*-parameter					*c*-parameter				
Content	Grade	Corr	RMSD	Mean Diff	SD Ratio	Rdiff	Corr	RMSD	Mean Diff	SD Ratio	Rdiff	Corr	RMSD	Mean Diff	SD Ratio	Rdiff
Mathematics	3	0.91	0.002	0.001	1.100	0.00	0.99	10.877	4.840	0.976	-0.24	0.67	0.078	0.032	0.725	-0.69
	4	0.95	0.002	0.001	0.814	-0.66	0.96	17.446	8.523	1.024	-0.05	0.58	0.080	0.041	0.611	-0.79
	5	0.90	0.002	0.001	1.334	0.33	0.94	20.268	2.290	0.864	-0.54	0.74	0.095	0.023	0.711	-0.70
	6	0.94	0.001	0.000	0.966	-0.27	0.94	24.213	3.009	0.758	-0.75	0.47	0.083	0.009	1.073	-0.46
	7	0.88	0.002	-0.001	1.100	-0.06	0.89	33.862	-3.538	0.985	-0.26	0.53	0.106	-0.019	0.786	-0.66
	8	0.89	0.002	-0.001	0.869	-0.50	0.86	22.618	-9.487	0.959	-0.34	0.78	0.104	-0.048	0.695	-0.72
Reading	3	0.94	0.001	0.000	0.958	-0.29	0.96	18.015	1.837	0.894	-0.50	0.72	0.079	0.019	0.782	-0.63
	4	0.95	0.001	0.000	0.878	-0.52	0.95	16.537	3.765	0.795	-0.70	-0.11	0.093	0.029	0.427	-0.93
	5	0.95	0.001	0.000	1.024	-0.09	0.94	19.291	2.608	0.979	-0.23	0.23	0.110	0.008	0.384	-0.93
	6	0.94	0.001	0.000	1.006	-0.15	0.92	21.510	4.006	1.050	-0.08	0.55	0.088	0.027	0.873	-0.58
	7	0.99	0.001	0.000	0.877	-0.71	0.93	20.147	8.534	0.971	-0.26	0.59	0.106	0.046	0.557	-0.83
	8	0.95	0.001	0.000	0.894	-0.47	0.92	23.288	8.487	0.828	-0.59	-0.22	0.103	0.038	0.782	-0.84

Table 52. Scale Score Statistics for Operational Test in Spring 2012 and Spring 2013

		Spring 2012					Spring 2013				
Content	Grade	N-Count	Alpha	Mean	SD	SEM	N-Count	Alpha	Mean	SD	SEM
Mathematics	3	45419	0.90	741.04	88.37	2.61	46316	0.91	739.00	88.33	2.62
	4	44146	0.90	745.79	87.78	2.66	45383	0.90	745.43	90.14	2.69
	5	43641	0.89	742.14	85.78	2.77	44295	0.89	740.71	86.98	2.81
	6	43415	0.90	734.25	79.51	2.95	43221	0.90	737.09	78.86	2.94
	7	42540	0.89	735.15	80.15	3.01	43146	0.89	732.30	80.70	3.02
	8	42076	0.90	726.02	83.07	3.03	41377	0.90	732.09	83.25	3.00
Reading	3	44742	0.90	742.68	82.26	2.82	45683	0.90	741.22	86.35	2.81
	4	43366	0.88	724.38	73.38	2.75	44704	0.89	729.59	77.54	2.85
	5	43073	0.89	734.25	79.60	2.69	43798	0.90	735.55	84.47	2.64
	6	43195	0.89	730.75	78.68	2.85	42971	0.89	731.18	77.53	2.78
	7	42500	0.87	738.85	70.35	2.70	43368	0.88	729.88	67.56	2.66
	8	42135	0.86	757.77	81.81	2.73	42341	0.88	750.16	82.06	2.72
Science	5	.	.	.	.	.	44805	0.86	695.10	72.00	2.84
	8	.	.	.	.	.	44209	0.85	694.21	57.11	2.98

Note: Spring 2012 statistics re-calculated by CTB based on the 2012 GRT; Spring 2013 Census Data. Suppressed items not included in data; Spring 2012 Science was not included to avoid scale confusion between Spring 2012 and Spring 2013.

Table 53. Spring 2013, Proficiency Levels Cut Scores and Scale Bounds

Content	Grade	LOSS	Cut 1 Perf. Level 2	Cut 2 Perf. Level 3	Cut 3 Perf. Level 4	HOSS
Mathematics	3	400	633	700	798	990
	4	400	639	700	805	990
	5	400	638	700	791	990
	6	400	664	700	795	990
	7	400	674	700	800	990
	8	400	642	700	774	990
Reading	3	400	649	700	891	990
	4	400	658	700	845	990
	5	400	641	700	830	990
	6	400	647	700	828	990
	7	400	668	700	802	990
	8	400	655	700	833	990
Science	5	400	648	700	765	990
	8	400	658	700	751	990

Note: LOSS = Lowest Obtainable Scale Score; HOSS = Highest Obtainable Scale Score;
Science Cuts are new and different from 2012

Table 54. Spring 2013, Proportion of Students Within Each Performance Level

Content	Grade	N-Count	Unsatisfactory	Limited Knowledge	Proficient	Advanced
Mathematics	3	46316	9.58	20.79	47.86	21.77
	4	45383	10.39	16.56	51.01	22.03
	5	44295	10.75	19.08	45.60	24.56
	6	43222	15.28	12.89	49.57	22.25
	7	43146	21.27	9.84	51.75	17.14
	8	41377	12.84	20.46	38.05	28.66
Reading	3	45683	11.86	16.47	67.63	4.04
	4	44704	14.11	17.47	62.63	5.80
	5	43798	11.42	19.91	60.33	8.35
	6	42971	12.90	21.01	55.77	10.31
	7	43368	15.16	13.16	61.24	10.44
	8	42341	10.68	12.22	63.13	13.97
Science	5	44805	20.99	28.01	35.15	15.85
	8	44209	19.81	27.77	38.39	14.03

Note: Census Data; Suppressed items are not included in data.

Table 55. Differences in Overall Pass Rate for Spring 2012 and Spring 2013

Content	Grade	Pass Rates (%, rounded) 2012	2013	Differences in Pass Rates 2013-2012
Mathematics	3	69.86	69.63	-0.23
	4	72.94	73.05	0.11
	5	70.16	70.17	0.01
	6	70.76	71.83	1.07
	7	69.74	68.89	-0.85
	8	67.99	66.71	-1.28
Reading	3	71.93	71.67	-0.26
	4	63.34	68.42	5.08
	5	67.67	68.67	1.00
	6	68.91	66.08	-2.83
	7	74.16	71.68	-2.48
	8	78.86	77.10	-1.76

Note: Spring 2013 Science Cut Scores are new and different than Spring 2012.

Table 56. Spring 2013, Summary of Range of P-Values and Item-Test Correlations Statistics for Operational and Field Test, by Item Type

Content	Grade	Item Type	Mean *P*-Values						Mean Item-Test Correlations					
			Operational Items			Field Test Items			Operational Items			Field Test Items		
			Low	Mean	High	Low	Mean	High	Low	Mean	High	Low	Mean	High
Mathematics	3	MC	0.38	0.76	0.93	0.09	0.62	0.96	0.25	0.44	0.60	0.03	0.31	0.57
	4	MC	0.51	0.76	0.93	0.22	0.59	0.96	0.26	0.42	0.57	0.06	0.32	0.51
	5	MC	0.41	0.73	0.91	0.17	0.55	0.89	0.23	0.40	0.57	0.02	0.32	0.50
	6	MC	0.36	0.66	0.89	0.09	0.51	0.91	0.20	0.41	0.58	0.03	0.30	0.48
	7	MC	0.24	0.62	0.88	0.08	0.46	0.85	0.22	0.39	0.55	0.00	0.28	0.50
	8	MC	0.42	0.65	0.88	0.19	0.45	0.76	0.28	0.42	0.55	0.02	0.29	0.57
Reading	3	CR	.	.	.	0.13	0.21	0.25	.	.	.	0.23	0.33	0.42
		MC	0.41	0.72	0.92	0.26	0.61	0.91	0.18	0.42	0.60	0.10	0.31	0.49
	4	CR	.	.	.	0.13	0.19	0.24	.	.	.	0.28	0.33	0.39
		MC	0.43	0.72	0.92	0.27	0.62	0.94	0.24	0.40	0.54	0.09	0.31	0.49
	5	CR	.	.	.	0.18	0.35	0.57	.	.	.	0.20	0.34	0.43
		MC	0.54	0.77	0.96	0.34	0.65	0.92	0.25	0.42	0.55	0.09	0.33	0.48
	6	CR	.	.	.	0.10	0.39	0.55	.	.	.	0.22	0.32	0.48
		MC	0.43	0.73	0.96	0.27	0.64	0.96	0.27	0.40	0.52	0.08	0.31	0.49
	7	CR	.	.	.	0.15	0.30	0.47	.	.	.	0.21	0.31	0.47
		MC	0.51	0.77	0.97	0.10	0.60	0.95	0.17	0.40	0.55	0.01	0.28	0.47
	8	CR	.	.	.	0.22	0.39	0.56	.	.	.	0.28	0.33	0.41
		MC	0.47	0.76	0.97	0.38	0.66	0.94	0.24	0.39	0.50	0.01	0.26	0.44
Science	5	MC	0.49	0.68	0.97	0.20	0.55	0.95	0.11	0.37	0.50	0.03	0.26	0.44
	8	MC	0.23	0.57	0.87	0.18	0.47	0.94	0.20	0.36	0.51	0.01	0.22	0.45
Writing	5	CR	0.51	0.55	0.59	0.39	0.53	0.62	.	.	.	.	.	.
	8	CR	0.55	0.56	0.56	0.51	0.57	0.60	.	.	.	.	.	.

Table 57. Spring 2012 and Spring 2013 Test Reliability Data

Content	Grade	Coefficient Alpha 2012	2013
Mathematics	3	0.90	0.91
	4	0.90	0.90
	5	0.89	0.89
	6	0.90	0.90
	7	0.89	0.89
	8	0.90	0.90
Reading	3	0.90	0.90
	4	0.88	0.89
	5	0.89	0.90
	6	0.89	0.89
	7	0.87	0.88
	8	0.86	0.88
Science	5	0.87	0.86
	8	0.86	0.85

Note: Spring 2012 Data calculated from the 2012 GRT.

Table 58. Mathematics, Raw Score to Scale Score Conversions & Standard Error of Measurement

Raw	Grade 3		Grade 4		Grade 5	
Score	Scale Score	SEM	Scale Score	SEM	Scale Score	SEM
0	400	165	400	172	400	185
1	400	165	400	172	400	185
2	400	165	400	172	400	185
3	400	165	400	172	400	185
4	400	165	400	172	400	185
5	400	165	400	172	400	185
6	400	165	400	172	400	185
7	400	165	400	172	400	185
8	400	165	400	172	400	185
9	400	165	400	172	400	185
10	400	165	400	172	400	185
11	442	124	400	172	400	185
12	479	87	433	139	427	159
13	504	65	473	99	474	111
14	522	53	501	75	505	81
15	538	45	523	62	527	64
16	551	40	541	52	546	54
17	563	36	556	46	561	47
18	574	34	570	41	575	42
19	584	32	582	37	587	39
20	593	30	592	34	598	36
21	602	28	602	32	609	33
22	611	27	612	30	618	31
23	619	26	620	28	627	30
24	627	26	629	27	636	28
25	**635**	**25**	637	26	**644**	**27**
26	642	24	**644**	**25**	652	26
27	649	24	652	24	660	26
28	657	23	659	23	668	25
29	664	22	666	23	675	24
30	671	22	673	22	682	24
31	677	21	679	22	690	23
32	684	21	686	22	697	23
33	691	20	693	21	**704**	**23**
34	697	20	**700**	**21**	711	23
35	**704**	**20**	707	21	718	23
36	711	19	714	21	726	23
37	718	20	721	22	733	23
38	725	20	728	22	741	23
39	732	20	736	22	750	23
40	740	21	744	23	758	24
41	748	22	753	24	767	24
42	757	23	762	25	777	25
43	767	24	773	26	788	27
44	779	26	784	28	**800**	**29**
45	792	29	798	30	815	32
46	**808**	**32**	**815**	**34**	832	36
47	828	37	836	40	855	43
48	856	46	866	52	889	56
49	903	64	921	80	949	86
50	990	126	990	131	990	112

Note: SEM = Standard Error of Measurement; BOLD = Scale Score at or closest to cut scores.

Table 58. Mathematics, Raw Score to Scale Score Conversions & Standard Error of Measurement

Raw	Grade 6		Grade 7		Grade 8	
Score	Scale Score	SEM	Scale Score	SEM	Scale Score	SEM
0	400	213	400	227	400	208
1	400	213	400	227	400	208
2	400	213	400	227	400	208
3	400	213	400	227	400	208
4	400	213	400	227	400	208
5	400	213	400	227	400	208
6	400	213	400	227	400	208
7	400	213	400	227	400	208
8	400	213	400	227	400	208
9	400	213	400	227	400	208
10	467	146	400	227	462	146
11	513	100	473	153	508	99
12	543	73	521	106	537	71
13	565	58	551	75	558	56
14	583	49	573	59	575	47
15	598	42	591	49	589	41
16	611	37	605	42	602	37
17	622	34	618	37	613	34
18	632	31	629	34	623	31
19	642	29	639	31	632	29
20	651	27	648	29	641	28
21	659	26	657	28	**649**	**26**
22	**666**	**24**	665	27	657	25
23	674	23	673	26	664	24
24	681	22	**680**	**25**	671	24
25	687	22	688	24	678	23
26	694	21	695	24	685	22
27	**700**	**21**	**702**	**23**	691	22
28	707	20	709	23	698	21
29	713	20	716	22	**704**	**21**
30	719	20	723	22	710	21
31	725	19	730	22	716	20
32	731	19	737	22	723	20
33	737	19	744	22	729	20
34	744	19	752	22	735	20
35	750	20	759	22	742	20
36	757	20	766	22	748	20
37	764	20	774	22	755	21
38	771	21	782	22	762	21
39	779	21	790	22	769	21
40	787	22	798	22	**777**	**22**
41	**796**	**23**	**807**	**23**	786	23
42	806	24	817	24	795	24
43	817	26	827	25	805	25
44	830	28	838	26	816	27
45	845	32	852	29	830	30
46	865	37	867	32	846	33
47	892	45	887	38	866	39
48	939	65	916	48	895	48
49	990	98	968	75	944	70
50	.	.	990	90	990	98

Note: SEM = Standard Error of Measurement; BOLD = Scale Score at or closest to cut scores.

Table 59. Reading, Raw Score to Scale Score Conversions & Standard Error of Measurement

Raw	Grade 3		Grade 4		Grade 5	
Score	Scale Score	SEM	Scale Score	SEM	Scale Score	SEM
0	400	200	400	182	400	182
1	400	200	400	182	400	182
2	400	200	400	182	400	182
3	400	200	400	182	400	182
4	400	200	400	182	400	182
5	400	200	400	182	400	182
6	400	200	400	182	400	182
7	400	200	400	182	400	182
8	400	200	400	182	400	182
9	400	200	400	182	400	182
10	400	200	440	142	400	182
11	450	150	485	97	400	182
12	498	102	513	69	443	139
13	527	73	533	55	487	95
14	548	57	549	45	513	69
15	564	47	562	39	532	54
16	578	41	574	35	548	45
17	590	36	584	32	560	39
18	601	33	593	30	571	34
19	610	30	602	28	581	31
20	619	28	610	26	590	29
21	627	27	618	25	598	27
22	635	26	625	24	606	25
23	643	25	632	23	613	24
24	**650**	**24**	638	23	620	23
25	657	23	645	22	626	22
26	663	23	651	22	633	21
27	670	22	**658**	**22**	639	21
28	677	22	664	21	**645**	**21**
29	683	22	671	21	651	20
30	690	22	677	21	658	20
31	696	22	683	21	664	20
32	**703**	**22**	690	21	670	21
33	710	22	697	21	676	21
34	717	22	**703**	**22**	683	21
35	724	22	710	22	690	21
36	732	23	718	22	697	22
37	740	23	725	23	**705**	**22**
38	748	24	733	23	712	23
39	757	24	742	24	721	24
40	766	25	751	25	730	24
41	776	26	760	26	739	25
42	787	27	771	28	749	27
43	799	29	783	30	761	28
44	812	31	797	32	774	30
45	828	33	813	35	788	32
46	846	37	832	39	806	36
47	870	43	**856**	**45**	828	42
48	**903**	**54**	890	57	**860**	**53**
49	961	79	951	86	915	80
50	990	96	990	111	990	136

Note: SEM = Standard Error of Measurement; BOLD = Scale Score at or closest to cut scores.

Table 59. Reading, Raw Score to Scale Score Conversions & Standard Error of Measurement

Raw	Grade 6		Grade 7		Grade 8	
Score	Scale Score	SEM	Scale Score	SEM	Scale Score	SEM
0	400	165	400	192	400	159
1	400	165	400	192	400	159
2	400	165	400	192	400	159
3	400	165	400	192	400	159
4	400	165	400	192	400	159
5	400	165	400	192	400	159
6	400	165	400	192	400	159
7	400	165	400	192	400	159
8	400	165	400	192	400	159
9	436	129	400	192	400	159
10	478	87	400	192	447	112
11	504	63	480	112	480	78
12	523	50	515	77	504	61
13	538	42	537	55	522	50
14	551	37	553	43	537	43
15	562	33	565	36	551	39
16	572	31	576	31	562	35
17	581	28	585	28	573	32
18	590	27	593	26	582	30
19	598	26	601	24	591	29
20	605	24	608	23	600	28
21	613	24	614	22	608	26
22	619	23	621	21	615	26
23	626	22	627	20	623	25
24	633	22	633	20	630	24
25	639	22	639	19	637	24
26	646	21	644	19	644	24
27	**652**	**21**	650	18	651	23
28	659	21	656	18	**658**	**23**
29	665	21	661	18	665	23
30	672	21	666	18	672	23
31	678	21	**672**	**17**	679	23
32	685	22	677	17	686	23
33	692	22	683	18	693	23
34	699	22	688	18	**701**	**24**
35	**706**	**22**	694	18	709	24
36	714	23	**700**	**18**	717	24
37	722	23	707	19	725	25
38	730	24	713	19	734	26
39	739	25	720	20	743	27
40	748	26	728	21	753	28
41	758	27	736	22	764	29
42	769	28	745	23	775	30
43	782	30	755	25	788	33
44	796	33	766	27	803	35
45	813	37	780	30	821	39
46	**833**	**42**	797	34	**842**	**44**
47	860	50	**818**	**40**	870	52
48	899	64	849	50	909	65
49	966	94	902	74	977	97
50	990	107	990	142	990	104

Note: SEM = Standard Error of Measurement; BOLD = Scale Score at or closest to cut scores.

Table 60. Science, Raw Score to Scale Score Conversions & Standard Error of Measurement

Raw	Grade 5		Grade 8	
Score	Scale Score	SEM	Scale Score	SEM
0	400	175	400	233
1	400	175	400	233
2	400	175	400	233
3	400	175	400	233
4	400	175	400	233
5	400	175	400	233
6	400	175	400	233
7	400	175	400	233
8	400	175	400	233
9	400	175	477	157
10	400	175	537	96
11	463	112	567	67
12	497	79	587	50
13	520	60	602	40
14	538	49	614	35
15	553	43	625	30
16	566	38	635	27
17	578	35	643	25
18	588	32	651	23
19	598	30	**658**	**22**
20	608	29	665	21
21	616	27	671	20
22	625	26	677	19
23	633	25	683	18
24	641	25	689	18
25	**648**	**24**	694	17
26	656	23	**700**	**17**
27	663	23	705	17
28	670	22	710	16
29	678	22	716	16
30	685	22	721	16
31	692	22	727	16
32	**700**	**22**	733	16
33	707	22	739	17
34	715	22	745	17
35	724	23	**751**	**17**
36	732	24	758	18
37	742	25	765	19
38	753	26	774	20
39	**765**	**28**	783	21
40	779	32	794	23
41	796	36	807	27
42	819	44	824	32
43	853	59	849	42
44	919	99	894	66
45	990	163	990	148

Note: SEM = Standard Error of Measurement; BOLD = Scale Score at or closest to cut scores.

Table 61. Spring 2013, Total Group Factor Analysis Results: Eigenvalues

Content	Grade	KMO Statistic	Initial Eigenvalue		Ratio 1st to 2nd Eigenvalue
			Total	% Variance	
Mathematics	3	0.97	12.46	86%	7.88
	4	0.97	10.65	92%	12.90
	5	0.97	9.26	97%	14.43
	6	0.97	10.21	97%	13.82
	7	0.96	9.15	90%	8.57
	8	0.97	10.30	92%	11.92
Reading	3	0.97	10.92	91%	7.86
	4	0.97	9.11	99%	11.73
	5	0.97	10.84	96%	10.79
	6	0.97	9.00	101%	14.21
	7	0.97	9.26	100%	11.44
	8	0.97	8.14	103%	15.99
Science	5	0.96	6.66	104%	14.35
	8	0.96	6.28	106%	12.39

Note: KMO = Kaiser's Measure of Sampling Adequacy

Table 62. Spring 2013, Mathematics Subgroup Factor Analysis Results: Eigenvalues

Grade	Subgroup	KMO Statistic	Initial Eigenvalue		Ratio 1st to 2nd Eigenvalue
			Total	% Variance	
3	Total Accommodated	0.95	11.00	81%	7.03
	ELL	0.95	11.63	81%	7.05
	Free Lunch	0.96	12.04	85%	7.52
	IEP	0.97	13.38	85%	8.38
4	Total Accommodated	0.95	9.10	86%	8.02
	ELL	0.96	10.66	87%	11.09
	Free Lunch	0.96	10.10	91%	12.13
	IEP	0.96	10.76	88%	9.25
5	Total Accommodated	0.94	7.67	88%	9.06
	ELL	0.94	8.75	86%	10.51
	Free Lunch	0.96	8.64	96%	13.17
	IEP	0.96	8.98	91%	10.23
6	Total Accommodated	0.94	7.54	88%	9.05
	ELL	0.94	8.87	86%	9.41
	Free Lunch	0.97	9.19	96%	11.92
	IEP	0.96	8.90	92%	11.08
7	Total Accommodated	0.92	6.77	84%	7.09
	ELL	0.91	7.26	79%	7.00
	Free Lunch	0.95	7.96	90%	7.90
	IEP	0.94	7.61	89%	8.03
8	Total Accommodated	0.92	7.10	82%	7.14
	ELL	0.93	8.10	82%	9.39
	Free Lunch	0.96	9.41	92%	11.37
	IEP	0.95	8.38	88%	8.95

Note: KMO = Kaiser's Measure of Sampling Adequacy; ELL = English Language Learners; IEP = Individualized Education Program.

Table 63. Spring 2013, Reading Subgroup Factor Analysis Results: Eigenvalues

Grade	Subgroup	KMO Statistic	Initial Eigenvalue		Ratio 1st to 2nd Eigenvalue
			Total	% Variance	
3	Total Accommodated	0.96	9.95	90%	11.90
	ELL	0.96	9.47	87%	7.88
	Free Lunch	0.97	10.47	91%	8.18
	IEP	0.97	13.12	92%	14.33
4	Total Accommodated	0.96	8.84	93%	12.75
	ELL	0.96	8.74	93%	11.52
	Free Lunch	0.97	8.83	99%	11.47
	IEP	0.97	11.24	95%	14.08
5	Total Accommodated	0.96	10.25	91%	13.32
	ELL	0.95	9.65	89%	10.28
	Free Lunch	0.97	10.57	95%	11.12
	IEP	0.97	12.79	93%	14.87
6	Total Accommodated	0.95	9.20	90%	13.49
	ELL	0.95	8.64	88%	11.31
	Free Lunch	0.97	8.72	100%	14.23
	IEP	0.97	10.56	95%	15.00
7	Total Accommodated	0.96	10.18	89%	11.14
	ELL	0.94	8.63	84%	7.91
	Free Lunch	0.97	9.13	99%	11.27
	IEP	0.97	10.86	94%	12.18
8	Total Accommodated	0.95	8.95	88%	12.59
	ELL	0.95	8.51	89%	13.21
	Free Lunch	0.97	8.21	102%	16.42
	IEP	0.97	9.55	95%	16.44

Note: KMO = Kaiser's Measure of Sampling Adequacy; ELL = English Language Learners; IEP = Individualized Education Program.

Table 64. Spring 2013, Science Subgroup Factor Analysis Results: Eigenvalues

Grade	Subgroup	KMO Statistic	Initial Eigenvalue		Ratio 1st to 2nd Eigenvalue
			Total	% Variance	
5	Total Accommodated	0.93	5.48	98%	10.79
	ELL	0.91	5.26	91%	10.04
	Free Lunch	0.96	6.18	104%	13.34
	IEP	0.95	6.80	100%	12.69
8	Total Accommodated	0.87	3.88	91%	8.38
	ELL	0.85	3.76	85%	6.42
	Free Lunch	0.95	5.54	106%	11.39
	IEP	0.92	4.80	97%	9.34

Note: KMO = Kaiser's Measure of Sampling Adequacy; ELL = English Language Learners; IEP = Individualized Education Program.

Table 65. Spring 2013, Proficiency Level Cut Scores and Standard Error of Measurement (SEM)

Content	Grade	Cut 1		Cut 2		Cut 3	
		Cut Score	SEM at Cut	Cut Score	SEM at Cut	Cut Score	SEM at Cut
Mathematics	3	633	25	700	20	798	32
	4	639	25	700	21	805	34
	5	638	27	700	23	791	29
	6	664	24	700	21	795	23
	7	674	25	700	23	800	23
	8	642	26	700	21	774	22
Reading	3	649	24	700	22	891	54
	4	658	22	700	22	845	45
	5	641	21	700	22	830	53
	6	647	21	700	22	828	42
	7	668	17	700	18	802	40
	8	655	23	700	24	833	44
Science	5	648	28	700	21	765	21
	8	658	25	700	20	751	19

Note: SEM at or closest above the cut scores.

Table 66. Estimates of Accuracy and Consistency of Performance Classification

Content	Grade	Accuracy	Consistency	False Positives	False Negatives	Kappa
Mathematics	3	0.79	0.71	0.10	0.11	0.59
	4	0.78	0.70	0.11	0.11	0.57
	5	0.76	0.67	0.11	0.12	0.54
	6	0.77	0.69	0.11	0.12	0.56
	7	0.76	0.68	0.12	0.12	0.54
	8	0.76	0.68	0.11	0.12	0.55
Reading	3	0.84	0.78	0.08	0.09	0.58
	4	0.80	0.73	0.10	0.10	0.55
	5	0.79	0.72	0.10	0.11	0.56
	6	0.79	0.71	0.10	0.11	0.54
	7	0.77	0.70	0.11	0.12	0.53
	8	0.78	0.71	0.10	0.11	0.53
Science	5	0.70	0.60	0.15	0.15	0.47
	8	0.70	0.60	0.15	0.15	0.46

Table 67. Accuracy & Consistency Estimates by Cut Score

Content	Grade	Accuracy			Consistency		
		U/L+P+A	U+L/P+A	U+L+P/A	U/L+P+A	U+L/P+A	U+L+P/A
Mathematics	3	0.95	0.92	0.92	0.93	0.88	0.89
	4	0.95	0.91	0.92	0.93	0.88	0.88
	5	0.95	0.91	0.91	0.92	0.87	0.87
	6	0.93	0.91	0.92	0.91	0.87	0.89
	7	0.92	0.90	0.92	0.88	0.86	0.89
	8	0.94	0.91	0.91	0.92	0.87	0.87
Reading	3	0.94	0.91	0.98	0.92	0.88	0.97
	4	0.93	0.90	0.96	0.90	0.86	0.95
	5	0.94	0.91	0.94	0.92	0.87	0.92
	6	0.94	0.90	0.95	0.91	0.86	0.93
	7	0.93	0.90	0.94	0.90	0.86	0.91
	8	0.94	0.91	0.93	0.92	0.87	0.90
Science	5	0.90	0.88	0.92	0.86	0.83	0.89
	8	0.90	0.87	0.92	0.85	0.82	0.89

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

Table 68. Accuracy & Consistency Estimates by Cut Score: False Positive and False Negative Rates

		U/L+P+A		U+L/P+A		U+L+P/A	
Content	Grade	False Positive	False Negative	False Positive	False Negative	False Positive	False Negative
Mathematics	3	0.02	0.03	0.04	0.05	0.04	0.03
	4	0.02	0.03	0.04	0.05	0.05	0.04
	5	0.02	0.03	0.04	0.05	0.05	0.04
	6	0.03	0.04	0.04	0.05	0.05	0.03
	7	0.04	0.05	0.05	0.05	0.05	0.03
	8	0.02	0.03	0.04	0.05	0.05	0.04
Reading	3	0.02	0.03	0.04	0.05	0.01	0.01
	4	0.03	0.04	0.05	0.05	0.02	0.01
	5	0.02	0.03	0.04	0.05	0.04	0.02
	6	0.02	0.04	0.05	0.05	0.03	0.02
	7	0.03	0.04	0.04	0.05	0.04	0.02
	8	0.02	0.04	0.04	0.05	0.04	0.03
Science	5	0.04	0.06	0.06	0.06	0.05	0.03
	8	0.04	0.06	0.06	0.06	0.05	0.03

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

Figures

Figure 1. Spring 2013 Grade 3 Mathematics operational scale score histogram

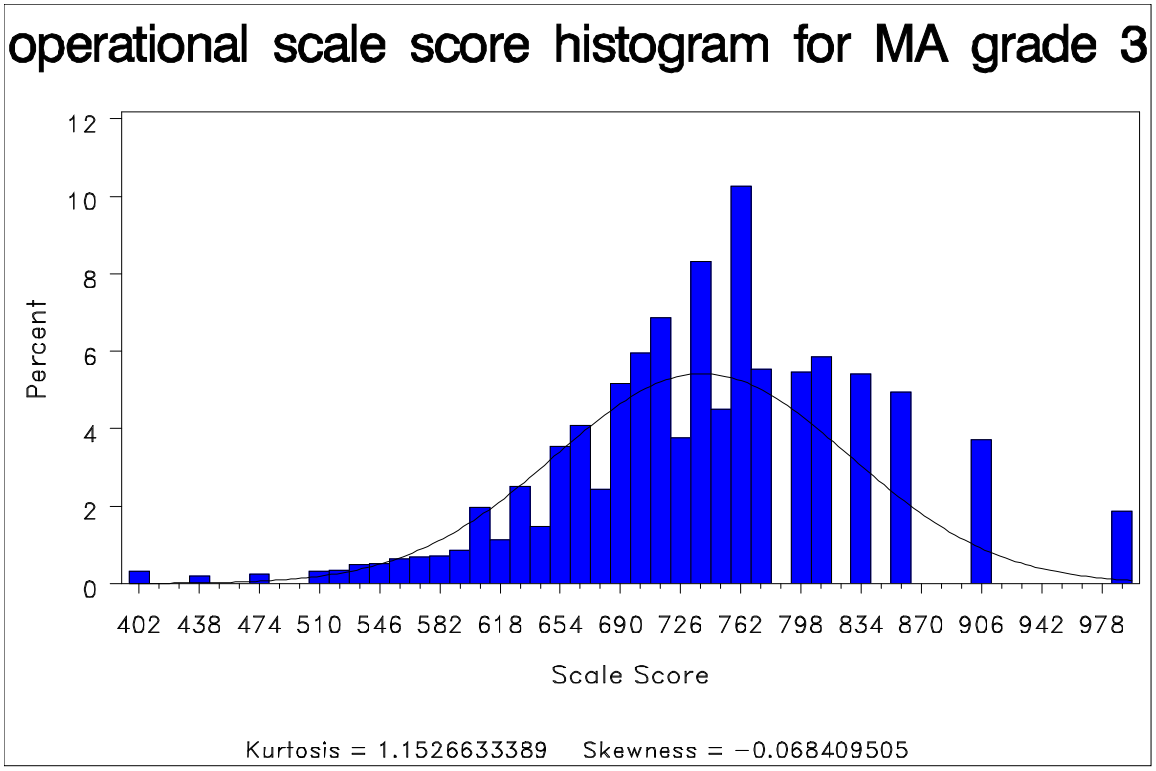


Figure 2. Spring 2013 Grade 4 Mathematics operational scale score histogram

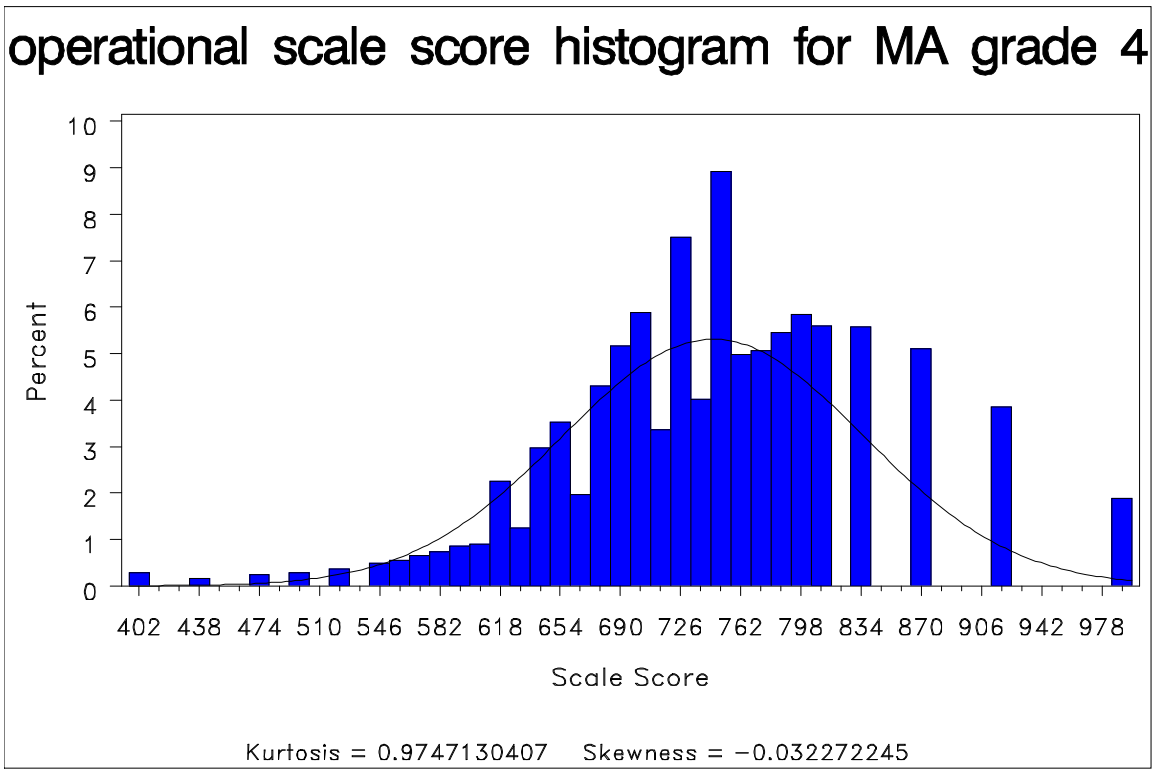


Figure 3. Spring 2013 Grade 5 Mathematics operational scale score histogram

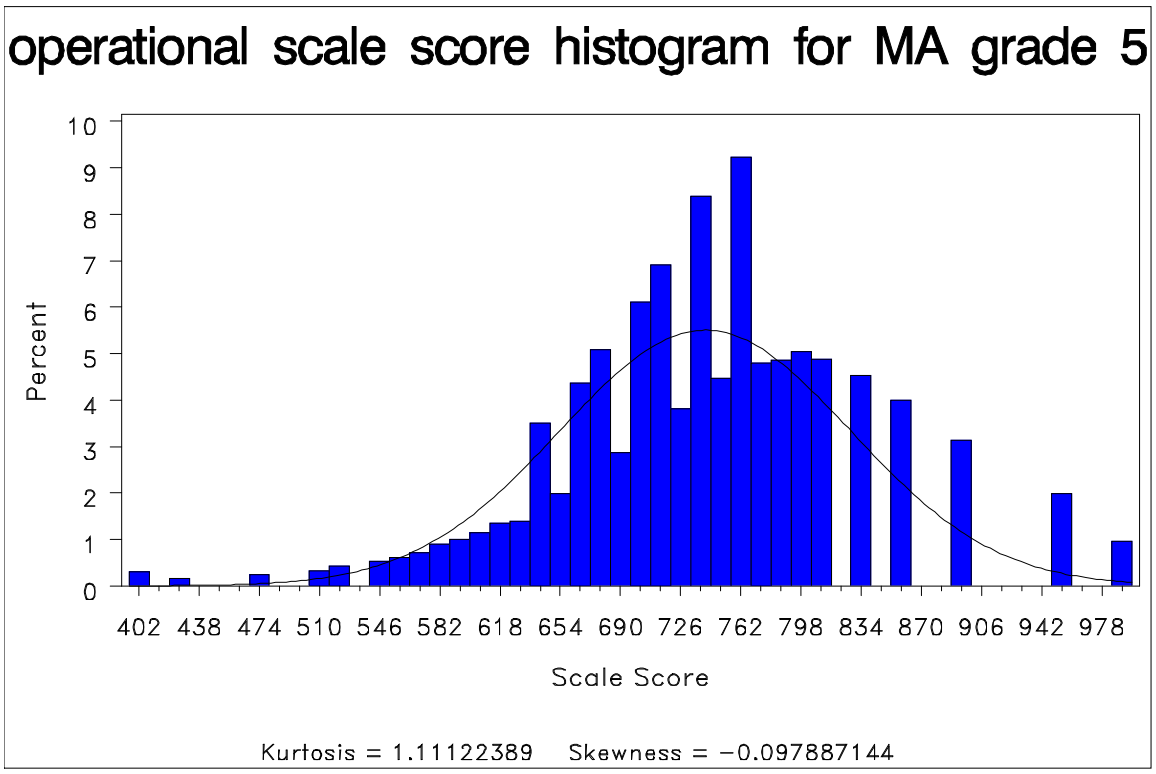


Figure 4. Spring 2013 Grade 6 Mathematics operational scale score histogram

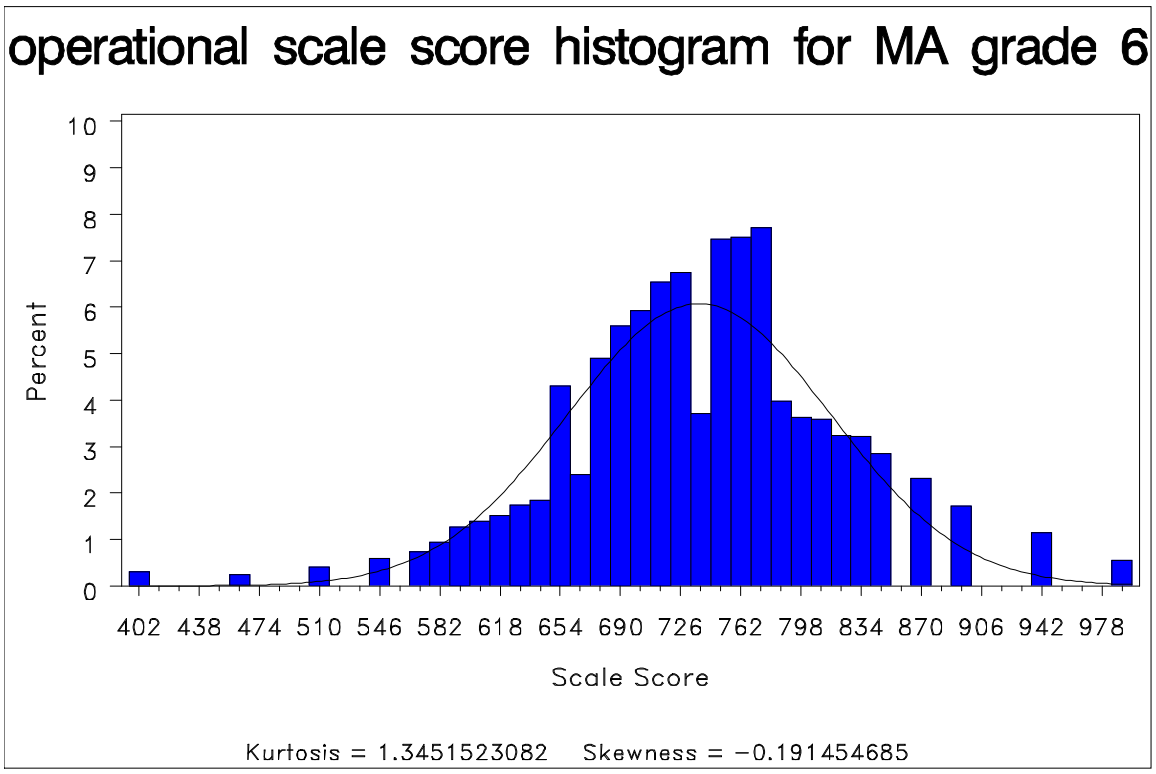


Figure 5. Spring 2013 Grade 7 Mathematics operational scale score histogram

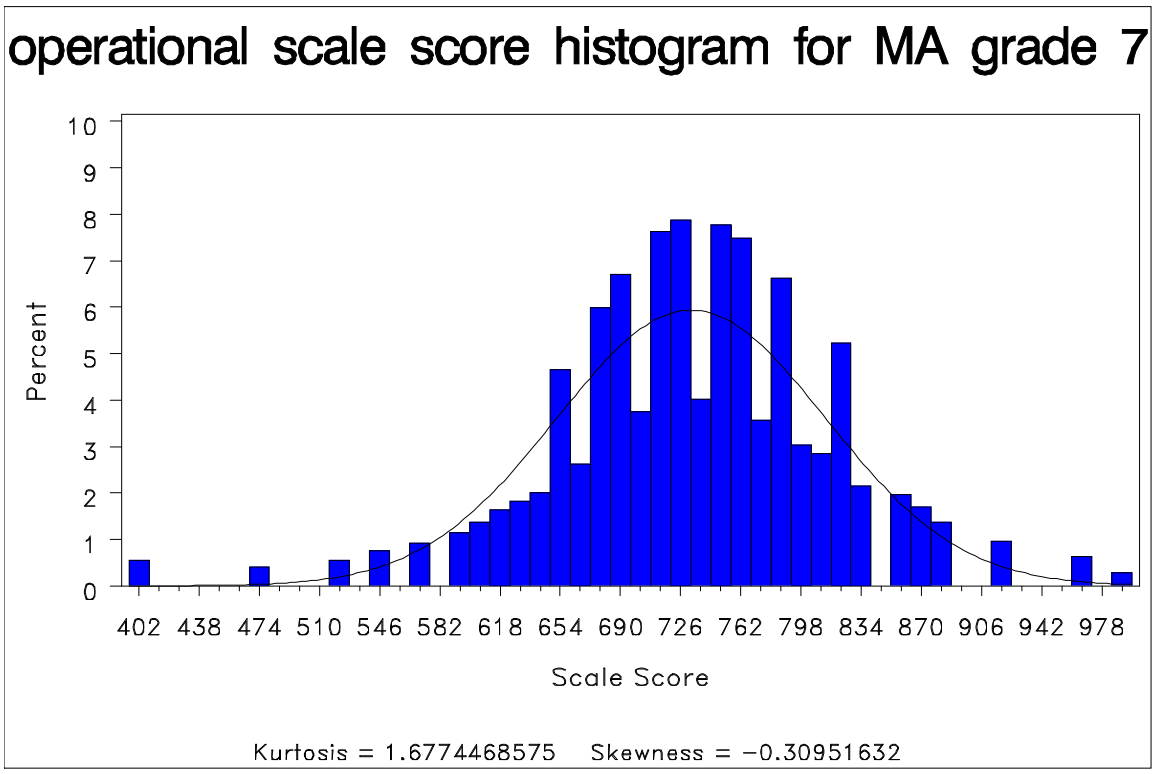


Figure 6. Spring 2013 Grade 8 Mathematics operational scale score histogram

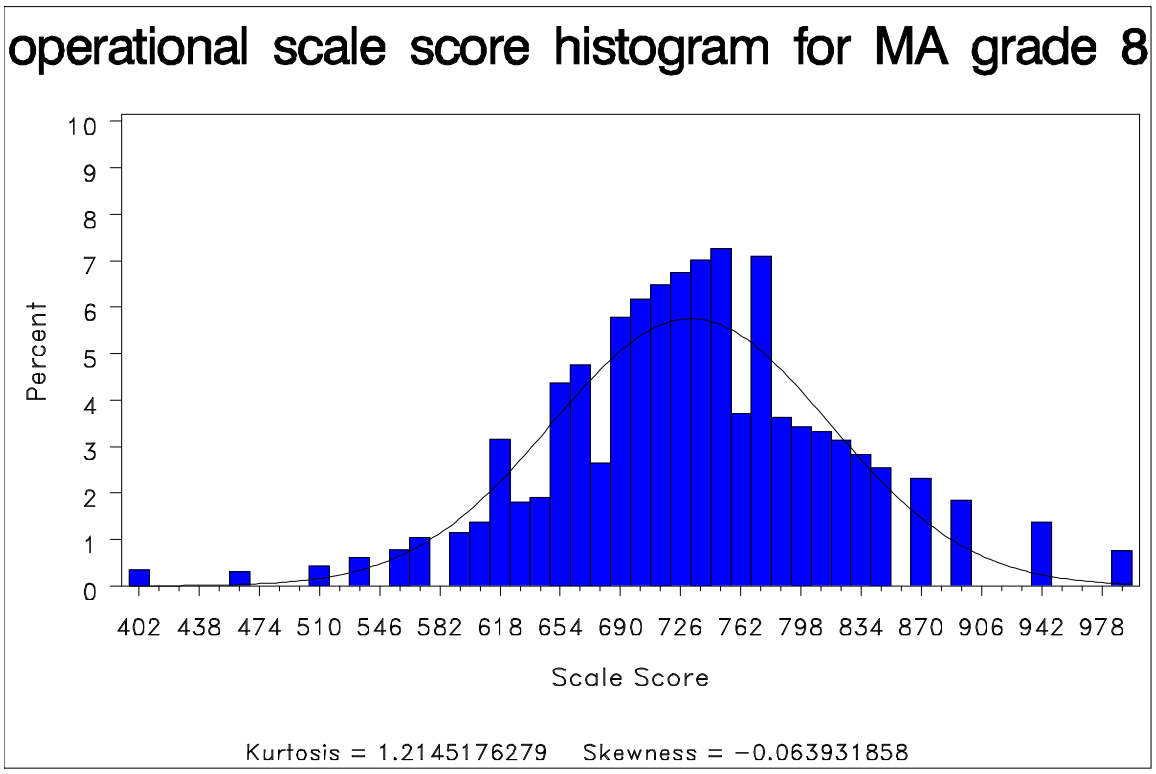


Figure 7. Spring 2013 Grade 3 Reading operational scale score histogram

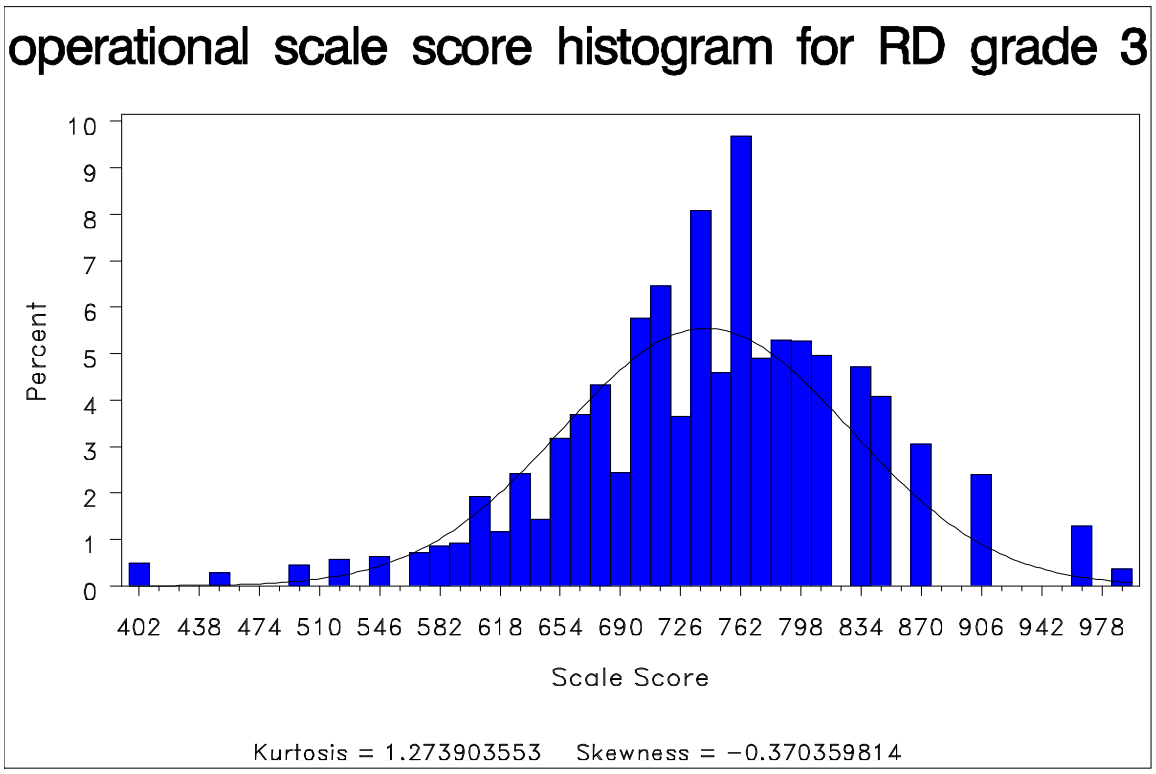


Figure 8. Spring 2013 Grade 4 Reading operational scale score histogram

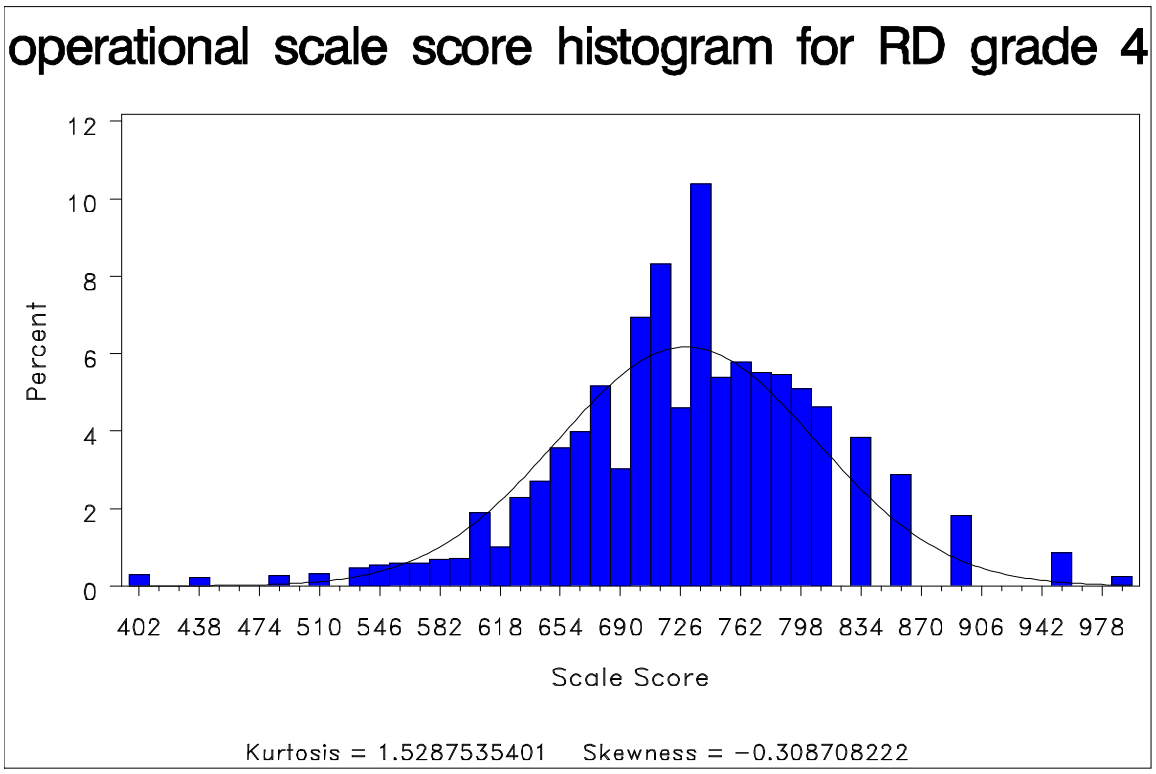


Figure 9. Spring 2013 Grade 5 Reading operational scale score histogram

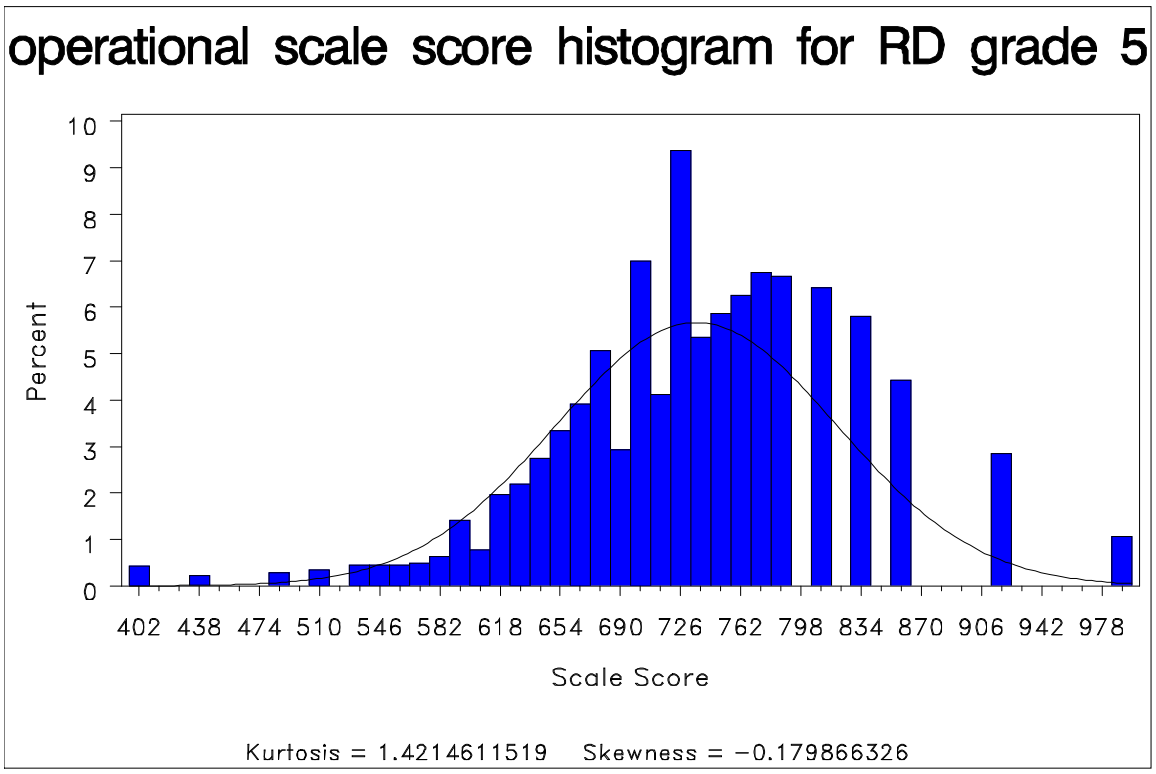


Figure 10. Spring 2013 Grade 6 Reading operational scale score histogram

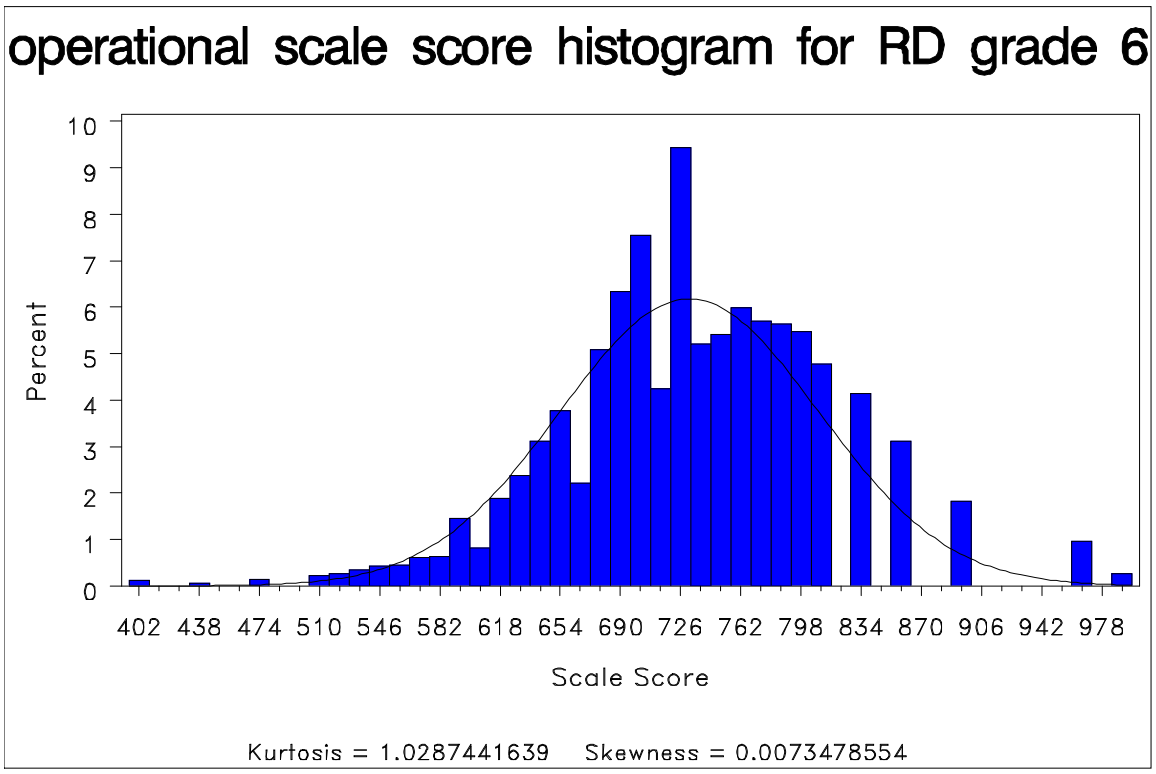


Figure 11. Spring 2013 Grade 7 Reading operational scale score histogram

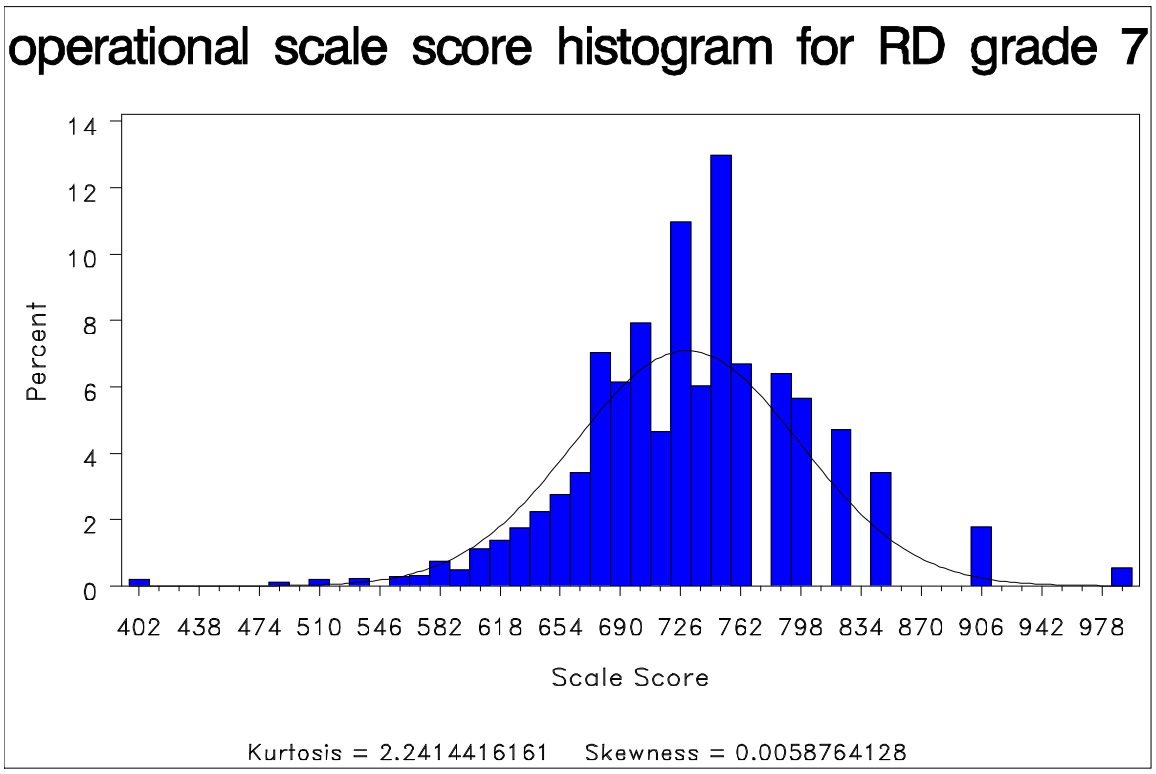


Figure 12. Spring 2013 Grade 8 Reading operational scale score histogram

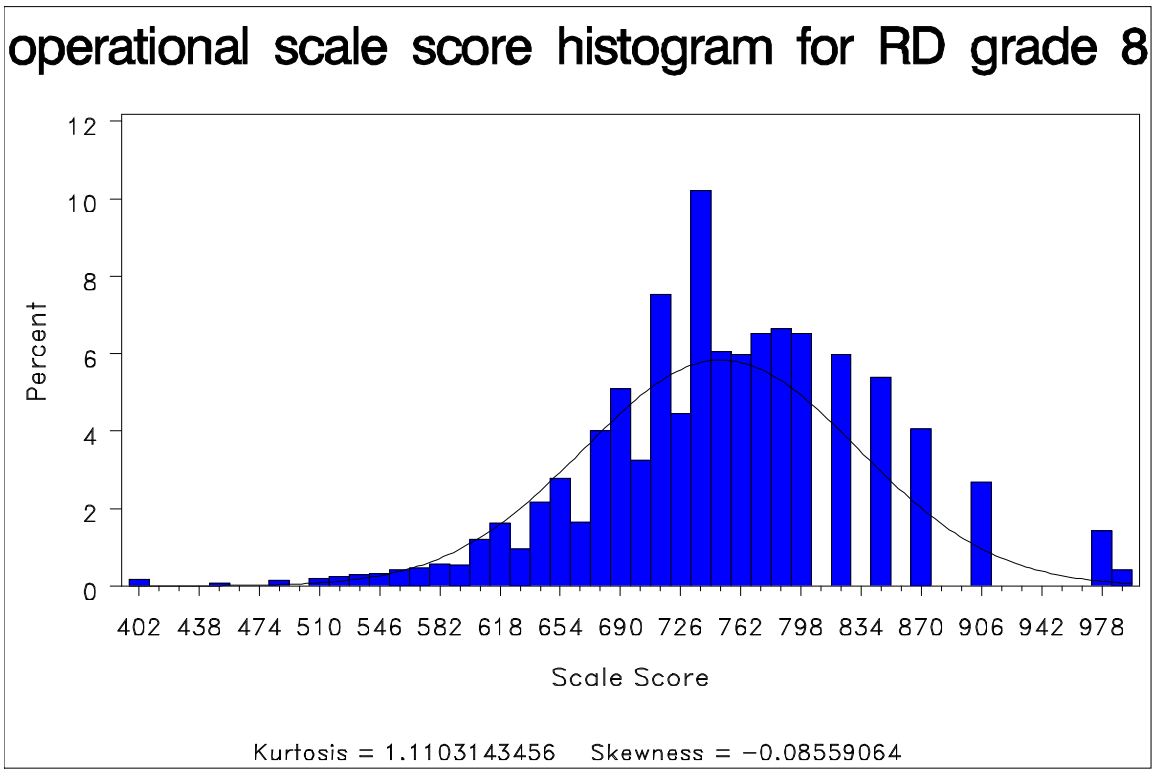


Figure 13. Spring 2013 Grade 5 Science operational scale score histogram

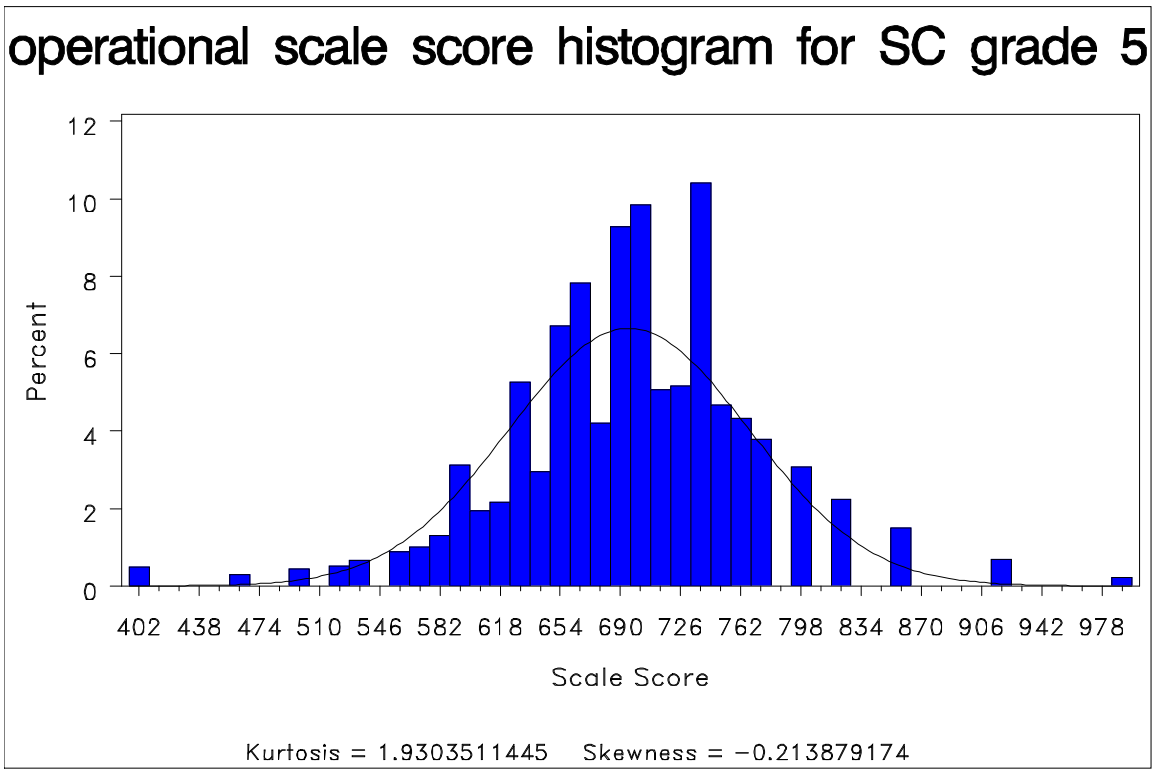


Figure 14. Spring 2013 Grade 8 Science operational scale score histogram

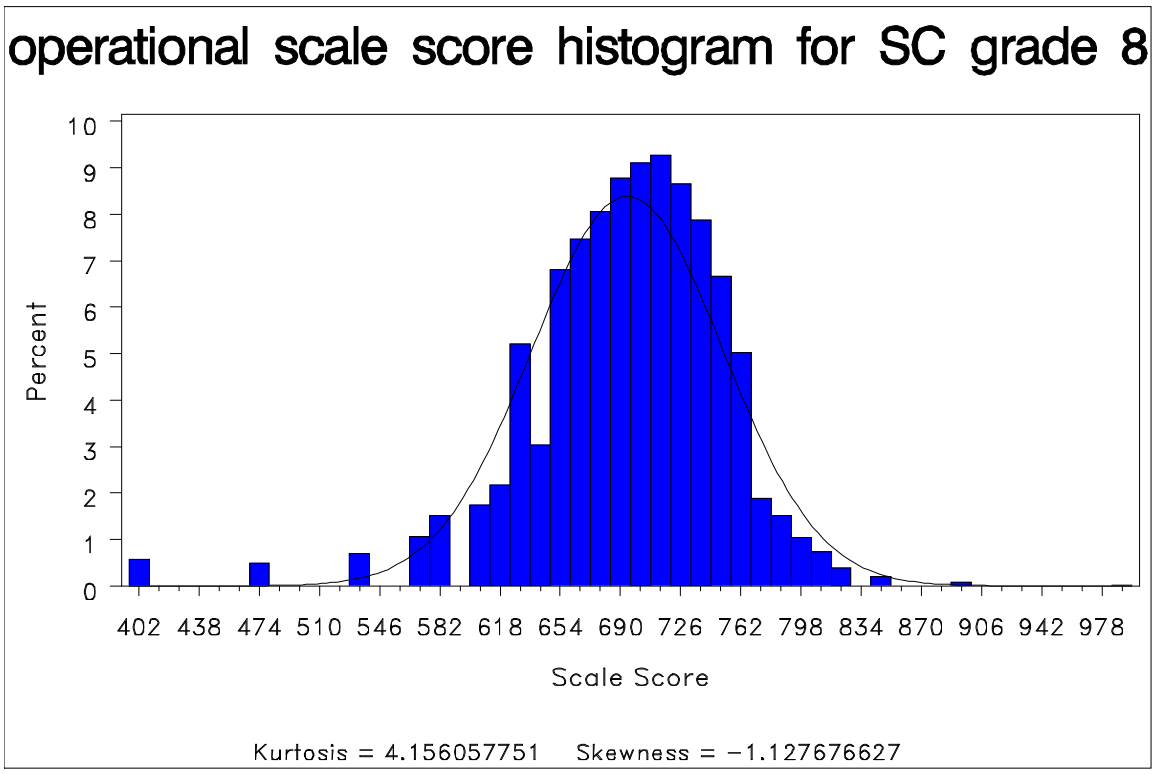
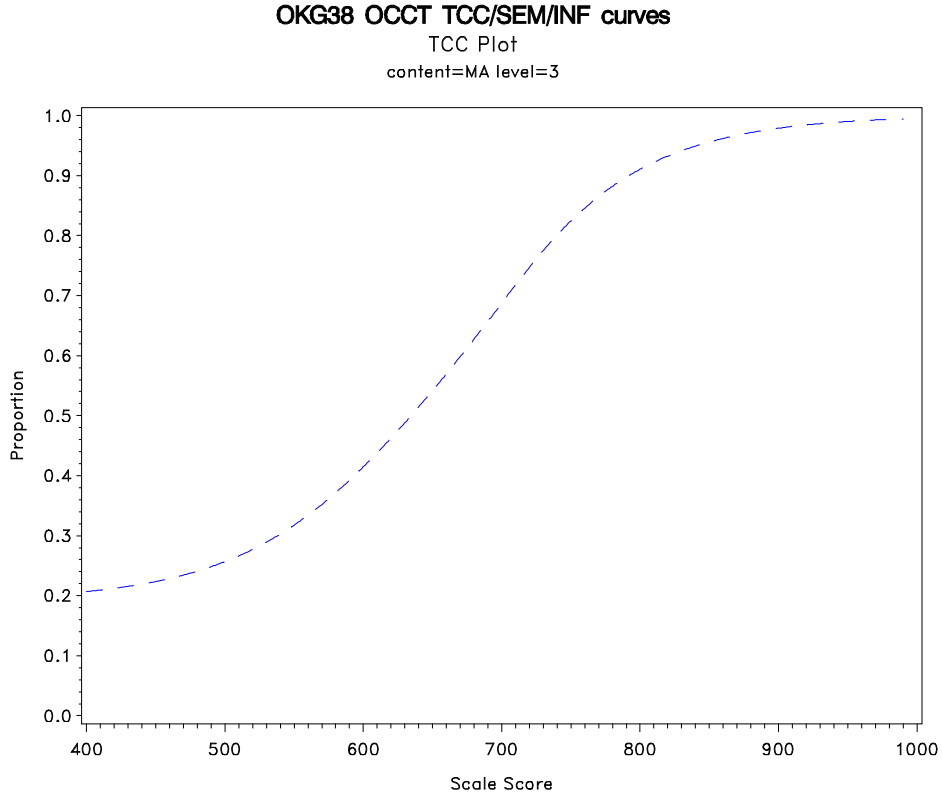


Figure 15. Spring 2013 Grade 3 Mathematics operational test characteristic curve and standard error of measurement curve



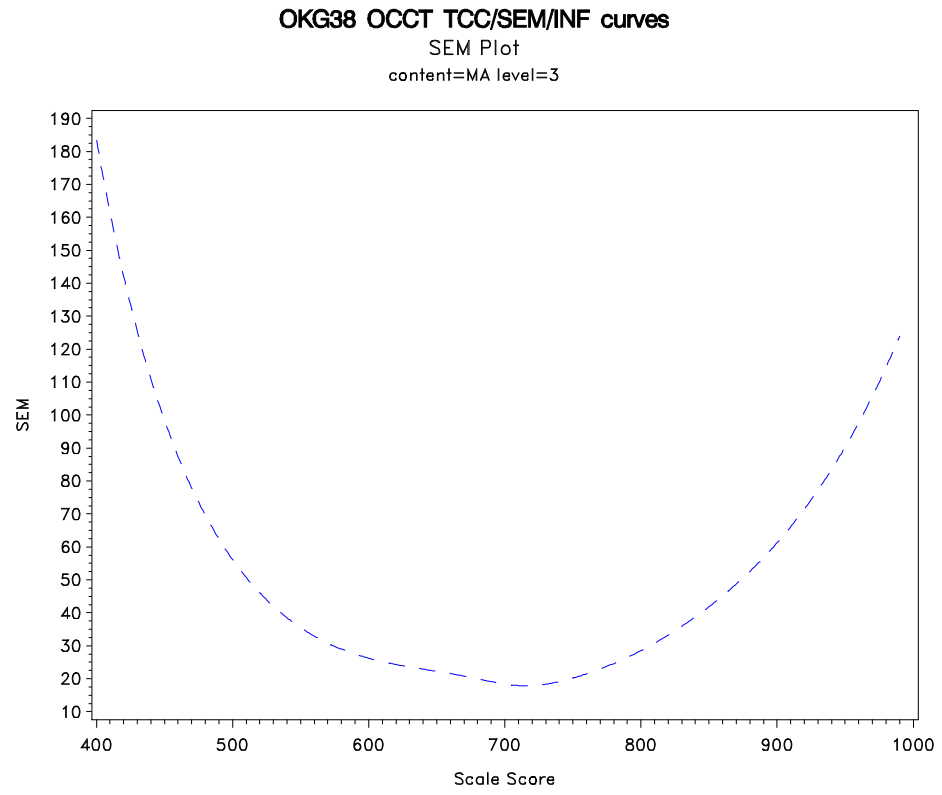
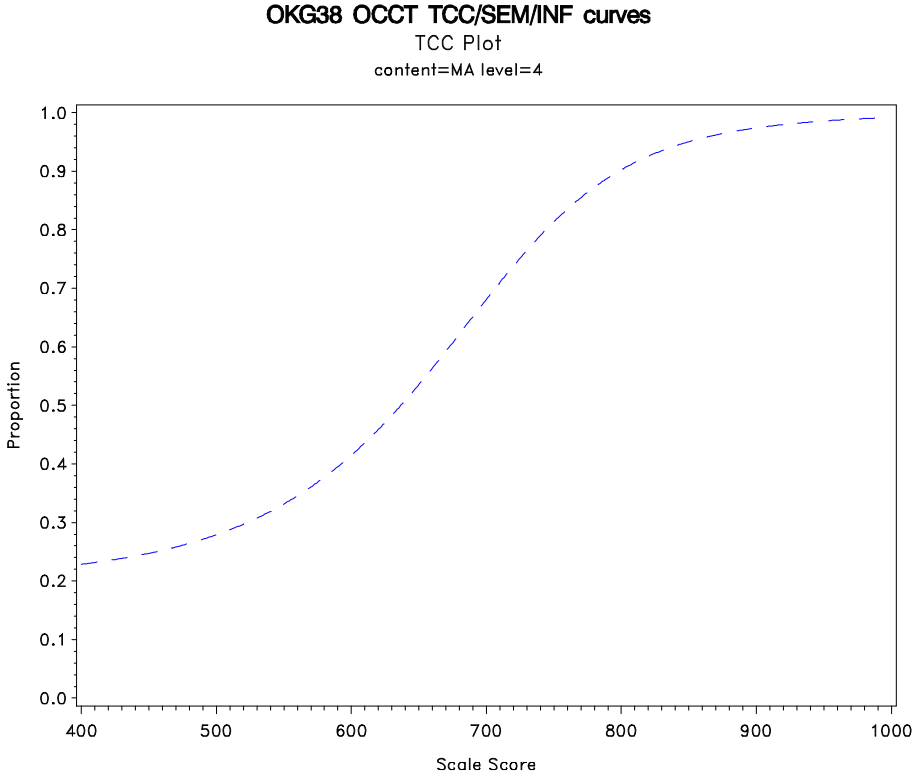


Figure 16. Spring 2013 Grade 4 Mathematics operational test characteristic curve and standard error of measurement curve



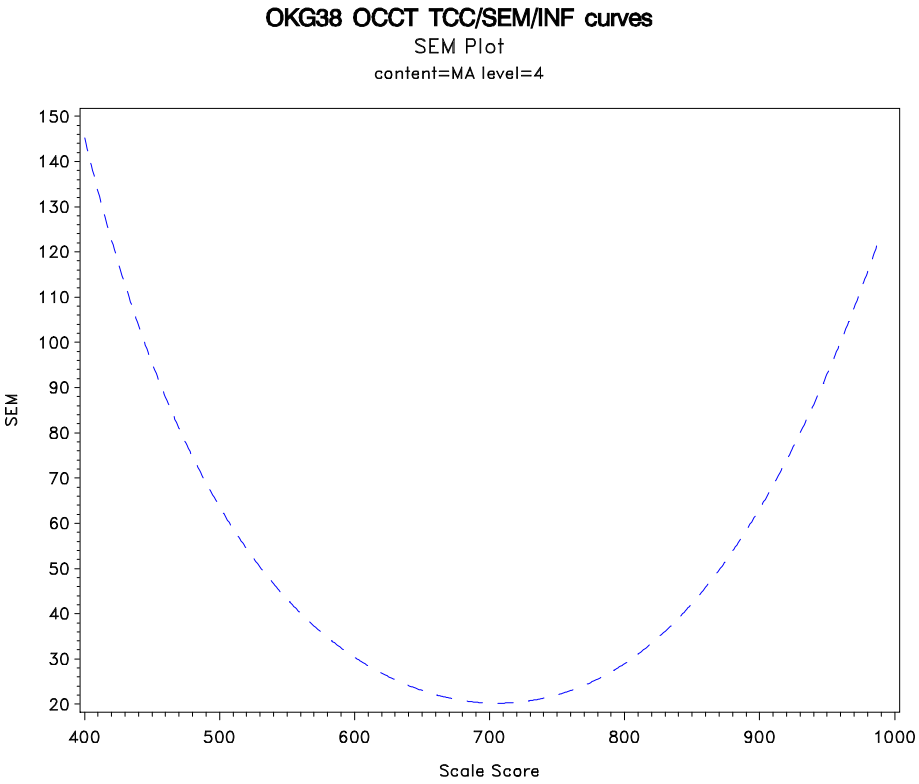


Figure 17. Spring 2013 Grade 5 Mathematics operational test characteristic curve and standard error of measurement curve

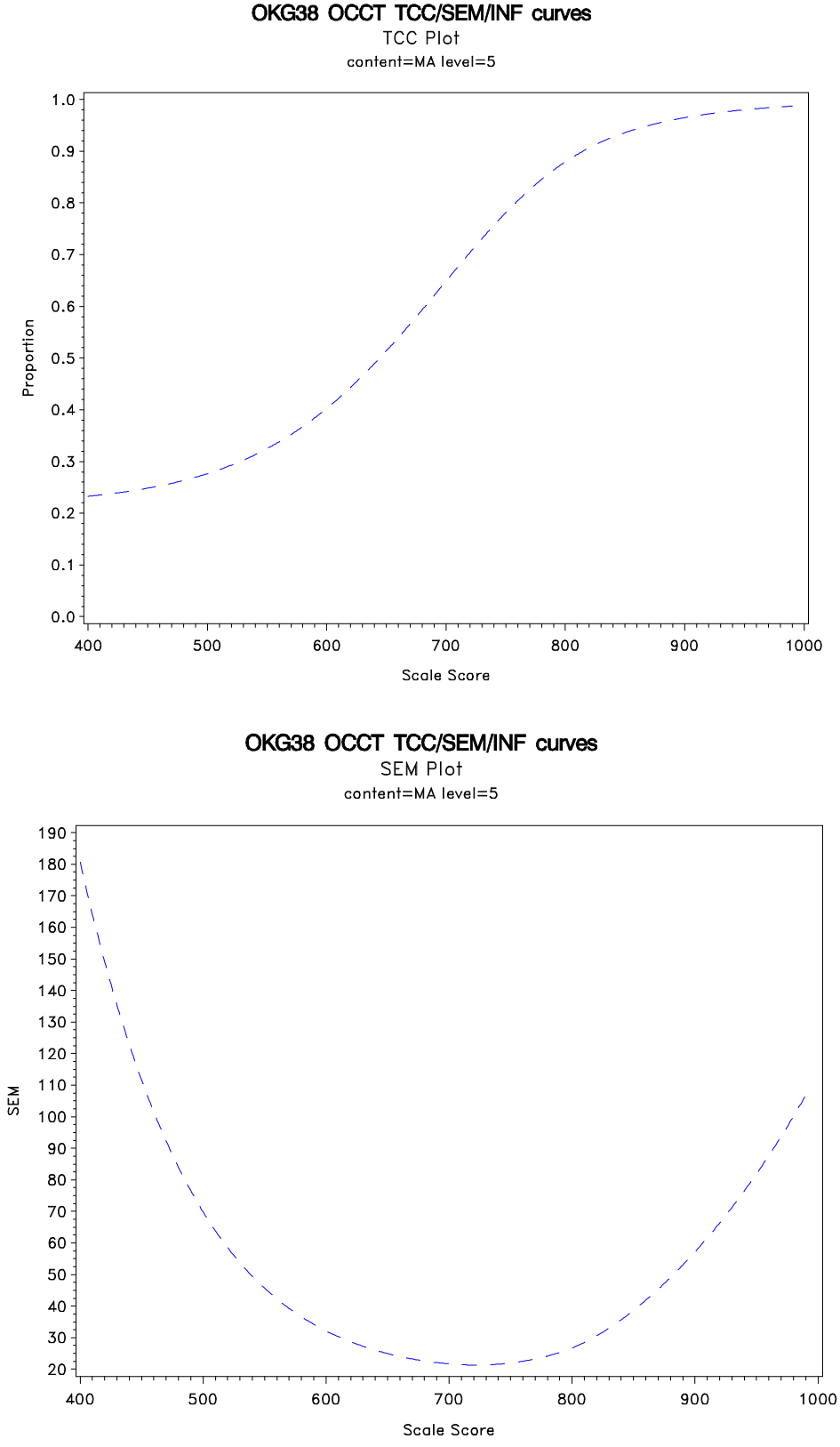
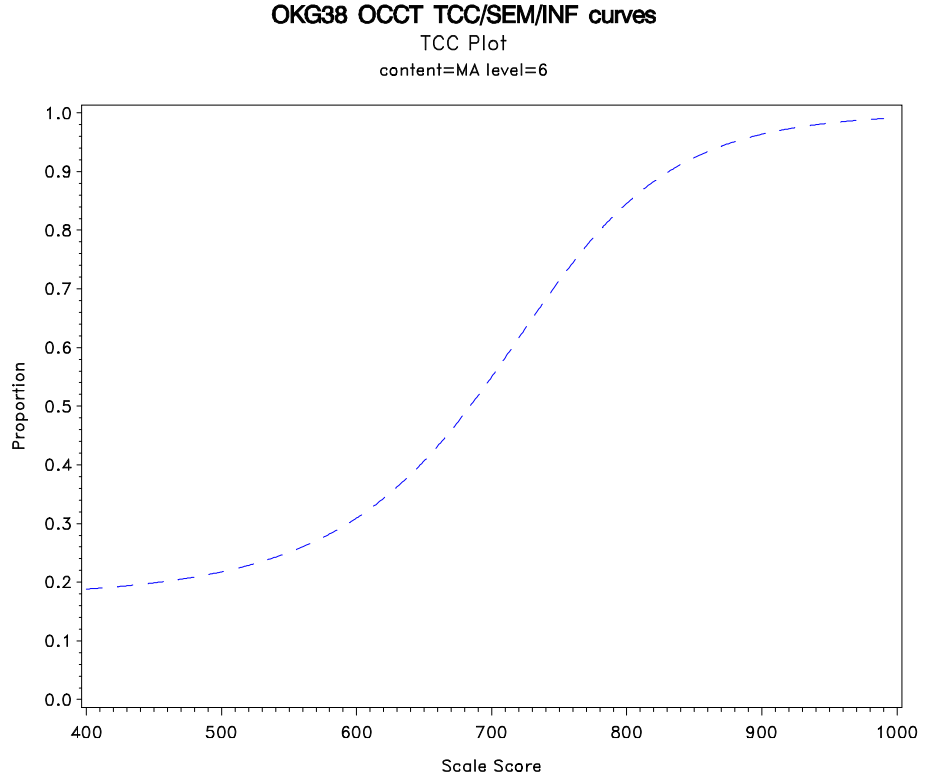


Figure 18. Spring 2013 Grade 6 Mathematics operational test characteristic curve and standard error of measurement curve



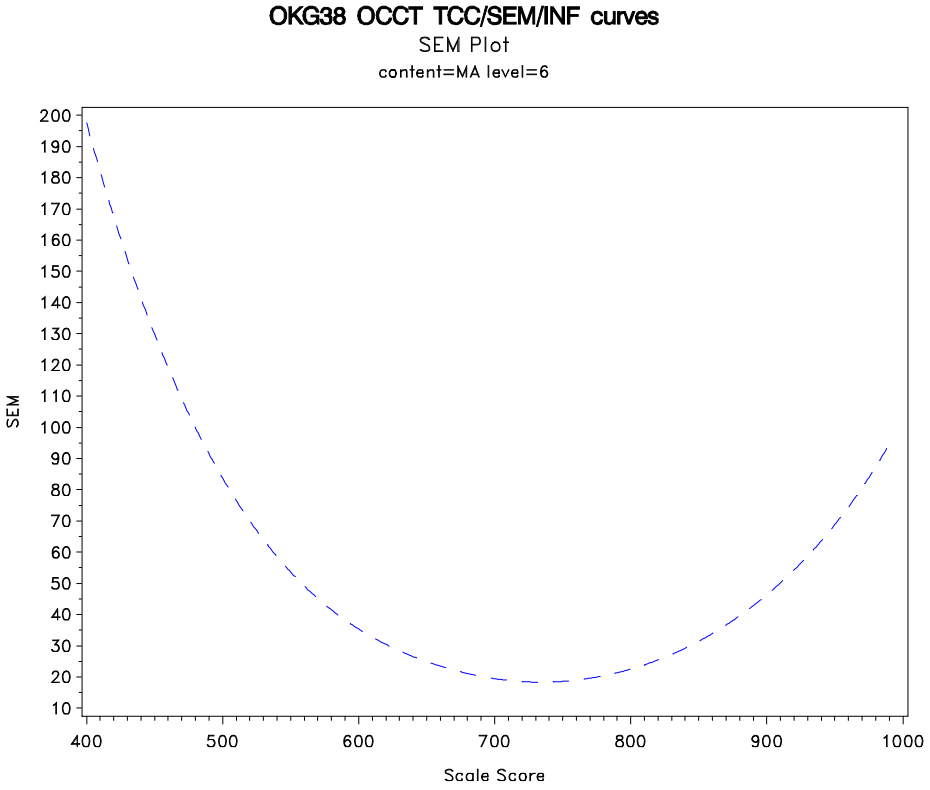
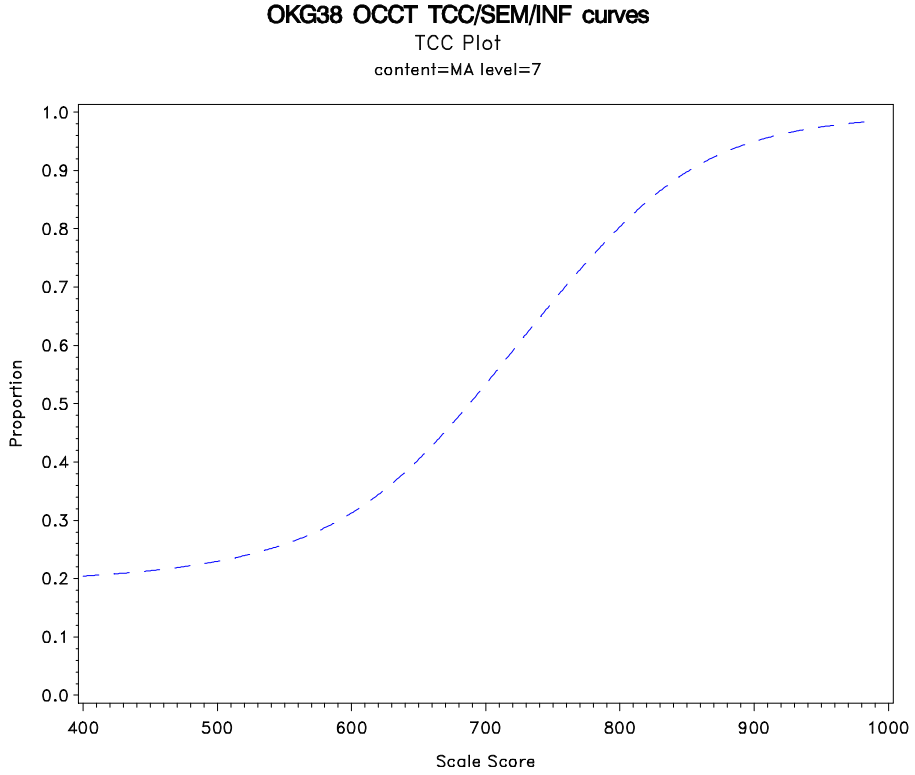


Figure 19. Spring 2013 Grade 7 Mathematics operational test characteristic curve and standard error of measurement curve



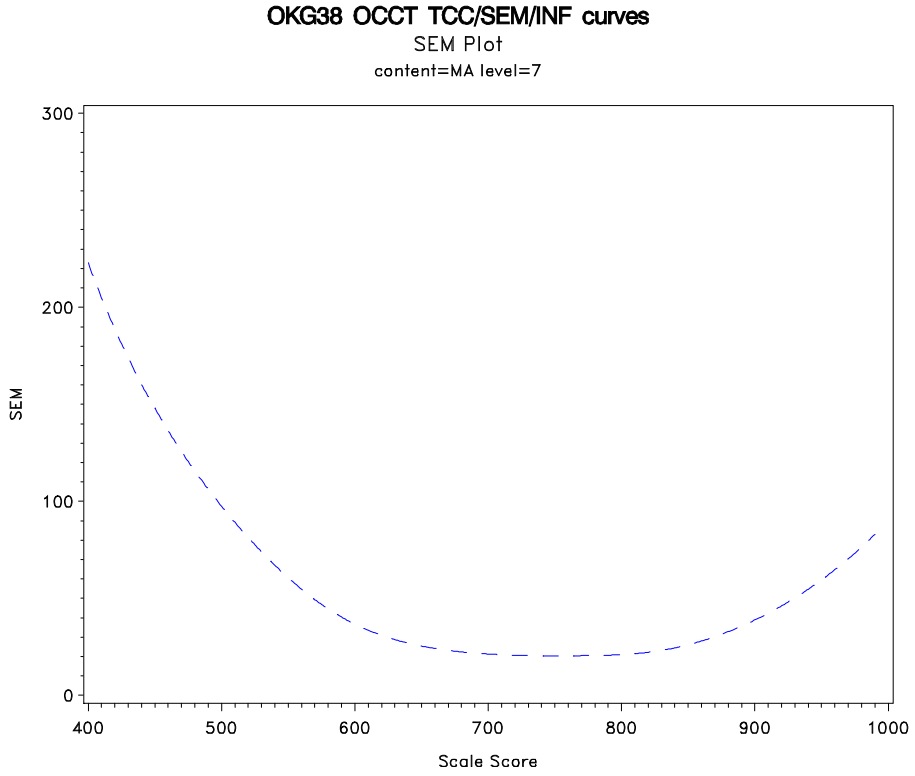
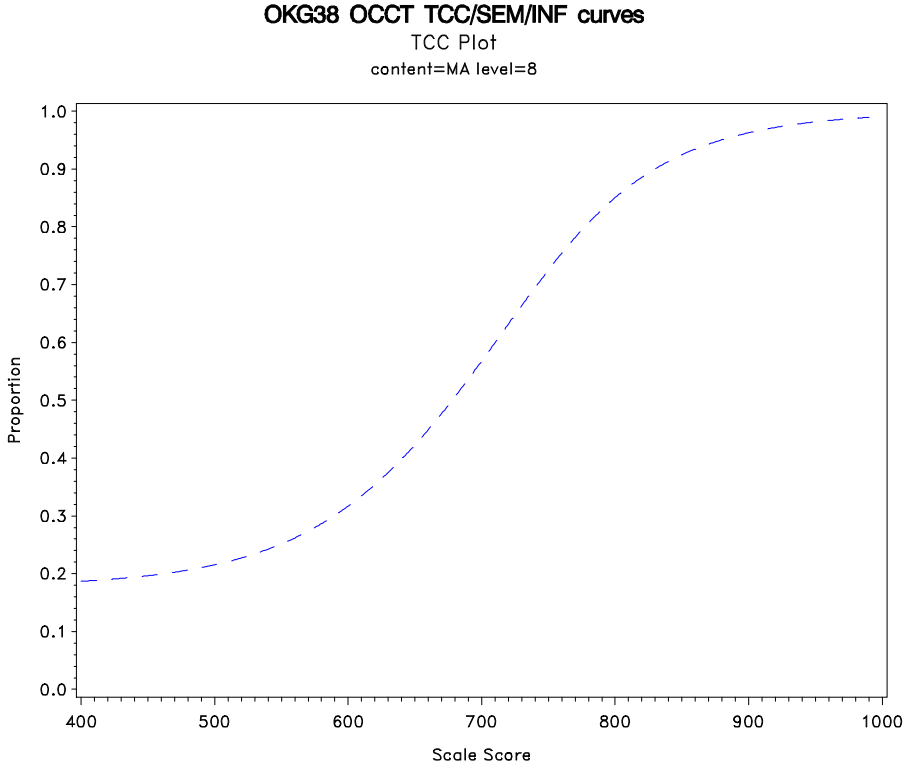


Figure 20. Spring 2013 Grade 8 Mathematics operational test characteristic curve and standard error of measurement curve



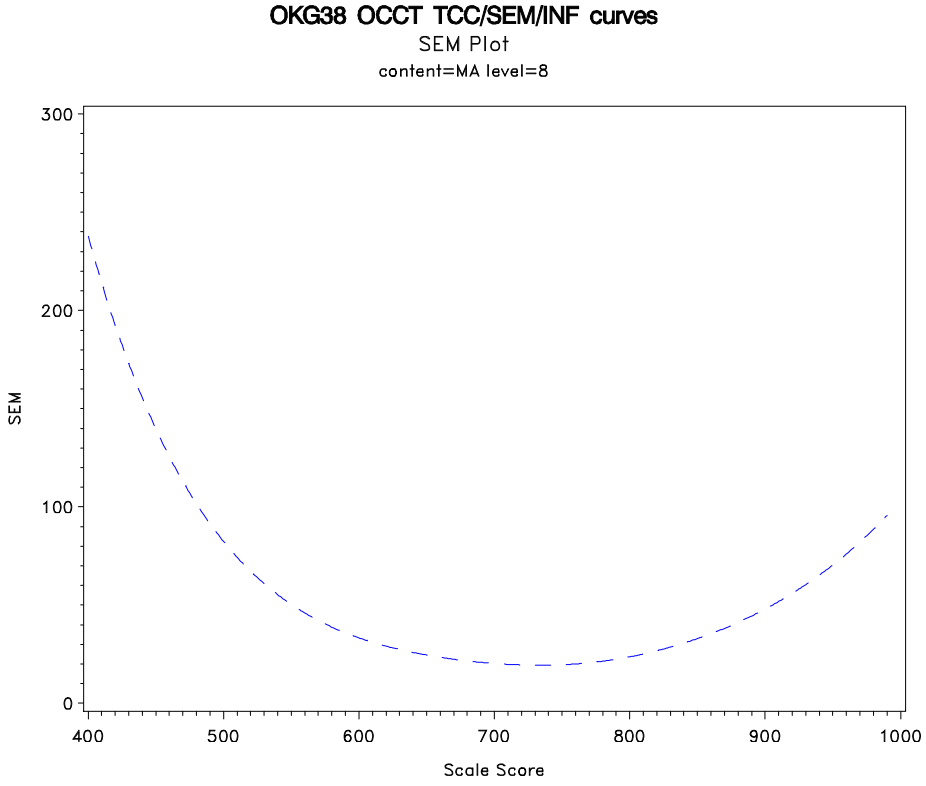
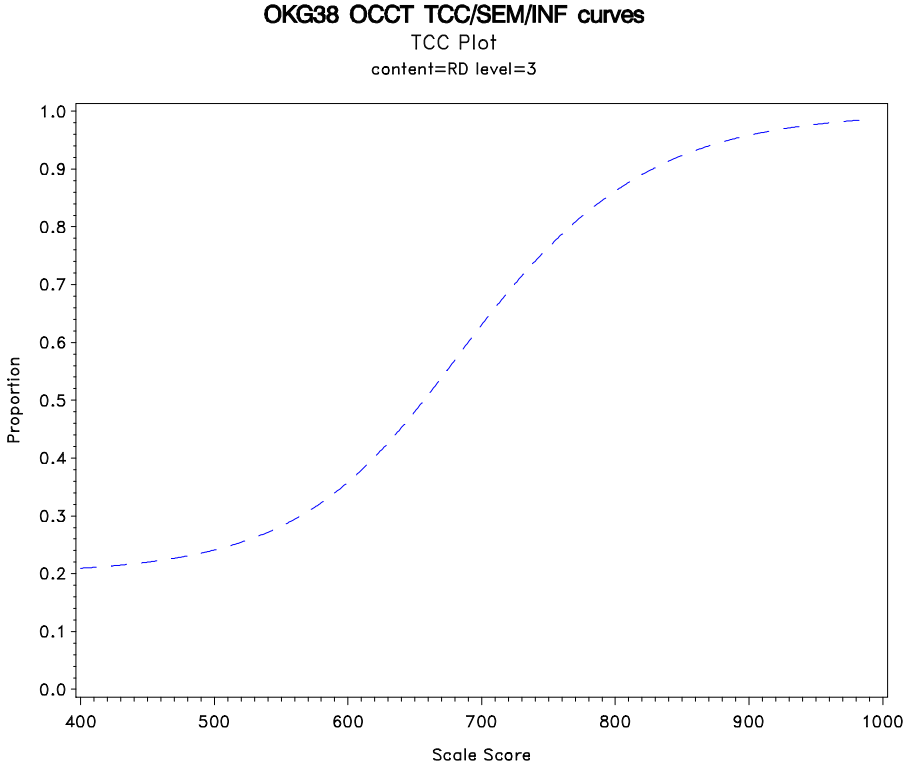


Figure 21. Spring 2013 Grade 3 Reading operational test characteristic curve and standard error of measurement curve



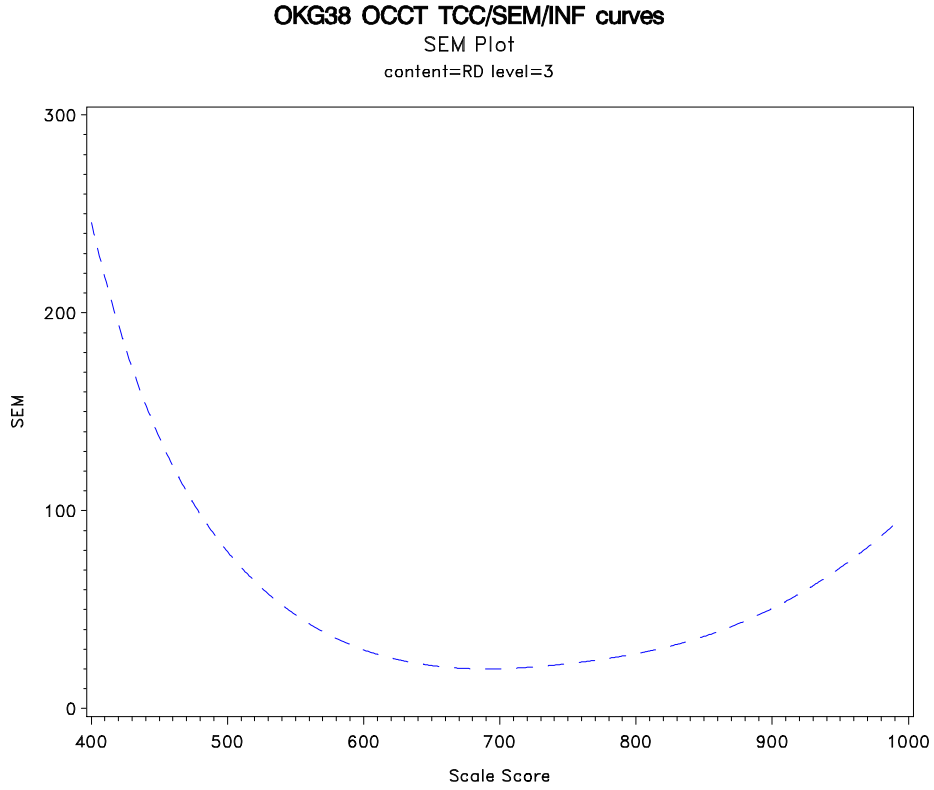
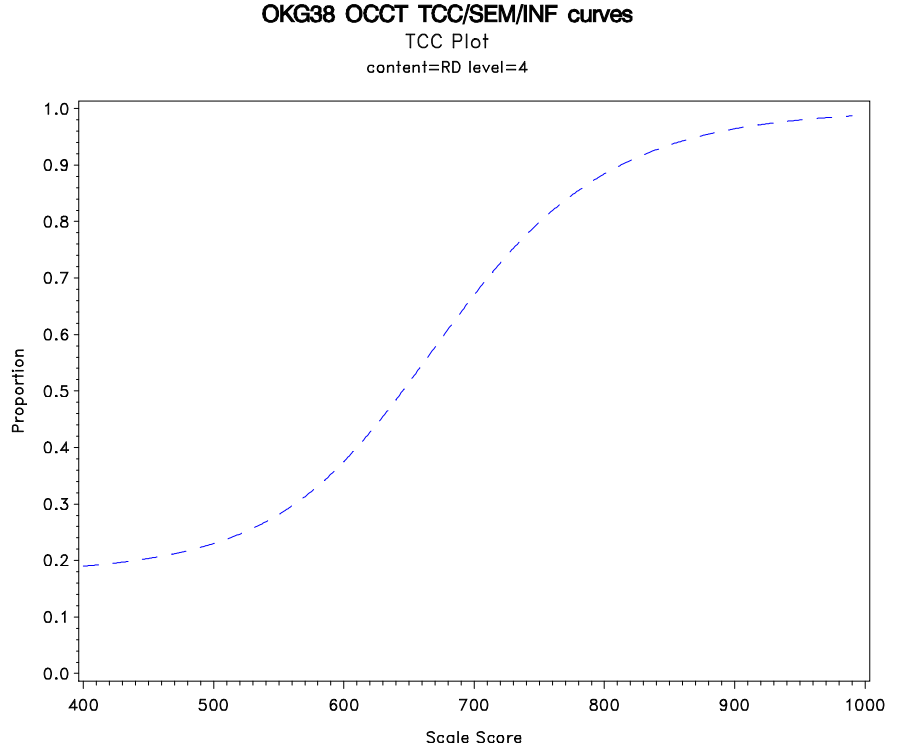


Figure 22. Spring 2013 Grade 4 Reading operational test characteristic curve and standard error of measurement curve



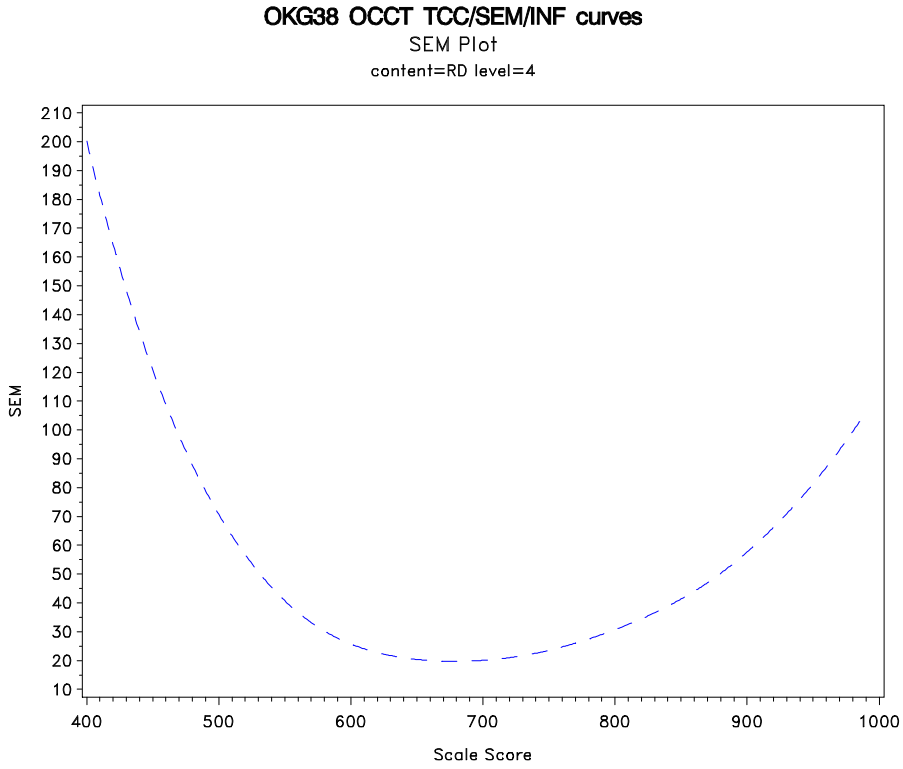
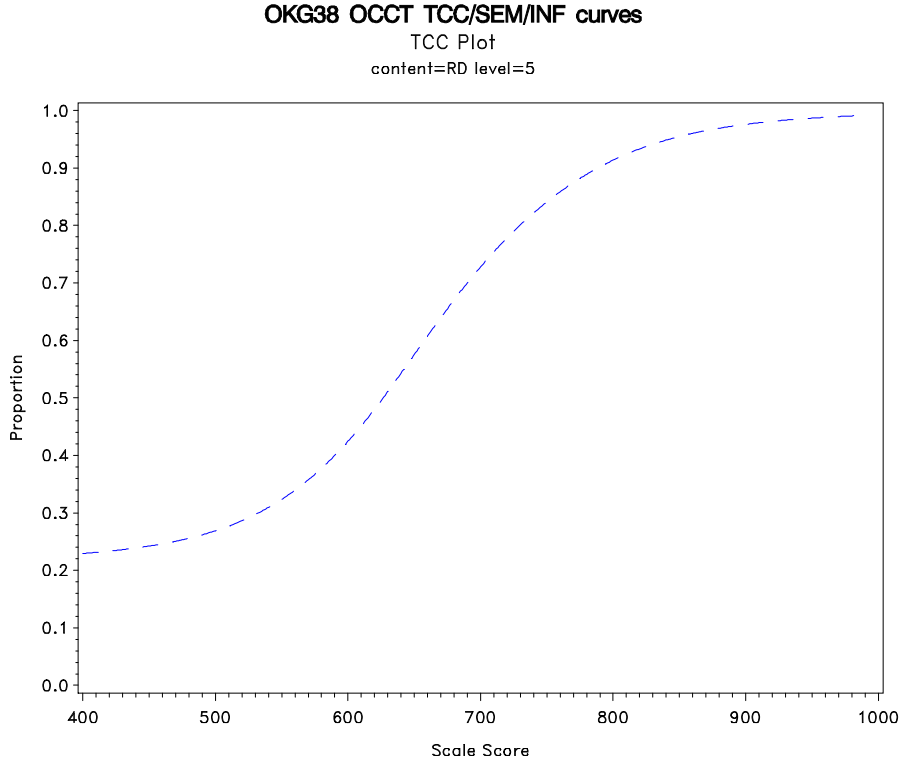


Figure 23. Spring 2013 Grade 5 Reading operational test characteristic curve and standard error of measurement curve



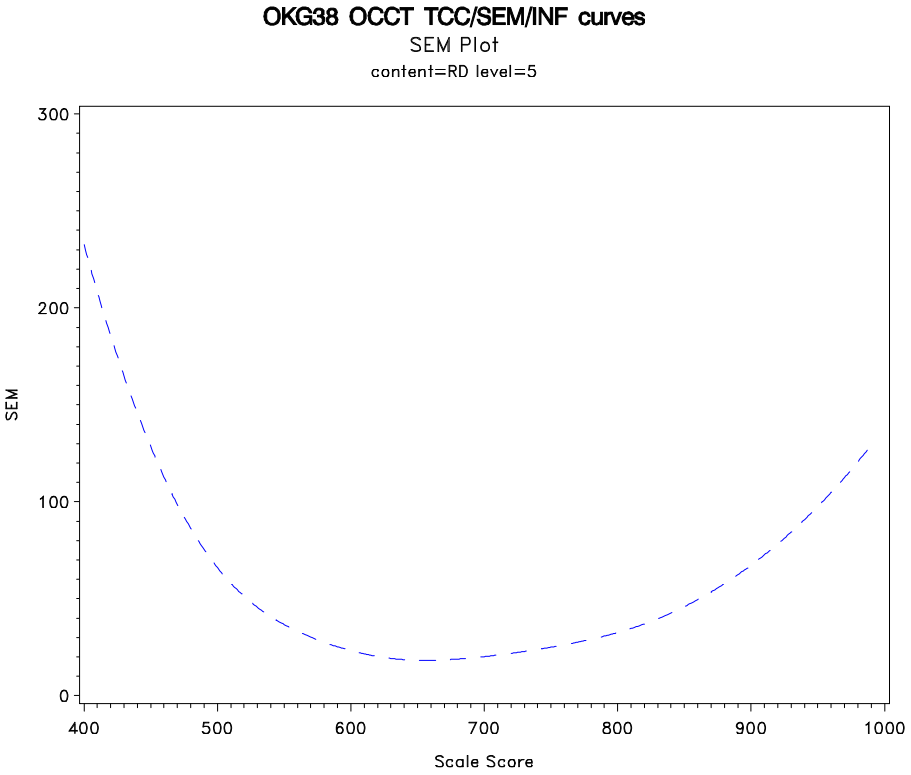
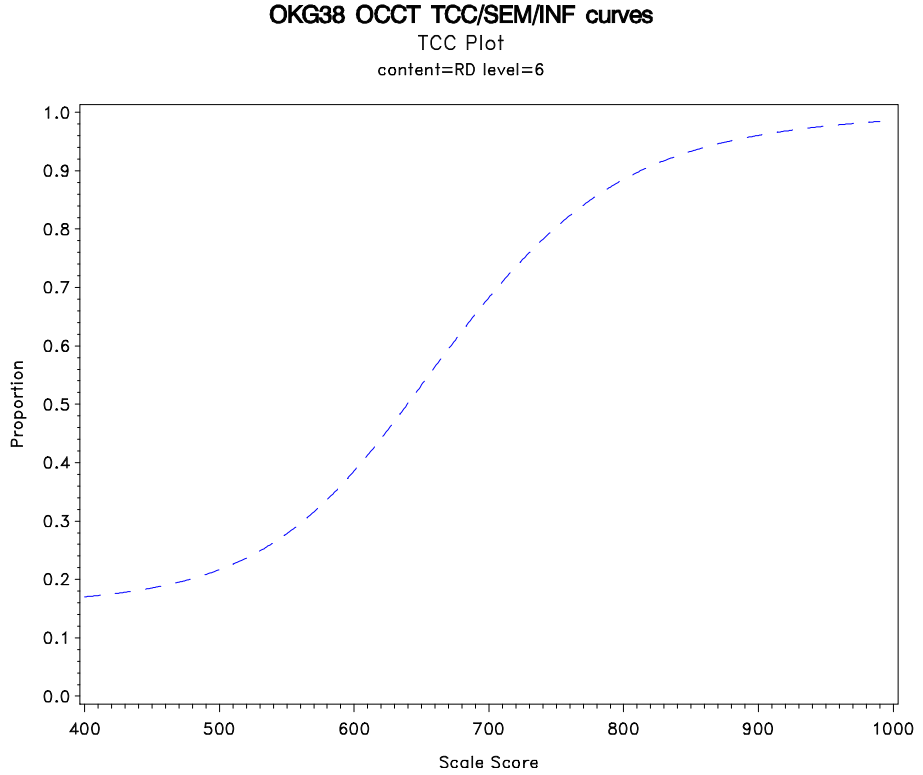


Figure 24. Spring 2013 Grade 6 Reading operational test characteristic curve and standard error of measurement curve



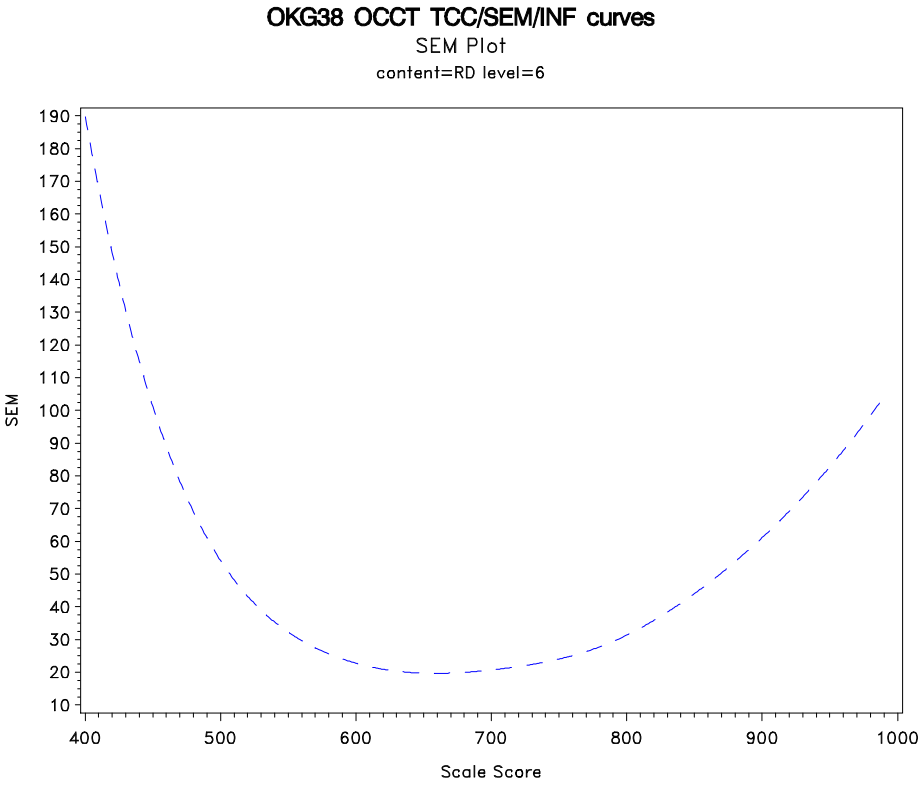


Figure 25. Spring 2013 Grade 7 Reading operational test characteristic curve and standard error of measurement curve

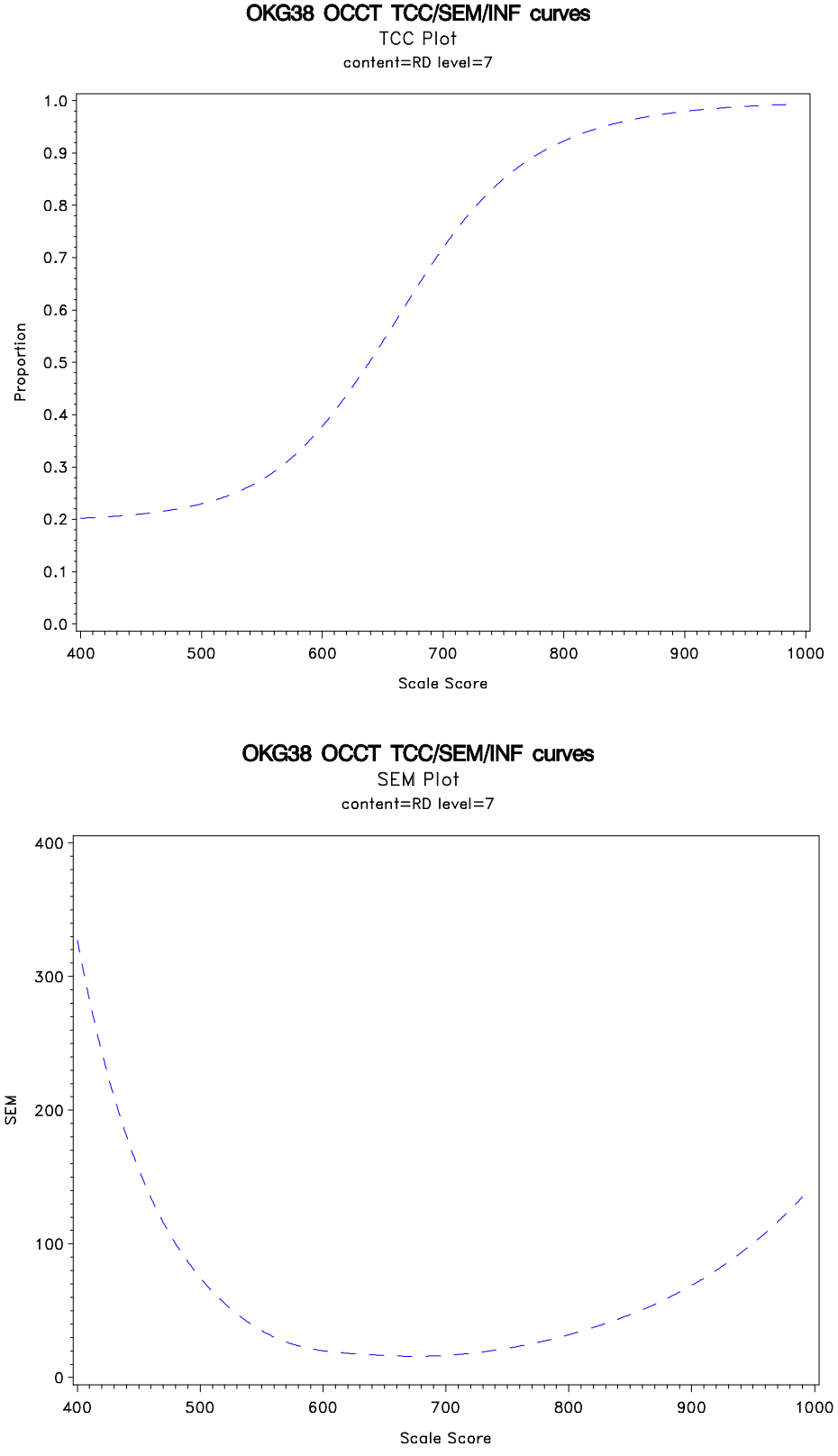
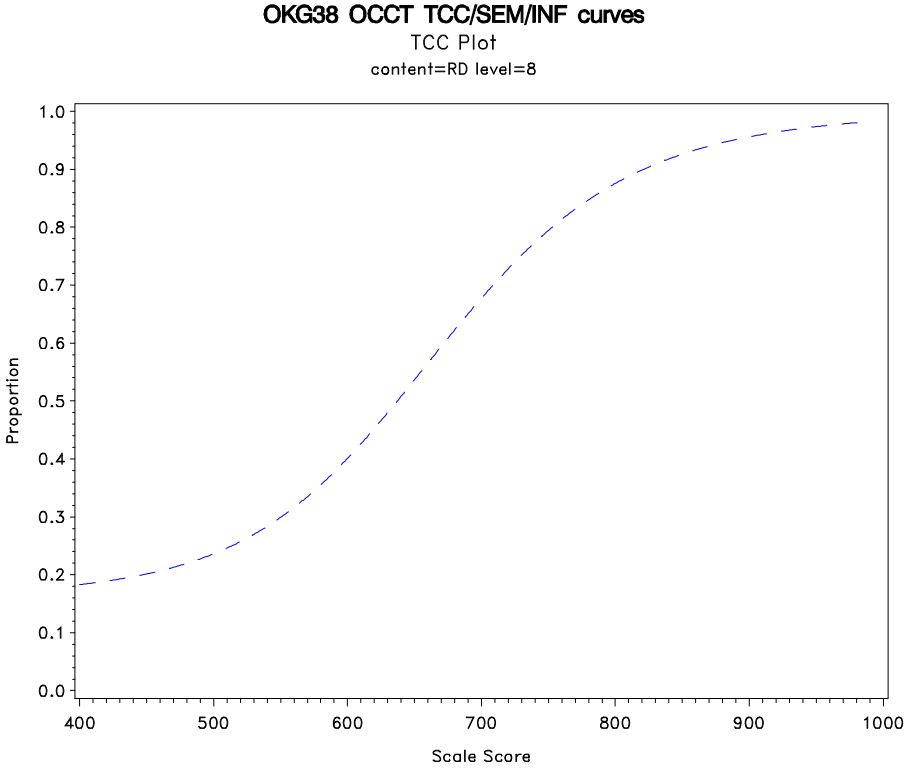


Figure 26. Spring 2013 Grade 8 Reading operational test characteristic curve and standard error of measurement curve



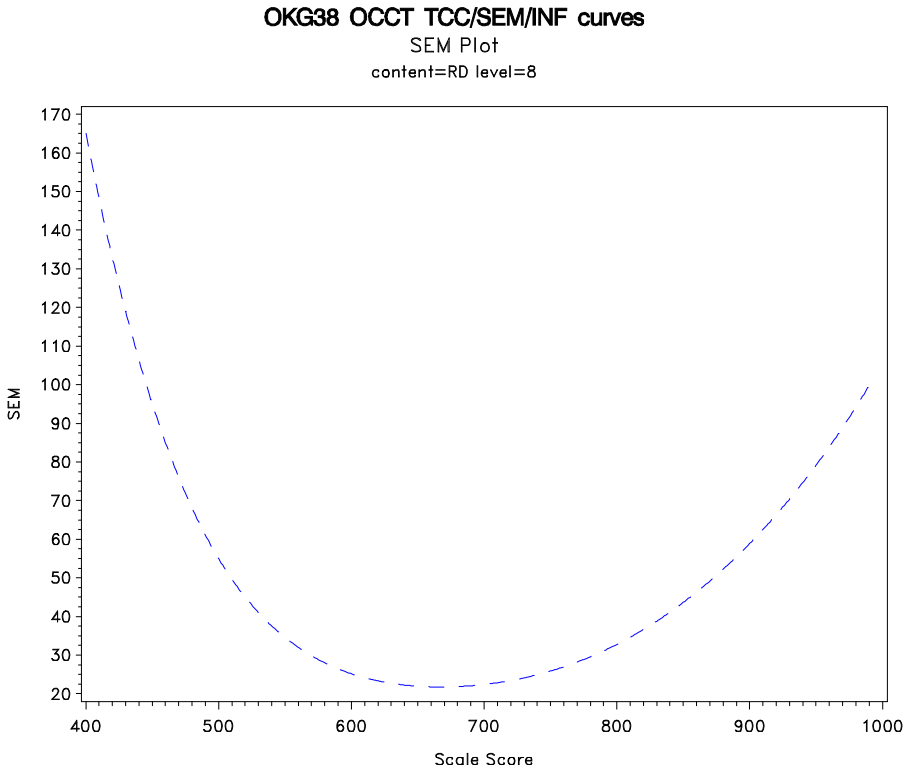
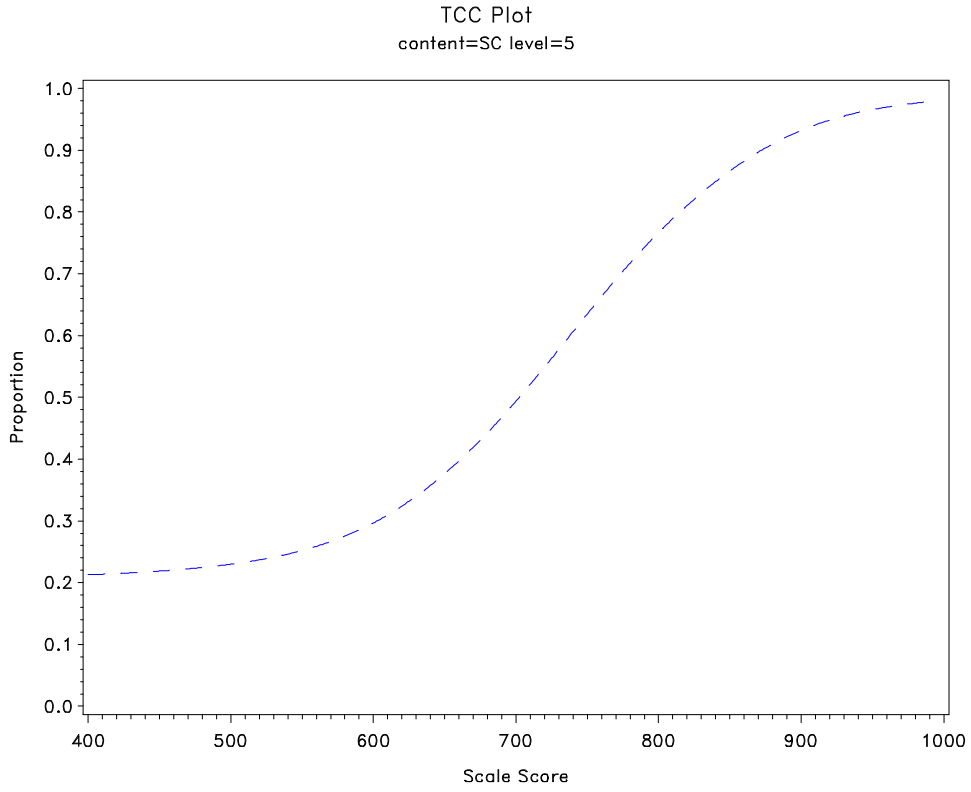


Figure 27. Spring 2013 Grades 5 Science operational test characteristic curve and standard error of measurement curve



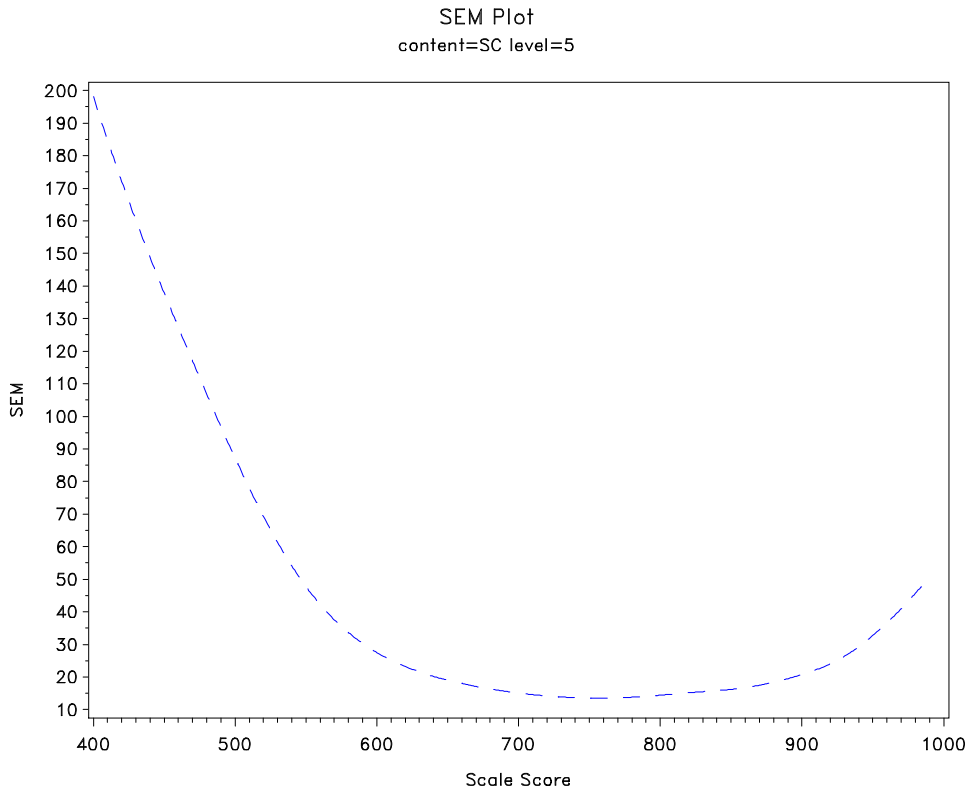


Figure 28. Spring 2013 Grades 8 Science operational test characteristic curve and standard error of measurement curve

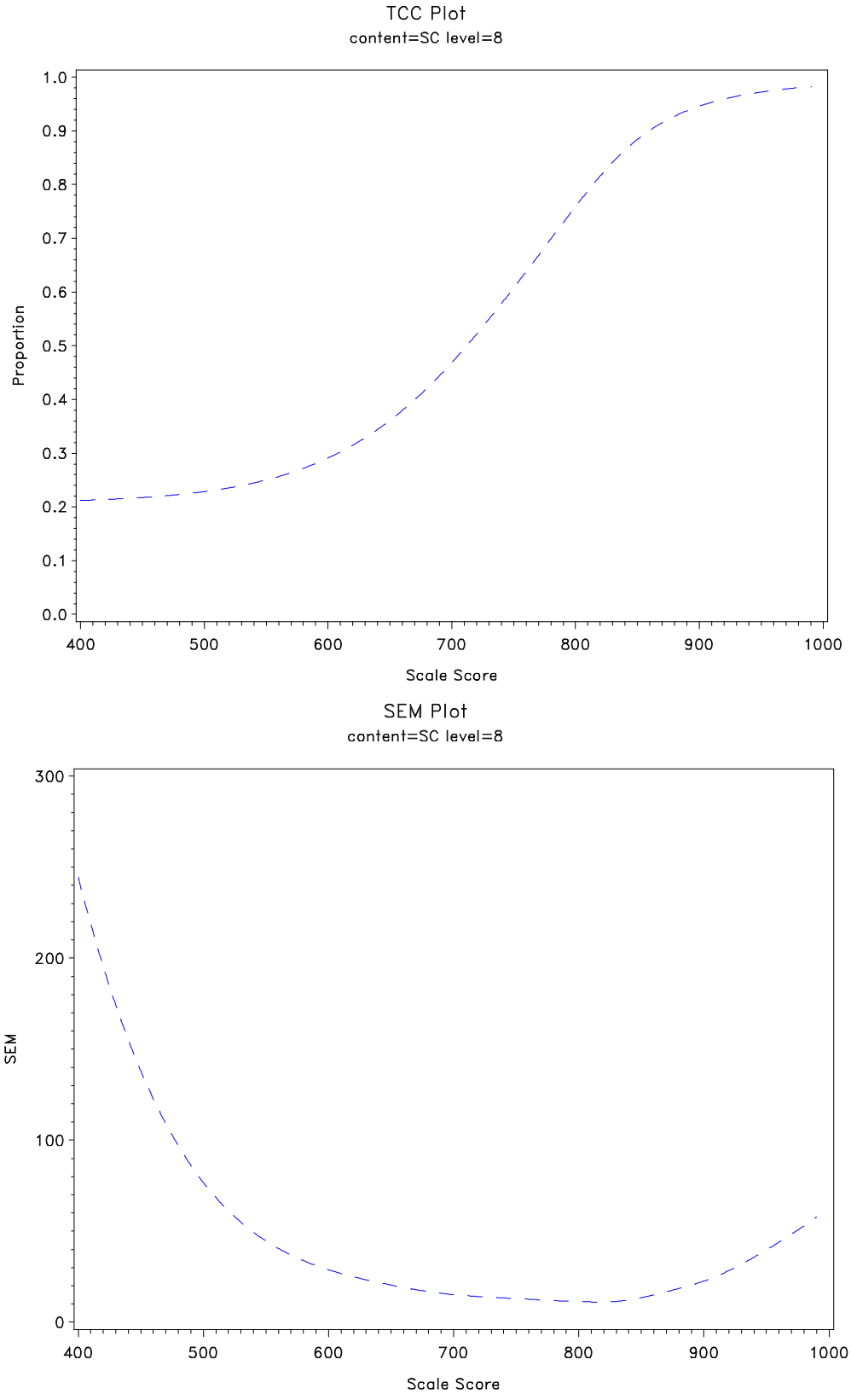


Figure 29. Spring 2013 Grade 3 Mathematics scree plot: Total

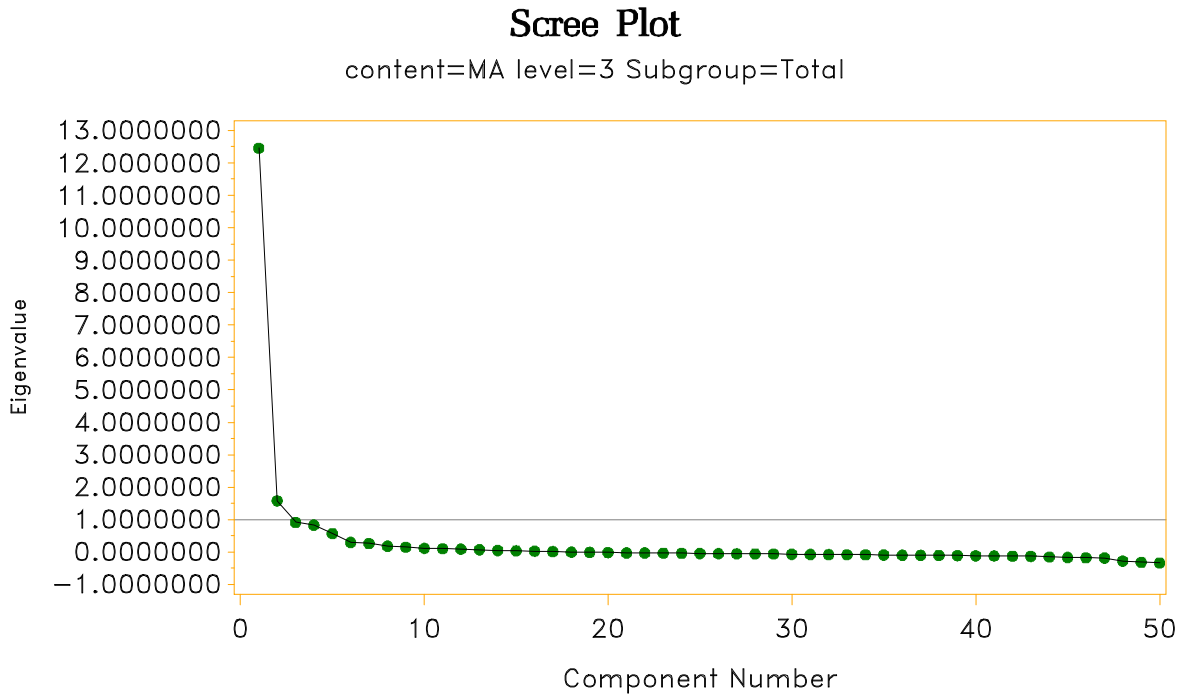


Figure 30. Spring 2013 Grade 3 Mathematics scree plot: Accommodated

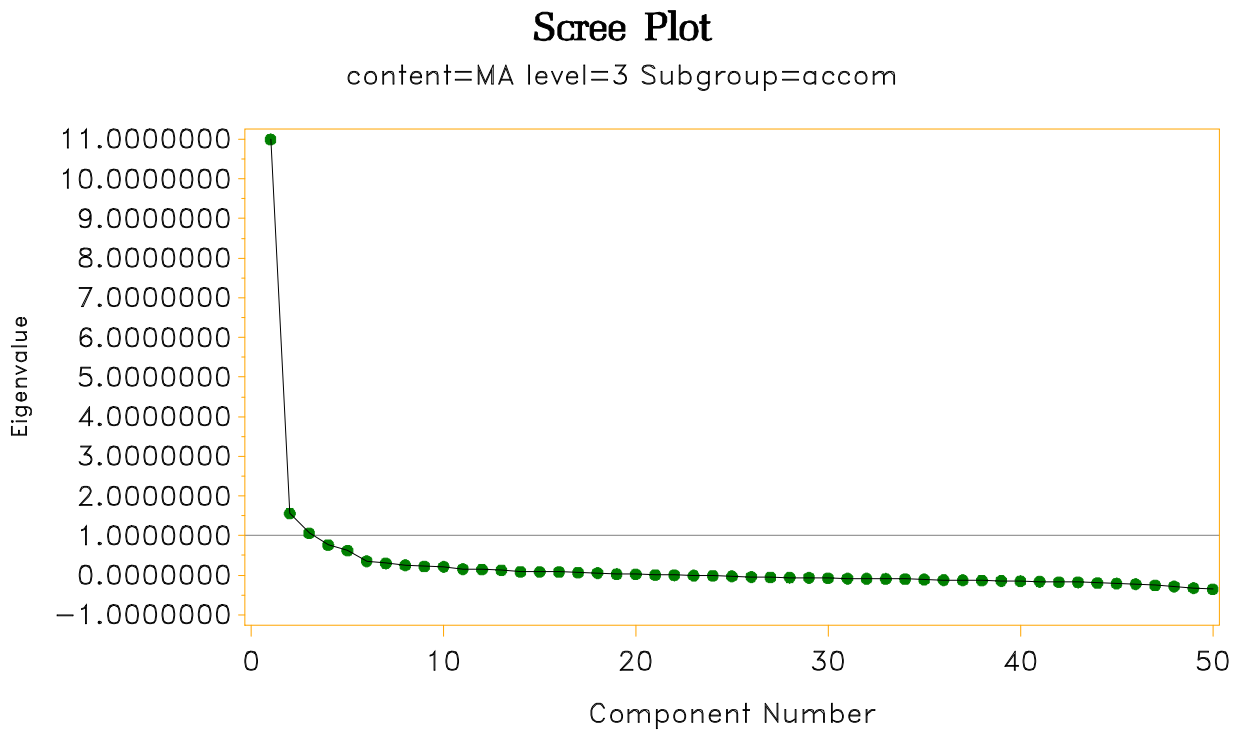


Figure 31. Spring 2013 Grade 3 Mathematics scree plot: English Language Learner

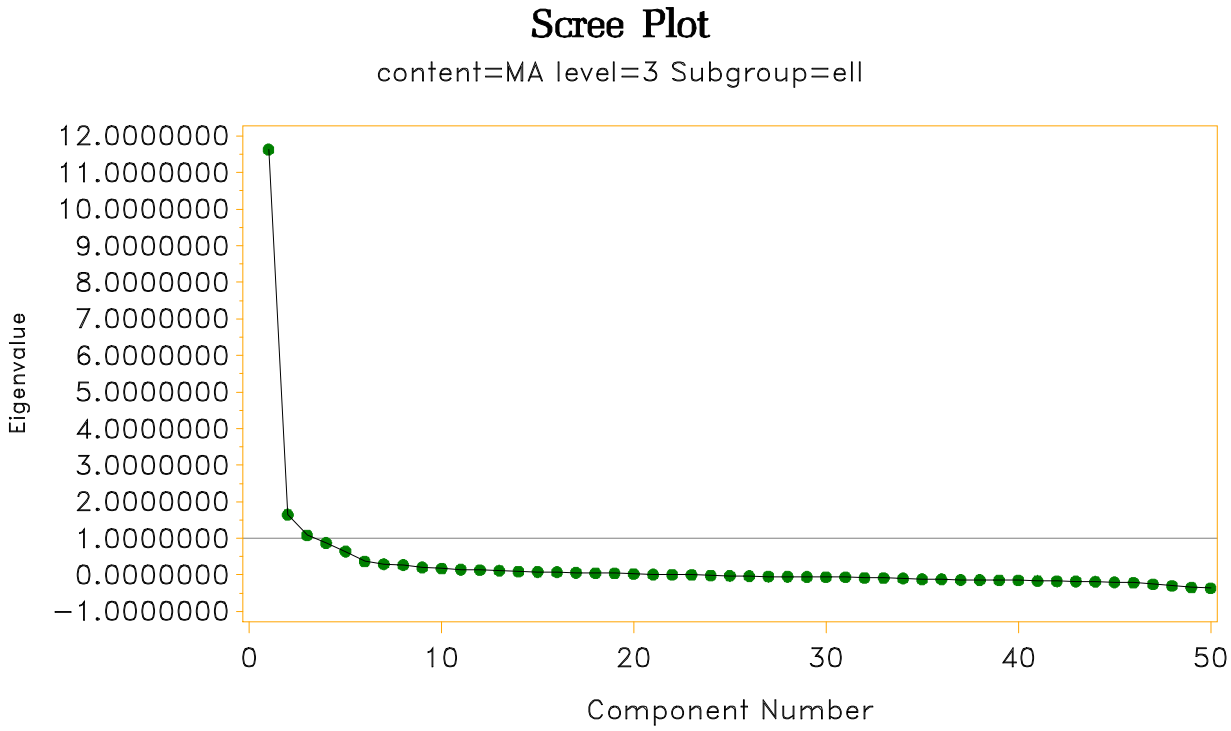


Figure 32. Spring 2013 Grade 3 Mathematics scree plot: Free Lunch

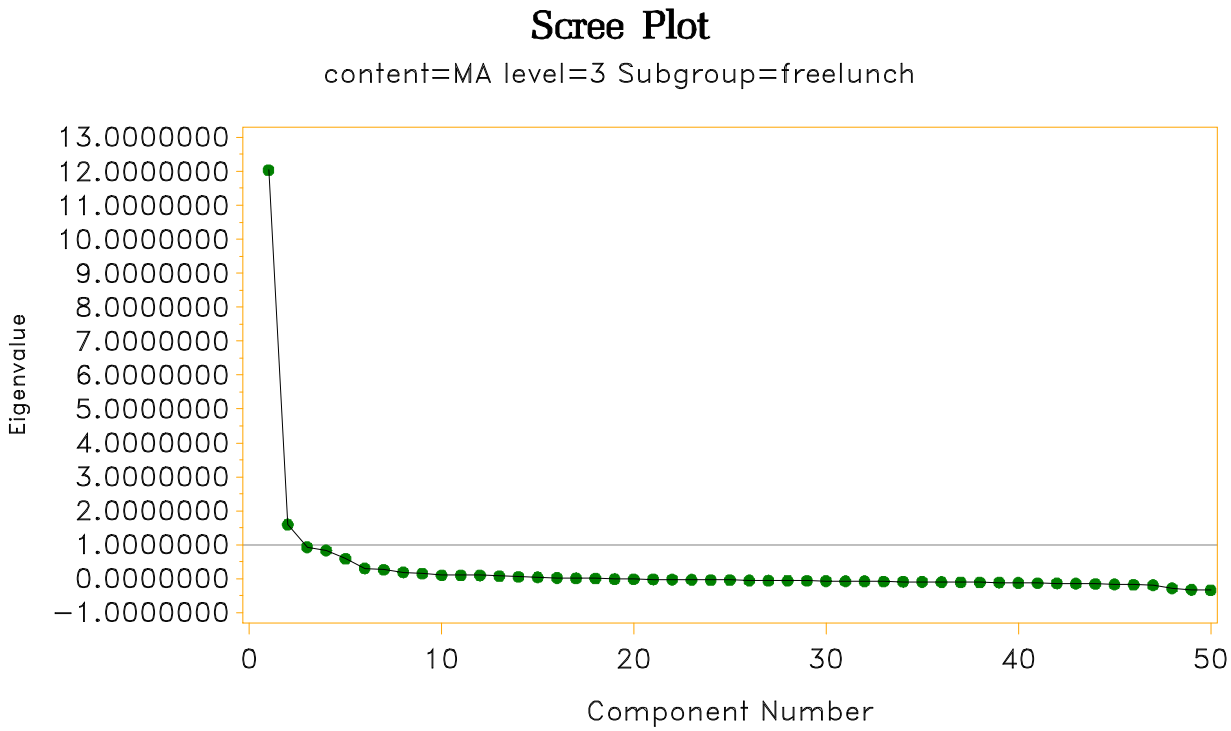


Figure 33. Spring 2013 Grade 3 Mathematics scree plot: Individualized Education Program

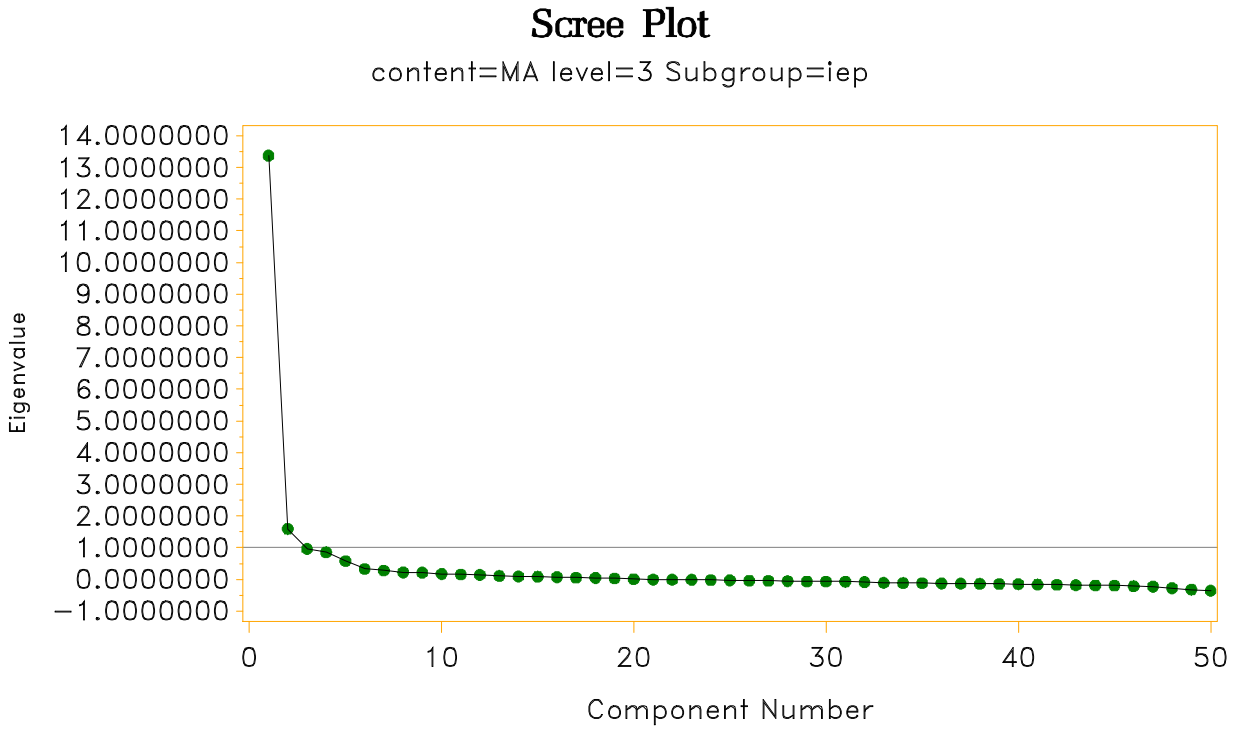


Figure 34. Spring 2013 Grade 4 Mathematics scree plot: Total

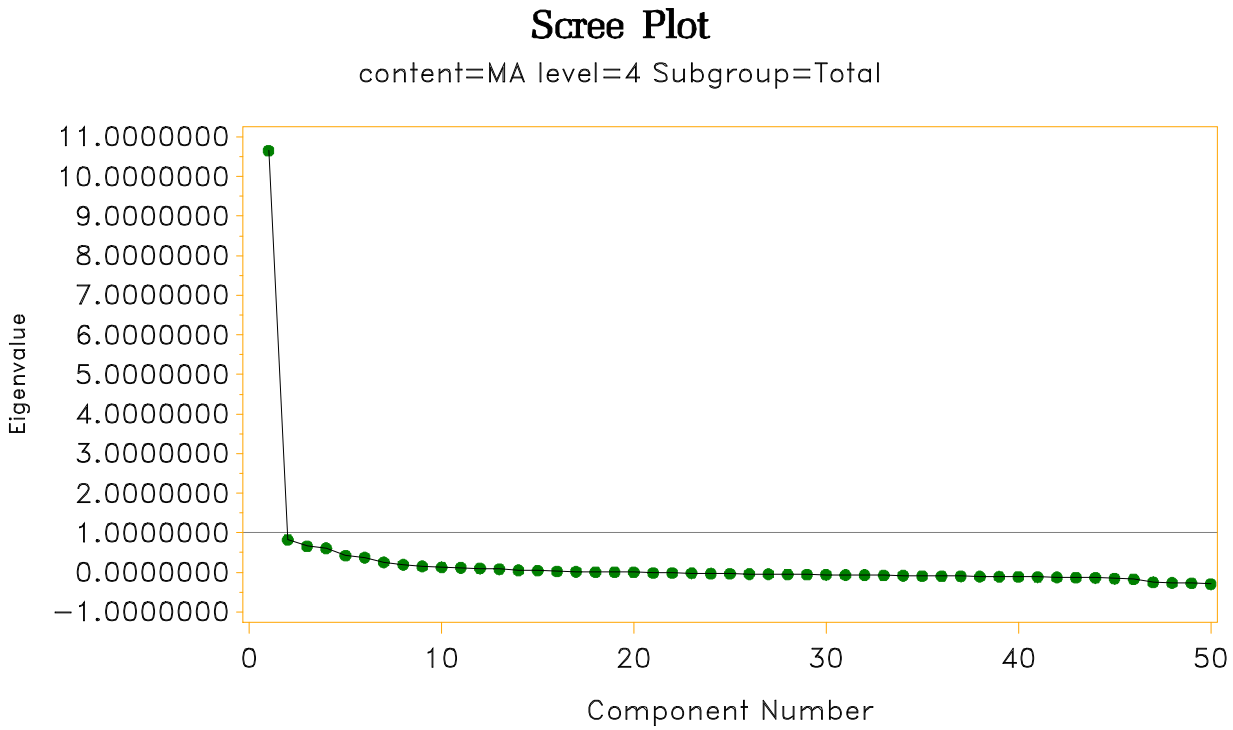


Figure 35. Spring 2013 Grade 4 Mathematics scree plot: Accommodated

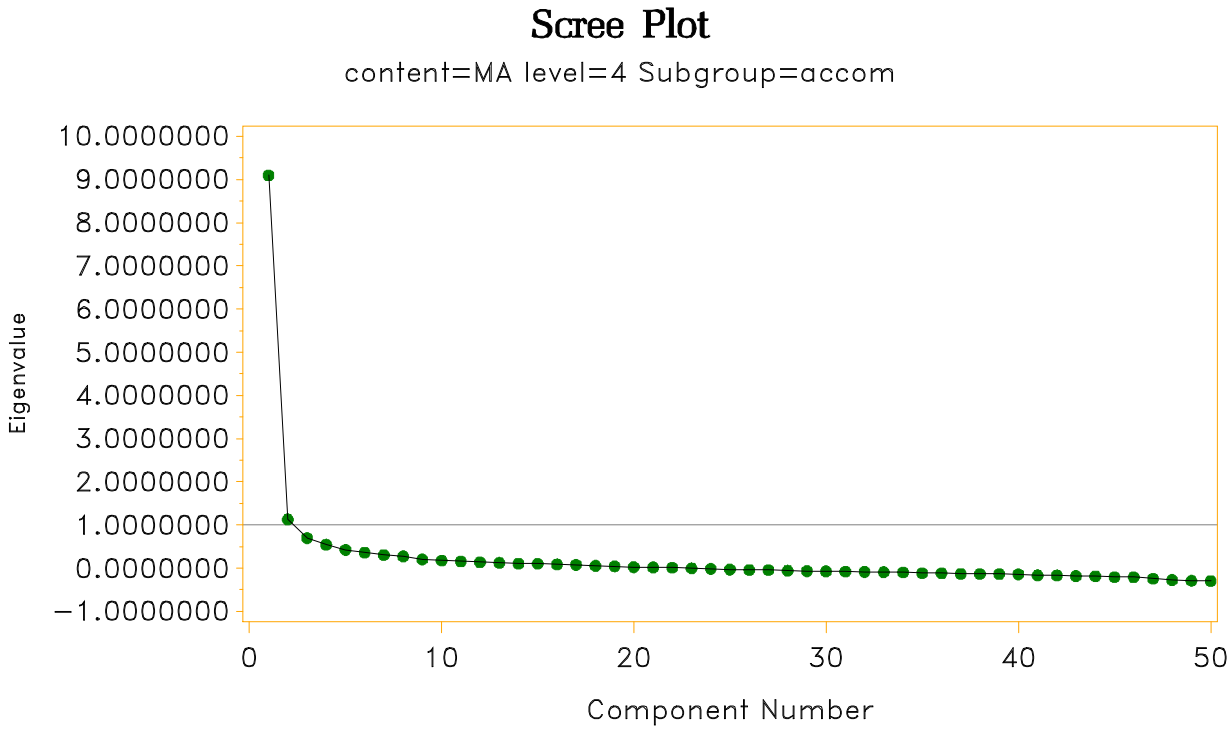


Figure 36. Spring 2013 Grade 4 Mathematics scree plot: English Language Learner

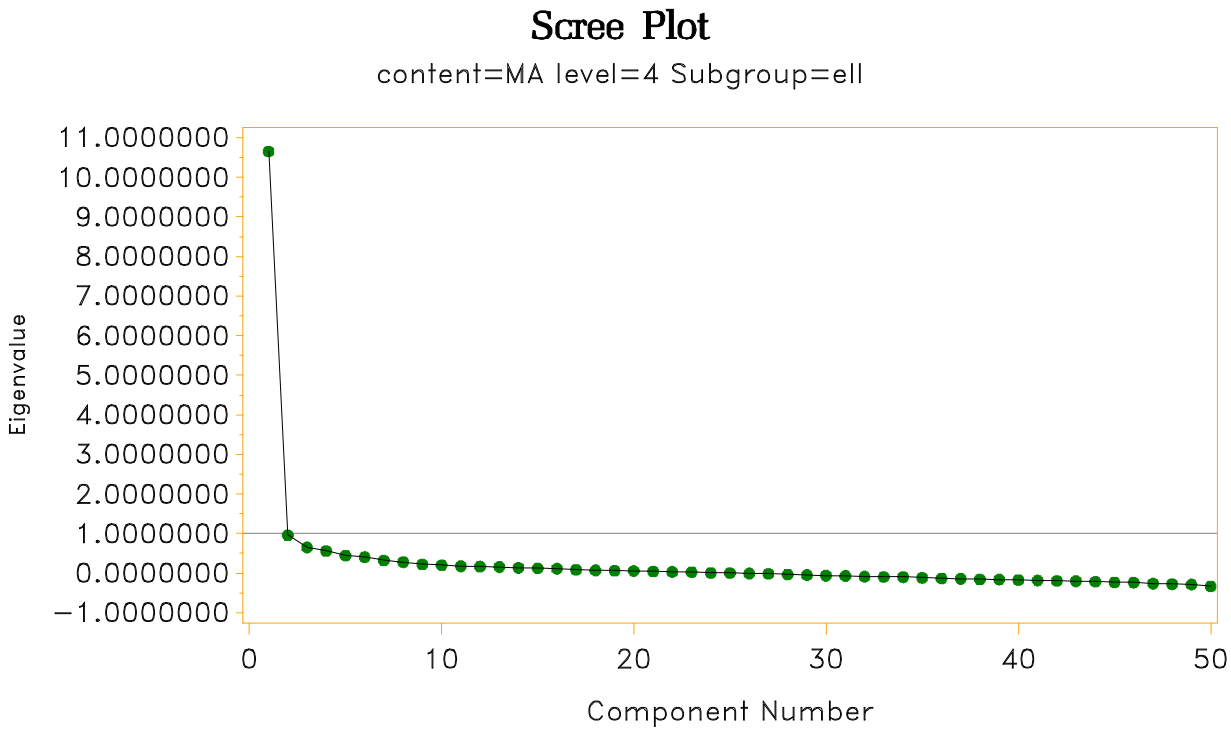


Figure 37. Spring 2013 Grade 4 Mathematics scree plot: Free Lunch

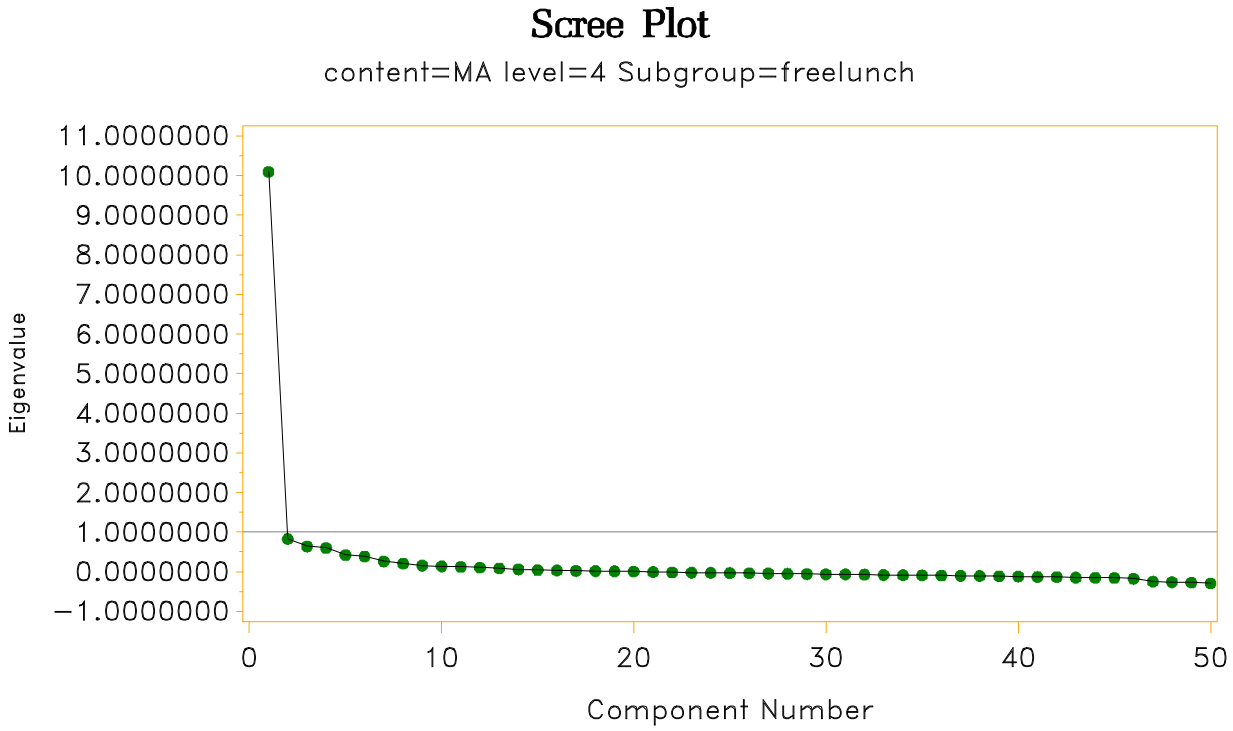


Figure 38. Spring 2013 Grade 4 Mathematics scree plot: Individualized Education Program

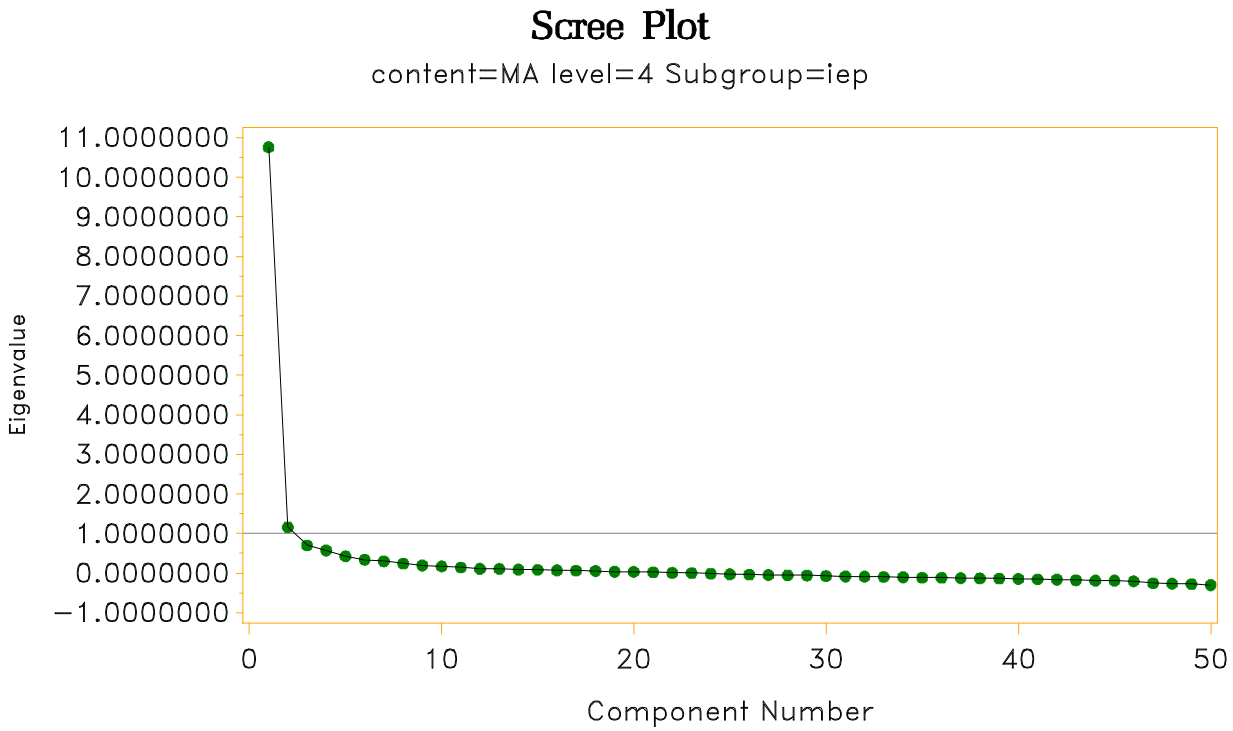


Figure 39. Spring 2013 Grade 5 Mathematics scree plot: Total

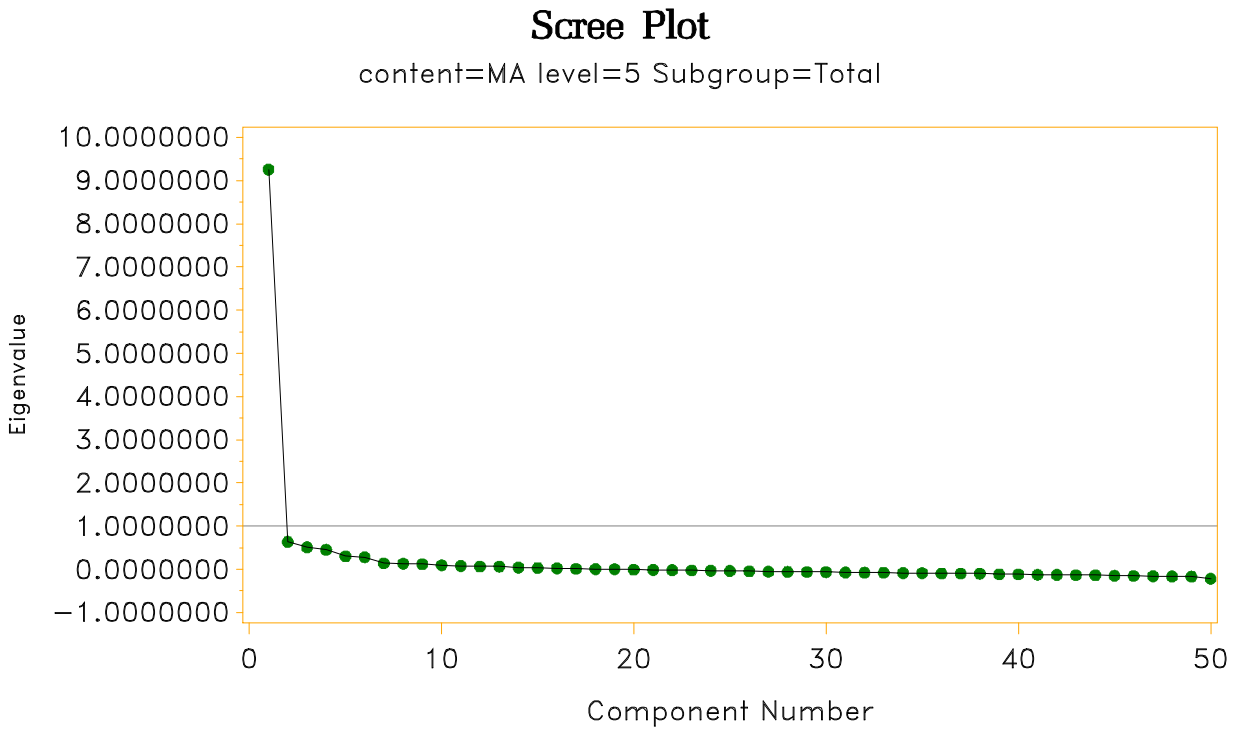


Figure 40. Spring 2013 Grade 5 Mathematics scree plot: Accommodated

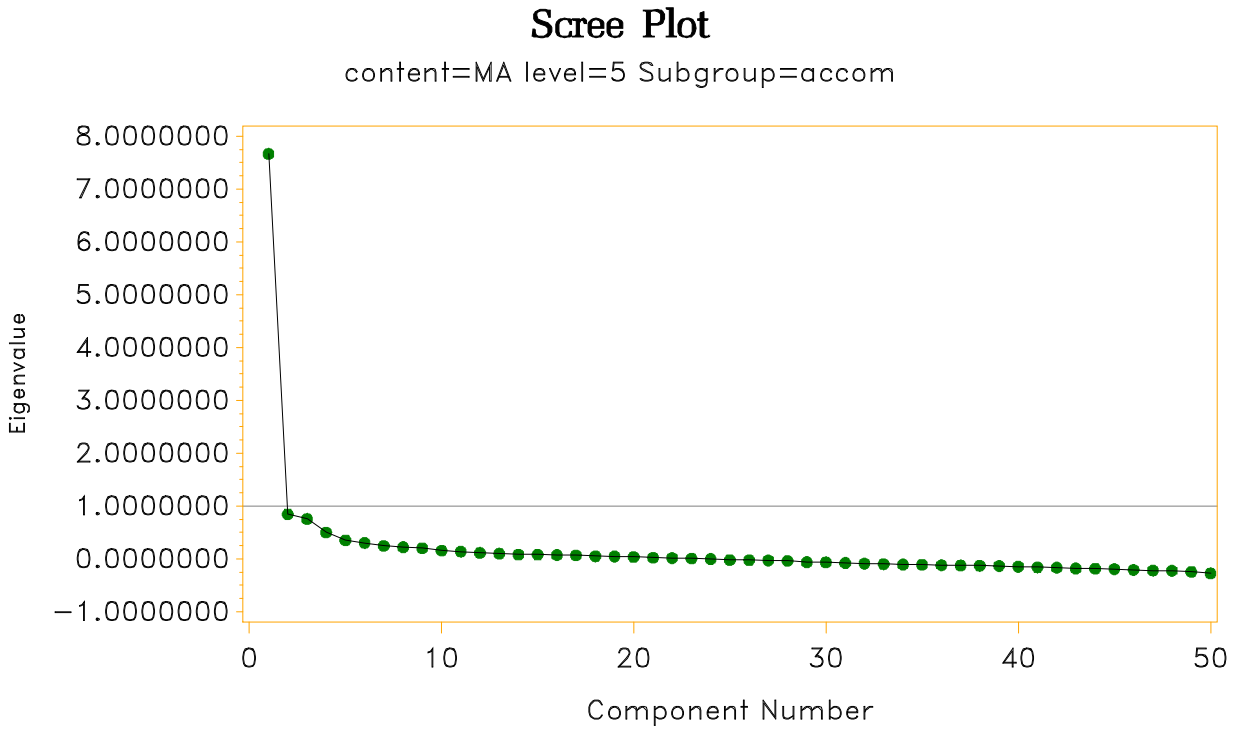


Figure 41. Spring 2013 Grade 5 Mathematics scree plot: English Language Learner

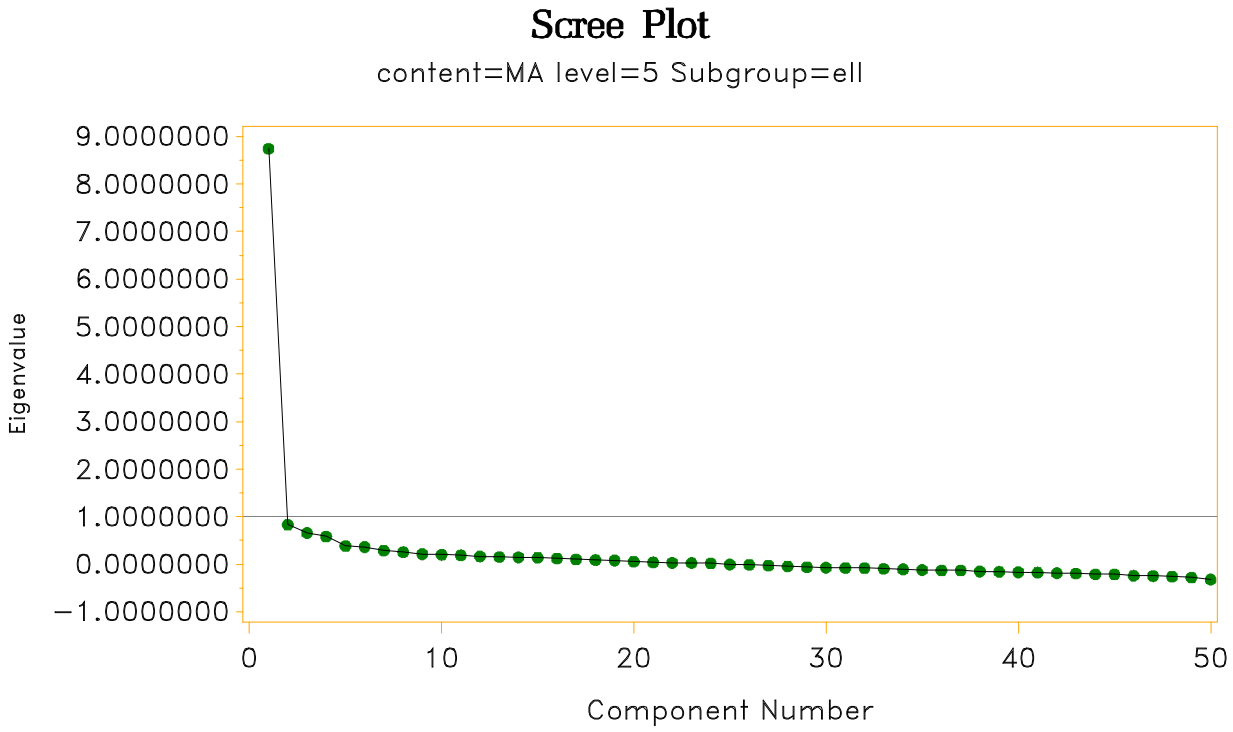


Figure 42. Spring 2013 Grade 5 Mathematics scree plot: Free Lunch

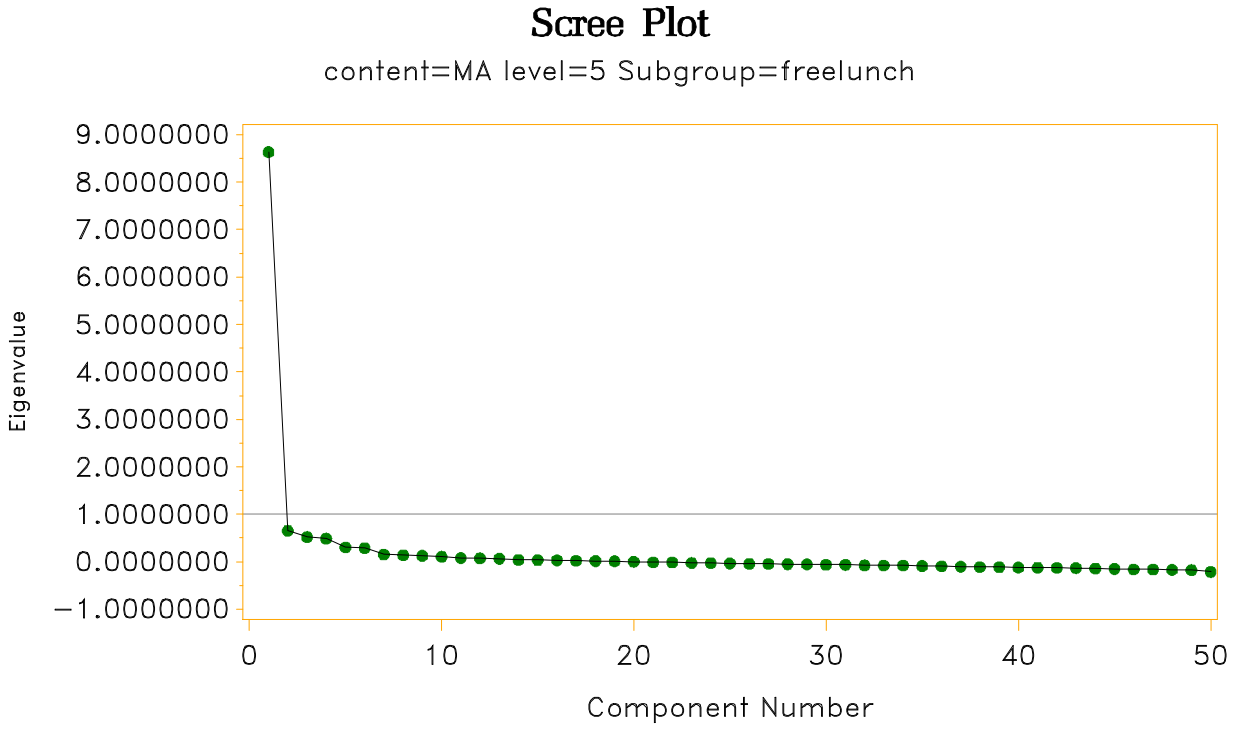


Figure 43. Spring 2013 Grade 5 Mathematics scree plot: Individualized Education Program

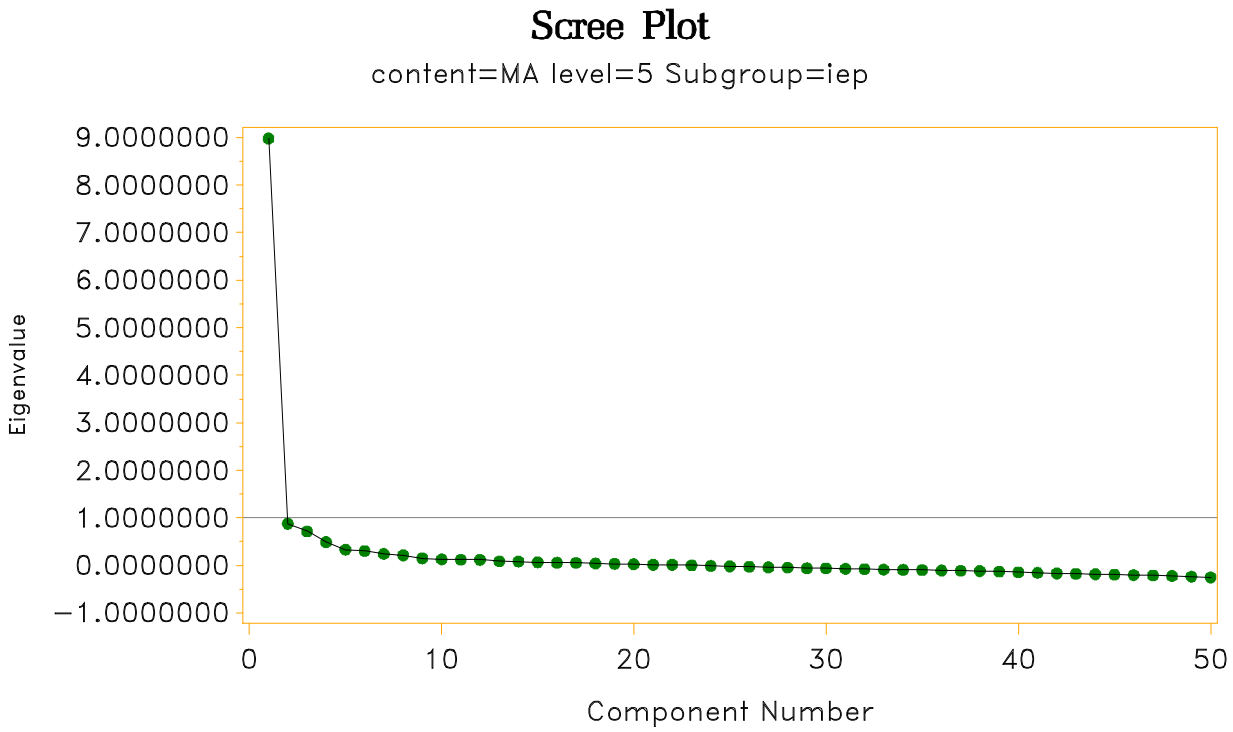


Figure 44. Spring 2013 Grade 6 Mathematics scree plot: Total

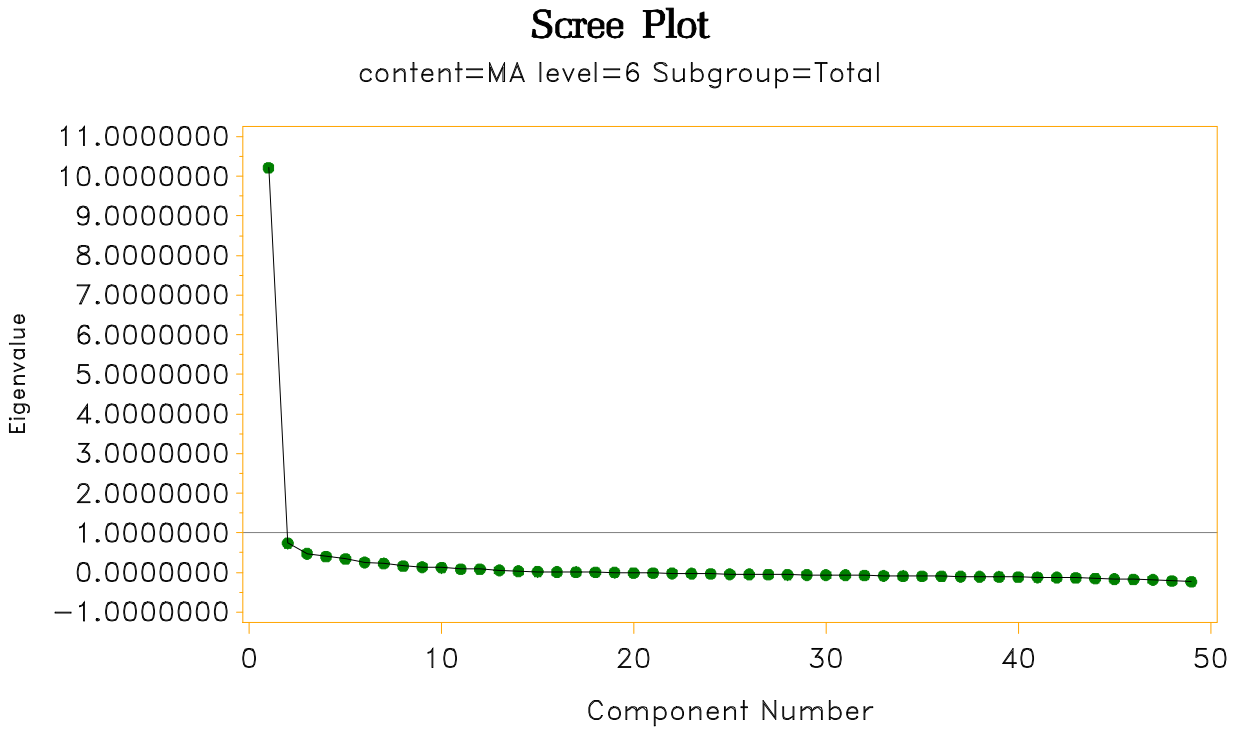


Figure 45. Spring 2013 Grade 6 Mathematics scree plot: Accommodated

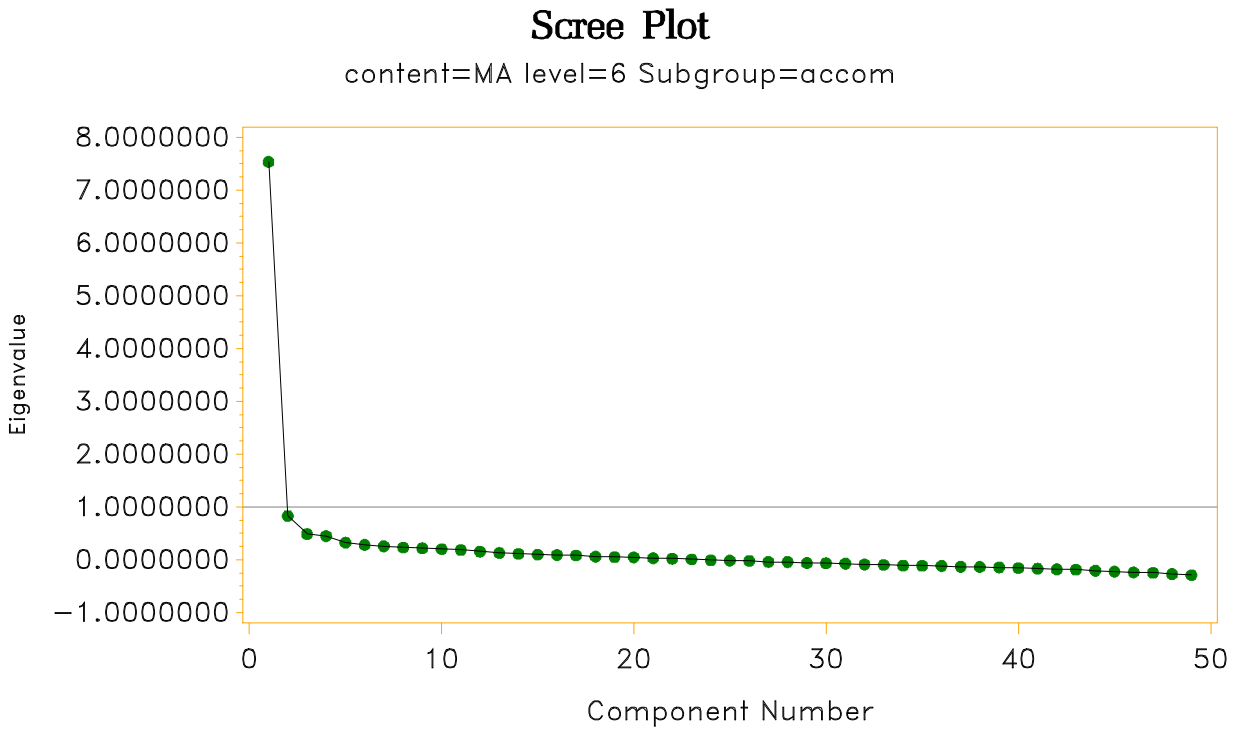


Figure 46. Spring 2013 Grade 6 Mathematics scree plot: English Language Learner

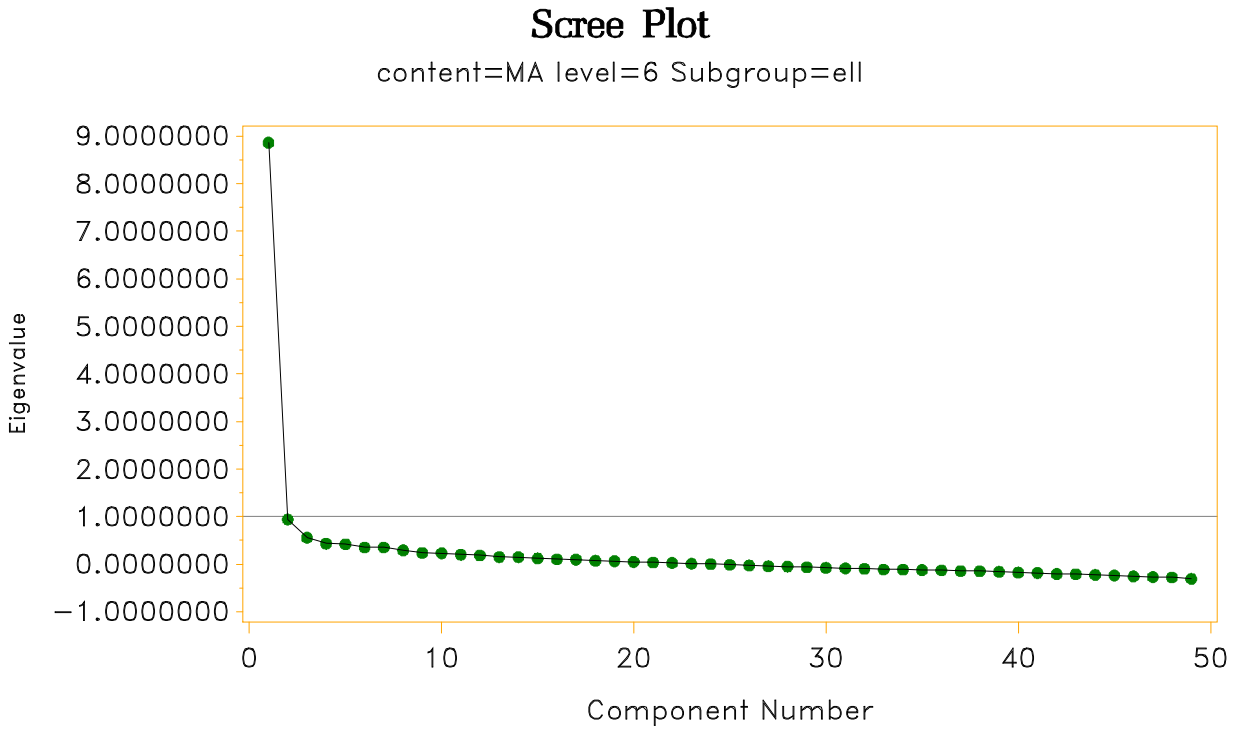


Figure 47. Spring 2013 Grade 6 Mathematics scree plot: Free Lunch

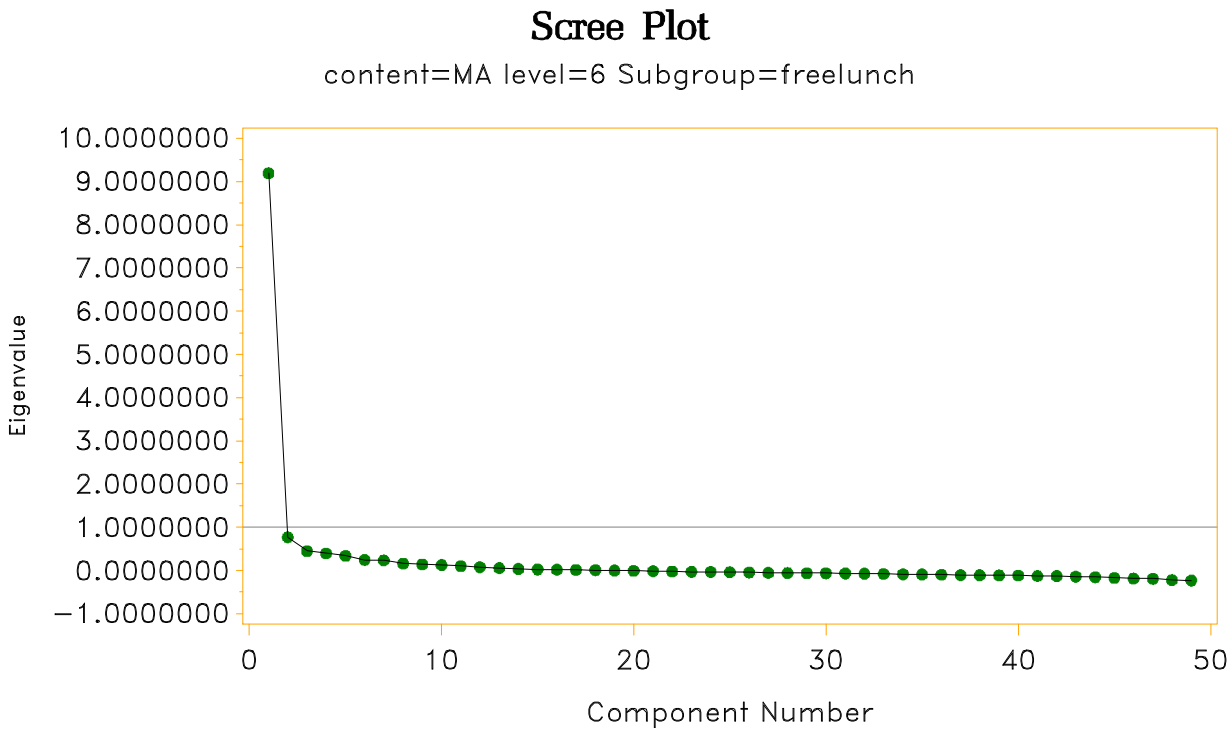


Figure 48. Spring 2013 Grade 6 Mathematics scree plot: Individualized Education Program

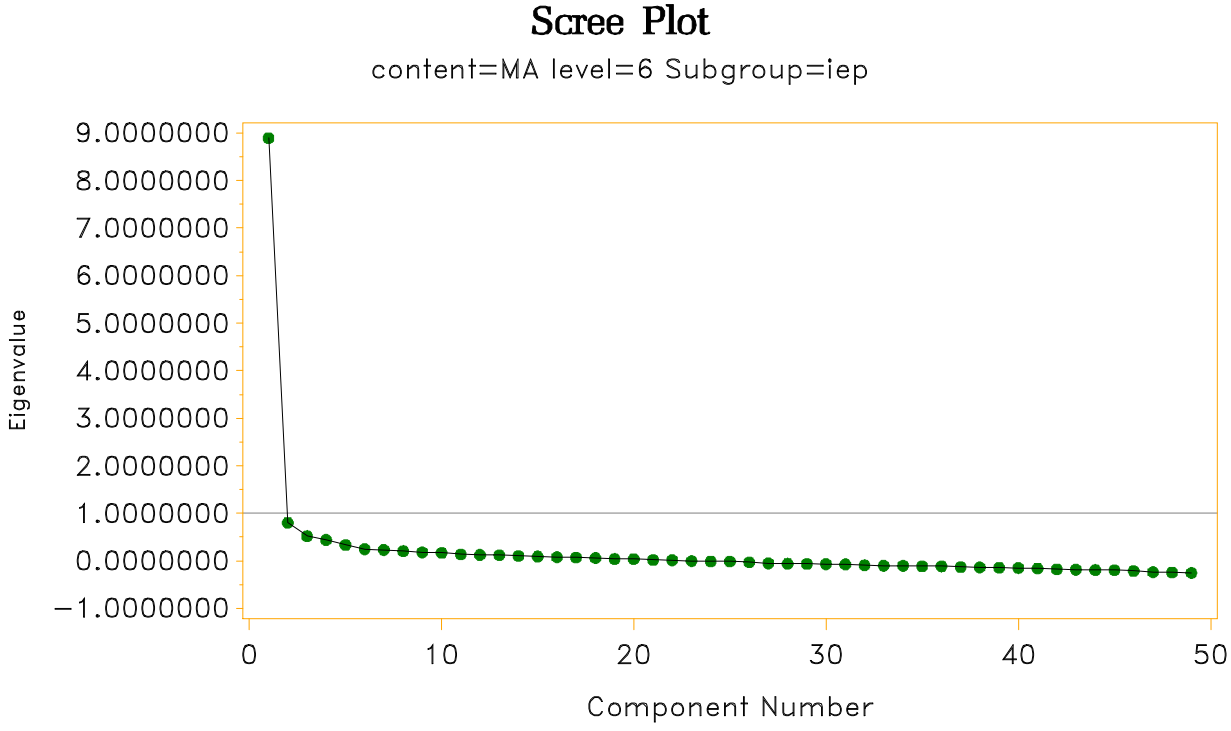


Figure 49. Spring 2013 Grade 7 Mathematics scree plot: Total

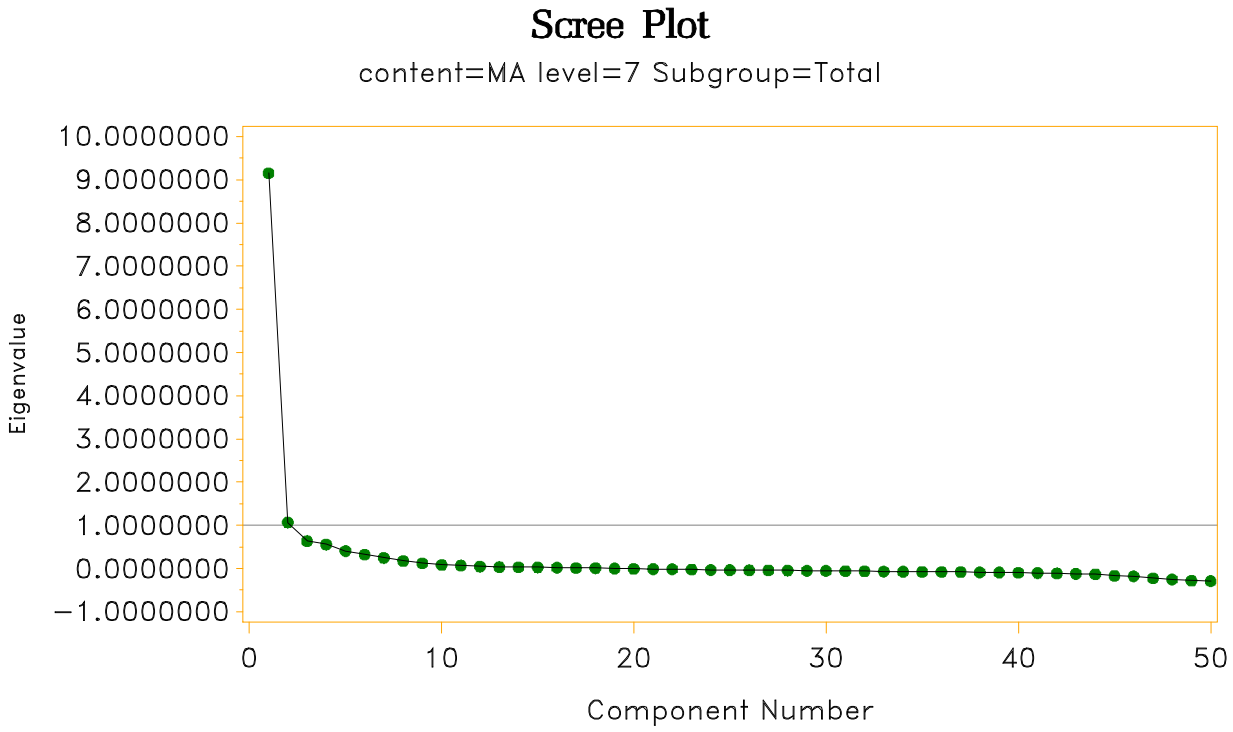


Figure 50. Spring 2013 Grade 7 Mathematics scree plot: Accommodated

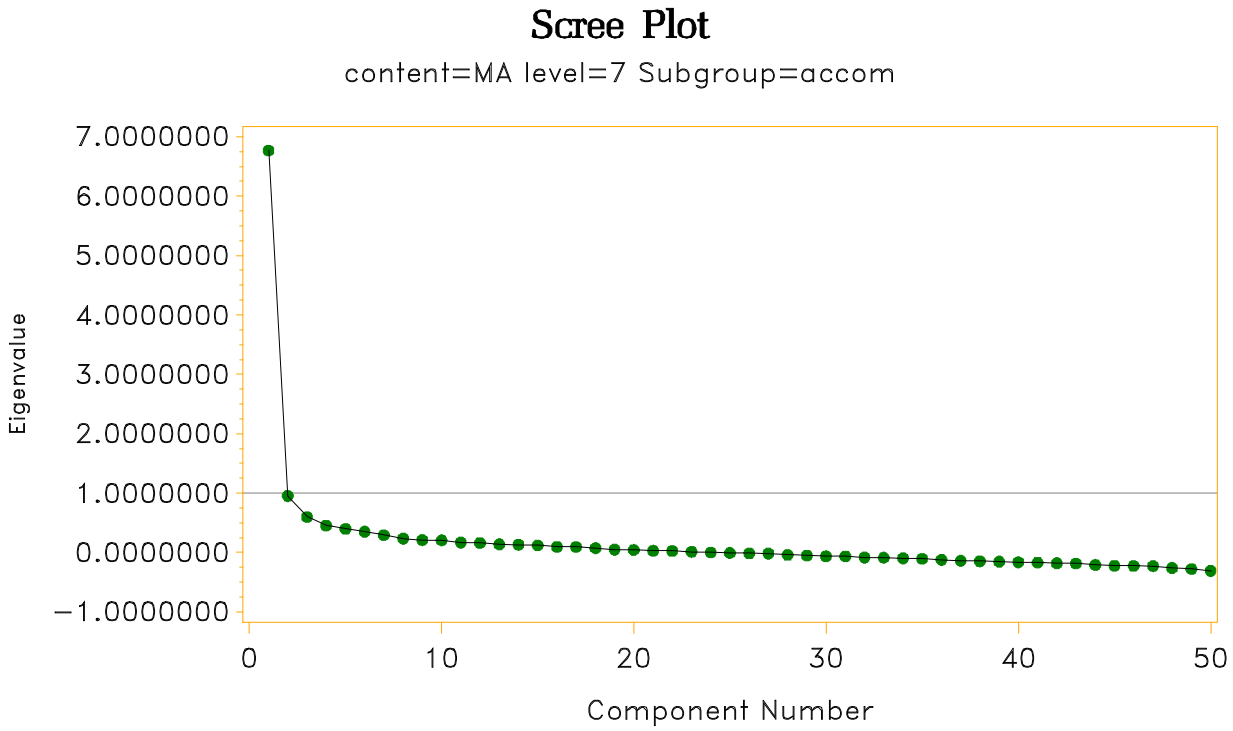


Figure 51. Spring 2013 Grade 7 Mathematics scree plot: English Language Learner

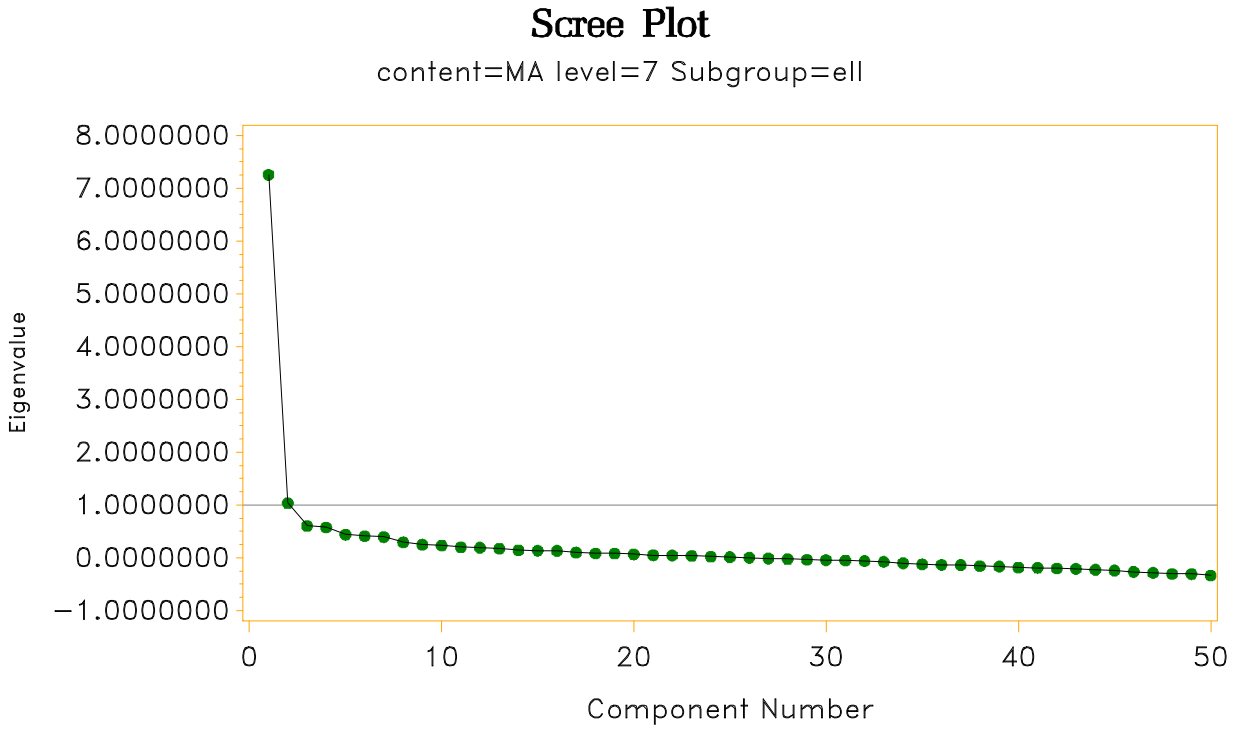


Figure 52. Spring 2013 Grade 7 Mathematics scree plot: Free Lunch

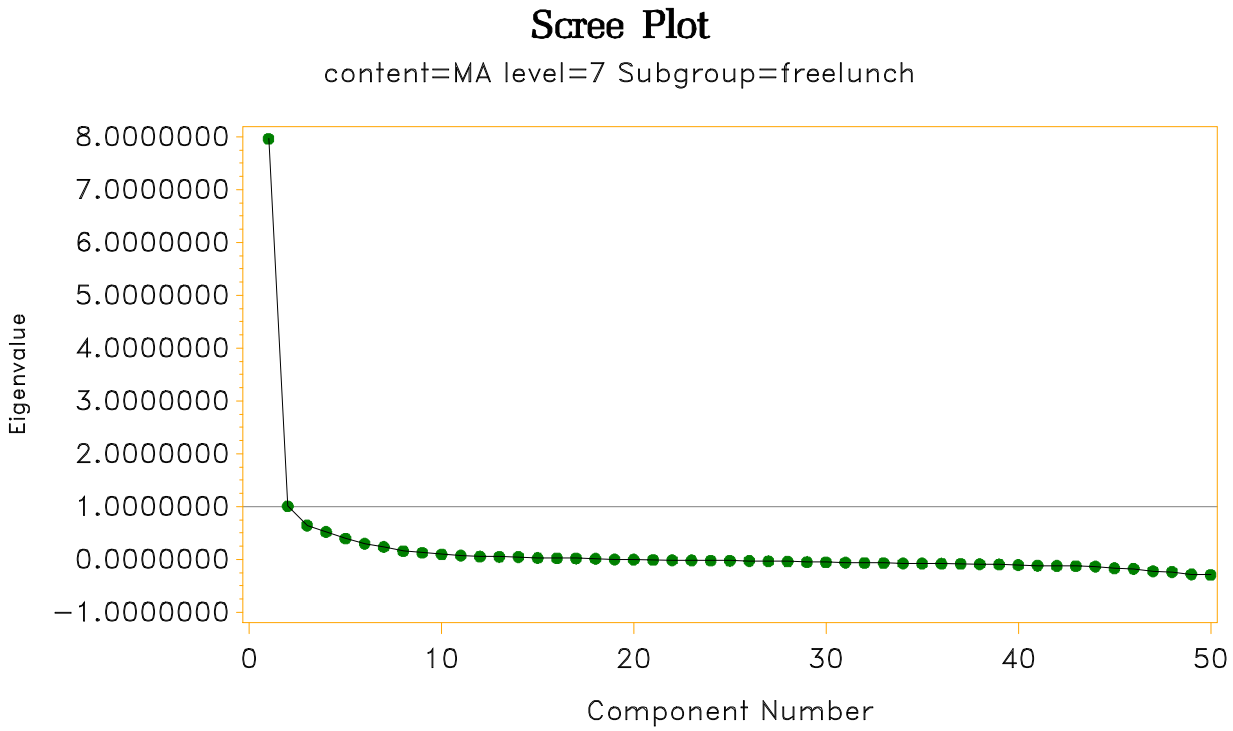


Figure 53. Spring 2013 Grade 7 Mathematics scree plot: Individualized Education Program

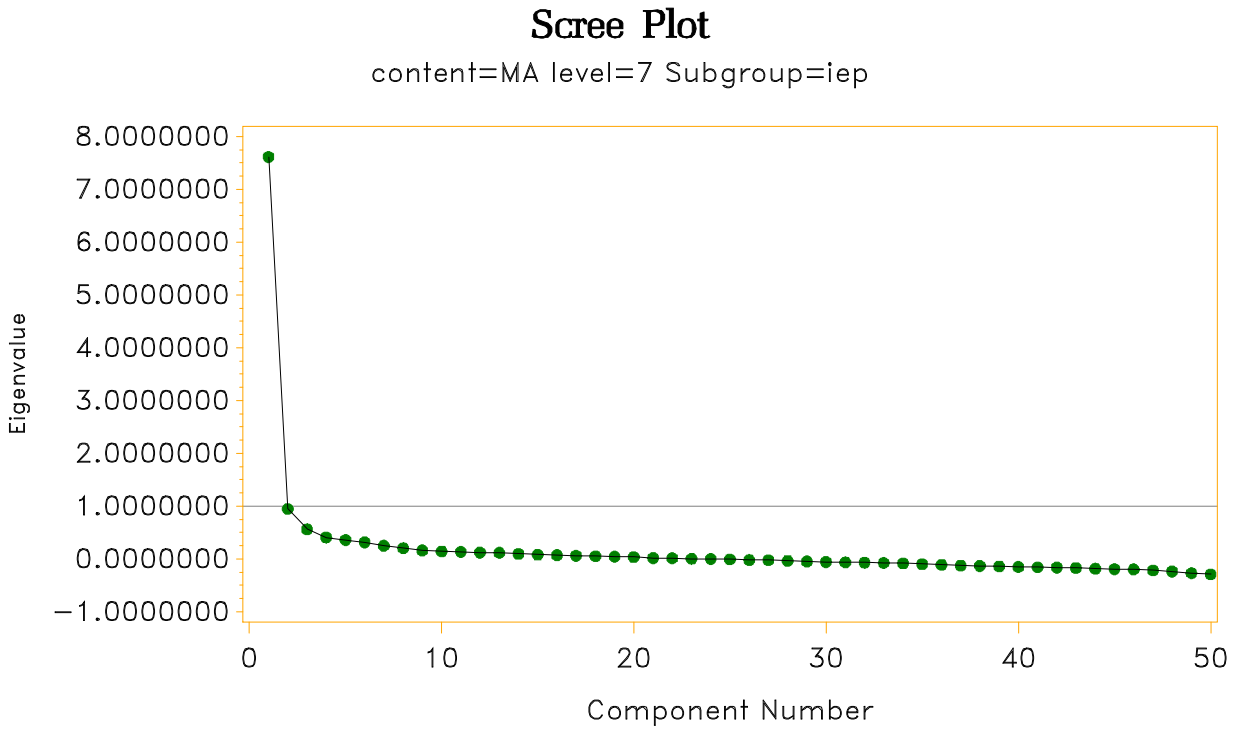


Figure 54. Spring 2013 Grade 8 Mathematics scree plot: Total

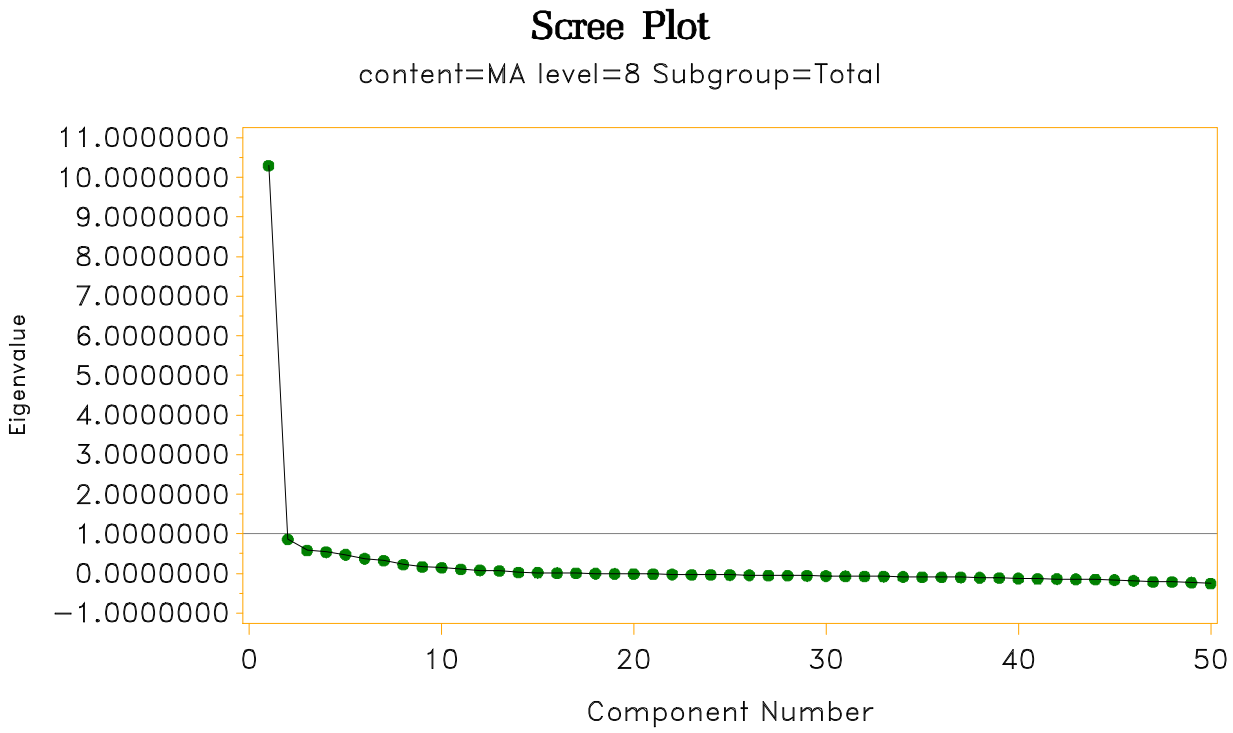


Figure 55. Spring 2013 Grade 8 Mathematics scree plot: Accommodated

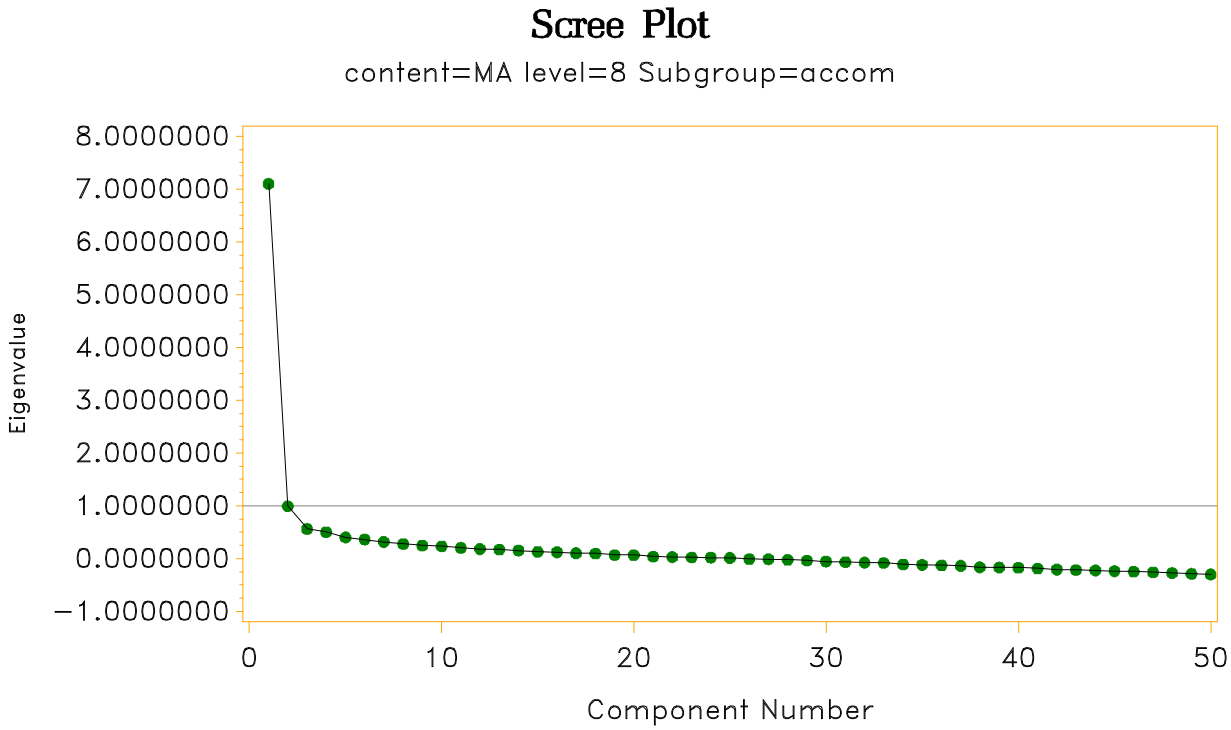


Figure 56. Spring 2013 Grade 8 Mathematics scree plot: English Language Learner

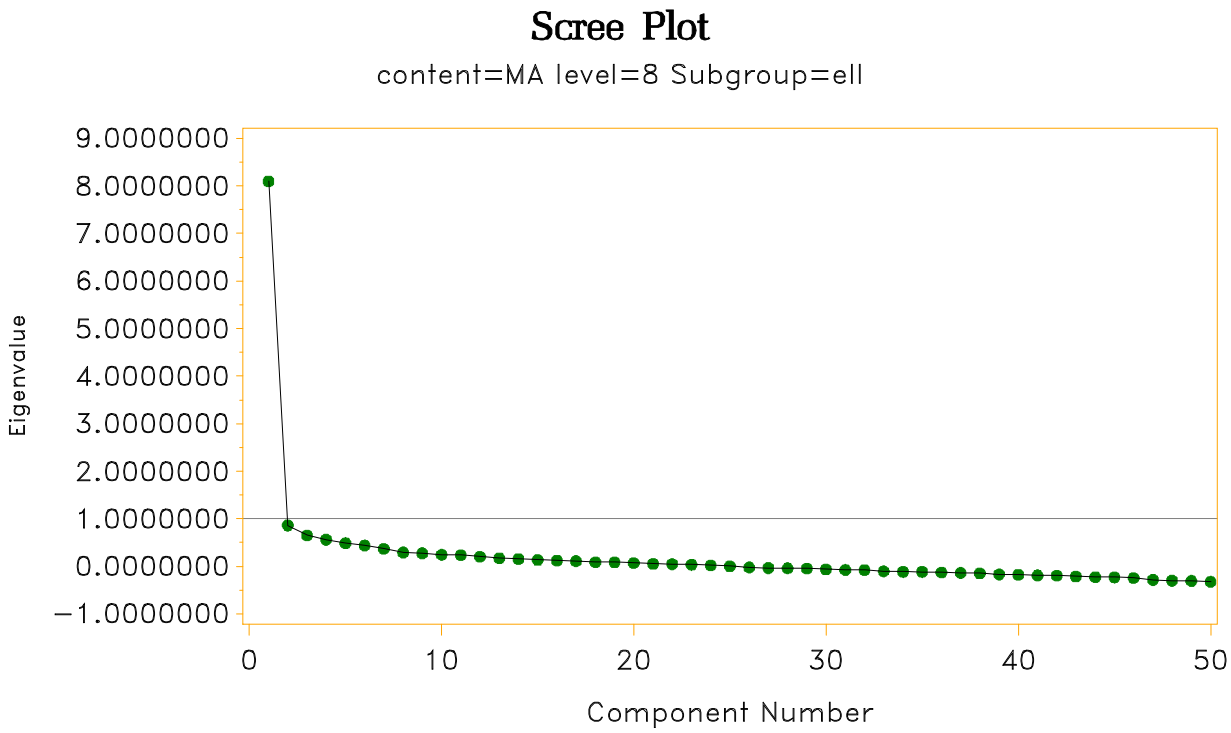


Figure 57. Spring 2013 Grade 8 Mathematics scree plot: Free Lunch

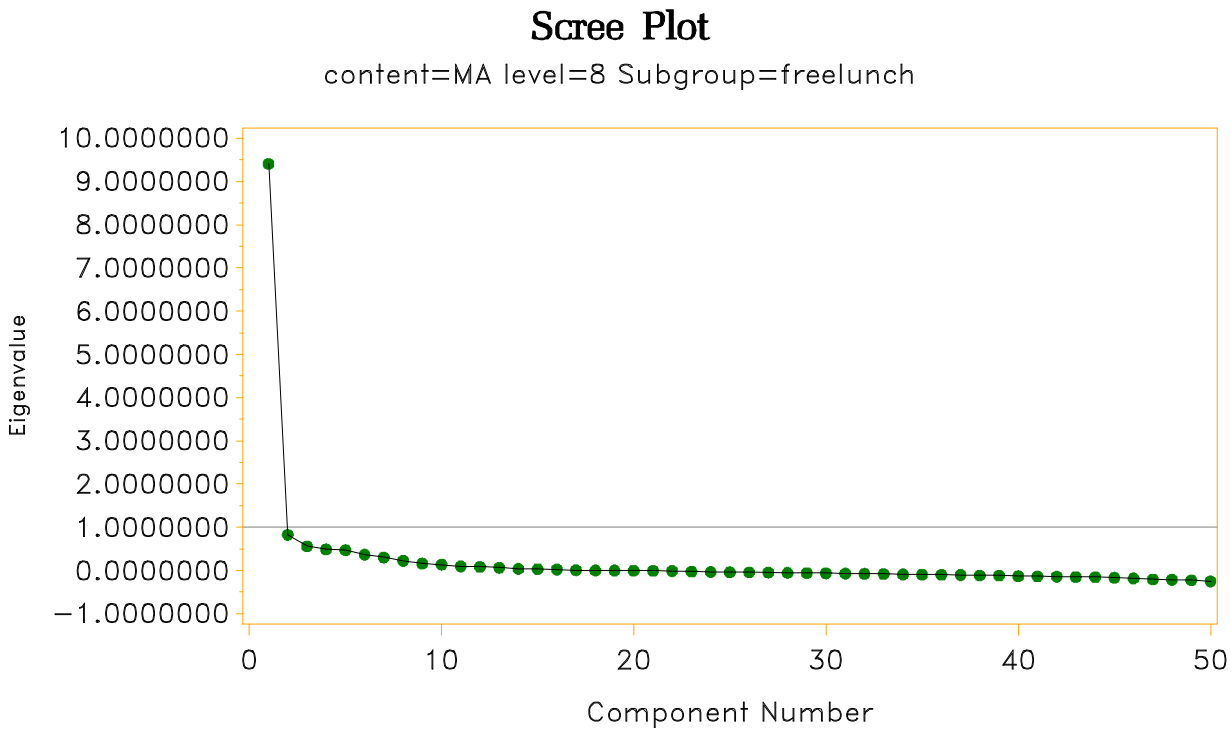


Figure 58. Spring 2013 Grade 8 Mathematics scree plot: Individualized Education Program

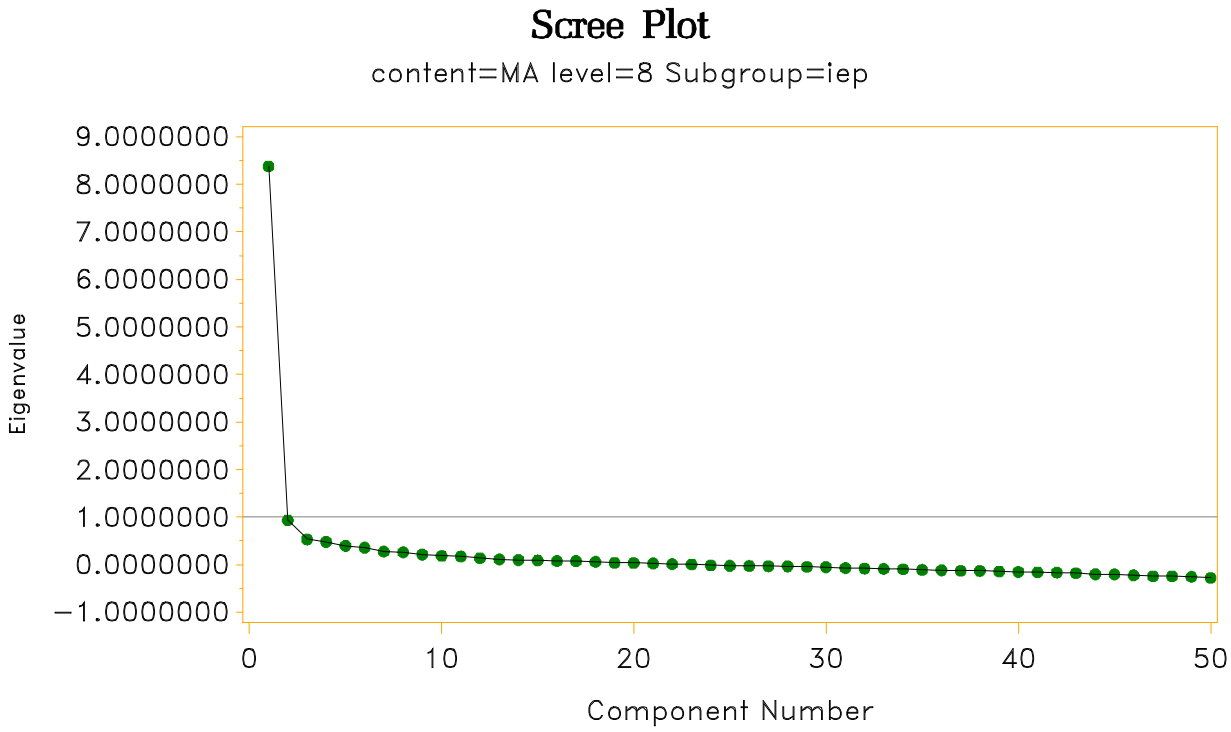


Figure 59. Spring 2013 Grade 3 Reading scree plot: Total

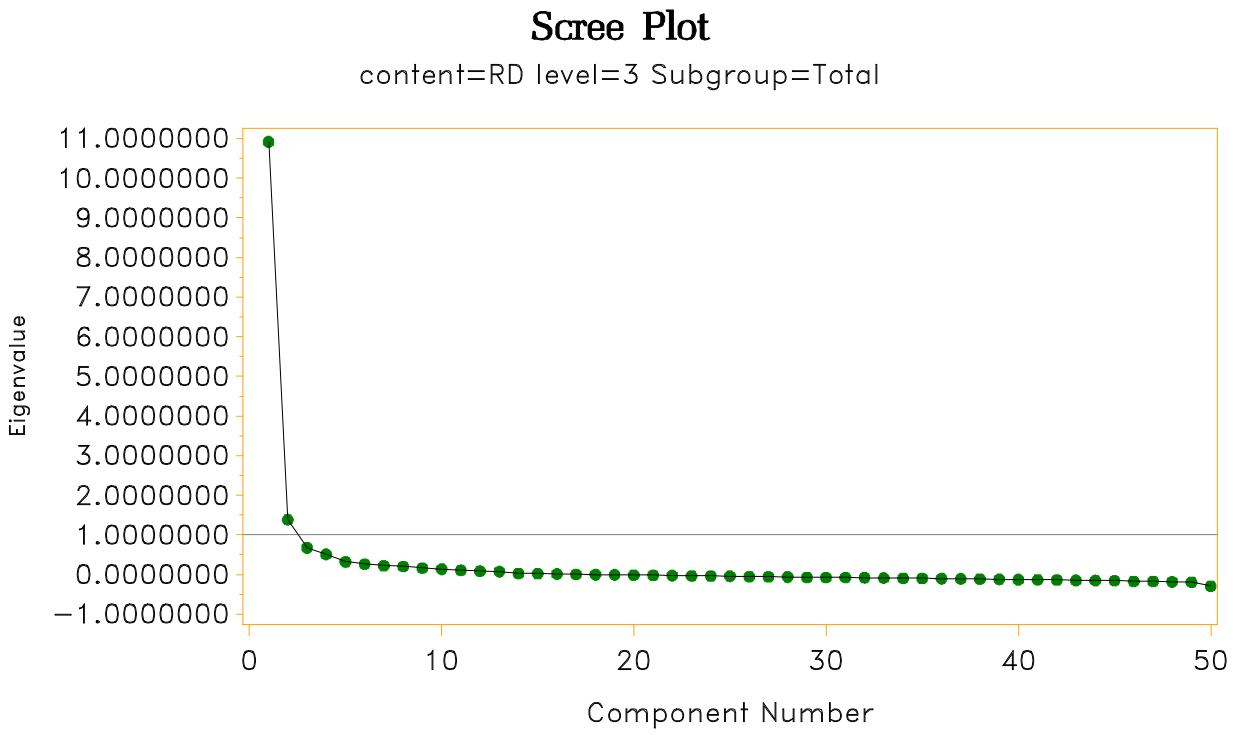


Figure 60. Spring 2013 Grade 3 Reading scree plot: Accommodated

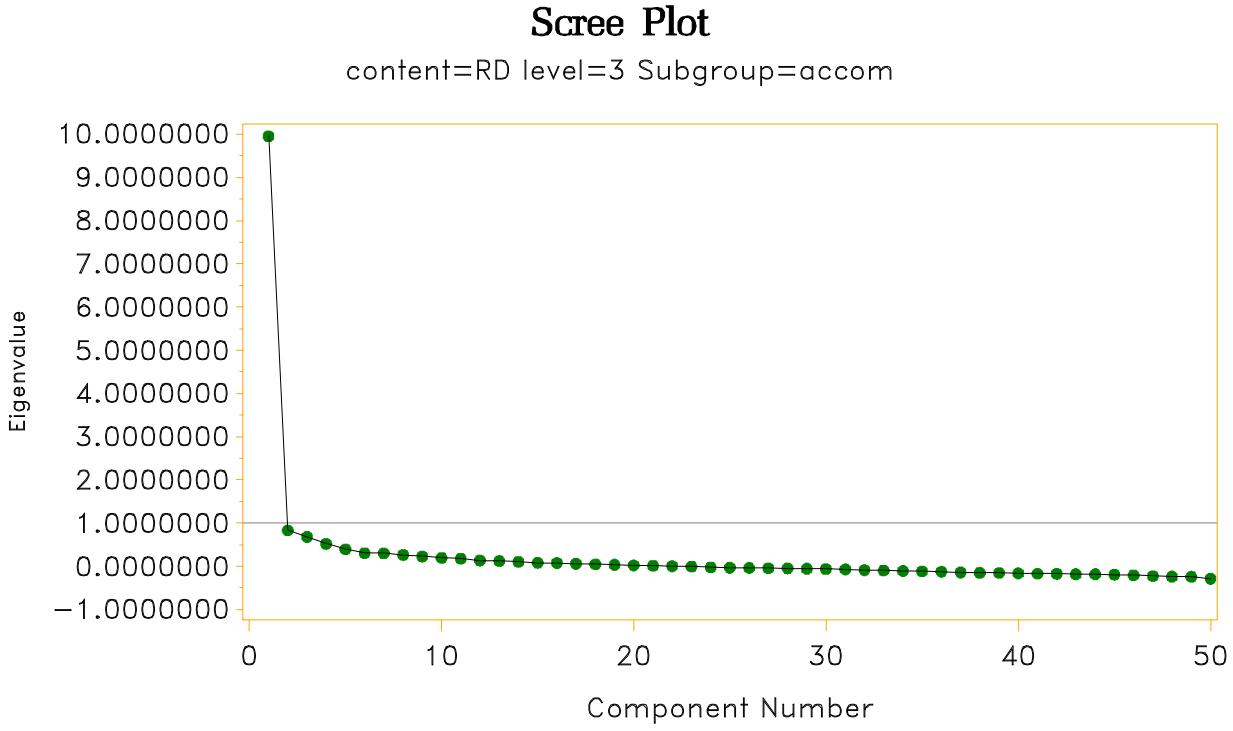


Figure 61. Spring 2013 Grade 3 Reading scree plot: English Language Learner

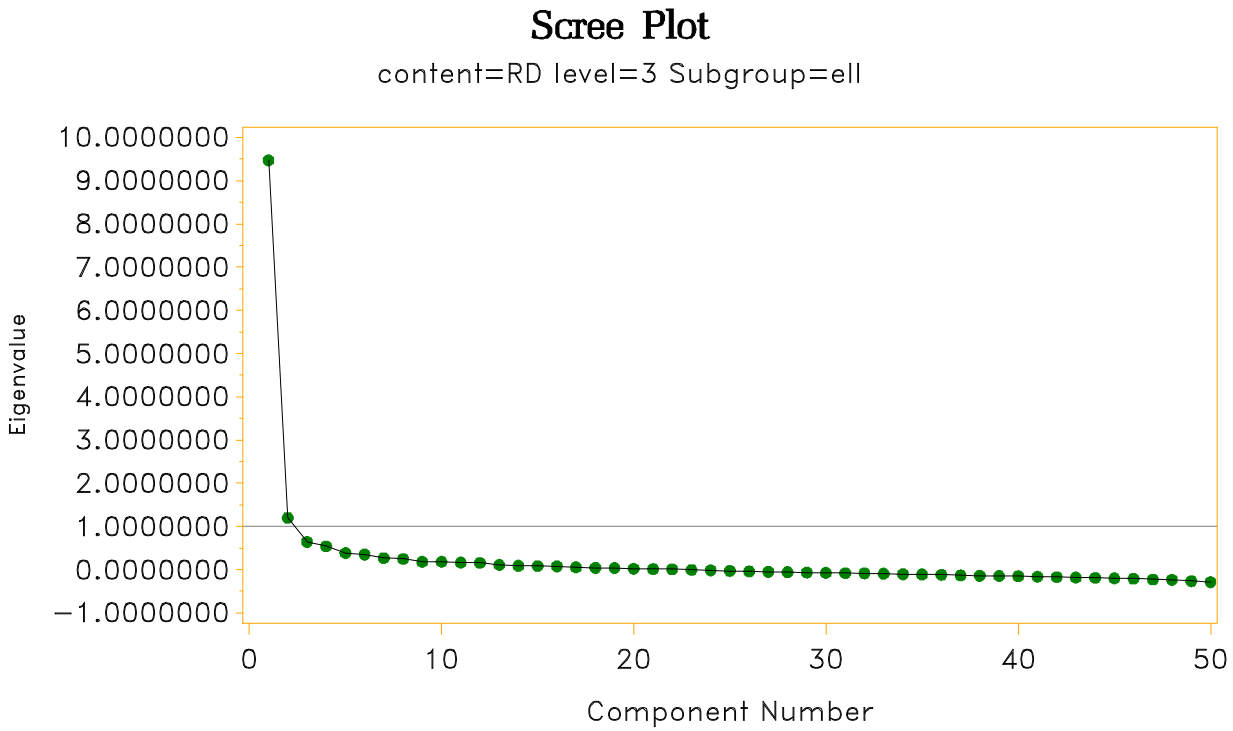


Figure 62. Spring 2013 Grade 3 Reading scree plot: Free Lunch

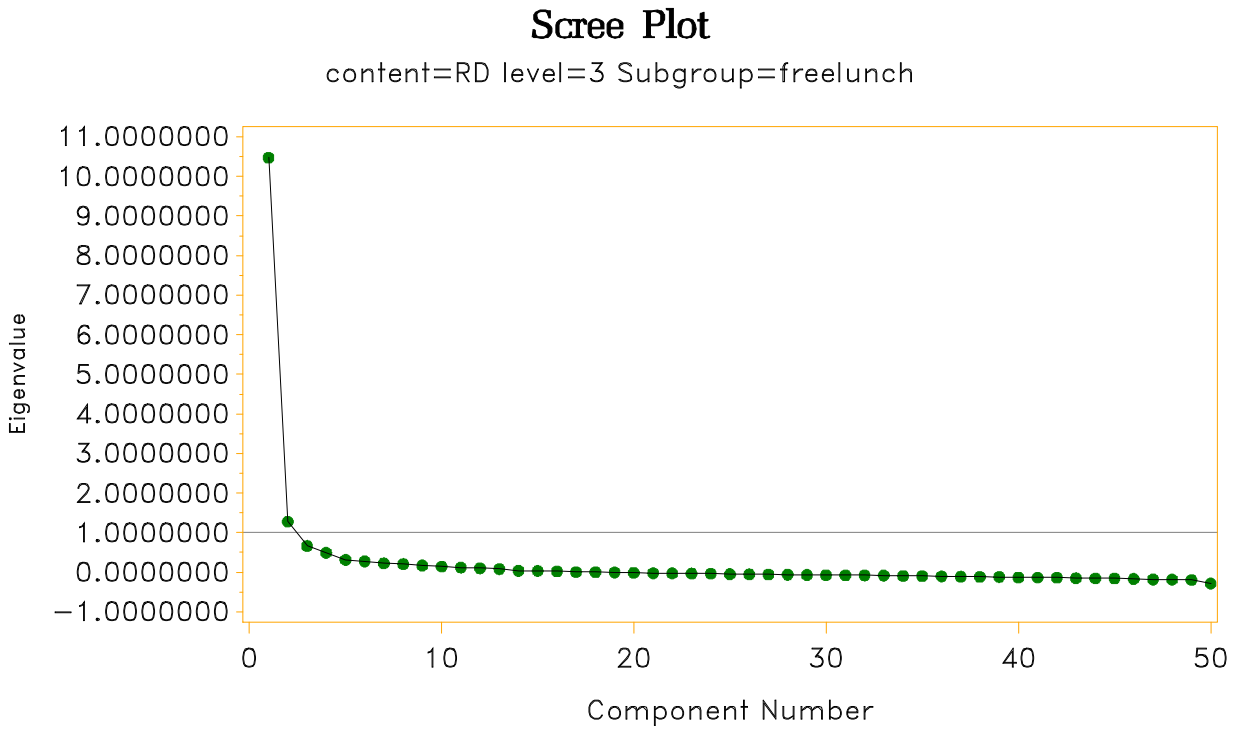


Figure 63. Spring 2013 Grade 3 Reading scree plot: Individualized Education Program

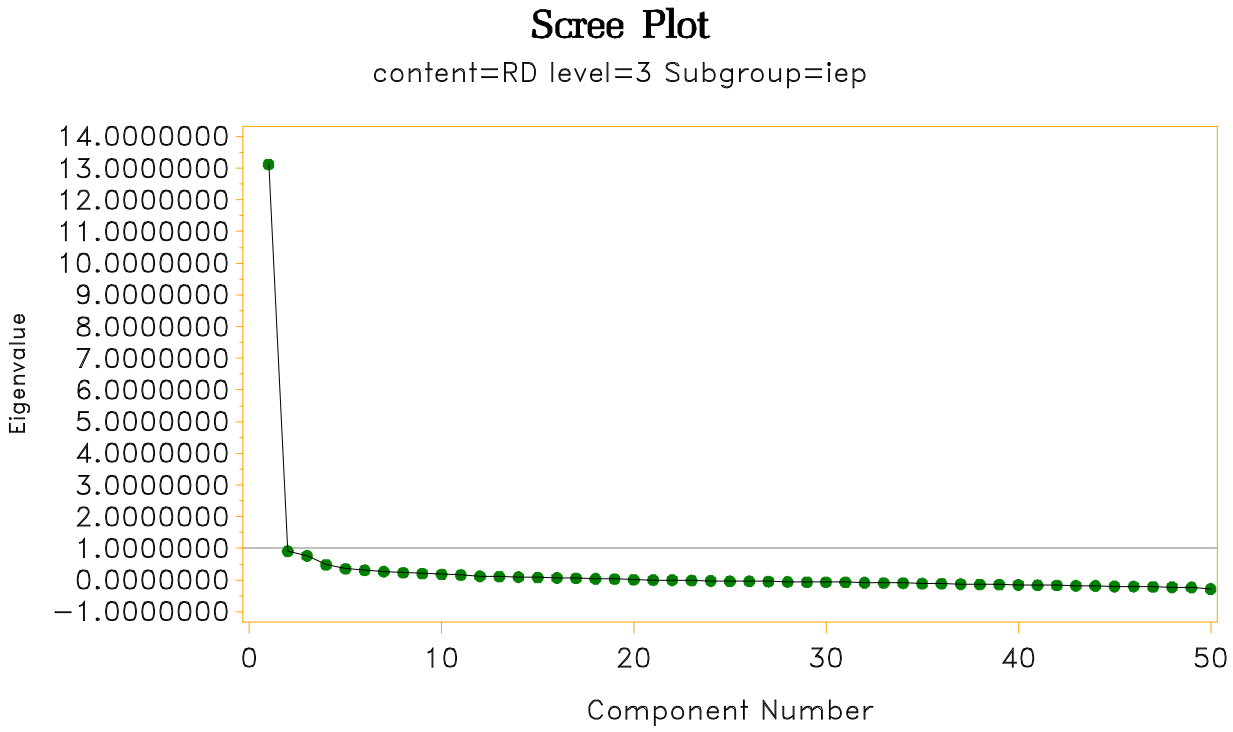


Figure 64. Spring 2013 Grade 4 Reading scree plot: Total

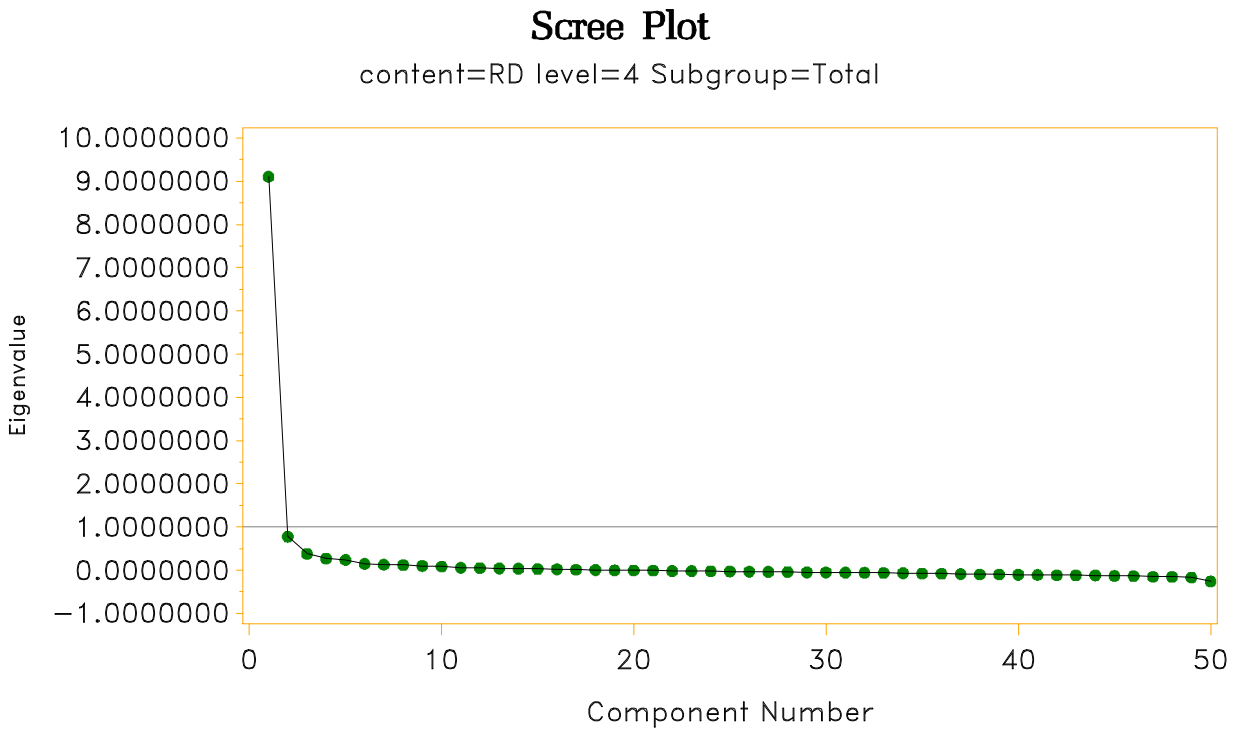


Figure 65. Spring 2013 Grade 4 Reading scree plot: Accommodated

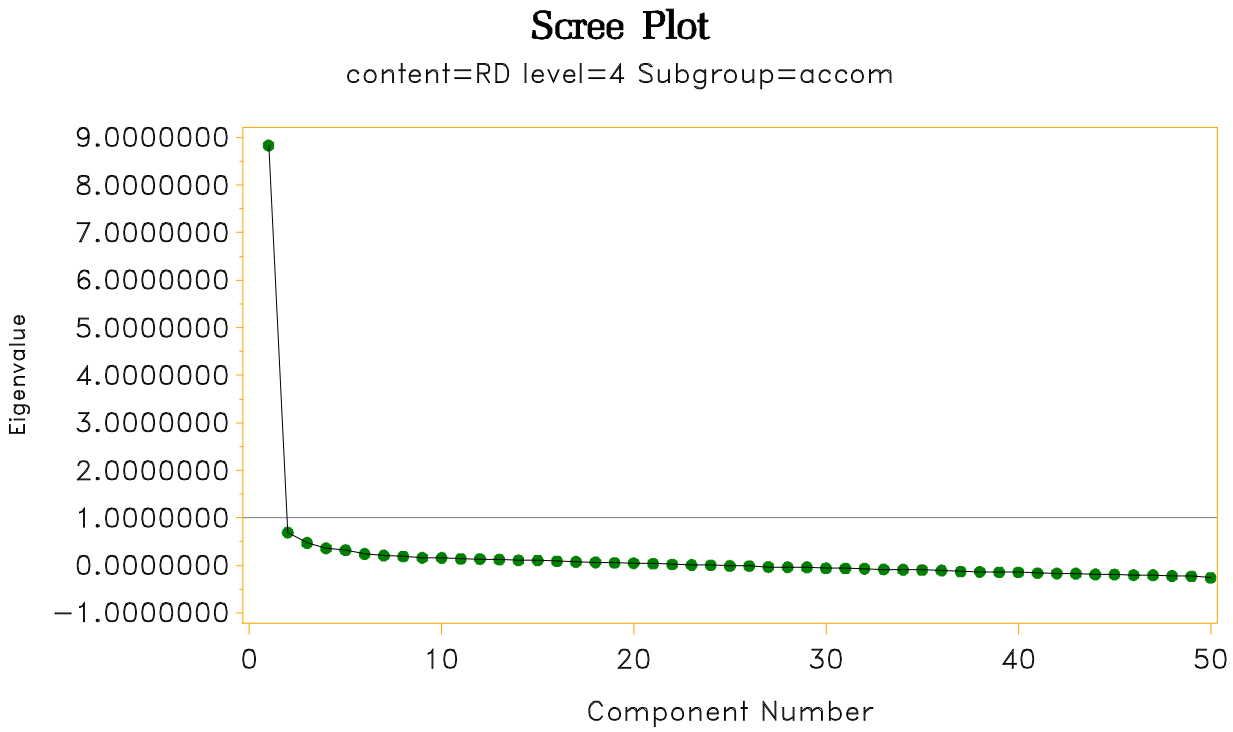


Figure 66. Spring 2013 Grade 4 Reading scree plot: English Language Learner

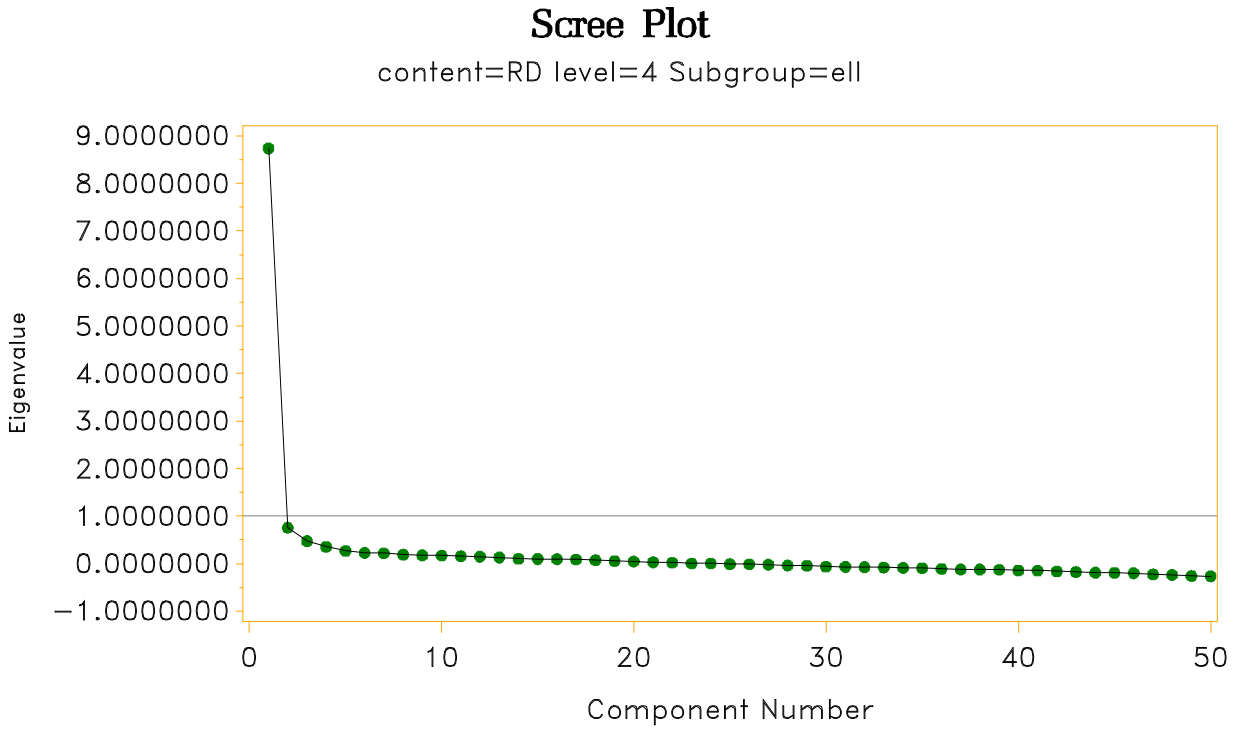


Figure 67. Spring 2013 Grade 4 Reading scree plot: Free Lunch

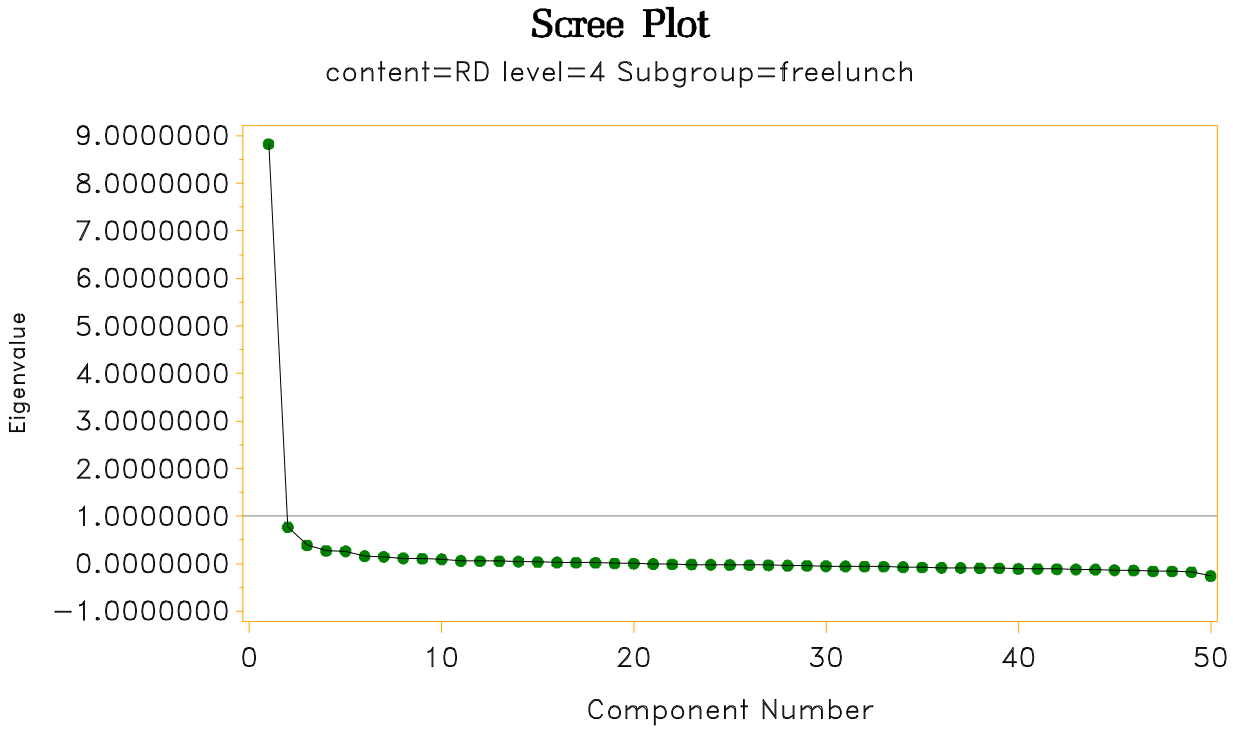


Figure 68. Spring 2013 Grade 4 Reading scree plot: Individualized Education Program

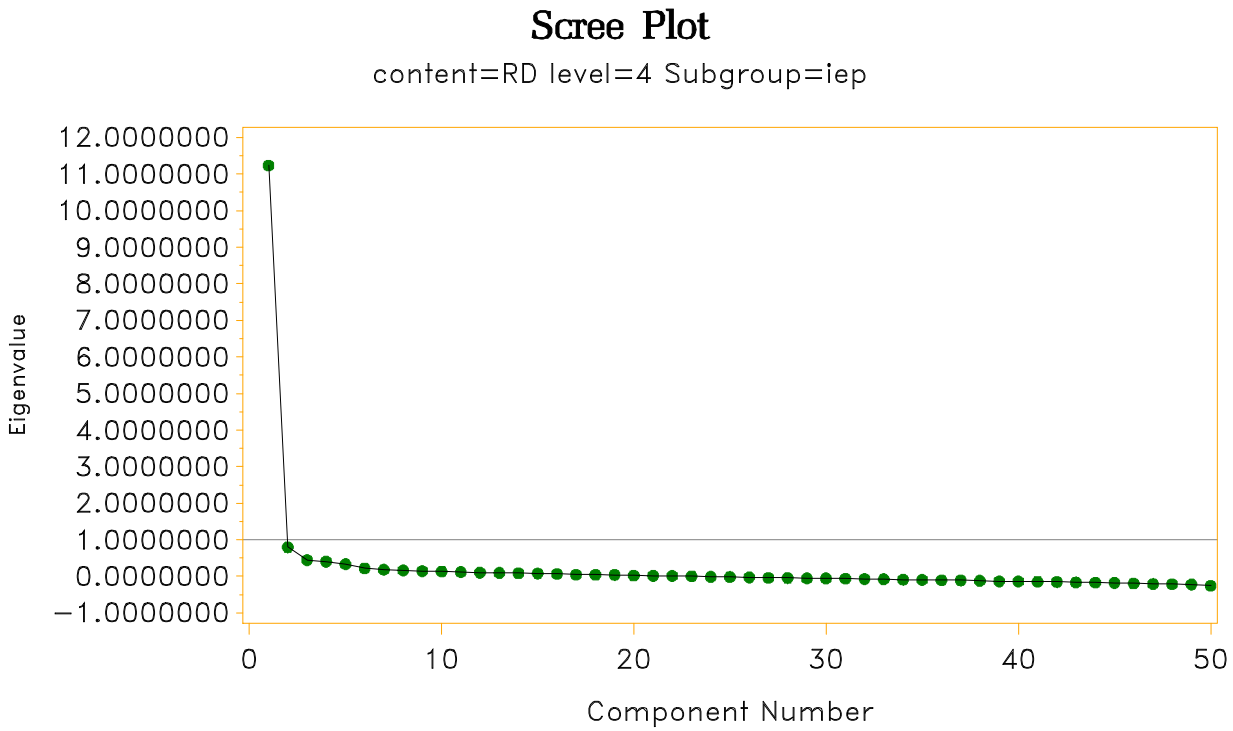


Figure 69. Spring 2013 Grade 5 Reading scree plot: Total

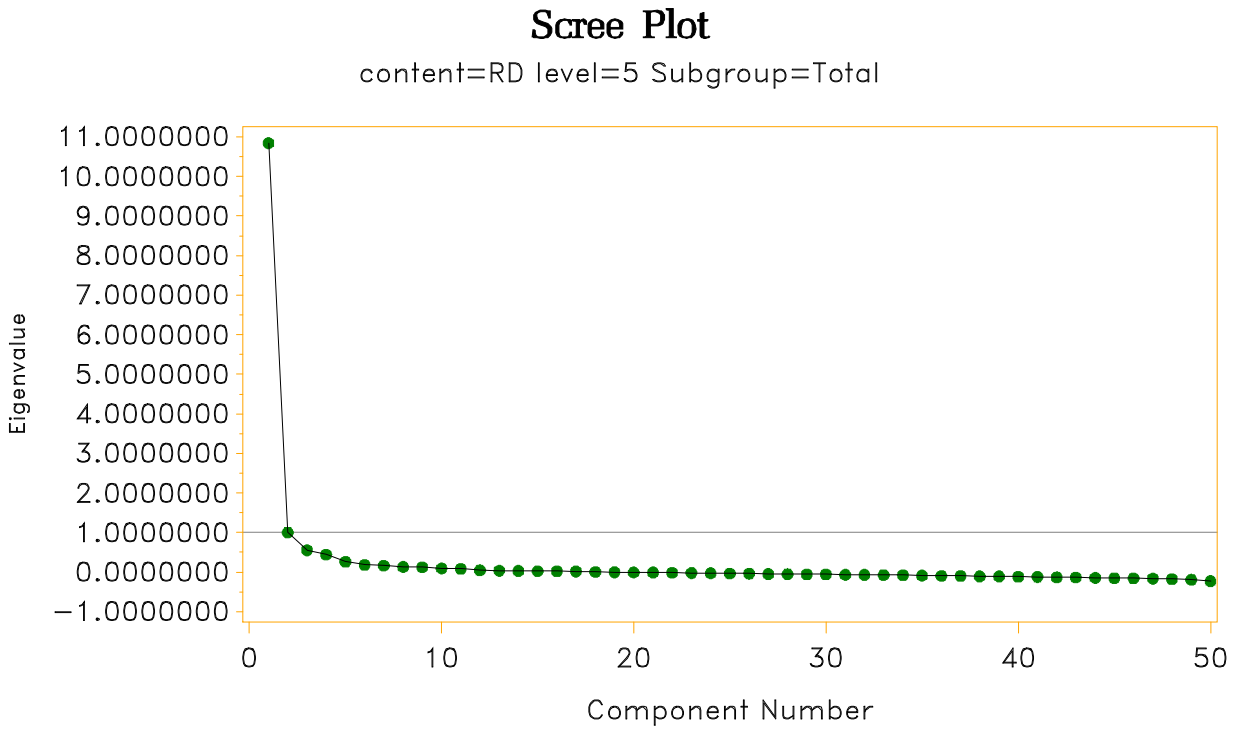


Figure 70. Spring 2013 Grade 5 Reading scree plot: Accommodated

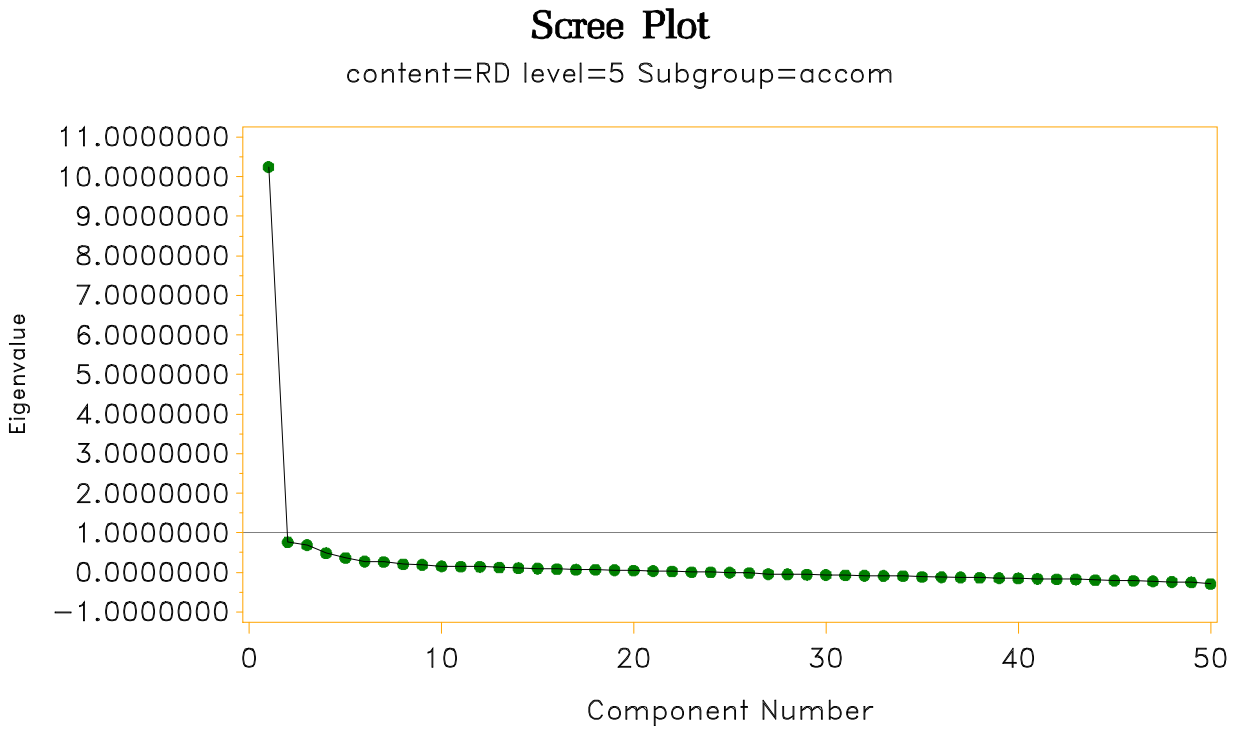


Figure 71. Spring 2013 Grade 5 Reading scree plot: English Language Learner

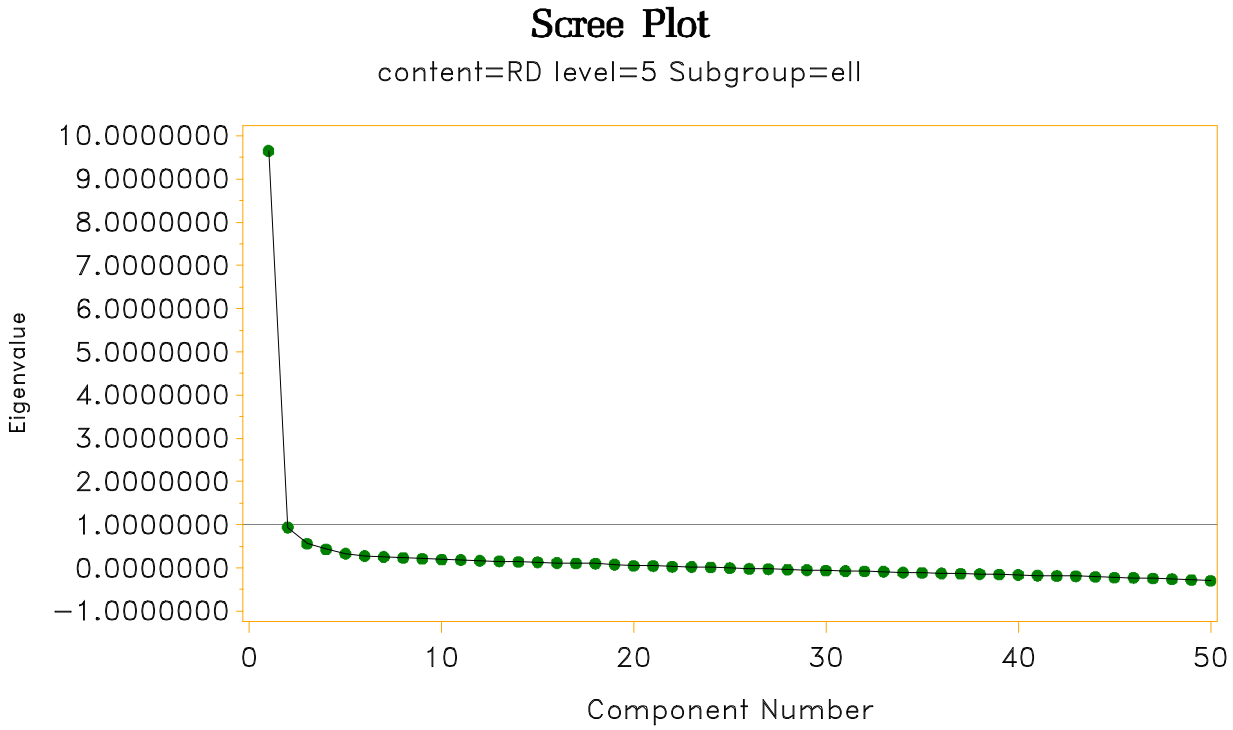


Figure 72. Spring 2013 Grade 5 Reading scree plot: Free Lunch

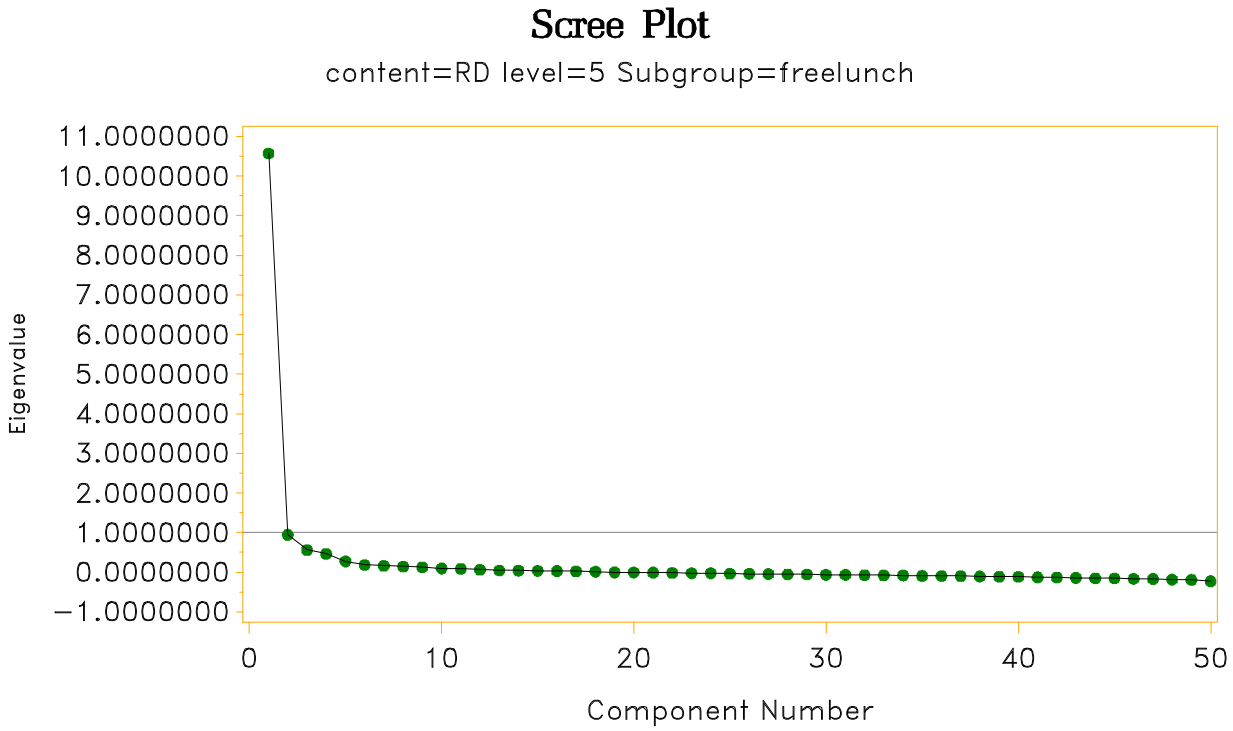


Figure 73. Spring 2013 Grade 5 Reading scree plot: Individualized Education Program

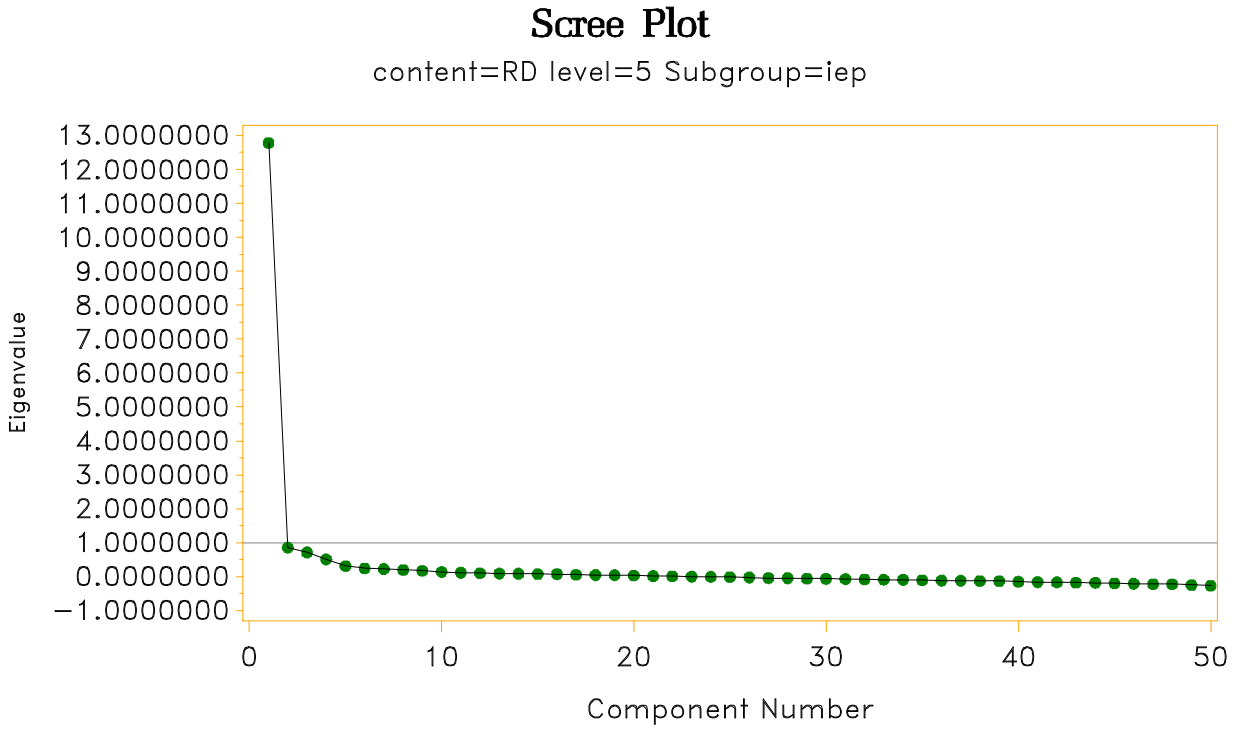


Figure 74. Spring 2013 Grade 6 Reading scree plot: Total

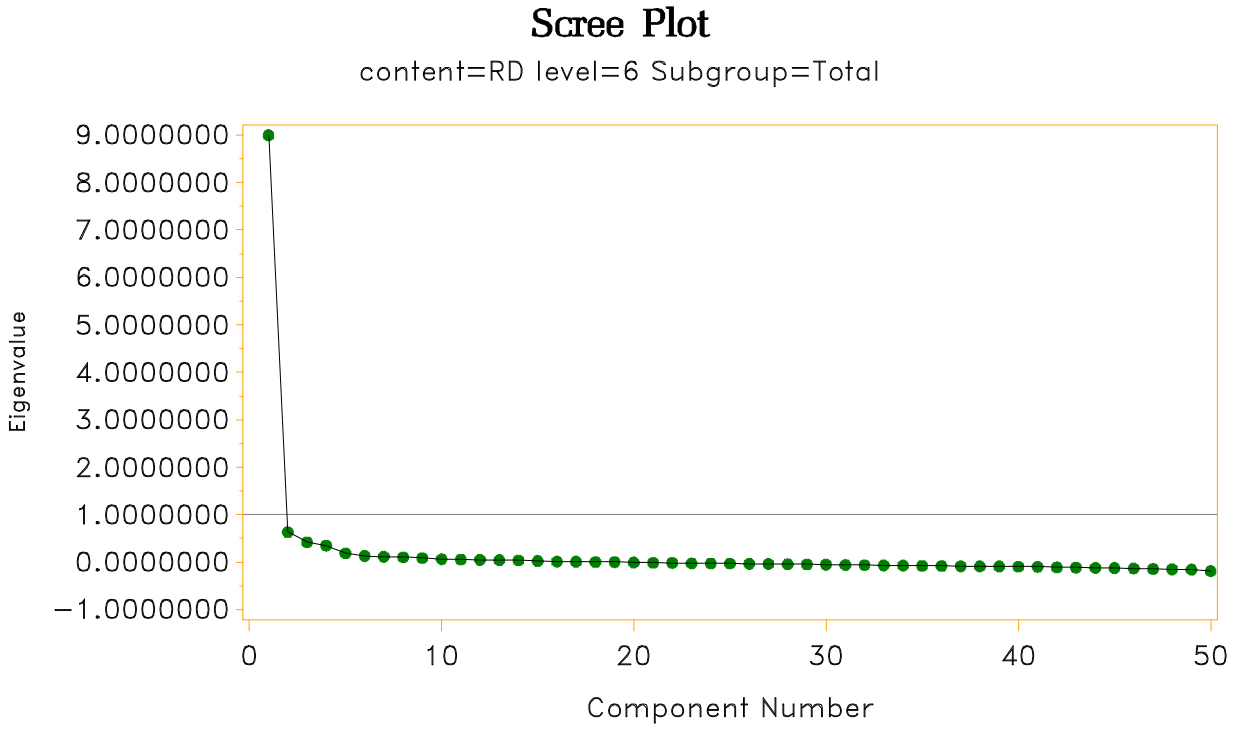


Figure 75. Spring 2013 Grade 6 Reading scree plot: Accommodated

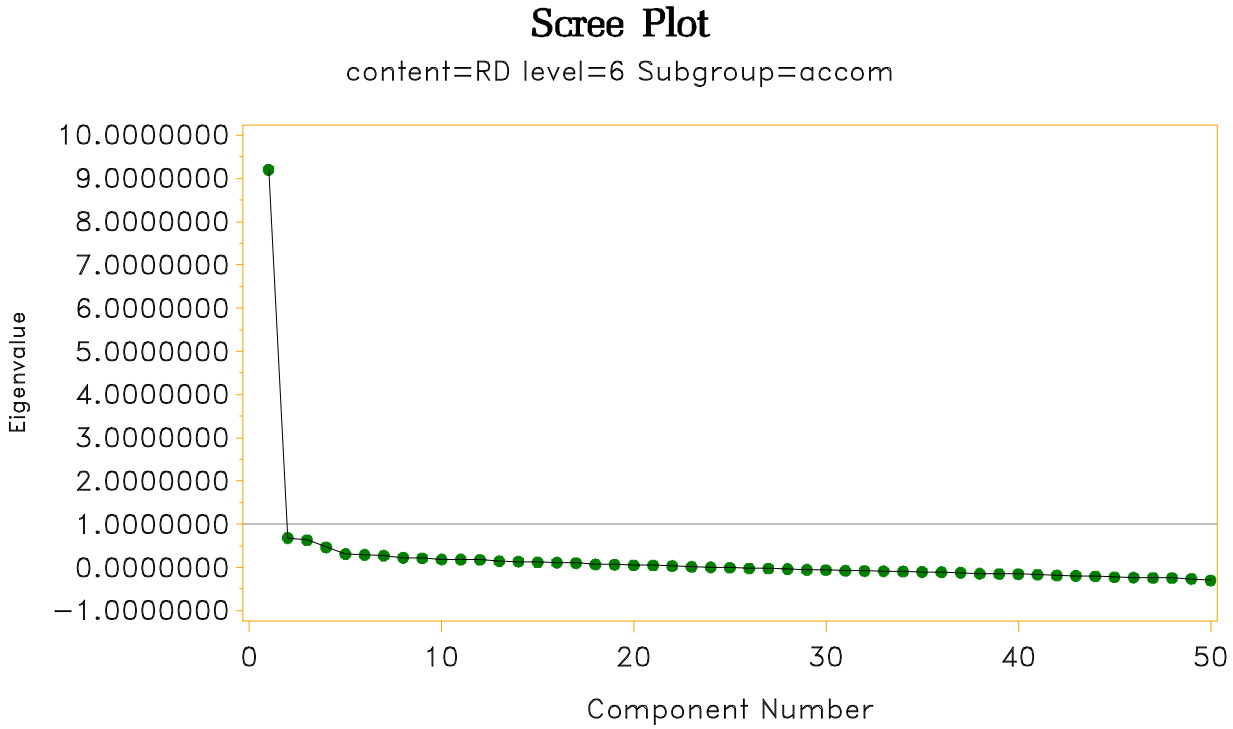


Figure 76. Spring 2013 Grade 6 Reading scree plot: English Language Learner

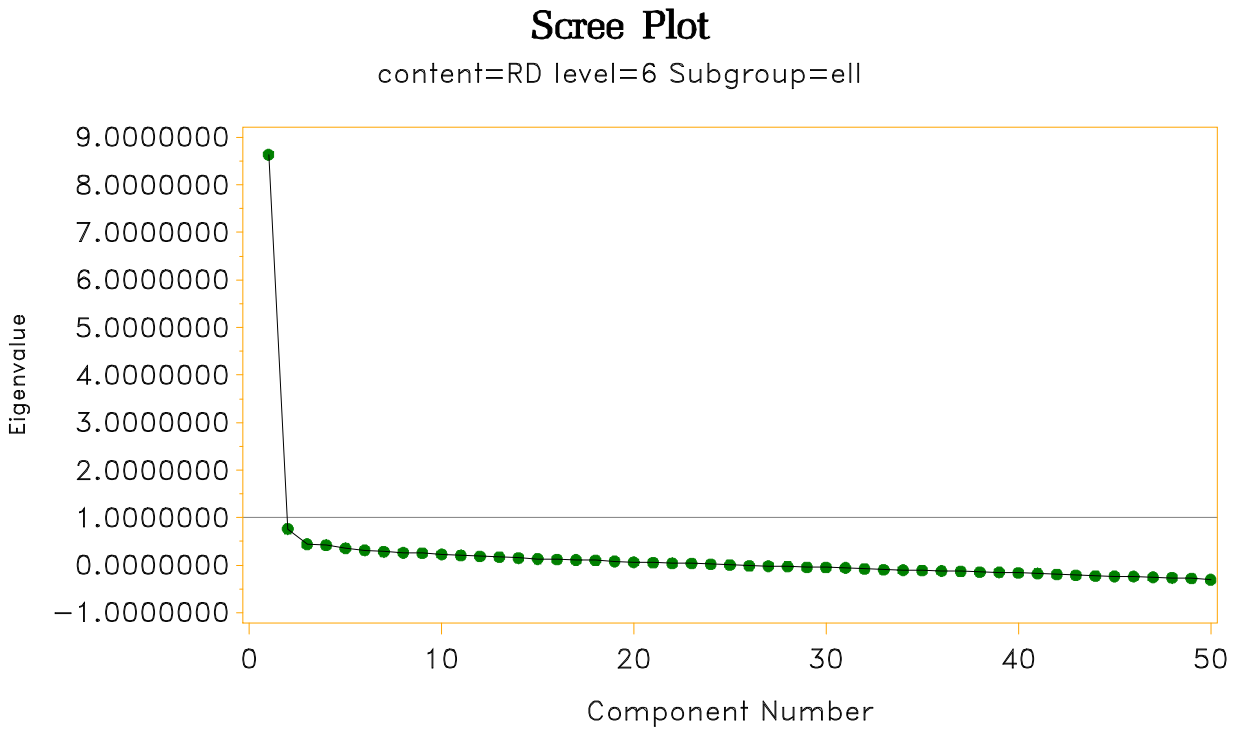


Figure 77. Spring 2013 Grade 6 Reading scree plot: Free Lunch

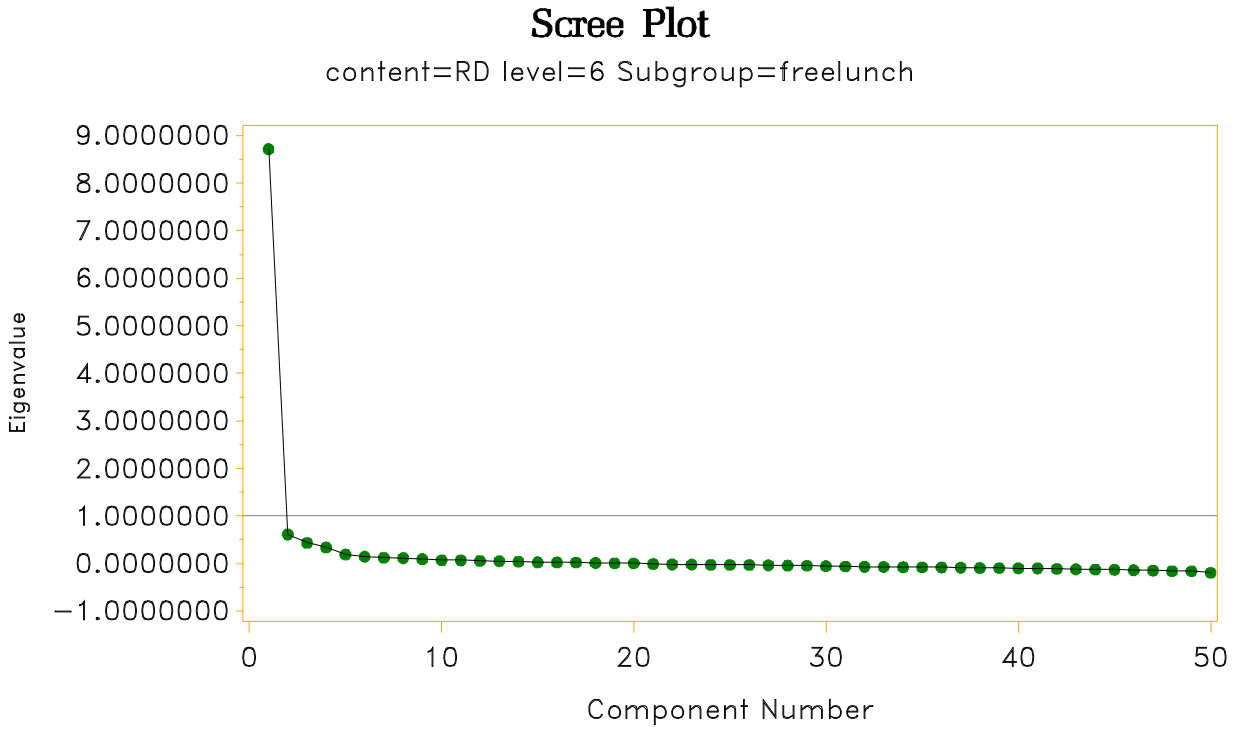


Figure 78. Spring 2013 Grade 6 Reading scree plot: Individualized Education Program

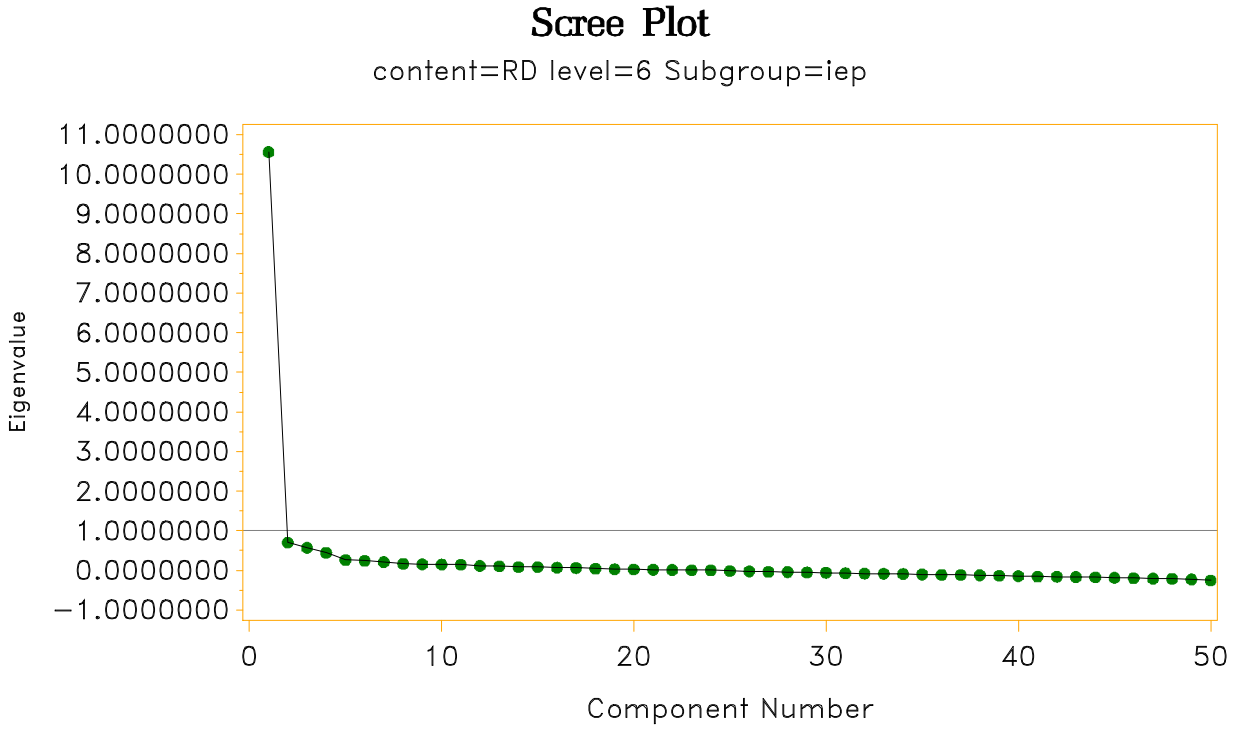


Figure 79. Spring 2013 Grade 7 Reading scree plot: Total

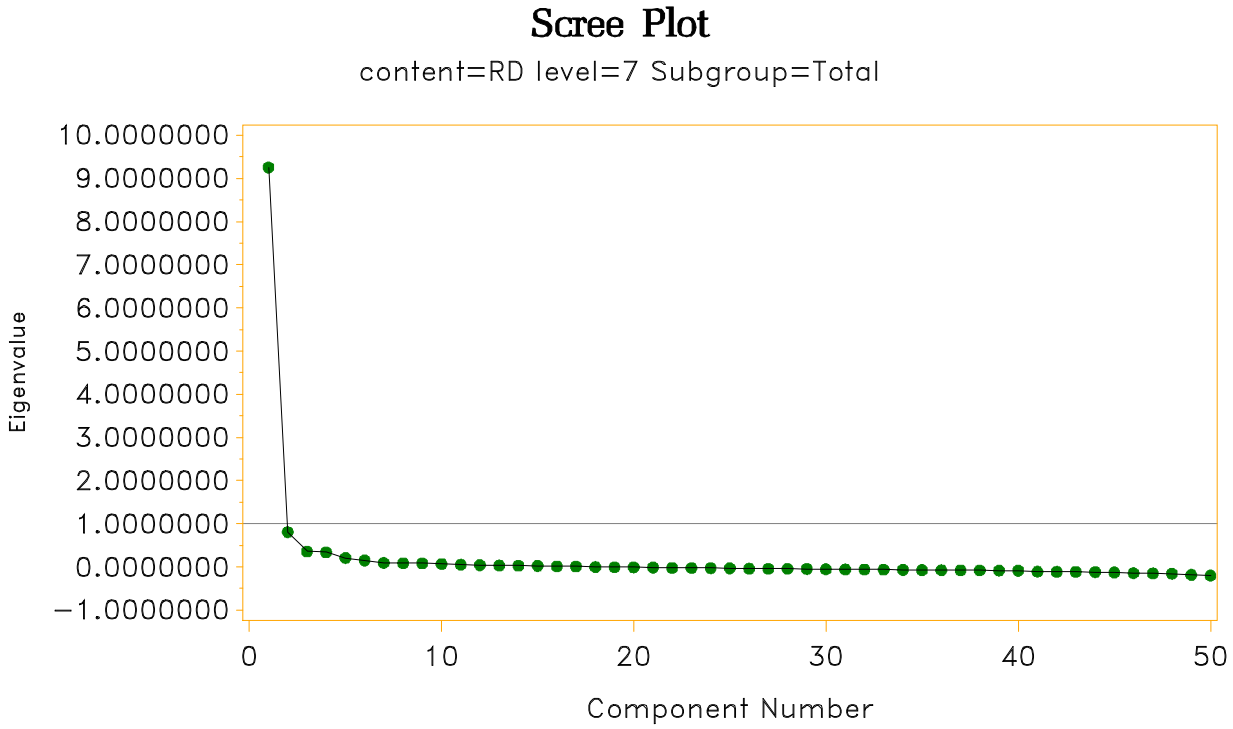


Figure 80. Spring 2013 Grade 7 Reading scree plot: Accommodated

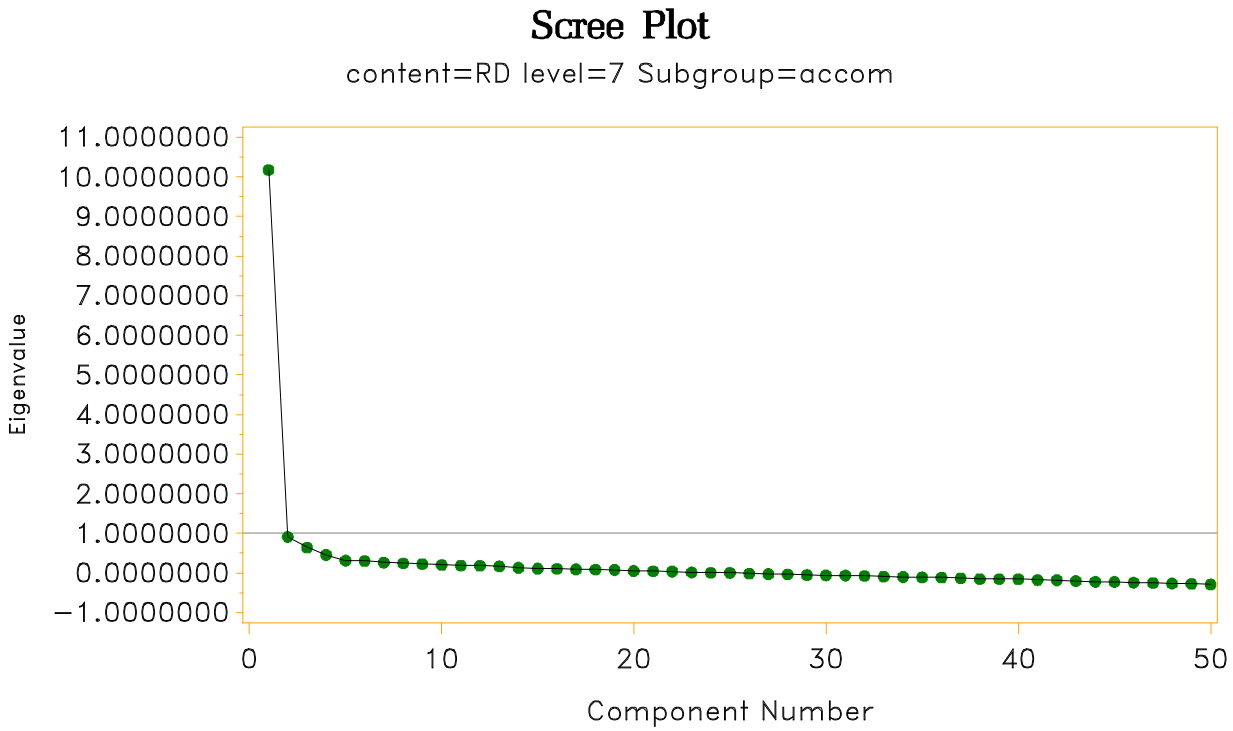


Figure 81. Spring 2013 Grade 7 Reading scree plot: English Language Learner

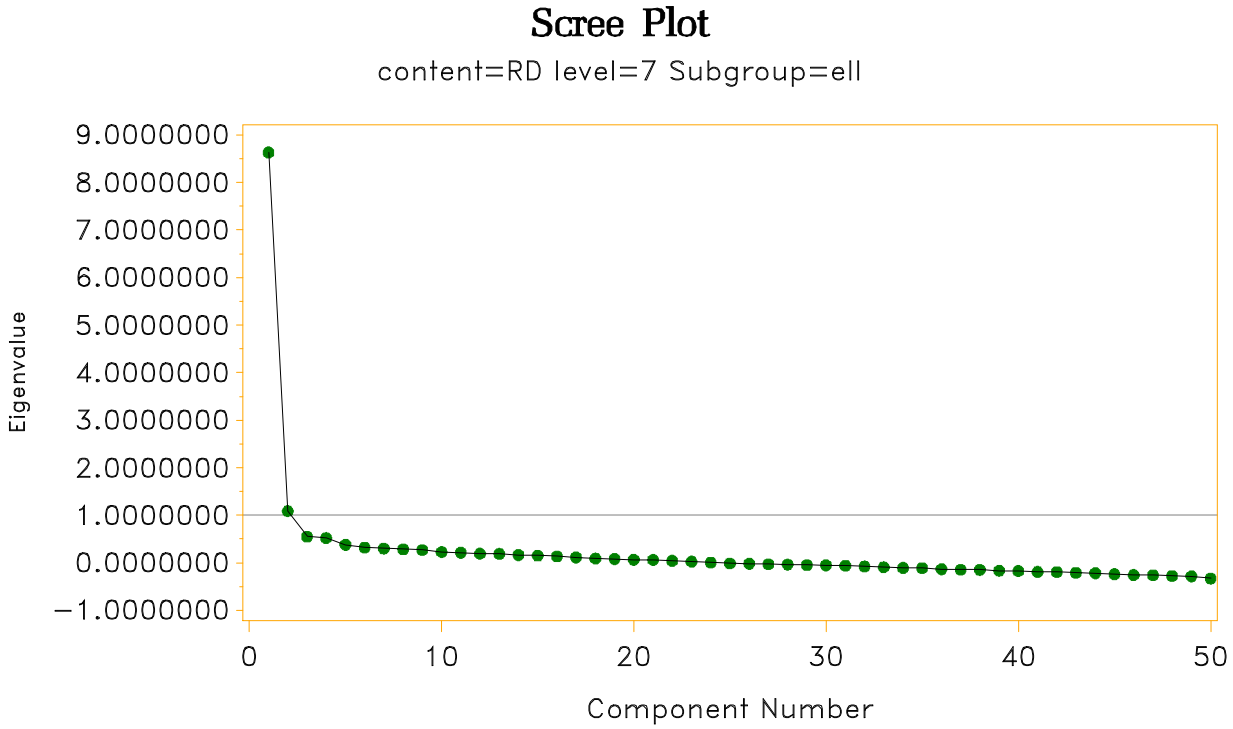


Figure 82. Spring 2013 Grade 7 Reading scree plot: Free Lunch

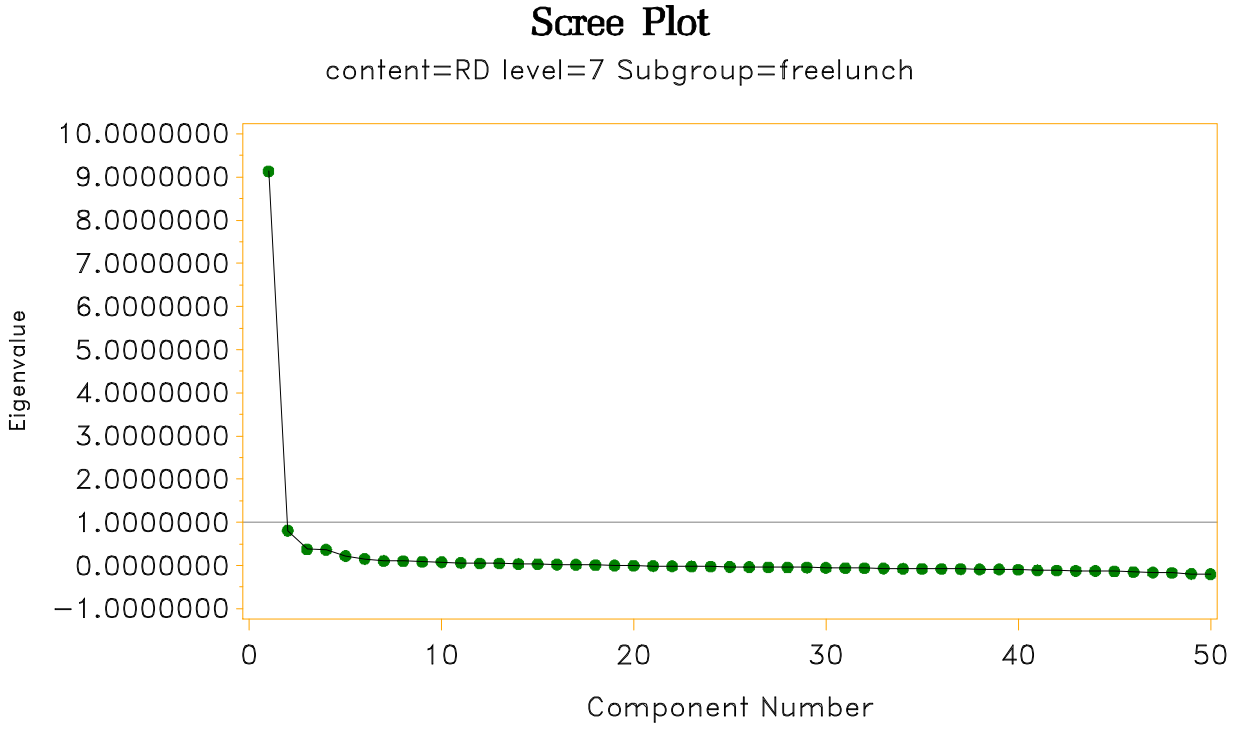


Figure 83. Spring 2013 Grade 7 Reading scree plot: Individualized Education Program

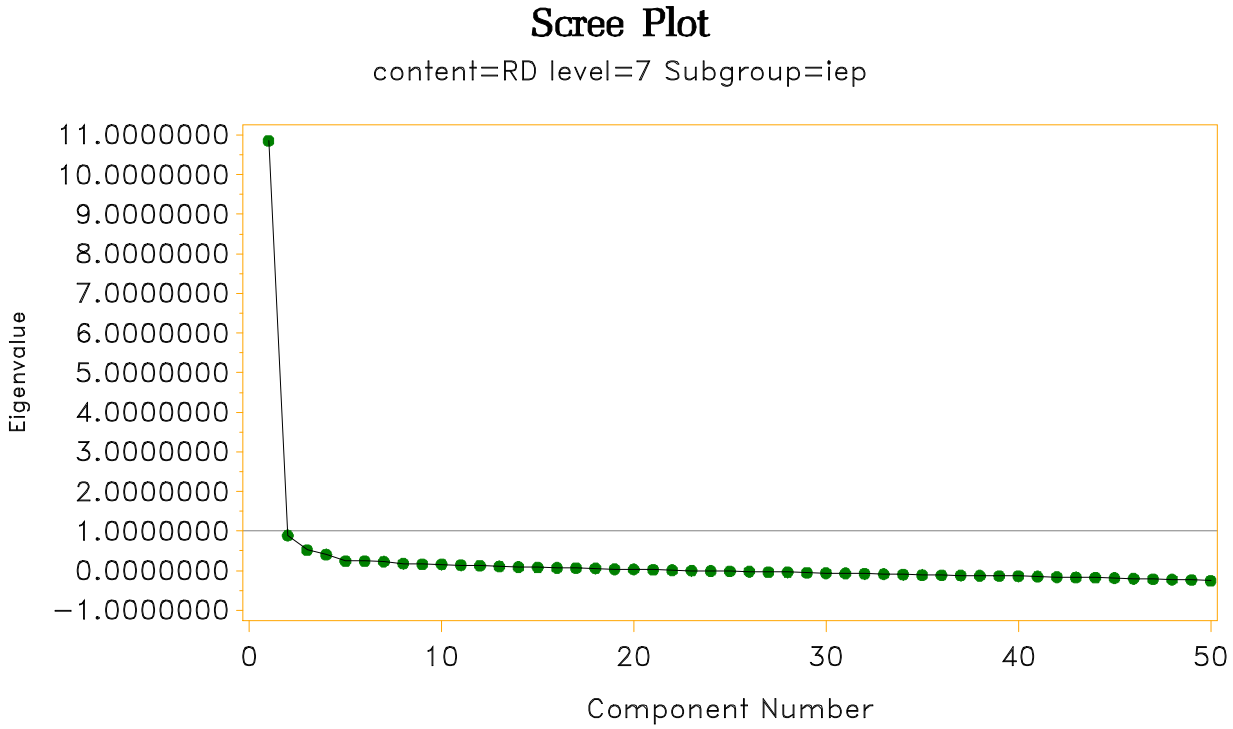


Figure 84. Spring 2013 Grade 8 Reading scree plot: Total

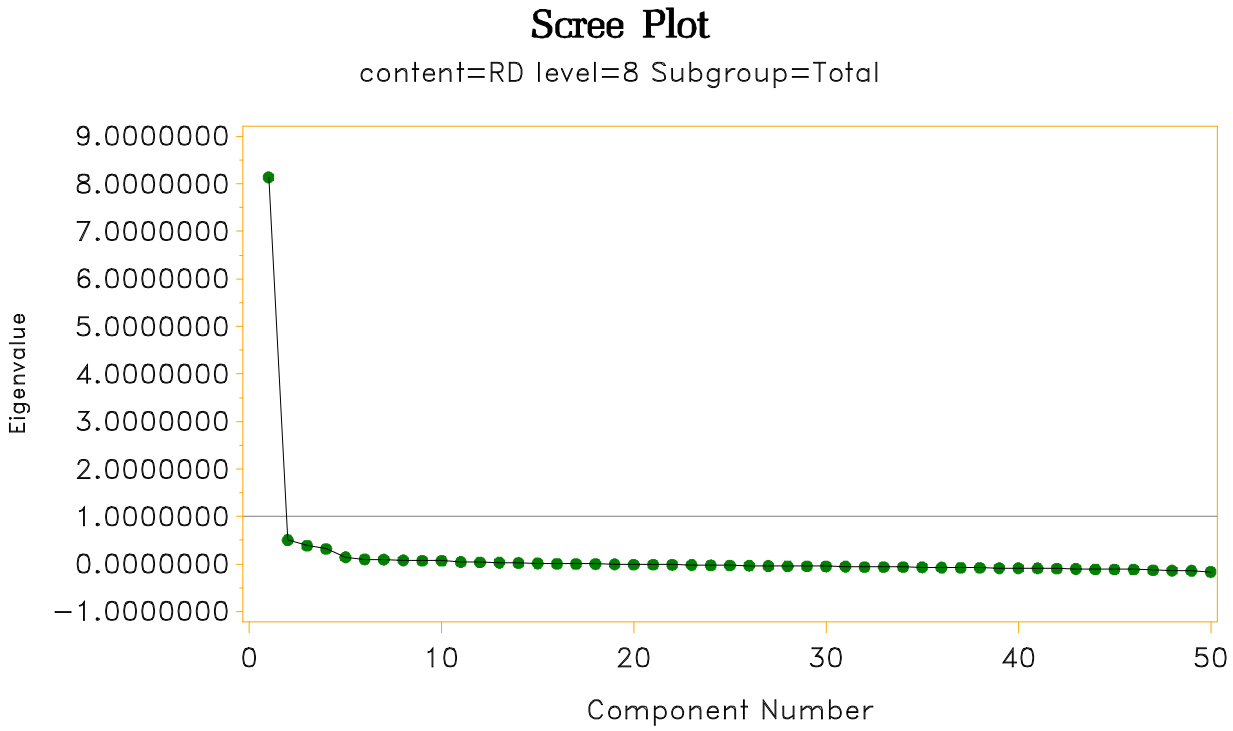


Figure 85. Spring 2013 Grade 8 Reading scree plot: Accommodated

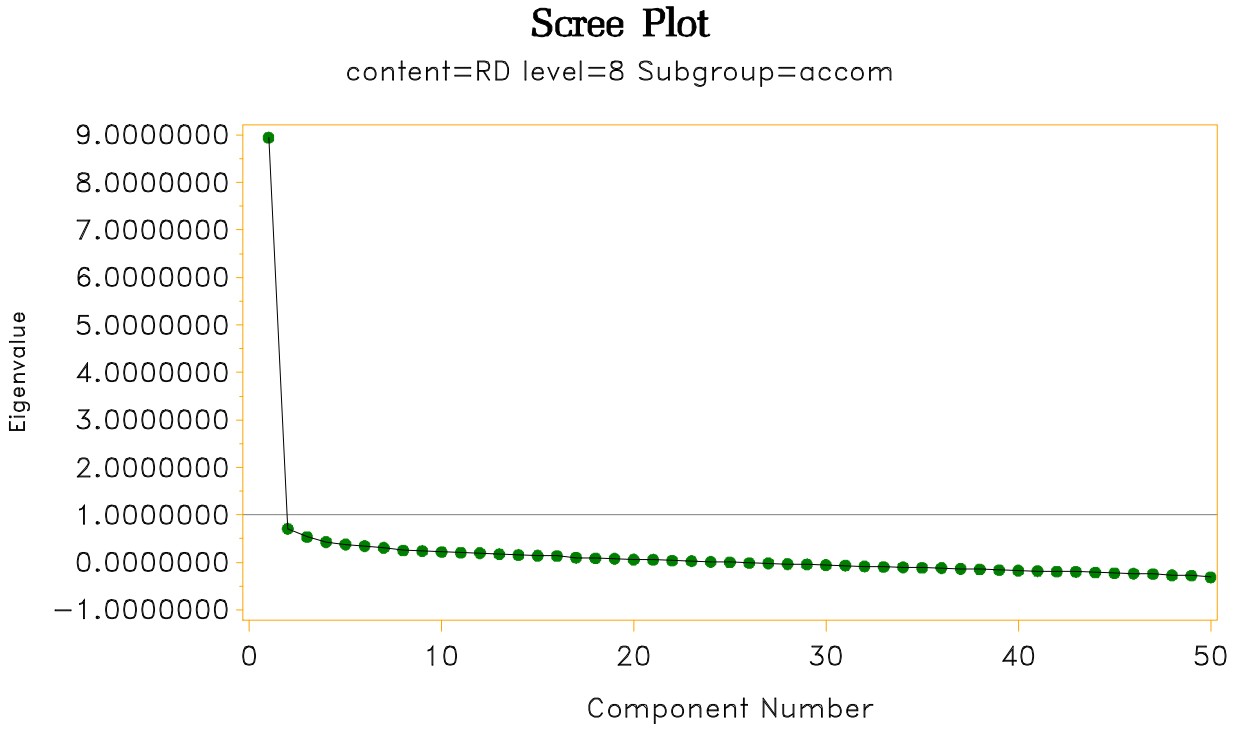


Figure 86. Spring 2013 Grade 8 Reading scree plot: English Language Learner

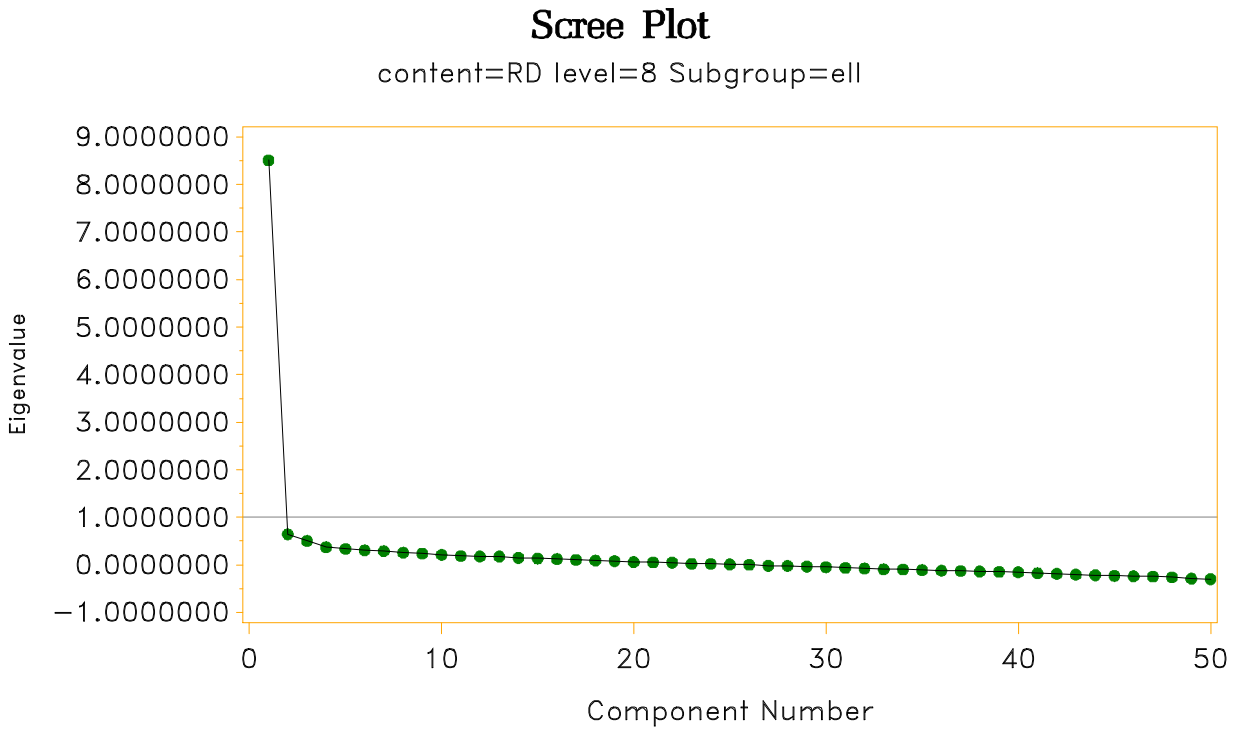


Figure 87. Spring 2013 Grade 8 Reading scree plot: Free Lunch

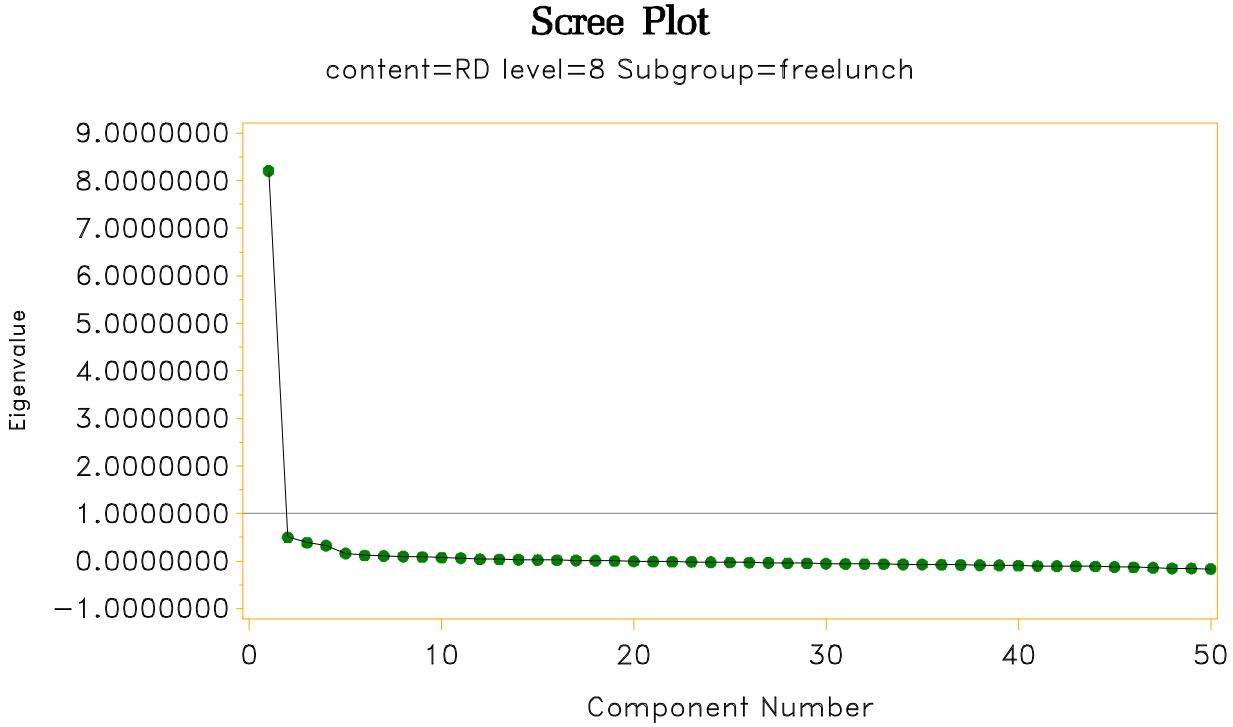


Figure 88. Spring 2013 Grade 8 Reading scree plot: Individualized Education Program

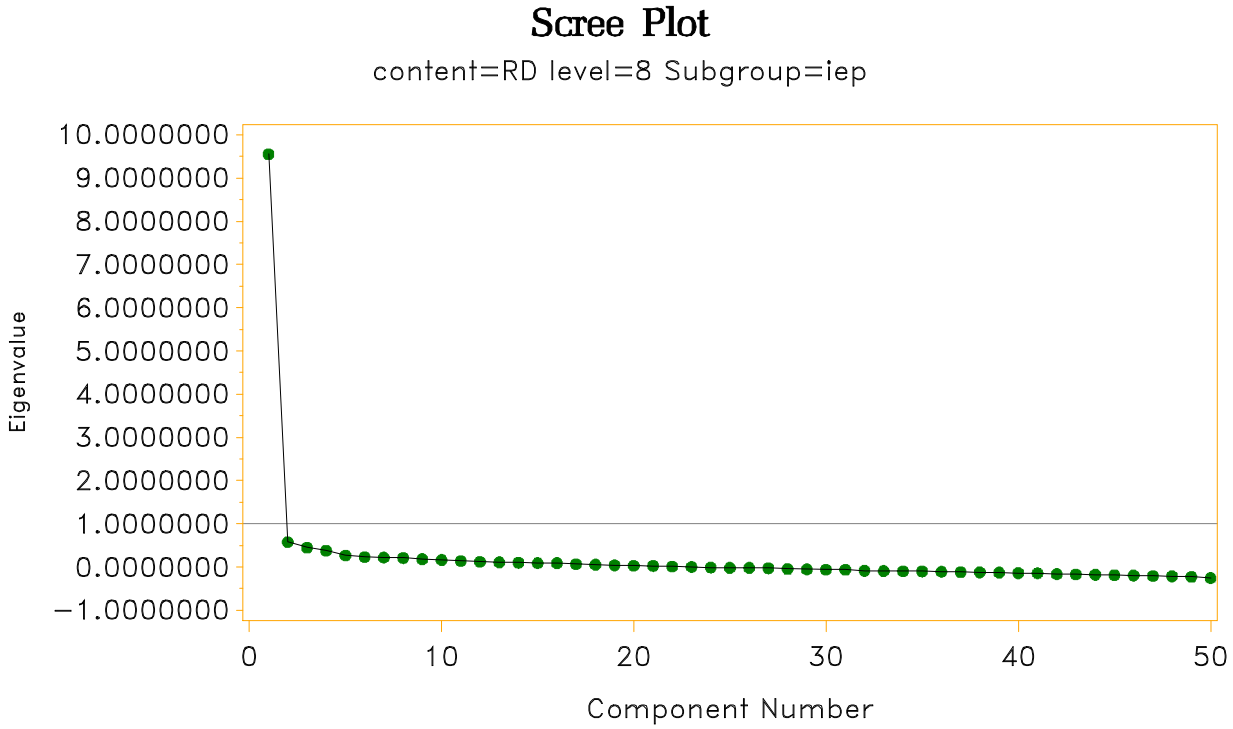


Figure 89. Spring 2013 Grade 5 Science scree plot: Total

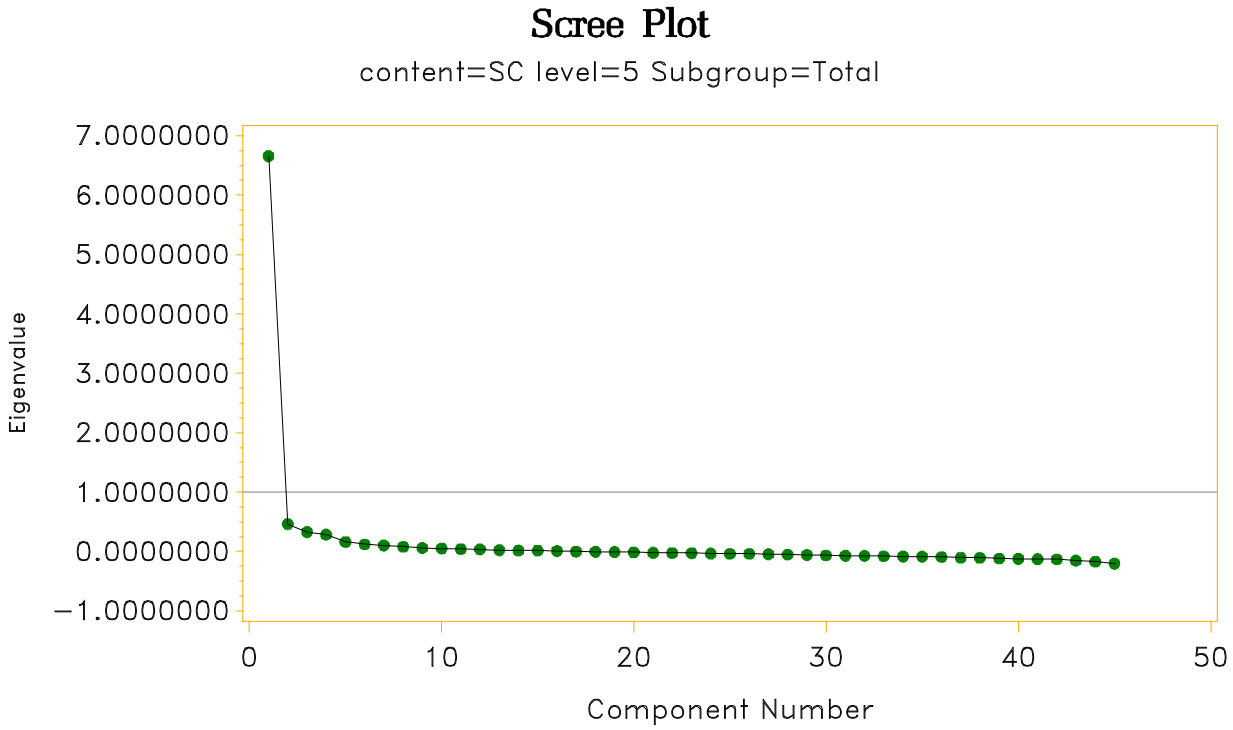


Figure 90. Spring 2013 Grade 5 Science scree plot: Accommodated

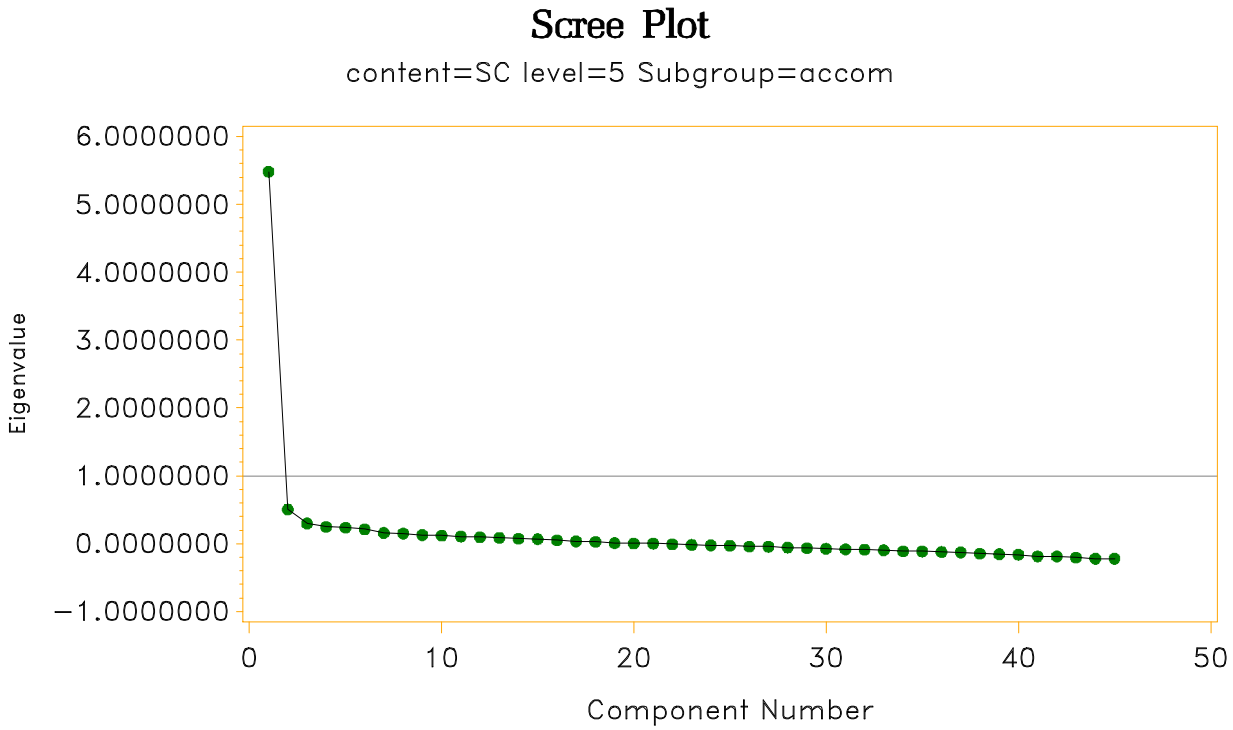


Figure 91. Spring 2013 Grade 5 Science scree plot: English Language Learner

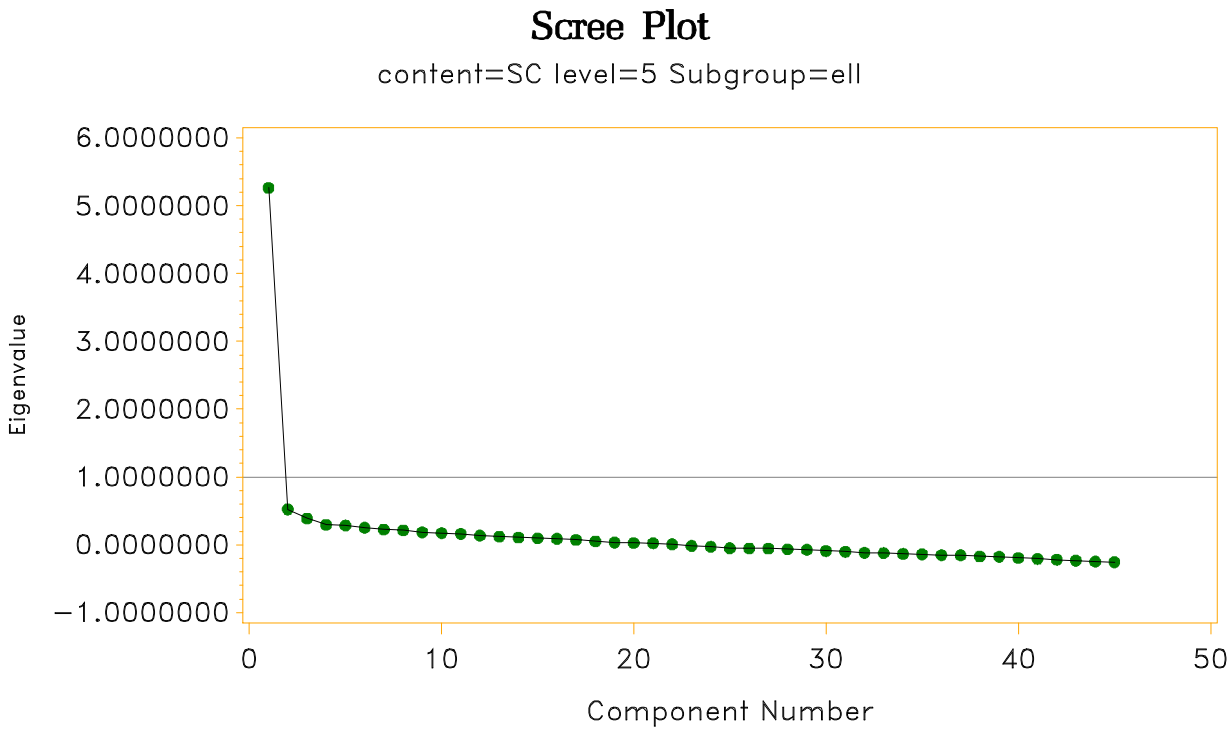


Figure 92. Spring 2013 Grade 5 Science scree plot: Free Lunch

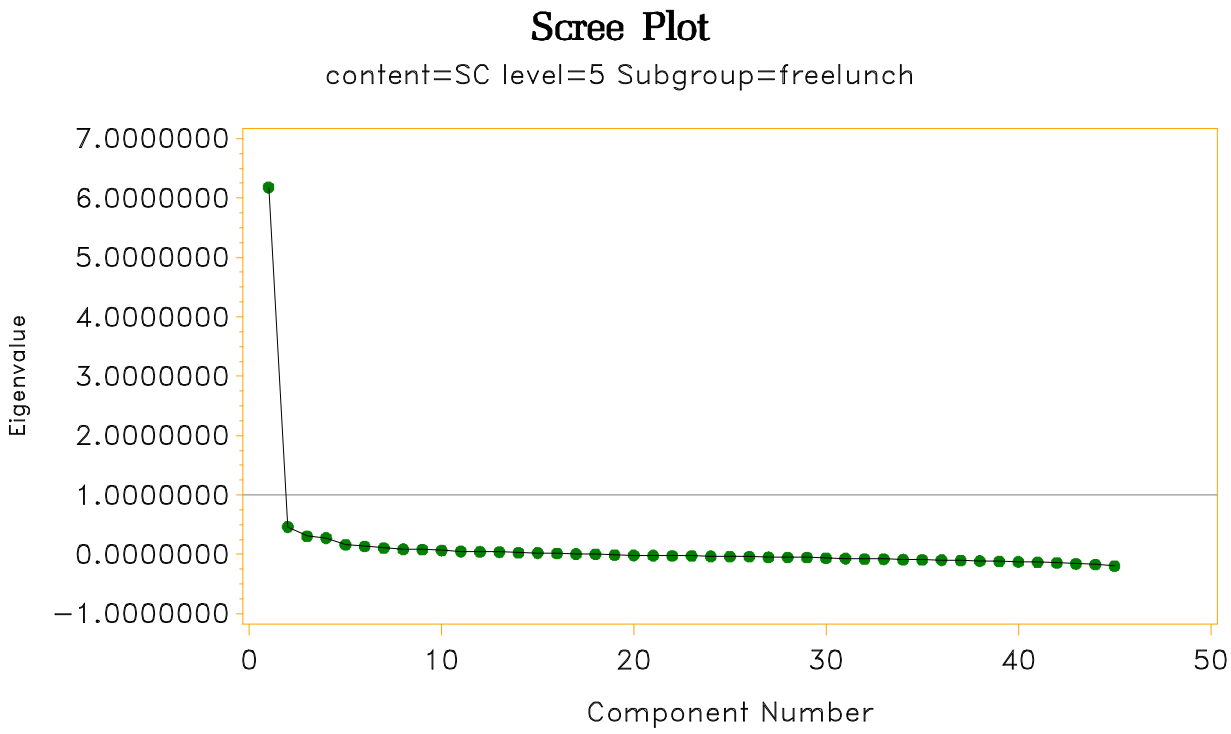


Figure 93. Spring 2013 Grade 5 Science scree plot: Individualized Education Program

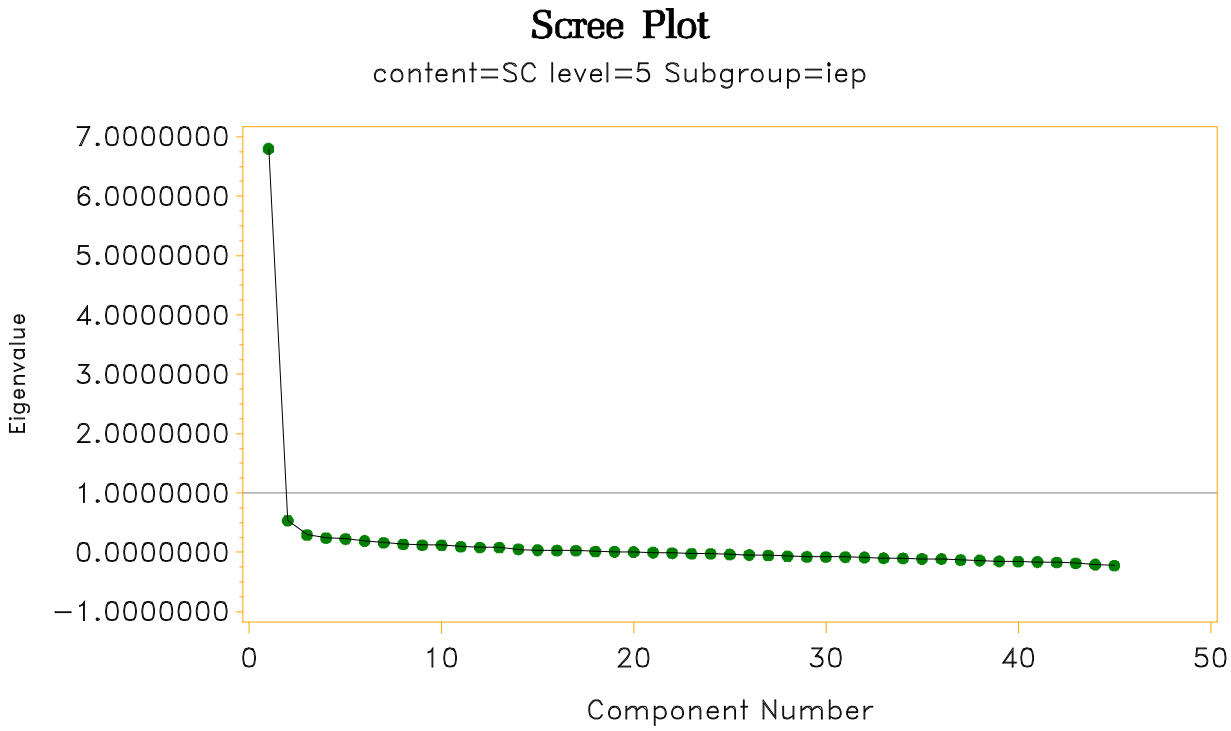


Figure 94. Spring 2013 Grade 8 Science scree plot: Total

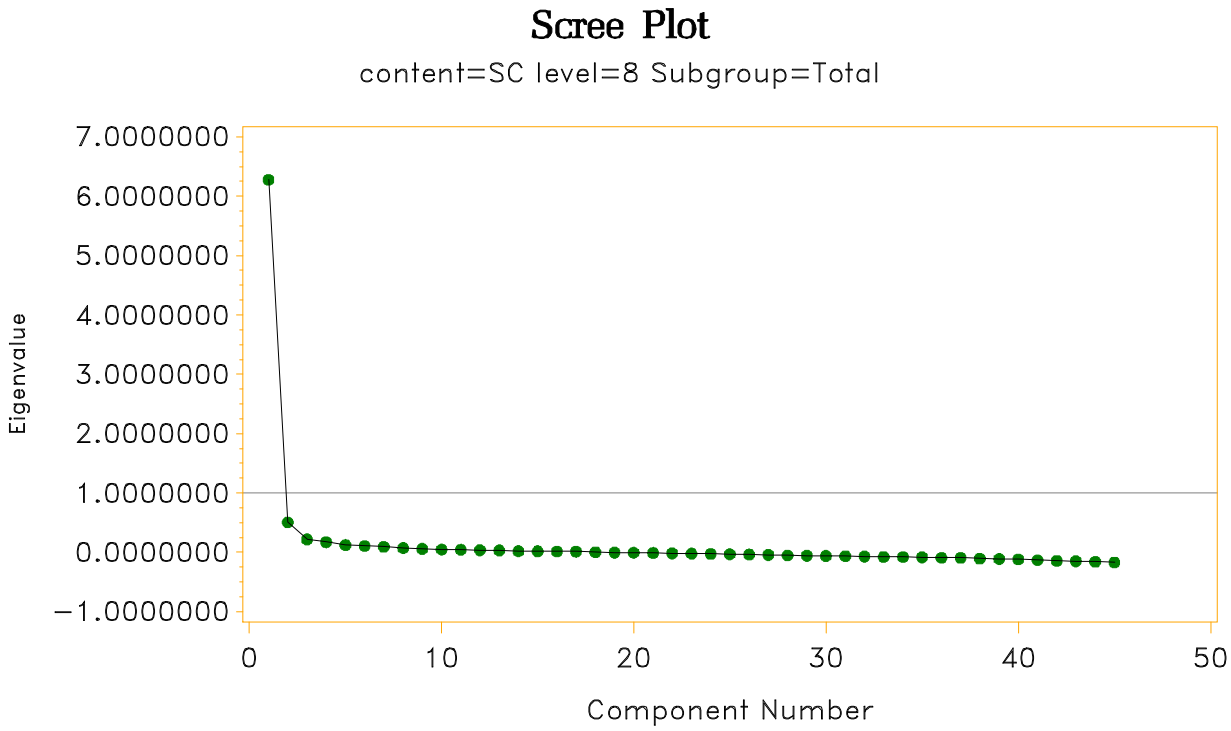


Figure 95. Spring 2013 Grade 8 Science scree plot: Accommodated

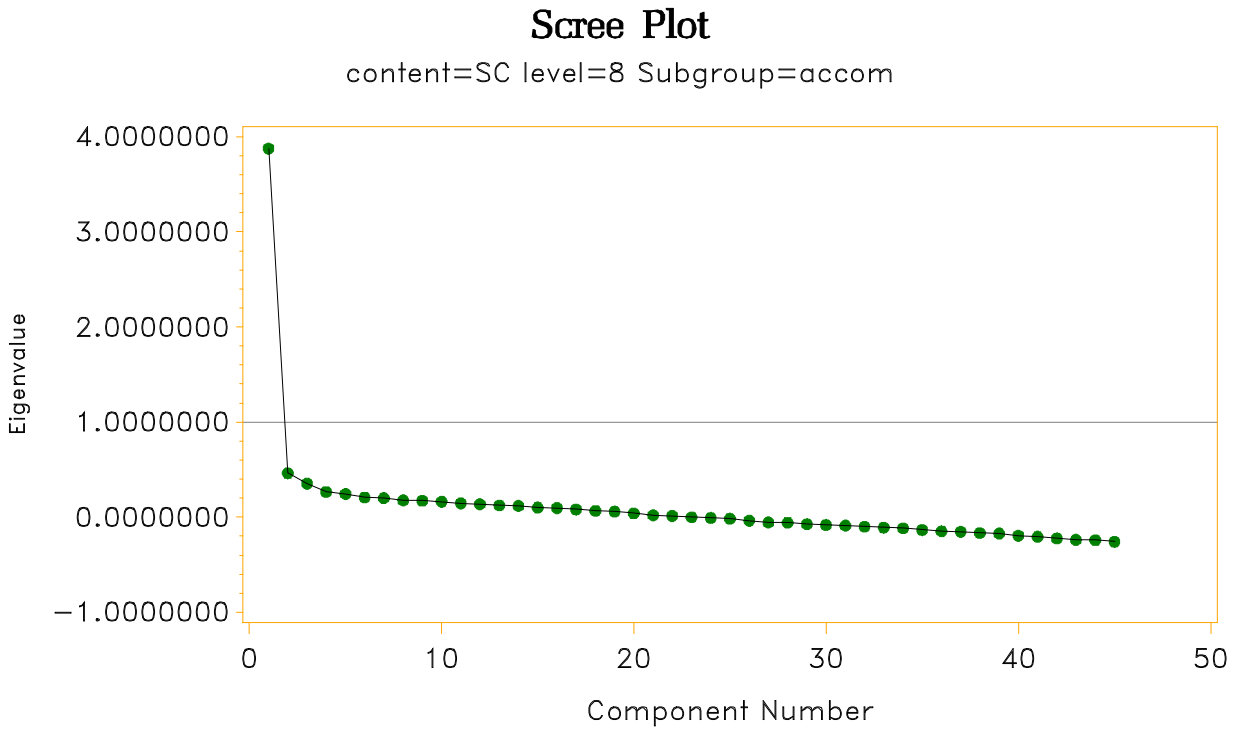


Figure 96. Spring 2013 Grade 8 Science scree plot: English Language Learner

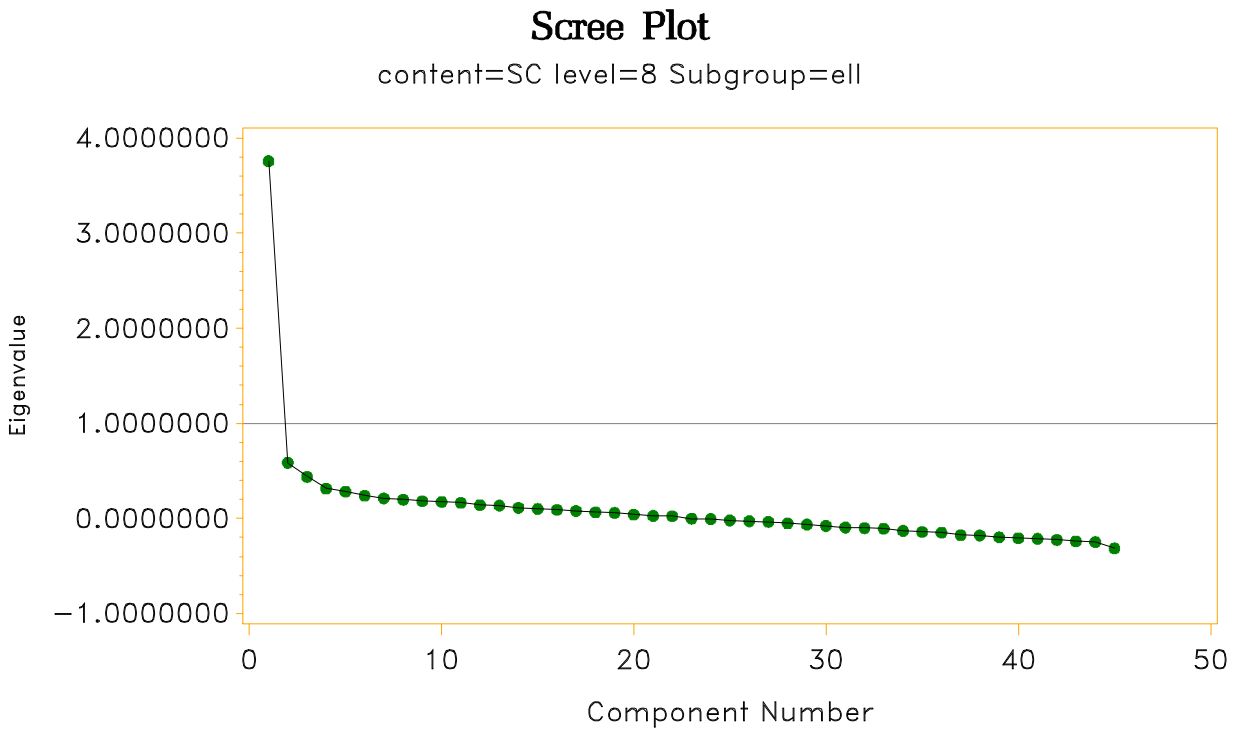


Figure 97. Spring 2013 Grade 8 Science scree plot: Free Lunch

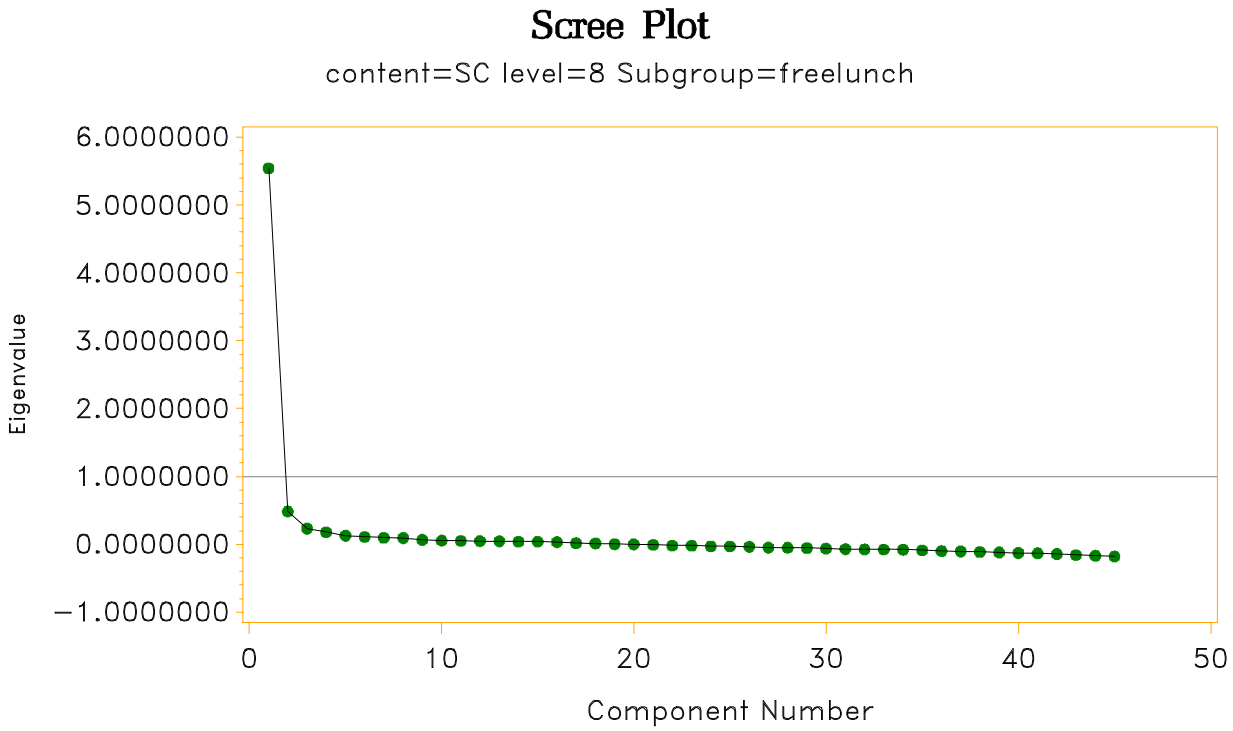
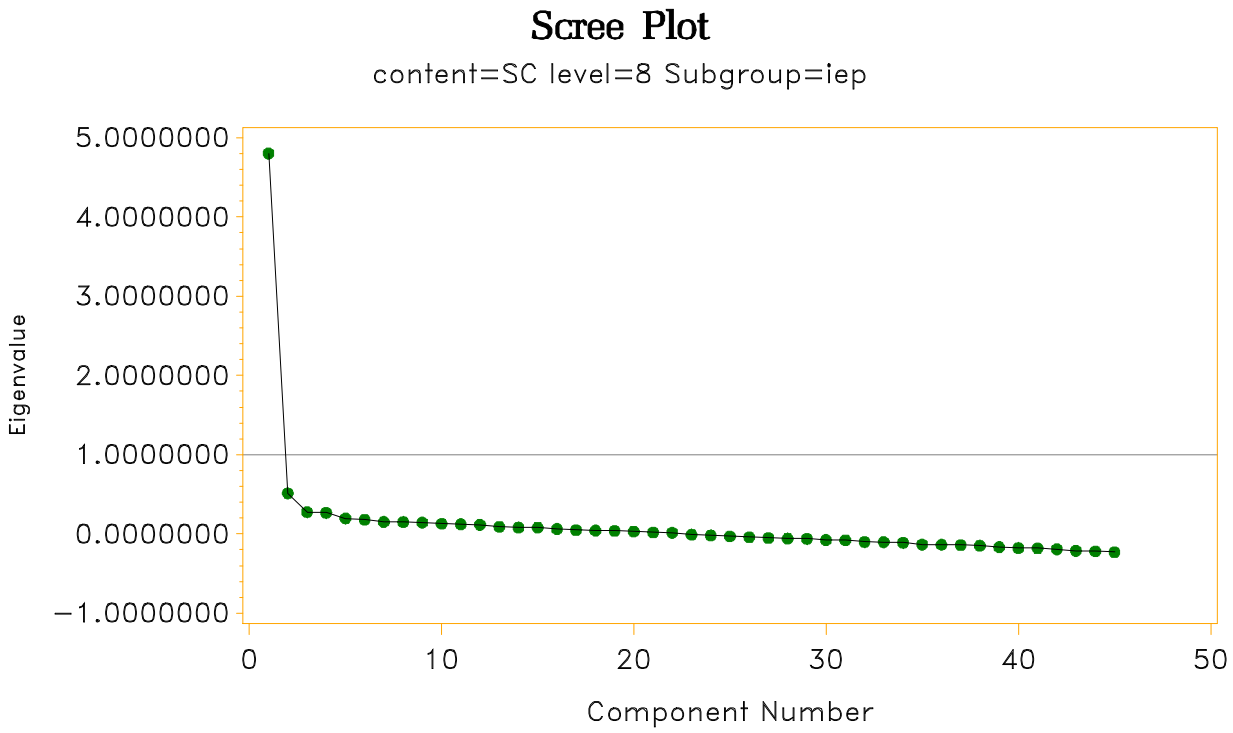


Figure 98. Spring 2013 Grade 8 Science scree plot: Individualized Education Program



Appendices

Appendix A
Standards, Objectives/Skills, and Processes Assessed by Subject

Note: In 2013, 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

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Algebraic Reasoning: Patterns and Relationships	**7**	**7**	**5**
Algebra Patterns, Equations (1.1)	2	2	0
Equations (1.2)	2	2	2
Number Properties (1.3)	3	3	3
Number Sense and Operation	**20**	**20**	**10**
Number Sense (2.1)	10	10	7
Number Operations (2.2)	10	10	3
Geometry	**7**	**7**	**5**
Properties of shapes (3.1)	3	3	2
Spatial Reasoning (3.2)	2	2	3
Coordinate Geometry (3.3)	2	2	0
Measurement	**9**	**9**	**6**
Measurement (4.1)	4	4	4
Time and Temperature (4.2)	2	2	0
Money (4.3)	3	3	2
Data Analysis	**7**	**7**	**4**
Data Analysis (5.1)	4	4	2
Probability (5.2)	3	3	2
Total Test	**50**	**50**	**30**

OCCT Test Blueprint and Actual Item Counts: Grade 4 Mathematics

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Algebraic Reasoning: Patterns and Relationships	**7**	**7**	**3**
Algebra Patterns (1.1)	3	3	0
Equations (1.2)	2	2	0
Number Properties (1.3)	2	2	3
Number Sense and Operation	**18**	**18**	**5**
Number Sense (2.1)	8	8	0
Number Operations (2.2)	10	10	5
Geometry	**9**	**9**	**5**
Lines (3.1)	2	2	0
Angles (3.2)	2	2	0
Polygons (3.3)	3	3	5
Transformations (3.4)	2	2	0
Measurement	**9**	**9**	**8**
Measurement (4.1)	5	5	3
Time and Temperature (4.2)	2	2	0
Money (4.3)	2	2	5
Data Analysis	**7**	**7**	**9**
Data Analysis (5.1)	2	2	0
Probability (5.2)	2	2	3
Central Tendency (5.3)	3	3	6
Total Test	**50**	**50**	**30**

OCCT Test Blueprint and Actual Item Counts: Grade 5 Mathematics

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Algebraic Reasoning: Patterns and Relationships	**13**	**13**	**3**
Algebra Patterns (1.1)	5	5	0
Equations (1.2)	4	4	0
Number Properties (1.3)	4	4	3
Number Sense and Operation	**16**	**16**	**14**
Number Sense (2.1)	8	8	9
Number Operations (2.2)	8	8	5
Geometry	**7**	**7**	**9**
Circles and Polygons (3.1)	4	4	6
Angles (3.2)	3	3	3
Measurement	**7**	**7**	**2**
Measurement (4.1)	5	5	0
Money (4.2)	2	2	2
Data Analysis	**7**	**7**	**2**
Data Analysis (5.1)	3	3	0
Probability (5.2)	2	2	0
Central Tendency (5.3)	2	2	2
Total Test	**50**	**49**	**30**

OCCT Test Blueprint and Actual Item Counts: Grade 6 Mathematics

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Algebraic Reasoning: Patterns and Relationships	**13**	**13**	**9**
Algebra Patterns (1.1)	4	4	0
Expressions and Equations (1.2)	4	4	5
Number Properties (1.3)	3	3	2
Solving Equations (1.4)	2	2	2
Number Sense and Operation	**15**	**15**	**5**
Number Sense (2.1)	5	5	0
Number Operations (2.2)	10	10	5
Geometry	**8**	**8**	**6**
Three Dimensional Figures (3.1)	2	2	1
Congruent and Similar Figures (3.2)	2	2	0
Coordinate Geometry (3.3)	4	3	5
Measurement	**7**	**7**	**4**
Circles (4.1)	4	4	3
Conversions (4.2)	3	3	1
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**	**49***	**30**

*One operational item aligned to objective 3.3 was suppressed because of an error.

OCCT Test Blueprint and Actual Item Counts: Grade 7 Mathematics

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Algebraic Reasoning: Patterns and Relationships	**15**	**15**	**16**
Linear Relationships (1.1)	5	5	5
Solving Equations (1.2)	5	5	5
Solving and Graphing Inequalities (1.3)	5	5	6
Number Sense and Operation	**11**	**11**	**5**
Number Sense (2.1)	5	5	4
Number Operations (2.2)	6	6	1
Geometry	**8**	**8**	**1**
Classifying Figures (3.1)	1-3	2	0
Lines and Angles (3.2)	1-3	2	1
Transformations (3.3)	4	4	0
Measurement	**9**	**9**	**4**
Perimeter and Area (4.1)	5	5	2
Circles (4.2)	2	2	2
Composite Figures (4.3)	2	2	0
Data Analysis	**7**	**7**	**4**
Data Analysis (5.1)	2	2	2
Probability (5.2)	2	2	0
Central Tendency (5.3)	3	3	2
Total Test	**50**	**50**	**30**

OCCT Test Blueprint and Actual Item Counts: Grade 8 Mathematics

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Algebraic Reasoning: Patterns and Relationships	**16**	**16**	**9**
Equations (1.1)	10-12	12	5
Inequalities (1.2)	4-6	4	4
Number Sense and Operation	**11**	**11**	**14**
Number Sense (2.1)	3-4	3	3
Number Operations (2.2)	7-8	8	11
Geometry	**9**	**9**	**3**
Three Dimensional Figures (3.1)	5	5	1
Pythagorean Theorem (3.2)	4	4	2
Measurement	**7**	**7**	**0**
Surface Area and Volume (4.1)	3	3	0
Ratio and Proportions (4.2)	2	2	0
Composite Figures (4.3)	2	2	0
Data Analysis	**7**	**7**	**4**
Data Analysis (5.1)	3	3	0
Central Tendency (5.3)	4	4	4
Total Test	**50**	**50**	**30**

OCCT Test Blueprint and Actual Item Counts: Grade 3 Reading

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Vocabulary	**12**	**12**	**6**
Words in Context (2.1)	2-4	3	3
Affixes, Roots, and Stems (2.2)	2-4	2	0
Synonyms, Antonyms, and Homonyms (2.3)	2-4	3	2
Using Resource Materials (2.4)	2-4	4	1
Comprehension/Critical Literacy	**24**	**26**	**19**
Literal Understanding (4.1)	5	5	5
Inferences and Interpretation (4.2)	7	9	3
Summary and Generalization (4.3)	6	5	3
Analysis and Evaluation (4.4)	6	7	7
Literature	**8**	**6**	**3**
Literary Elements (5.2)	3-4	3	1
Figurative Language/Sound Devices (5.3)	4-5	3	2
Research and Information	**6**	**6**	**2**
Accessing Information (6.1)	6	6	2
Total Test	**50**	**50**	**29***

OCCT Test Blueprint and Actual Item Counts: Grade 4 Reading

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Vocabulary	**12**	**12**	**4**
Words in Context (1.1)	4	5	1
Affixes, Roots, and Stems (1.2)	4	3	2
Synonyms, Antonyms and Homonyms (1.3)	4	4	1
Comprehension/Critical Literacy	**23**	**23**	**17**
Literal Understanding (3.1)	4	4	3
Inferences and Interpretation (3.2)	6	4	5
Summary and Generalization (3.3)	7	7	4
Analysis and Evaluation (3.4)	6	8	5
Literature	**9**	**9**	**3**
Literary Elements (4.2)	5	5	2
Figurative Language/Sound Devices (4.3)	4	4	1
Research and Information	**6**	**6**	**6**
Accessing Information (5.1)	6	6	6
Total Test	**50**	**50**	**30**

OCCT Test Blueprint and Actual Item Counts: Grade 5 Reading

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Vocabulary	**12**	**11**	**4**
Words in Context (1.1)	4	3	2
Affixes, Roots, and Stems (1.2)	4	5	2
Synonyms, Antonyms, and Homonyms (1.3)	4	3	0
Comprehension/Critical Literacy	**20**	**22**	**12**
Literal Understanding (3.1)	4	7	3
Inferences and Interpretation (3.2)	4-6	6	5
Summary and Generalization (3.3)	4-6	4	3
Analysis and Evaluation (3.4)	4-6	5	1
Literature	**12**	**11**	**9**
Literary Genre (4.1)	4	3	2
Literary Elements (4.2)	4	3	5
Figurative Language/Sound Devices (4.3)	4	5	2
Research and Information	**6**	**6**	**5**
Accessing Information (5.1)	2-4	4	2
Interpreting Information (5.2)	2-4	2	3
Total Test	**50**	**50**	**30**

OCCT Test Blueprint and Actual Item Counts: Grade 6 Reading

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Vocabulary	**8**	**8**	**3**
Words in Context (1.1)	4	2	2
Word Origins (1.2)	4	6	1
Comprehension/Critical Literacy	**20**	**22**	**10**
Literal Understanding (3.1)	4	8	1
Inferences and Interpretation (3.2)	4-6	5	3
Summary and Generalization (3.3)	4-6	4	3
Analysis and Evaluation (3.4)	4-6	5	3
Literature	**14**	**12**	**11**
Literary Genres (4.1)	4	7	2
Literary Elements (4.2)	4-6	2	4
Figurative Language/Sound Devices (4.3)	4-6	3	5
Research and Information	**8**	**8**	**6**
Accessing Information (5.1)	4	4	3
Interpreting Information (5.2)	4	4	3
Total Test	**50**	**50**	**30**

OCCT Test Blueprint and Actual Item Counts: Grade 7 Reading

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Vocabulary	**10**	**10**	**5**
Words in Context (1.1)	3-4	6	1
Word Origins (1.2)	3-4	1	2
Idioms and Comparisons (1.3)	3-4	3	2
Comprehension/Critical Literacy	**20**	**20**	**11**
Literal Understanding (3.1)	4-5	5	2
Inferences and Interpretation (3.2)	4-6	6	1
Summary and Generalization (3.3)	4-6	5	4
Analysis and Evaluation (3.4)	4-6	4	4
Literature	**12**	**12**	**6**
Literary Genres (4.1)	4	5	2
Literary Elements (4.2)	4	3	0
Figurative Language/Sound Devices (4.3)	4	4	4
Research and Information	**8**	**8**	**8**
Accessing Information (5.1)	4	5	5
Interpreting Information (5.2)	4	3	3
Total Test	**50**	**50**	**30**

OCCT Test Blueprint and Actual Item Counts: Grade 8 Reading

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Vocabulary	**6**	**7**	**6**
Words in Context (1.1)	2	2	4
Word Origins (1.2)	2	2	0
Idioms and Comparisons (1.3)	2	3	2
Comprehension/Critical Literacy	**21**	**21**	**10**
Literal Understanding (3.1)	4-5	4	3
Inferences and Interpretation (3.2)	4-6	6	1
Summary and Generalization (3.3)	5-7	4	3
Analysis and Evaluation (3.4)	6-8	7	3
Literature	**15**	**14**	**8**
Literary Genre (4.1)	4-5	5	2
Literary Elements (4.2)	5-7	5	1
Figurative Language/Sound Devices (4.3)	4-6	4	5
Research and Information	**8**	**8**	**4**
Accessing Information (5.1)	4	4	3
Interpreting Information (5.2)	4	4	1
Total Test	**50**	**50**	**30**

OCCT Test Blueprint and Actual Item Counts: Grade 5 Science

OKC^3 Standard and Objective	Ideal Number of Items for Alignment to *OKC*^{3}*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Process Standards			
Observe and Measure	**10**	**11**	**11**
SI Metric (P1.1)	4-6	4	5
Similar/different characteristics (P1.2)	4	7	6
Classify	**10**	**9**	**14**
Observable properties (P2.1)	4-6	5	9
Serial order (P2.2)	4-5	4	5
Experiment	**13-15**	**13**	**13**
Experimental design (P3.2)	9-11	9	7
Hazards/practice safety (P3.4)	4	4	6
Interpret and Communicate	**12-14**	**12**	**22**
Data tables/line/bar/trend and circle graphs (P4.2)	4-6	4	9
Prediction based on data (P4.3)	4-6	5	7
Explanations based on data (P4.4)	4-6	3	6
Total Test	**45**	**45**	**60**
Content Standards			
Properties of Matter and Energy	**16-18**	**17**	**24**
Matter has physical properties (1.1)	4-5	4	6
Physical properties can be measured (1.2)	4-5	5	6
Energy can be transferred (1.3)	4-5	4	6
Potential/Kinetic Energy (1.4)	4-5	4	6
Organisms and Environments	**10-13**	**12**	**12**
Organisms dependence (2.1)	5-7	6	6
Individual organism and species survival (2.2)	5-7	6	6
Structures of the Earth and the Solar System	**12-15**	**12**	**18**
Properties of Soils (3.1)	4-6	4	6
Weather patterns (3.2)	4-6	4	6
Earth as a planet (3.3)	4	4	6
Total Test	**41**	**41**	**54**

* 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

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Process Standards			
Observe and Measure	**8-11**	**8**	**9**
Qualitative/quantitative observations/changes (P1.1)	4-6	4	3
SI (metrics) units/appropriate tools (P1.2 and P1.3)	4-5	4	6
Classify	**7-9**	**9**	**14**
Classification system (P2.1)	4-6	4	7
Properties ordered (P2.2)	3-5	5	7
Experiment	**15-17**	**16**	**19**
Experimental design (P3.2)	6-7	7	8
Identify variables (P3.3)	6-7	6	9
Hazards/practice safety (P3.6)	3-4	3	2
Interpret and Communicate	**12-14**	**12**	**18**
Data tables/line/bar/trend and circle graphs (P4.2)	6-7	6	12
Explanations/prediction (P4.3)	6-7	6	6
Total Test	**45**	**45**	**60**
Content Standards			
Properties and Chemical Changes in Matter	**8**	**8**	**11**
Chemical reactions (1.1)	4	4	5
Conservation of matter (1.2)	4	4	6
Motion and Forces	**8**	**8**	**10**
Motion of an object (2.1)	4	4	4
Object subjected to a force (2.2)	4	4	6
Diversity and Adaptations of Organisms	**7**	**7**	**8**
Classification (3.1)	3	3	5
Internal and external structures (3.2)	4	4	3
Structures/Forces of the Earth/Solar System	**11**	**11**	**20**
Landforms result from constructive and destructive forces (4.1)	4	4	5
Rock cycle (4.2)	3-4	4	4
Global Weather Patterns (4.3)	3-4	3	11
Earth's History	**7-8**	**8**	**9**
Catastrophic events (5.1)	3-4	4	5
Fossil evidence (5.2)	3-4	4	4
Total Test	**41-42**	**41**	**58**

* 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

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
James Towne Settlement and Plimoth Plantation Exploration	**8**	**0**	**25**
James Towne Settlement (1.1, 1.2, 1.3, 1.4)	4	0	19
Plimoth Plantation (1.5)	4	0	6
Colonial America	**10**	**0**	**24**
Colonial economics, trade/migration, perspectives (2.1, 2.3, 2.6)	4-6	0	14
Self-government, role of religion, leaders, and British and Native American Relationships (2.2, 2.4, 2.5)	4-6	0	10
American Revolution	**18**	**0**	**20**
Causes and effects of American Revolution (3.1)	4-6	0	5
Founding Documents of the Revolutionary Era (3.2, 3.3, 3.4)	4-5		10
Events of the Revolutionary War (3.5)	4-5	0	3
Key individuals of the Revolutionary Era (3.6)	4-5	0	2
Early Federal Period	**14**	**0**	**18**
Causes, leaders, and issues of the Constitutional Convention (4.1, 4.2)	4-5	0	17
Purposes and principles of the U.S. Constitution (4.3)	4-6	0	1
Ratification of the U.S. Constitution and the Bill of Rights (4.4, 4.5)	4-5	0	0
Total Test	**50**	**0**	**87**

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

OKC[3] Standard and Objective	Ideal Number of Items for Alignment to *OKC*[3]*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Geographic Tools/Geography Skills	**6**	**0**	**5**
Human and Physical Characteristics of Regions	**12**	**0**	**14**
Political and Physical/Cultural Regions (2.1, 2.2)	4-6	0	3
Physical and Human Characteristics Linking/Dividing Regions (2.3, 2.5)	4-6	0	9
Conflict and Cooperation (2.4)	4-6	0	2
Physical Systems of the Earth	**6**	**0**	**14**
Visual Information, Landforms and Bodies of Water (3.1)	2-4	0	12
Impact of Natural Disasters on Human Populations (3.2)	4-5	0	2
Human Systems: People and Cultures	**16**	**0**	**33**
Cultural Traits, Major World Religions, and Major Political Systems (4.1, 4.2, 4.5)	6-8	0	11
Economic Systems, Economic Interdependence and Trade (4.4, 4.6)	4-5	0	12
Human Characteristics of Developing and Developed Countries and Population Issues (4.3, 4.7)	4-5	0	10
Human Interaction with the Environment	**10**	**0**	**11**
Distribution of Resources (5.1)	4-6	0	4
Human Modification and Regional Problems (5.2, 5.3)	4-6	9	7
Total Test	**50**	**0**	**77**

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

OKC^3 Standard and Objective	Ideal Number of Items for Alignment to *OKC*^{3}*	Actual Number of Items on 2013 Test	Number of Items Field Tested in 2013
Causes and Events of the American Revolution	**8**	**0**	**13**
Consequences of the French and Indian War, British Imperial Policies (1.1, 1.2)	4	0	7
Ideological War, *Declaration of Independence's* Grievances, Ideals, and Social Contract Selection (1.3, 1.4, 1.5)	4	0	6
The Revolutionary Era (2.0)	**6**	**0**	**14**
Articles of Confederation, Motivations & Choices, Key Military & Diplomatic Events (2.1, 2.2, 2.3)	6	0	14
Developing the American Government System (3.0)	**10**	**0**	**3**
Causes for the Constitutional Convention, and Ratification (3.1, 3.2, 3.3)	4-6	0	2
Constitutional Principles and the *Bill of Rights* (3.4, 3.5)	4-6	0	1
The Transformation of the United States to the Mid-1800s	**16**	**0**	**39**
Major Events and Issues of Early Presidential Administrations (4.1)	4-6	0	13
Jacksonian Era and Westward Expansion (4.2, 4.6)	4-6	0	12
Sectional Economic Systems, African American Experiences, and Reform Movements/Leaders (4.3, 4.4, 4.5)	4-6	0	14
Causes, Events, and Leadership in the Civil War	**10**	**0**	**23**
Causes of the Civil War: 1850s through the 1860s Presidential Elections (5.1, 5.2)	4-6	0	13
Advantages/Disadvantages, Leadership, Major Turning Points of the War (5.3, 5.4)	4-6	0	10
Total Test	**50**	**0**	**80**

Appendix B
A Statistical Investigation of Oklahoma Computer Disruptions

2013 No. 053

A Statistical Investigation of Oklahoma Computer Disruptions

Final Report

Prepared for: **Oklahoma Department of Education**
2500 N. Lincoln Blvd.
Oklahoma City, OK 73105-4599

Authors: **Matthew S. Swain**
Arthur A. Thacker
Bethany H. Bynum

Date: **August 6, 2013**

Human Resources Research Organization (HumRRO)
66 Canal Center Plaza, Suite 700, Alexandria, Virginia 22314-1578
Phone: 703.549.3611 | Fax: 703.549.9661 | www.humrro.org

HumRRO
Human Resources Research Organization

2013 No. 053

A Statistical Investigation of Oklahoma Computer Disruptions

Final Report

Prepared for: **Oklahoma Department of Education**
2500 N. Lincoln Blvd.
Oklahoma City, OK 73105-4599

Authors: **Matthew S. Swain**
Arthur A. Thacker
Bethany H. Bynum

Date: **August 6, 2013**

A Statistical Investigation of Oklahoma Computer Disruptions

Table of Contents

List of Tables

A Statistical Investigation of Oklahoma Computer Disruptions

Abstract

On April 29th and April 30th, 2013, students in Oklahoma experienced technological delays or "interruptions" while completing their statewide Math and Reading assessments. The potential impact of those disruptions on test scores is the focus of this investigation. Multiple analytical approaches were conducted to determine the presence and magnitude of any effects. Additionally, analyses were conducted on multiple "cohorts" of students to examine if the effects of disruption on these two days were disparate from typical delays or interruptions during testing.

The testing vendor, CTB McGraw/Hill (CTB), conducted their own investigation of the impact of the test interruptions. The Oklahoma Department of Education (ODE) requested that an independent investigation of the interruption also be conducted. The Human Resources Research Organization (HumRRO) was selected to conduct the independent investigation[1]. Data were provided by CTB (for the 2013 cohort of students) and ODE (test data from the prior vendor). CTB also shared their investigation methodology and results. This investigation verified CTB's initial data (e.g. percent of students impacted) but used independently derived methodology to investigate the impact of the interruptions and draw conclusions. HumRRO primarily compared students' predicted scores to their actual interrupted scores to investigate the presence and magnitude of the interruption effect.

HumRRO's analyses detected a small effect of the interruption. This effect was not consistent across grades and/or subjects, however. Students in some grade/subjects appear to have been disadvantaged by the interruption, while others performed better than expected. Even when statistically significant differences in mean scores were found within a grade/subject, the difference was not consistent across the distribution of students taking the test. Therefore, HumRRO does not recommend a mathematical adjustment of test scores.

Methodology

The primary goal of this investigation was to determine if the score a disrupted student received would have been different if the student had not been disrupted. That is, if two students who were very similar on all available variables that predicted 2013 test scores differed only in whether or not they were disrupted, any differences in those test scores could be more confidently stated to be an effect of disruption. Propensity score matching provides the ability to match students with similar distributions on a set of variables (Connelly, Sackett, & Waters, 2013). Each student in the disrupted group for each grade and subject test was matched with a student from the non-disrupted group. This matching was done using all available variables except for pertinent 2013 scores (i.e., the dependent variable). Multiple grades and subject tests were investigated leading to formation of four "cohorts."

Description of Four Cohorts

Students in grades 6 through 8 completing their Math and Reading tests and students in high school completing their end of instruction (EOI) exams were included in the investigation. The

[1] A similar study was carried out by Hill (2013) to investigate this same issue in Indiana.

OK testing window for the grades 3-8 tests runs from April 10 to April 30. Cohort A was defined first as all students in grades 6 through 8 who were interrupted during their Math and Reading tests. Because the interruptions on 4/29 and 4/30 were of particular interest, Cohort B was further refined from Cohort A to only include students who were interrupted on those two days. Tables 1 and 2 provide the frequencies of interruption by day and grade for Math and Reading tests, respectively. Clearly, interruptions on 4/29 and 4/30 were atypical.

Table 1. Number of Students Interrupted by Day and Grade on Math Test

	Grade			
Day	6	7	8	Total
4/10/2013	139	23	12	174
4/11/2013	192	96	40	328
4/12/2013	122	55	26	203
4/15/2013	97	92	14	203
4/16/2013	95	75	45	215
4/17/2013	275	158	43	476
4/18/2013	108	69	61	238
4/19/2013	45	38	41	124
4/22/2013	48	23	31	102
4/23/2013	59	27	154	240
4/24/2013	66	46	167	279
4/25/2013	35	24	43	102
4/26/2013	14	12	12	38
4/29/2013	**263**	**244**	**427**	**934**
4/30/2013	**490**	**344**	**676**	**1510**
5/1/2013	10	5	15	30
5/2/2013	9	12	8	29
5/3/2013	1	4	1	6
5/6/2013	0	0	3	3
5/7/2013	0	1	0	1
Total	**2068**	**1348**	**1819**	**5235**

Table 2. Number of Students Interrupted by Day and Grade on Reading Test

	Grade			
Day	6	7	8	Total
4/10/2013	174	86	78	338
4/11/2013	161	160	265	586
4/12/2013	276	146	191	613
4/15/2013	142	338	115	595
4/16/2013	286	216	124	626
4/17/2013	177	77	122	376
4/18/2013	84	127	93	304
4/19/2013	27	30	62	119
4/22/2013	45	52	71	168
4/23/2013	109	95	169	373
4/24/2013	116	68	106	290
4/25/2013	95	19	92	206
4/26/2013	9	20	82	111
4/29/2013	**220**	**173**	**384**	**777**
4/30/2013	**357**	**269**	**264**	**890**
5/1/2013	52	10	11	73
5/2/2013	7	16	11	34
5/3/2013	1	3	1	5
5/6/2013	1	0	3	4
5/7/2013	0	1	11	12
Total	**2339**	**1906**	**2255**	**6500**

Interruptions also occurred during the EOI exams completed primarily by high school students. Using the available data, two additional cohorts were constructed to explore the presence of disruption effects on EOI exams. These cohorts were chosen primarily based on what data were available. More specifically, the goal was to choose exams that had sufficient sample size to detect disruption effects, if any. Cohort C focused on students interrupted during their Algebra I test whereas Cohort D focused on students interrupted during their English II exam.

The report that follows will focus first on Cohorts A and B followed by Cohorts C and D.

Computer Disruption Data Overview

Interruption data provided to HumRRO included the test subject and grade level that each student was completing at the time of interruption for Cohorts A and B. Archival data regarding the student's prior year performance, demographic information, as well as current year performance on other exams was also provided. A total of 130,429 valid, unique IDs with complete Math test scores were recorded and 130,373 valid, unique IDs with complete Reading test scores were also recorded. After integrating the interruption data provided to HumRRO, it appears that 5,235 students were interrupted during their Math test *across all days* out of 130,429 students. For the Reading test, 6,500 students were interrupted *across all days* out of 130,373 students. These students formed Cohort A. An interruption rate was calculated to

examine the prevalence of interruption. The numbers of disrupted and non-disrupted students before matching, separated by grade level, are provided in Table 3.

Table 3. Interruption Data by Subject and Grade

	Math			Reading		
Grade	All Students	Disrupted	Interruption Rate	All Students	Disrupted	Interruption Rate
6	43,976	2,068	4.6%	43,957	2,339	5.3%
7	43,728	1,348	3.0%	43,716	1,906	4.3%
8	42,725	1,819	4.2%	42,700	2,255	5.2%
Total	**130,429**	**5,235**	**4.0%**	**130,373**	**6,500**	**5.0%**

The numbers of students in Table 3 are fairly consistent with those provided by CTB, for both non-disrupted and disrupted students. For those students who were disrupted, some were interrupted or delayed during their testing session multiple times. Although most students who were disrupted only experienced one delay, some students experienced many delays. Table 2 provides the frequency of disruption for each student in the disrupted group by test.

Table 4. Frequency of Interruption by Test

Interruption Number	Math	Reading
1	4,455	5,446
2	423	681
3	186	171
4	80	88
5	42	52
6	25	26
7	13	19
8	8	7
9	1	3
10	2	5
11	0	1
12	0	0
13	0	1

Propensity Score Matching

In an attempt to isolate the potential effect of disruption, propensity score matching was employed to select a sample of students from the non-disrupted group that closely resembles the disrupted group on all variables available that relate to 2013 scores. In one sense, propensity score matching attempts to "control for" these variables. With these groups established, differences in 2013 scores are more likely to be due to an effect of disruption.

The following variables were used as matching variables for Cohorts A and B:

- 2012 Math and Reading scaled scores
- Other subject 2013 scale scores (Reading score used if interrupted on 2013 Math test)
- Gender (0 = Female, 1 = Male)
- Ethnicity (Dummy coded)
- Limited English proficiency
- Student-level free/reduced lunch
- School-level proportion of free/reduced lunch students
- School-level achievement (Average of 2012 student-level Reading and Math scores)

Differences between disrupted and non-disrupted students on these matching variables were examined prior to matching. These differences were calculated using Cohen's *d* with "rule of thumb" cutoff values suggested by Harder, Stuart, & Anthony (2010). Across grades and subjects, the disrupted students seemed to be disproportionately African American and from schools with more free and reduced lunch students and lower achievement means. Given the relationships between these variables and the test scores on the disrupted test, it is clear why propensity score matching is a necessary method.

Next, using logistic regression, we regressed group membership (disrupted or not disrupted) on to the matching variables. Generally, the pseudo R-square values of the logistic regression were small, ranging from .0042 to .0242. Overall, the small values suggest that the combination of prior year student achievement, demographics, SES, school-level achievement or school-level SES had little relationship to the likelihood that a student experienced disruptions.

To match the two samples, the predicted disruption probabilities from the logistic regression analyses were saved. The predicted probabilities represent the probability that a student was in the disrupted sample. We used the *nearest neighbor method* to match the two samples. That is, the predicted disruption probability for each student in the disrupted sample was matched to the student with the closest predicted disruption probability in the non-disrupted sample. The sampling was done without replacement so that each student in the disrupted sample was matched with a unique student in the non-disrupted sample. The average difference between a disrupted student's predicted probability and the matched non-disrupted student's predicted probability was .000003 for Reading and .000003 for Math. The largest difference was .0016. Differences no larger than .20 have been shown to reduce bias and produce accurate group difference estimates (Austin, 2009; Connelly, Sackett, & Waters, 2013). The results suggest that everyone in the disrupted sample was matched with a student in the non-disrupted sample with a very similar predicted disruption probability. To further evaluate the closeness of the matched sample, we examined the mean difference of the matching variables. The average Cohen's *d* between the two samples was .002 and ranged from -.060 to .075 after matching. All of the differences greater than .10 prior to matching were reduced. Near zero effect sizes suggest that

our samples have been effectively balanced on prior year achievement, school-level achievement, gender, ethnicity, and school-level free-or-reduced percentage. A summary of the mean, standard deviations, and effect sizes before and after matching can be found in the Appendix.

2013 Score Differences

Using the matched samples, we examined whether students' scores for tests impacted by computer interruptions differed from students' scores that were not impacted by computer disruptions. By matching the samples on variables that are likely to predict student scores, any difference between the two samples can be better attributed to the computer disruptions. We used a number of analyses to examine differences in scores. Scale scores for Math and Reading for the matched groups were examined four ways:

1. Mean differences on 2013 scores
2. R^2 change when combing groups and adding dichotomous disruption variable
3. R^2 differences when predicting 2013 scores separately
4. Applying Non-Disrupted regression equation from step 3 onto Disrupted group as well as 5^{th}, 10^{th}, 90^{th}, and 95^{th} percentile cuts

Cohort A Analyses

Differences in average 2013 test scores

Because these two matched groups are alike on variables that are typically predictive of test scores, if testing disruptions had no overall impact, then the averages of the 2013 test scores for the two groups would be expected to be nearly identical. Differences in the average 2013 test scores would be evidence that computer disruption impacted test performance. We examined mean differences between the two samples using a t-test and Cohen's *d* effect size. Table 5 presents the results by grade and subject for Cohort A.

Table 5. Descriptive and Inferential Statistics of 2013 Test Scores After Matching – Cohort A

		Disrupted		Non-Disrupted			
	N	Mean	SD	Mean	SD	*t*	*d*
Grade 6 Reading	2,002	733.7	77.87	731.9	73.84	-0.75	0.024
Grade 7 Reading	1,625	724.6	66.99	726.4	68.69	0.76	-0.027
Grade 8 Reading	1,900	757.8	84.37	758.6	78.62	0.32	-0.010
Grade 6 Math	1,710	744.9	77.77	743.6	75.24	-0.49	0.017
Grade 7 Math	1,085	726.5	84.46	730.8	80.98	1.22	-0.052
Grade 8 Math	1,533	727.8	79.72	732.0	86.09	1.42	-0.051

Note. No *p*-values below .05.

For Reading, the effect sizes ranged from -.027 to .024, suggesting no overall directional effect. For Math, the effect sizes ranged from -.052 to .017. All effect sizes indicate very small mean differences between scale scores of the matched samples.

Examining the predictability of 2013 test scores

First, to examine predictability of 2013 test scores, we examined the incremental variance that disruption accounted for when added to the prediction of the 2013 test scores beyond other known indicators of performance (including prior year achievement, ethnicity, gender, school-level achievement, and school-level percentage of free-or-reduced lunch). If the inclusion of disruption in the multiple regression models adds to the estimation of 2013 scores, then this would supply evidence that disruption impacted 2013 test scores. For this model, the two groups were included in the regression models together.

Table 6 reports the R-square values for each model and the R-square change between the two models. With the inclusion of disruption in the model, the R-square value changed very little or not at all, indicating that disruption did not add to the prediction of 2013 scores.

Table 6. Combined Regression Equation Adding Disruption – Cohort A

	N	Covariates Only R^2	Covariates + Disrupted R^2	ΔR^2
Grade 6 Reading	4,004	0.5759	0.5760	<0.001
Grade 7 Reading	3,250	0.5331	0.5339	0.001
Grade 8 Reading	3,800	0.5105	0.5105	<0.001
Grade 6 Math	3,420	0.5788	0.5791	<0.001
Grade 7 Math	2,170	0.5550	0.5552	<0.001
Grade 8 Math	3,066	0.5162	0.5167	0.001

Next, using all of the data we have available to create equations that predict students' 2013 test scores, we can statistically determine whether disrupted students scored differently than expected. Prediction equations were estimated for the disrupted students and separately for the matched sample of non-disrupted students. In addition to prediction equations, this technique gives us multiple regression coefficients (R-square) that can be interpreted like a correlation coefficient to tell us how well 2013 test scores can actually be predicted from our available data. If students' performance was affected, the strength of the prediction for the disrupted students should be less than the non-disrupted students as shown by lower multiple regression coefficients. A lower coefficient means that students' performance in the disrupted group was not as predictable as students' performance in the non-disrupted group. This would supply another piece of evidence about the impact of the computer disruptions. Table 7 presents the multiple regression coefficients (R-square) for the disrupted and non-disrupted groups. Overall, 2013 test scores were well predicted for both samples, with 52% to 59% of the variance accounted for by the predictor variables for Reading, and 55% to 59% for Math. Generally, there were slightly higher R-square values for the non-disrupted group, although the difference in variance accounted for was practically small ranging from 0.9% to 6%.

Table 7. Separate Regression Equations for Disrupted and Non-Disrupted Groups – Cohort A

	N	Disrupted R^2	Non-Disrupted R^2	Δ R^2
Grade 6 Reading	2,002	0.574	0.589	-0.015
Grade 7 Reading	1,625	0.521	0.551	-0.030
Grade 8 Reading	1,900	0.533	0.494	0.039
Grade 6 Math	1,710	0.578	0.587	-0.009
Grade 7 Math	1,085	0.548	0.577	-0.029
Grade 8 Math	1,533	0.488	0.547	-0.059

Examining distributions of predicted student scores

The prediction equations for the non-disrupted students give us a statistical statement about what to expect normally for students testing under non-disrupted conditions. We know that the prediction is not perfect, but given the high R-square values we can use the prediction equation to calculate how disrupted students might have scored had they not been disrupted. For each disrupted student, we computed their predicted score using the regression equation computed for the non-disrupted students. Next, we took the difference between the observed score and predicted score, where positive values indicate higher observed scores than predicted and negative values indicate higher predicted scores than observed--imagine a distribution of observed minus predicted score differences. "*Large numbers*" of students with "*notable differences*" between obtained and predicted scores provides another piece of evidence about the impact of the computer disruptions. Table 8 presents the distribution of observed and predicted scores for the disrupted and non-disrupted sample for Reading and Math. The difference between the observed and predicted score is also reported.

Finally, we assessed "*large number*" and a "*notable difference*" between obtained and predicted scores for interrupted students by comparing the difference in observed and predicted scores for the non-disrupted group to the difference for the disrupted group. Since our non-disrupted group of students represents what would be expected under normal testing conditions, on average, the difference between the observed and predicted score can be considered a baseline. That is, the average difference for non-disrupted students is what is expected to naturally occur based on the fact that our prediction, inherently, can never be perfect.

We evaluated differences in two ways. First, we examined the difference in the standard deviation of differences for the disrupted and non-disrupted group using an *F*-test. This provides information on whether the spread of differences in predicted and observed scores is statistically different for the two groups (i.e., students are more affected, on average, in the disrupted group than the non-disrupted group). Results suggest that the standard deviation of the difference scores is larger in the disrupted group for grade 6 Reading and grades 6 and 7 Math (see bolded SDs in Table 8).

Next, we took the difference in observed and predicted scores at the 5th, 10th, 90th and 95th percentile for the non-disrupted group and determined the number of students in the disrupted group that were at or below the same cut point for the 5th and 10th percentile and those that were at or above the cut point for the 90th and 95th percentile. If there are a large number of students below the 5th and 10th percentile of the non-disrupted group than what would be expected (5% and 10%, respectively) then more students in the disrupted group scored lower than expected. On the other hand, if there are a large number of students above the 90th and

95th percentile of the non-disrupted group (10% and 5%, respectively) then more students in the disrupted group scored higher than expected. Either case would provide evidence that the computer disruption had an impact on scores. Table 9 presents the percent of students in the disrupted group with score differences less than the 5th and 10th percentile and higher than the 90th and 95th percentile of the non-disrupted group.

Table 8. Distribution of the Difference between Predicted and Observed Scores for Non-disrupted and Disrupted Groups – Cohort A

	Disrupted						Non-Disrupted						
	N	Mean	SD	5th	10th	90th	95th	*N*	Mean	SD	5th	10th	90th	95th
Grade 6 Reading														
Difference	2002	1.81	**52.02**	-73.94	-55.14	59.77	84.53	2002	0.00	47.36	-71.02	-52.14	56.69	78.04
Predicted Score	2002	731.94	**61.38**	634.72	660.52	805.18	826.03	2002	731.95	56.65	642.62	665.18	801.51	821.08
Observed Score	2002	733.75	**77.87**	613	639	833	860	2002	731.95	73.84	613	639	813	860
Grade 7 Reading														
Difference	1625	-3.87	46.91	-77.33	-58.60	49.90	67.02	1625	-0.07	46.01	-63.31	-48.98	50.19	74.55
Predicted Score	1625	728.44	51.25	645.67	667.32	788.63	806.22	1625	726.45	51.02	644.46	666.61	784.50	801.89
Observed Score	1625	724.56	66.99	621	650	797	849	1625	726.38	68.69	621	644	797	849
Grade 8 Reading														
Difference	1900	-0.64	58.57	-88.05	-68.67	68.58	98.49	1900	0.08	55.94	-86.26	-65.72	67.76	104.63
Predicted Score	1900	758.45	55.38	670.25	690.90	823.10	845.99	1900	758.57	55.23	670.35	692.53	822.76	843.49
Observed Score	1900	757.81	**84.37**	623	652.5	842	909	1900	758.65	78.62	630	658	842	909
Grade 6 Math														
Difference	1710	2.97	**51.22**	-72.99	-54.18	58.81	83.88	1710	0.04	48.37	-73.80	-53.87	59.80	80.12
Predicted Score	1710	741.94	**61.18**	649.31	671.46	815.01	841.66	1710	743.58	57.64	649.82	676.08	813.13	834.56
Observed Score	1710	744.91	77.77	622	651	833	865	1710	743.62	75.24	622	651	830	865
Grade 7 Math														
Difference	1085	-2.92	**58.45**	-89.32	-61.70	59.89	76.58	1085	0.00	52.70	-76.44	-57.62	56.04	73.67
Predicted Score	1085	729.39	64.43	623.49	654.87	802.54	828.87	1085	730.80	61.49	636.73	655.83	800.52	822.44
Observed Score	1085	726.46	84.46	605	629	827	852	1085	730.80	80.98	605	639	827	852
Grade 8 Math														
Difference	1533	-3.94	57.72	-89.46	-69.96	62.50	87.98	1533	0.01	57.97	-86.19	-66.31	68.29	95.63
Predicted Score	1533	731.72	57.56	636.86	666.24	800.67	824.81	1533	732.02	63.64	627.90	656.37	809.02	831.84
Observed Score	1533	727.78	79.72	602	632	816	854	1533	732.02	86.09	602	632	830	866

Note. Bolded values indicate significantly larger standard deviation for Disrupted group at $p < .05$.

Compare predictions for interrupted students to non-interrupted students

Table 9. Percent of Disrupted Students with Predicted and Observed Score Differences at the 5th, 10th, 90th and 95th Percentile of Non-disrupted Students – Cohort A

		5th	10th	90th	95th
	N	Larger than 5% or 10% = Disadvantaged		Larger than 5% or 10% = Advantaged	
Grade 6 Reading	2,002	5.29%	11.24%	10.94%	5.89%
Grade 7 Reading	1,625	8.74%	13.48%	9.97%	3.75%
Grade 8 Reading	1,900	5.32%	11.00%	10.05%	4.63%
Grade 6 Math	1,710	4.85%	10.23%	9.82%	5.79%
Grade 7 Math	1,085	6.54%	11.43%	11.24%	5.81%
Grade 8 Math	1,533	5.54%	11.02%	8.61%	3.78%

For Reading, the results show a slightly higher percent of students above the 5th and 10th cut point for grade 7. These results suggest that for grade 7 Reading, approximately 3.74% of students had lower observed scores than predicted scores based on the non-disrupted sample. In Math, the differences around the 5th and 10th percentiles were less pronounced, although the actual percentages did bounce around what would be expected.

On the upper end of the distribution, it can be noted that some students in the disrupted group scored higher than what would be expected. For example, some students in the disrupted sample scored high in Grade 7 Math than what was expected, but there were other students in this same group that were disadvantaged.

Cohort B Analyses

Because the disruptions on 4/29 and 4/30 appeared to be more pronounced, the above analyses were replicated using *only students disrupted on 4/29 and 4/30* as the disrupted sample. In this way, any effect of disruption, on average, may be more pronounced on this sample than by looking at students disrupted on any day.

Table 10. Cohort A and Cohort B Comparison

	Math			Reading		
Grade	Cohort A	Cohort B	Proportion of Cohort A	Cohort A	Cohort B	Proportion of Cohort A
6	2,068	753	0.36	2,339	577	0.25
7	1,348	588	0.44	1,906	442	0.23
8	1,819	1,103	0.61	2,255	648	0.29
Total	**5,235**	**2,444**		**6,500**	**1,667**	

Interestingly, the proportions of students in Cohort A who are also in Cohort B vary by grade and subject. Grade 8 Math students appear to have primarily been interrupted on 4/29 or 4/30.

Propensity score matching was again used to match the reduced sample of disrupted students to similar students in the non-disrupted group. The same four analyses conducted on Cohort A were also conducted on Cohort B.

Differences in average 2013 test scores

Table 11. Descriptive and Inferential Statistics of 2013 Test Scores After Matching – Cohort B

		Disrupted		Non-Disrupted			
	N	Mean	SD	Mean	SD	*t*	*d*
Grade 6 Reading	468	756.3	82.39	750.2	77.78	-1.16	0.08
Grade 7 Reading	617	743.2	78.27	744.9	74.07	0.39	-0.02
Grade 8 Reading	369	726.7	73.87	726.1	67.22	-0.12	0.01
Grade 6 Math	455	728.5	89.21	726.8	72.02	-0.33	0.02
Grade 7 Math	543	755.5	93.14	762.5	86.29	1.28	-0.08
Grade 8 Math	932	719.5	81.63	729.0	86.01	2.44*	-0.11

* $p < .05$.

Mean scores between the matched disrupted and non-disrupted students were neither statistically nor practically significant except for Grade 8 Math. Although the mean score of disrupted students was lower than the non-disrupted students at a statistical level, the practical significance was still small. Additionally, with 932 students in both samples, the power of this statistical test is very high so even a small difference may be statistically significant.

Examining the predictability of 2013 test scores

Regression equations were formed using the same predictor variables as in Cohort A. Again, one regression combined the disrupted and non-disrupted groups to examine the increase in predictive validity when adding "disruption" to the equation. Overall, the changes in R^2 values were very small.

Table 12. Combined Regression Equation Adding Disruption – Cohort B

	N	Covariates Only R^2	Covariates + Disrupted R^2	ΔR^2
Grade 6 Reading	936	0.6005	0.6009	<0.001
Grade 7 Reading	738	0.5319	0.5321	<0.001
Grade 8 Reading	1,086	0.5444	0.5444	<0.001
Grade 6 Math	1,234	0.5292	0.5293	<0.001
Grade 7 Math	910	0.4892	0.4894	<0.001
Grade 8 Math	1,864	0.5051	0.5080	0.003

Like Cohort A, two more regression equations were built to predict 2013 scores from the same predictor variables in the previous model. These models were run on both groups separately to examine if the difference in R^2 was substantively smaller in the disrupted group than the non-disrupted group. This may signal unexplained variance possibly attributable to a disruption effect. The R^2 values were smaller only in the Grade 8 Math exam, although this change is small, accounting for about 1% of variance explained.

Table 13. Separate Regression Equations for Disrupted and Non-Disrupted Groups – Cohort B

	N	Disrupted R^2	Non-Disrupted R^2	Δ R^2
Grade 6 Reading	468	0.619	0.604	0.015
Grade 7 Reading	369	0.555	0.532	0.024
Grade 8 Reading	543	0.574	0.530	0.045
Grade 6 Math	617	0.534	0.534	<0.000
Grade 7 Math	455	0.504	0.498	0.006
Grade 8 Math	932	0.461	0.560	-0.099

Examining distributions of predicted student scores

By looking at the distribution of difference scores (i.e., observed minus predicted scores), we can determine, as best as our models predict, the scores students who were disrupted would have received if they were not disrupted. Given that Cohort B is a refined sample of Cohort A, any "*notable difference*" or effect of disruption could be expected to be more extreme in this sample.

An *F*-test was conducted comparing the variance of observed, predicted, and difference scores, of both groups. The significantly larger standard deviations for the disrupted group are bolded in Table 14. Interestingly, the Grade 8 Math disrupted group did not have statistically significantly larger variance than its matched non-disrupted group.

Table 14. Distribution of the Difference between Predicted and Observed Scores for Non-disrupted and Disrupted Groups – Cohort B

	Disrupted						Non-Disrupted						
	N	Mean	SD	5th	10th	90th	95th	*N*	Mean	SD	5th	10th	90th	95th
Grade 6 Reading														
Difference	468	3.28	53.38	-87.02	-58.63	60.40	89.60	468	0.00	48.97	-69.55	-53.25	62.78	81.71
Predicted Score	468	753.03	53.73	659.27	684.55	820.58	838.81	468	750.23	60.43	653.23	678.31	829.04	849.02
Observed Score	468	756.31	82.39	619	659	862	894	468	750.23	77.78	633	652	833	860
Grade 7 Reading														
Difference	369	-1.86	**51.87**	-86.94	-63.94	51.56	71.41	369	-0.01	46.01	-65.14	-53.04	55.97	79.64
Predicted Score	369	728.55	47.56	649.69	668.17	785.31	799.94	369	726.09	49.01	646.52	665.73	781.42	797.76
Observed Score	369	726.69	73.87	593	633	818	849	369	726.08	67.22	621	650	797	849
Grade 8 Reading														
Difference	543	-1.92	63.31	-100.42	-70.16	71.13	98.62	543	-0.03	59.18	-86.42	-65.67	75.25	108.75
Predicted Score	543	757.46	66.52	639.76	672.18	837.02	865.69	543	762.54	62.79	666.74	691.77	837.14	858.30
Observed Score	543	755.53	93.14	591	637	870	909	543	762.51	86.29	623	658	870	909
Grade 6 Math														
Difference	617	0.04	50.59	-75.67	-57.21	57.35	81.22	617	1.69	54.40	-73.58	-60.63	60.76	84.91
Predicted Score	617	744.82	54.10	654.18	679.22	809.00	825.66	617	741.46	59.81	655.87	672.37	807.55	839.04
Observed Score	617	744.86	74.07	622	659	830	865	617	743.15	78.27	632	651	830	865
Grade 7 Math														
Difference	455	-0.94	**65.38**	-110.29	-66.03	71.21	81.95	455	-0.04	51.00	-79.98	-59.48	53.93	78.50
Predicted Score	455	729.48	53.50	637.79	659.40	791.01	812.24	455	726.81	50.85	646.49	670.06	784.81	799.42
Observed Score	455	728.54	**89.21**	597	639	827	852	455	726.76	72.02	618	648	807	838
Grade 8 Math														
Difference	932	-8.80	61.36	-99.62	-77.95	60.03	87.78	932	0.02	57.07	-86.13	-61.29	67.03	93.85
Predicted Score	932	728.32	55.36	638.45	667.91	790.78	811.54	932	728.98	64.35	622.63	652.67	802.95	825.13
Observed Score	932	719.52	81.63	588	623	816	846	932	729.00	86.01	589	623	830	866

Note. Bolded values indicate significantly larger standard deviation for Disrupted group at $p < .05$.

Compare predictions for interrupted students to non-interrupted students

Table 15. Percent of Disrupted Students with Predicted and Observed Score Differences at the 5th, 10th, 90th and 95th Percentile of Non-disrupted Students – Cohort B

		5th	10th	90th	95th
	N	Larger than 5% or 10% = Disadvantaged		Larger than 5% or 10% = Advantaged	
Grade 6 Reading	468	8.55%	11.97%	9.40%	6.62%
Grade 7 Reading	369	9.76%	13.55%	8.13%	4.07%
Grade 8 Reading	543	6.81%	12.52%	8.84%	4.05%
Grade 6 Math	617	4.70%	10.86%	11.02%	5.67%
Grade 7 Math	455	7.47%	11.43%	16.04%	6.15%
Grade 8 Math	932	6.97%	14.81%	8.80%	4.08%

According to Table 15, some disrupted students scored lower than expected whereas others scored higher than expected. The largest group of disadvantaged students appears to be Grade 8 Math at the 10th percentile mark; however, other groups have larger proportions of students at the 5th percentile mark.

Cohort C Analyses

In addition to the interruptions in Grades 6 through 8 exams, HumRRO was provided student scores and interruption data for EOI exams. These exams are primarily completed by high school students and are often taken predominantly in one grade level. This is relevant in selecting variables to predict the current year scores. Since prior performance is the best predictor of future performance, this fact limited our investigation to certain exams with available prior performance data.

Algebra I test scores were selected for Cohort C because most students who complete that exam are 9th graders and their Grade 8 Math exam is available as a sufficient indicator of prior performance. However, some 7th and 8th graders also completed the Algebra 2013 test. These students also completed their Math exams in 2013 and 2012, which can be used as prior performance indicators. Given that these students are likely advanced in this subject (i.e., upper end of the ability distribution), we can examine these students separately to determine if disruption had a differential effect on higher ability students.

To be clear, Cohort C is comprised of students who completed the Algebra I test in 2013 matched with 2012 Math scores; therefore, this sample should consist primarily of students who were 9th graders in spring 2013. The same 2013 Algebra I scores were matched to 2013 Math scores to result in a sample that primarily consists of students who were 7th and 8th graders in spring 2013. The same interruption data on the Algebra I test was used for both groups. The number of valid, unique Algebra I score records is 31,021; however, grade information was only provided in the interruption file. The number of interrupted cases during the 2013 Algebra I test by grade is displayed in Table 16.

Table 16. Interrupted Cases on 2013 Algebra I Test by Grade

	N	Percentage
Grade 7	59	2.07%
Grade 8	602	21.08%
Grade 9	1,759	61.59%
Grade 10	347	12.15%
Grade 11	72	2.52%
Grade 12	17	0.60%
Total	**2,856**	**100%**

Note that only the Grades 7 and 8 records from the interruption file were matched with the 2013 Algebra I score file to form the subset "2013 Algebra (7^{th} & 8^{th})." All analyses were conducted on all groups after propensity score matching. The matching formulas were identical for Cohort C; however, Math 2013 scores were used for the grades 7 and 8 subset.

Differences in average 2013 test scores

After propensity score matching, mean difference tests were performed on the matched samples. Note that the performance of the 7^{th} and 8^{th} grade group is higher than the group overall. Additionally, both Disrupted samples seem to have lower scores than the Non-Disrupted matched samples, overall.

Table 17. Descriptive and Inferential Statistics of 2013 Test Scores After Matching – Cohort C

		Disrupted		Non-Disrupted			
	N	Mean	SD	Mean	SD	*t*	*d*
2013 Algebra	2,179	739.5	50.17	744.7	49.41	3.40*	0.10
2013 Algebra (7^{th} & 8^{th})	621	763.5	55.45	771.5	48.81	2.70*	0.15

*$p < .05$.

Examining the predictability of 2013 test scores

The regression equations were performed on both groups in Cohort C. R^2 values were lower for the 2013 Algebra group likely due to only Math 2012 scores being available for a prior performance indicator. Including Math and Reading 2013 scores would have reduced the sample to the 7^{th} and 8^{th} grade group since 9^{th} graders likely did not complete the 2013 Math and Reading tests meant for 8^{th} graders. Importantly, the change in R^2 values were not practically large, indicating a near-zero effect of disruption on Algebra score variance.

Table 18. Combined Regression Equation Adding Disruption – Cohort C

	N	Covariates Only R^2	Covariates + Disrupted R^2	ΔR^2
2013 Algebra	4,358	0.3445	0.3460	-0.002
2013 Algebra (7^{th} & 8^{th})	621	0.5553	0.5587	-0.003

The separate regression equations indicated less variance explained in the Disrupted group than the Non-Disrupted group by about 2 or 3%. Although this difference is likely practically small, it is in the direction of concern. Further analyses may signal a consistent disadvantage in one of the Cohort C groups.

Table 19. Separate Regression Equations for Disrupted and Non-Disrupted Groups – Cohort C

	N	Disrupted R^2	Non-Disrupted R^2	ΔR^2
2013 Algebra	2,179	0.332	0.360	-0.029
2013 Algebra (7^{th} & 8^{th})	621	0.557	0.577	-0.021

Examining distributions of predicted student scores

The distributions of observed, predicted, and difference scores were examined for Cohort C focusing on the 7^{th} and 8^{th} grade students who were at the higher end of the ability distribution. The *F*-tests to compare the variances of the scores between the groups was again conducted. Those results may be found in Table 20.

The 7^{th} and 8^{th} grade group had a statistically significantly larger variance in difference scores for the Disrupted group than the Non-Disrupted group. This result foreshadows larger proportions of students below (and perhaps above) the percentile cuts.

Table 20. Distribution of the Difference between Predicted and Observed Scores for Non-disrupted and Disrupted Groups – Cohort C

	Disrupted						Non-Disrupted						
	N	Mean	SD	5th	10th	90th	95th	*N*	Mean	SD	5th	10th	90th	95th
Algebra														
Difference	2179	-3.77	41.22	-58.05	-44.34	37.99	54.30	2179	-0.03	39.51	-56.61	-42.93	42.08	54.70
Predicted Score	2179	743.32	29.03	706.32	717.52	775.22	785.03	2179	744.71	29.66	704.66	716.15	777.28	787.29
Observed Score	2179	739.55	50.17	670	687	793	815	2179	744.67	49.41	670	687	798	826
Algebra (7th & 8th)														
Difference	621	-6.02	**38.20**	-52.27	-39.94	32.08	51.54	621	0.01	31.74	-41.43	-32.17	32.84	44.01
Predicted Score	621	769.52	38.77	705.60	723.70	816.70	843.30	621	771.49	37.08	714.00	730.40	818.90	838.10
Observed Score	621	763.50	**55.45**	691	704	826	843	621	771.51	48.81	708	722	826	842

Note. Bolded values indicate significantly larger standard deviation for Disrupted group at $p < .05$.

Compare predictions for interrupted students to non-interrupted students

Table 21. Percent of Disrupted Students with Predicted and Observed Score Differences at the 5th, 10th, 90th and 95th Percentile of Non-disrupted Students – Cohort C

		5th	10th	90th	95th
	N	Larger than 5% or 10% = Disadvantaged		Larger than 5% or 10% = Advantaged	
2013 Algebra	2,179	5.42	10.65	8.08	4.91
2013 Algebra (7^{th} & 8^{th})	621	9.34	17.07	9.66	6.60

Table 21 contains the percentages of students in the Disrupted group above or below the percentile cuts of the Non-Disrupted Students. The 7^{th} and 8^{th} grade students who were interrupted while completing the Algebra I test in 2013 were proportionately more disadvantaged than advantaged. In other words, it appears that 7^{th} and 8^{th} grade students were differentially affected by interruption, on average, than others completing the same exam.

Cohort D Analyses

Another set of students completing an EOI test was examined. The English II test was chosen due to the available data in the EOI dataset. A measure of prior performance was desired for creating the matched sample. However, because students typically take the English II test in Grade 10, the nearest other assessment where most students also had scores was the US History EOI. Using the US History EOI was not a perfect solution, however. The content was obviously different, but highly correlated. The bigger issue was that not all English II EOI interrupted student also had a US History score.

After examining the students who were disrupted on the English II test who had US History test scores, another approach was developed to increase the sample size. More data was provided to HumRRO that enabled us to use 2011 reading test scores (from when the Grade 10 student were in Grade 8) as a prior performance indicator. This approach increased the sample size of non-disrupted students and provided an improved prediction model for the 2013 English II scores. Therefore, both groups are presented as Cohort D, much like Cohort C had two groups focused on the same exam.

For clarity, “2013 English Group 1” is all students who had US History exam scores as a “prior performance” indicator and “2013 English Group 2” is all students who had 2011 Reading scores as a predictor of performance. The prediction models differed primarily in these variables. Group 2 also included 2011 Math scores since the data was available for nearly every student.

Differences in average 2013 test scores

After propensity score matching, mean difference tests were performed on the matched samples. Disrupted students in Group 1 did better than matched Non-Disrupted students; however, this difference was neither statistically nor practically different. Group 2 mean difference was in the opposite direction, but this was also non-significant and small.

Table 22. Descriptive and Inferential Statistics of 2013 Test Scores After Matching – Cohort D

		Disrupted		Non-Disrupted			
	N	Mean	SD	Mean	SD	*t*	*d*
2013 English Group 1	427	748.4	74.80	746.1	73.47	-0.44	-0.03
2013 English Group 2	2,688	763.6	64.10	765.2	67.92	0.93	0.03

Note. No *p*-values below .05.

Examining the predictability of 2013 test scores

The regression equations were performed on both groups in Cohort D. R^2 values were lower for the Group 2 equations because the 2011 scores were not a recent test score, limiting the predictive power. However, for both groups, the "Disrupted" dichotomous variable does not additionally explain a practically large amount of English II test score variance.

Table 23. Combined Regression Equation Adding Disruption – Cohort D

	N	Covariates Only R^2	Covariates + Disrupted R^2	ΔR^2
2013 English Group 1	854	0.5187	0.5188	<0.001
2013 English Group 2	5,376	0.3503	0.3504	<0.001

The separate regression equations indicated *more* variance explained in the Disrupted group than the Non-Disrupted group in the Group 1 analyses, with a difference of about 4%. This is not a large difference and is of less concern that if Disrupted students had less variance explained. The difference in Group 2 equations is practically very small.

Table 24. Separate Regression Equations for Disrupted and Non-Disrupted Groups – Cohort D

	N	Disrupted R^2	Non-Disrupted R^2	ΔR^2
2013 English Group 1	427	0.545	0.503	0.042
2013 English Group 2	2,688	0.351	0.354	-0.003

Examining distributions of predicted student scores

The distributions of observed, predicted, and difference scores were examined for Cohort D and the *F*-tests to compare the variances of the scores between the groups were again conducted. Those results may be found in Table 25. Only one variable had larger variance for the Disrupted group at a significant level.

Table 25. Distribution of the Difference between Predicted and Observed Scores for Non-disrupted and Disrupted Groups – Cohort D

	Disrupted							Non-Disrupted					
	N	Mean	SD	5th	10th	90th	95th	*N*	Mean	SD	5th	10th	90th	95th
English Group 1														
Difference	427	1.26	51.41	-77.54	-54.29	57.77	74.86	427	-0.04	51.82	-87.26	-59.88	60.36	79.94
Predicted Score	427	747.09	52.14	653.61	683.34	814.11	834.58	427	746.15	52.09	655.99	680.26	817.72	834.51
Observed Score	427	748.36	74.80	615	656	838	859	427	746.11	73.47	622	656	838	864
English Group 2														
Difference	2688	-1.78	52.00	-80.89	-61.72	60.27	80.57	2688	-0.05	55.00	-82.57	-60.67	59.01	80.11
Predicted Score	2688	765.34	**43.00**	704.70	725.90	808.70	822.30	2688	765.29	40.00	709.20	724.60	807.50	821.90
Observed Score	2688	763.56	64.00	662	693	838	864	2688	765.24	68.00	654	687	840	859

Note. Bolded values indicate significantly larger standard deviation for Disrupted group at $p < .05$.

Compare predictions for interrupted students to non-interrupted students

Table 26. Percent of Disrupted Students with Predicted and Observed Score Differences at the 5th, 10th, 90th and 95th Percentile of Non-disrupted Students – Cohort D

		5th	10th	90th	95th
	N	Larger than 5% or 10% = Disadvantaged		Larger than 5% or 10% = Advantaged	
2013 English Group 1	427	3.75	8.20	9.84	3.75
2013 English Group 2	2,688	4.84	10.27	10.42	5.21

The percentages of students in the Disrupted group beyond the cutoff of the Non-Disrupted group are provided in Table 26. It seems that fewer students were both advantaged and disadvantaged in Group 1, but these differences do not appear large. The percentages for Group 2 are very close to the expected values and there appear to be no strong directional concerns.

Summary

This report provides a statistical investigation of computer disruptions on student test scores. A number of analyses were conducted on many groups of students completing different tests to investigate whether computer disruptions impacted scores. The only group that seemed to be disadvantaged was the 7th and 8th graders completing the Algebra I test, and the disadvantage was not consistent throughout the distribution of student scores. These differences rely on the predictability of the model, which is inherently imperfect. While there is some evidence to suggest that there were effects from the disruption, nothing emerged in a systematic way across grades, subjects, or methodologies that merit a statistical adjustment. While we cannot know for an individual student that the computer disruption did not impact his or her test score, we conclude that for the overall population of Oklahoma students the computer disruptions had little impact on student test scores.

References

Austin, P. C. (2009). Some methods of propensity-score matching had superior performance to others: Results of an empirical investigation and Monte Carlo simulation. *Biometrical Journal, 51*, 171-184. doi: 10.1002/bimj.200810488

Connelly, B. S., Sackett, P.R., & Waters, S. D. (2013). Balancing treatment and control groups in quasi-experiments: An Introduction to propensity scoring. *Personnel Psychology, 66*, 407-442. doi: 10.1111/peps.12020

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods, 15*(3), 234-249. doi: 10.1037/a0019623

Hill, R. (2013). An analysis of the Impact of Interruptions on the 2013 administration of the Indiana Statewide Testing for Educational Progress—Plus (ISTEP+). The National Center for the Improvement of Educational Assessment, Inc. Downloaded July 28, 2013 from http://www.doe.in.gov/sites/default/files/news/hill-report.pdf.

Appendix

Mean Differences between Non-Disrupted and Disrupted Samples before and after Propensity Matching

Table A1. Mean Covariate Differences for Grade 6 Reading – Cohort A

	Before Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	737.90	78.11	40872	734.47	82.55	2283	0.043
2012 Math Scale	742.12	99.21	38109	736.55	97.65	2145	0.057
2012 Reading Scale	732.91	95.92	38012	726.85	101.34	2136	0.062
School Free-Lunch	0.56	0.20	38896	0.58	0.19	2161	-0.088
School Achievement	738.61	26.44	38896	736.18	27.10	2161	0.091
African American	0.11	0.31	41444	0.15	0.35	2317	**-0.117**
American Indian	0.23	0.42	41444	0.21	0.41	2317	0.043
Asian	0.02	0.16	41444	0.03	0.17	2317	-0.034
Hispanic	0.14	0.34	41442	0.14	0.34	2317	0.005
Pacific Islander	0.01	0.07	41444	0.01	0.07	2317	0.000
Caucasian	0.70	0.46	41444	0.68	0.47	2317	0.044
Male	0.50	0.50	41428	0.50	0.50	2314	0.011
	After Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	737.68	75.80	2002	739.20	79.82	2002	-0.020
2012 Math Scale	740.15	93.20	2002	739.21	95.92	2002	0.010
2012 Reading Scale	731.14	87.70	2002	729.53	98.76	2002	0.017
School Free-Lunch	0.59	0.20	2002	0.58	0.19	2002	0.023
School Achievement	736.03	26.64	2002	736.34	26.95	2002	-0.012
African American	0.15	0.36	2002	0.15	0.36	2002	0.011
American Indian	0.21	0.41	2002	0.21	0.41	2002	-0.007
Asian	0.02	0.15	2002	0.03	0.17	2002	-0.042
Hispanic	0.13	0.34	2002	0.14	0.34	2002	-0.016
Pacific Islander	0.01	0.08	2002	0.01	0.07	2002	0.013
Caucasian	0.68	0.47	2002	0.68	0.47	2002	-0.014
Male	0.48	0.50	2002	0.49	0.50	2002	-0.015

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A2. Mean Covariate Differences for Grade 6 Math – Cohort A

	Before Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	731.49	77.29	41178	732.79	82.16	1977	-0.016
2012 Math Scale	741.98	98.90	38635	736.42	94.67	1857	0.057
2012 Reading Scale	732.71	96.48	38153	731.60	90.90	1833	0.012
School Free-Lunch	0.57	0.20	39992	0.57	0.17	1896	-0.044
School Achievement	737.76	26.56	39992	738.05	24.20	1896	-0.011
African American	0.11	0.31	41965	0.16	0.36	2011	**-0.152**
American Indian	0.22	0.42	41965	0.22	0.41	2011	0.018
Asian	0.02	0.16	41965	0.04	0.19	2011	-0.067
Hispanic	0.14	0.35	41964	0.14	0.35	2011	-0.014
Pacific Islander	0.01	0.07	41965	0.00	0.07	2011	0.013
Caucasian	0.70	0.46	41965	0.66	0.47	2011	0.083
Male	0.50	0.50	41864	0.48	0.50	2011	0.050
	After Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	739.53	75.52	1710	738.15	79.91	1710	0.018
2012 Math Scale	742.62	92.83	1710	739.51	94.26	1710	0.033
2012 Reading Scale	736.69	89.18	1710	733.03	91.75	1710	0.041
School Free-Lunch	0.57	0.20	1710	0.57	0.17	1710	-0.002
School Achievement	738.20	25.85	1710	738.25	23.42	1710	-0.002
African American	0.16	0.37	1710	0.16	0.37	1710	0.000
American Indian	0.22	0.41	1710	0.21	0.41	1710	0.017
Asian	0.04	0.19	1710	0.04	0.19	1710	0.000
Hispanic	0.15	0.36	1710	0.15	0.35	1710	0.020
Pacific Islander	0.00	0.03	1710	0.00	0.06	1710	-0.060
Caucasian	0.66	0.47	1710	0.66	0.47	1710	-0.006
Male	0.49	0.50	1710	0.48	0.50	1710	0.033

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A3. Mean Covariate Differences for Grade 7 Reading – Cohort A

Before Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	733.18	80.19	41134	723.82	85.05	1847	**0.113**
2012 Math Scale	733.69	95.85	38422	723.88	102.50	1732	0.099
2012 Reading Scale	729.60	96.57	38486	719.00	102.50	1731	**0.106**
School Free-Lunch	0.55	0.19	39075	0.56	0.18	1765	-0.043
School Achievement	732.65	23.81	39056	728.23	25.35	1764	**0.180**
African American	0.11	0.31	41815	0.16	0.36	1887	**-0.129**
American Indian	0.23	0.42	41815	0.21	0.41	1887	0.030
Asian	0.02	0.15	41815	0.02	0.15	1887	0.000
Hispanic	0.13	0.34	41813	0.12	0.32	1887	0.046
Pacific Islander	0.00	0.06	41815	0.01	0.08	1887	-0.029
Caucasian	0.69	0.46	41815	0.67	0.47	1887	0.048
Male	0.50	0.50	41799	0.50	0.50	1886	0.008
After Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	724.80	77.89	1625	728.01	82.05	1625	-0.040
2012 Math Scale	721.00	102.66	1625	724.99	102.34	1625	-0.039
2012 Reading Scale	716.65	109.24	1625	719.01	103.90	1625	-0.022
School Free-Lunch	0.56	0.19	1625	0.55	0.18	1625	0.038
School Achievement	728.14	24.37	1625	728.30	25.57	1625	-0.006
African American	0.16	0.37	1625	0.15	0.36	1625	0.024
American Indian	0.22	0.41	1625	0.21	0.41	1625	0.019
Asian	0.03	0.17	1625	0.02	0.15	1625	0.027
Hispanic	0.12	0.32	1625	0.12	0.32	1625	-0.004
Pacific Islander	0.01	0.07	1625	0.01	0.08	1625	-0.016
Caucasian	0.66	0.47	1625	0.67	0.47	1625	-0.030
Male	0.50	0.50	1625	0.49	0.50	1625	0.033

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A4. Mean Covariate Differences for Grade 7 Math – Cohort A

	Before Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	730.70	67.27	41702	721.07	74.39	1276	**0.136**
2012 Math Scale	733.52	95.92	39121	721.55	103.24	1191	**0.120**
2012 Reading Scale	729.45	97.02	38842	720.16	99.00	1169	0.095
School Free-Lunch	0.55	0.19	40363	0.60	0.16	1227	**-0.290**
School Achievement	731.88	23.93	40343	725.46	25.27	1227	**0.261**
African American	0.11	0.31	42421	0.16	0.36	1307	**-0.136**
American Indian	0.23	0.42	42421	0.23	0.42	1307	-0.003
Asian	0.02	0.15	42421	0.02	0.13	1307	0.042
Hispanic	0.13	0.34	42421	0.17	0.37	1307	-0.099
Pacific Islander	0.00	0.07	42421	0.01	0.07	1307	-0.016
Caucasian	0.69	0.46	42421	0.66	0.47	1307	0.060
Male	0.50	0.50	42408	0.51	0.50	1306	-0.018
	After Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	729.50	68.49	1085	727.81	71.04	1085	0.024
2012 Math Scale	727.52	92.31	1085	724.73	100.98	1085	0.029
2012 Reading Scale	725.59	97.65	1085	722.12	96.18	1085	0.036
School Free-Lunch	0.60	0.18	1085	0.61	0.16	1085	-0.054
School Achievement	724.77	24.25	1085	725.15	25.27	1085	-0.015
African American	0.18	0.38	1085	0.17	0.37	1085	0.022
American Indian	0.20	0.40	1085	0.22	0.41	1085	-0.043
Asian	0.02	0.15	1085	0.02	0.14	1085	0.032
Hispanic	0.20	0.40	1085	0.18	0.38	1085	0.075
Pacific Islander	0.00	0.06	1085	0.00	0.07	1085	-0.014
Caucasian	0.66	0.47	1085	0.65	0.48	1085	0.021
Male	0.49	0.50	1085	0.50	0.50	1085	-0.011

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A5. Mean Covariate Differences for Grade 8 Reading – Cohort A

Before Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	731.00	83.07	39830	732.46	87.45	2189	-0.017
2012 Math Scale	734.49	97.70	37151	732.62	94.76	2041	0.019
2012 Reading Scale	737.40	91.04	37350	733.56	95.21	2048	0.041
School Free-Lunch	0.52	0.20	37910	0.54	0.18	2070	**-0.102**
School Achievement	736.91	24.41	37910	734.38	23.52	2070	**0.106**
African American	0.11	0.31	40681	0.15	0.35	2229	**-0.105**
American Indian	0.22	0.42	40681	0.21	0.41	2229	0.034
Asian	0.02	0.15	40681	0.02	0.15	2229	0.006
Hispanic	0.13	0.33	40679	0.14	0.34	2228	-0.037
Pacific Islander	0.00	0.07	40681	0.01	0.08	2229	-0.027
Caucasian	0.70	0.46	40681	0.68	0.47	2229	0.041
Male	0.50	0.50	40654	0.50	0.50	2228	-0.001
After Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	738.54	83.13	1900	737.43	86.79	1900	0.013
2012 Math Scale	734.74	106.92	1900	733.75	94.06	1900	0.010
2012 Reading Scale	733.97	103.24	1900	735.68	93.53	1900	-0.017
School Free-Lunch	0.54	0.20	1900	0.54	0.18	1900	0.010
School Achievement	734.67	24.33	1900	734.68	23.68	1900	0.000
African American	0.14	0.35	1900	0.15	0.36	1900	-0.025
American Indian	0.21	0.41	1900	0.21	0.41	1900	0.013
Asian	0.03	0.16	1900	0.02	0.15	1900	0.014
Hispanic	0.15	0.36	1900	0.14	0.34	1900	0.040
Pacific Islander	0.01	0.08	1900	0.01	0.08	1900	-0.007
Caucasian	0.68	0.47	1900	0.67	0.47	1900	0.002
Male	0.49	0.50	1900	0.49	0.50	1900	-0.009

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A6. Mean Covariate Differences for Grade 8 Math – Cohort A

	Before Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	750.65	82.08	40262	749.63	79.88	1760	0.013
2012 Math Scale	734.59	97.70	37628	727.36	96.68	1644	0.074
2012 Reading Scale	737.67	91.72	37406	731.69	84.76	1632	0.068
School Free-Lunch	0.53	0.20	38969	0.58	0.19	1694	**-0.280**
School Achievement	736.40	24.65	38969	727.38	22.13	1694	**0.386**
African American	0.11	0.31	40939	0.14	0.34	1786	-0.083
American Indian	0.22	0.42	40939	0.19	0.40	1786	0.068
Asian	0.02	0.15	40939	0.01	0.12	1786	0.071
Hispanic	0.13	0.33	40939	0.17	0.38	1786	**-0.132**
Pacific Islander	0.00	0.07	40939	0.01	0.08	1786	-0.018
Caucasian	0.69	0.46	40939	0.70	0.46	1786	-0.004
Male	0.50	0.50	40929	0.49	0.50	1784	0.013
	After Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	756.06	83.99	1533	755.45	78.36	1533	0.008
2012 Math Scale	728.61	109.66	1533	728.40	94.36	1533	0.002
2012 Reading Scale	733.60	104.04	1533	733.15	83.63	1533	0.005
School Free-Lunch	0.58	0.19	1533	0.58	0.19	1533	0.022
School Achievement	727.34	23.99	1533	727.22	22.16	1533	0.005
African American	0.15	0.35	1533	0.14	0.35	1533	0.017
American Indian	0.20	0.40	1533	0.19	0.40	1533	0.008
Asian	0.02	0.12	1533	0.02	0.12	1533	-0.005
Hispanic	0.17	0.38	1533	0.18	0.38	1533	-0.014
Pacific Islander	0.01	0.07	1533	0.01	0.07	1533	0.000
Caucasian	0.69	0.46	1533	0.69	0.46	1533	-0.016
Male	0.49	0.50	1533	0.48	0.50	1533	0.014

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A7. Mean Covariate Differences for Grade 6 Reading – Cohort B

Before Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	737.52	78.25	42606	752.91	84.37	549	**-0.189**
2012 Math Scale	741.58	99.22	39749	761.13	89.66	505	**-0.207**
2012 Reading Scale	732.38	96.37	39644	749.29	83.11	504	**-0.188**
School Free-Lunch	0.56	0.20	40538	0.48	0.22	519	**0.419**
School Achievement	738.31	26.38	40538	751.51	30.66	519	**-0.463**
African American	0.11	0.31	43201	0.13	0.34	560	-0.062
American Indian	0.23	0.42	43201	0.17	0.38	560	**0.140**
Asian	0.02	0.16	43201	0.05	0.21	560	**-0.119**
Hispanic	0.14	0.34	43199	0.13	0.34	560	0.021
Pacific Islander	0.01	0.07	43201	0.00	0.06	560	0.025
Caucasian	0.69	0.46	43201	0.70	0.46	560	-0.013
Male	0.50	0.50	43185	0.49	0.50	557	0.023
After Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	757.31	82.91	468	758.79	79.65	468	-0.018
2012 Math Scale	762.20	100.46	468	766.50	83.36	468	-0.047
2012 Reading Scale	746.76	105.49	468	753.35	75.84	468	-0.073
School Free-Lunch	0.48	0.22	468	0.47	0.22	468	0.011
School Achievement	751.82	28.10	468	752.24	30.34	468	-0.014
African American	0.12	0.33	468	0.12	0.33	468	0.000
American Indian	0.17	0.38	468	0.17	0.38	468	0.006
Asian	0.07	0.26	468	0.05	0.22	468	0.081
Hispanic	0.15	0.36	468	0.13	0.34	468	0.067
Pacific Islander	0.00	0.07	468	0.00	0.07	468	0.000
Caucasian	0.71	0.46	468	0.72	0.45	468	-0.028
Male	0.46	0.50	468	0.49	0.50	468	-0.051

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A8. Mean Covariate Differences for Grade 6 Math – Cohort B

	Before Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	731.54	77.39	42443	732.05	84.63	712	-0.006
2012 Math Scale	741.81	98.78	39827	736.71	94.85	665	0.053
2012 Reading Scale	732.67	96.39	39329	731.86	86.03	657	0.009
School Free-Lunch	0.57	0.20	41208	0.57	0.17	680	-0.023
School Achievement	737.74	26.49	41208	739.33	24.83	680	-0.062
African American	0.11	0.31	43257	0.19	0.40	719	**-0.246**
American Indian	0.22	0.42	43257	0.21	0.41	719	0.030
Asian	0.03	0.16	43257	0.04	0.19	719	-0.064
Hispanic	0.14	0.35	43256	0.15	0.35	719	-0.021
Pacific Islander	0.01	0.07	43257	0.00	0.05	719	0.041
Caucasian	0.70	0.46	43257	0.63	0.48	719	**0.143**
Male	0.50	0.50	43156	0.48	0.50	719	0.046
	After Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	742.25	71.63	617	740.30	81.34	617	0.026
2012 Math Scale	745.58	98.78	617	739.02	96.46	617	0.067
2012 Reading Scale	736.91	99.64	617	733.21	85.08	617	0.040
School Free-Lunch	0.56	0.20	617	0.57	0.17	617	-0.046
School Achievement	740.61	24.75	617	739.55	24.79	617	0.043
African American	0.21	0.41	617	0.20	0.40	617	0.016
American Indian	0.19	0.40	617	0.21	0.41	617	-0.040
Asian	0.04	0.20	617	0.04	0.19	617	0.008
Hispanic	0.15	0.36	617	0.15	0.36	617	-0.014
Pacific Islander	0.01	0.08	617	0.00	0.06	617	0.047
Caucasian	0.64	0.48	617	0.63	0.48	617	0.034
Male	0.49	0.50	617	0.47	0.50	617	0.023

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A9. Mean Covariate Differences for Grade 7 Reading – Cohort B

Before Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	732.88	80.36	42557	722.22	86.01	424	**0.128**
2012 Math Scale	733.40	96.02	39757	720.21	109.55	397	**0.128**
2012 Reading Scale	729.22	96.80	39822	721.26	102.03	395	0.080
School Free-Lunch	0.55	0.19	40436	0.52	0.18	404	**0.150**
School Achievement	732.46	23.85	40416	732.63	28.40	404	-0.007
African American	0.11	0.32	43268	0.10	0.30	434	0.055
American Indian	0.23	0.42	43268	0.21	0.40	434	0.051
Asian	0.02	0.15	43268	0.02	0.13	434	0.052
Hispanic	0.13	0.34	43266	0.12	0.33	434	0.026
Pacific Islander	0.00	0.07	43268	0.01	0.08	434	-0.035
Caucasian	0.69	0.46	43268	0.76	0.43	434	**-0.156**
Male	0.50	0.50	43252	0.50	0.50	433	-0.007
After Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	721.38	84.57	369	728.33	82.75	369	-0.083
2012 Math Scale	721.03	100.50	369	722.94	110.59	369	-0.018
2012 Reading Scale	721.72	110.46	369	721.48	102.69	369	0.002
School Free-Lunch	0.51	0.20	369	0.52	0.18	369	-0.031
School Achievement	732.72	23.51	369	732.57	28.85	369	0.006
African American	0.10	0.30	369	0.09	0.29	369	0.028
American Indian	0.19	0.39	369	0.21	0.40	369	-0.041
Asian	0.01	0.09	369	0.02	0.13	369	-0.075
Hispanic	0.14	0.35	369	0.13	0.34	369	0.039
Pacific Islander	0.00	0.05	369	0.01	0.09	369	-0.076
Caucasian	0.77	0.42	369	0.78	0.42	369	-0.019
Male	0.48	0.50	369	0.50	0.50	369	-0.049

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A10. Mean Covariate Differences for Grade 7 Math – Cohort B

	Before Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	730.47	67.39	42436	725.55	75.69	542	0.069
2012 Math Scale	733.33	95.94	39806	720.66	111.47	506	**0.122**
2012 Reading Scale	729.28	96.99	39510	720.96	104.39	501	0.083
School Free-Lunch	0.55	0.19	41088	0.63	0.13	502	**-0.517**
School Achievement	731.83	23.92	41068	720.38	27.21	502	**0.448**
African American	0.11	0.32	43178	0.15	0.36	550	**-0.114**
American Indian	0.23	0.42	43178	0.21	0.40	550	0.048
Asian	0.02	0.15	43178	0.03	0.17	550	-0.037
Hispanic	0.13	0.34	43178	0.23	0.42	550	**-0.271**
Pacific Islander	0.00	0.07	43178	0.00	0.04	550	0.046
Caucasian	0.69	0.46	43178	0.69	0.46	550	0.007
Male	0.50	0.50	43165	0.50	0.50	549	0.004
	After Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	727.78	66.09	455	731.12	73.73	455	-0.048
2012 Math Scale	719.35	113.09	455	724.81	105.35	455	-0.050
2012 Reading Scale	720.98	100.63	455	723.56	96.83	455	-0.026
School Free-Lunch	0.62	0.17	455	0.64	0.13	455	-0.094
School Achievement	721.78	25.97	455	720.23	26.76	455	0.059
African American	0.14	0.35	455	0.17	0.37	455	-0.079
American Indian	0.19	0.39	455	0.19	0.40	455	-0.017
Asian	0.04	0.20	455	0.03	0.17	455	0.048
Hispanic	0.33	0.47	455	0.26	0.44	455	**0.160**
Pacific Islander	0.00	0.00	455	0.00	0.05	455	-0.094
Caucasian	0.69	0.46	455	0.67	0.47	455	0.042
Male	0.48	0.50	455	0.49	0.50	455	-0.018

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A11. Mean Covariate Differences for Grade 8 Reading – Cohort B

Before Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	731.06	83.17	41404	732.58	91.92	615	-0.017
2012 Math Scale	734.45	97.53	38615	730.06	98.97	577	0.045
2012 Reading Scale	737.27	91.09	38817	732.36	102.29	581	0.051
School Free-Lunch	0.52	0.20	39393	0.55	0.16	587	**-0.165**
School Achievement	736.76	24.34	39393	738.05	26.04	587	-0.051
African American	0.11	0.31	42280	0.17	0.38	630	**-0.178**
American Indian	0.22	0.42	42280	0.19	0.39	630	0.089
Asian	0.02	0.15	42280	0.03	0.18	630	-0.056
Hispanic	0.13	0.33	42277	0.16	0.37	630	**-0.104**
Pacific Islander	0.00	0.07	42280	0.01	0.11	630	-0.089
Caucasian	0.69	0.46	42280	0.66	0.47	630	0.077
Male	0.50	0.50	42253	0.47	0.50	629	0.058
After Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Math Scale	745.03	84.42	543	737.56	92.54	543	0.084
2012 Math Scale	740.94	99.16	543	731.78	99.41	543	0.092
2012 Reading Scale	739.54	100.90	543	734.72	102.88	543	0.047
School Free-Lunch	0.55	0.20	543	0.55	0.16	543	-0.012
School Achievement	740.80	27.80	543	738.86	26.29	543	0.072
African American	0.16	0.37	543	0.17	0.38	543	-0.025
American Indian	0.21	0.41	543	0.19	0.39	543	0.046
Asian	0.02	0.14	543	0.03	0.18	543	-0.081
Hispanic	0.14	0.35	543	0.16	0.37	543	-0.067
Pacific Islander	0.01	0.10	543	0.01	0.10	543	0.000
Caucasian	0.67	0.47	543	0.66	0.47	543	0.027
Male	0.49	0.50	543	0.47	0.50	543	0.048

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A12. Mean Covariate Differences for Grade 8 Math – Cohort B

	Before Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	750.75	82.06	40960	745.13	78.98	1062	0.070
2012 Math Scale	734.53	97.66	38272	725.10	97.69	1000	0.097
2012 Reading Scale	737.56	91.71	38044	732.04	80.73	994	0.064
School Free-Lunch	0.53	0.20	39641	0.60	0.20	1022	**-0.349**
School Achievement	736.32	24.58	39641	724.72	23.27	1022	**0.485**
African American	0.11	0.31	41647	0.16	0.37	1078	**-0.142**
American Indian	0.22	0.42	41647	0.20	0.40	1078	0.059
Asian	0.02	0.15	41647	0.01	0.12	1078	0.062
Hispanic	0.13	0.33	41647	0.19	0.39	1078	**-0.171**
Pacific Islander	0.00	0.07	41647	0.01	0.09	1078	-0.033
Caucasian	0.69	0.46	41647	0.67	0.47	1078	0.060
Male	0.50	0.50	41637	0.50	0.50	1076	-0.005
	After Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
2013 Reading Scale	750.33	86.83	932	749.99	77.36	932	0.004
2012 Math Scale	724.17	107.83	932	725.62	95.75	932	-0.014
2012 Reading Scale	729.19	95.85	932	733.32	80.97	932	-0.047
School Free-Lunch	0.60	0.19	932	0.60	0.20	932	0.034
School Achievement	724.72	25.69	932	724.47	23.17	932	0.010
African American	0.17	0.38	932	0.17	0.37	932	0.011
American Indian	0.16	0.37	932	0.19	0.40	932	-0.084
Asian	0.02	0.14	932	0.02	0.13	932	0.032
Hispanic	0.20	0.40	932	0.20	0.40	932	0.011
Pacific Islander	0.00	0.06	932	0.01	0.08	932	-0.047
Caucasian	0.68	0.47	932	0.66	0.47	932	0.043
Male	0.47	0.50	932	0.49	0.50	932	-0.043

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A13. Mean Covariate Differences for 2013 Algebra – Cohort C

Before Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
Predicted Disruption	0.07	0.01	30721	0.07	0.02	2179	**-0.218**
Algebra Scale	746.45	50.08	31021	739.28	50.20	2196	**0.143**
2012 Math Scale	745.28	91.82	31021	736.73	98.00	2196	0.090
2012 Reading Scale	761.30	87.63	30804	755.68	89.18	2181	0.064
School Free-Lunch	0.47	0.13	30941	0.46	0.14	2195	0.033
School Achievement	752.89	29.75	30941	750.51	30.65	2195	0.079
African American	0.10	0.30	31021	0.13	0.33	2196	-0.081
American Indian	0.20	0.40	31021	0.22	0.41	2196	-0.049
Asian	0.03	0.16	31021	0.02	0.15	2196	0.023
Hispanic	0.00	0.00	31021	0.00	0.00	2196	0.000
Pacific Islander	0.01	0.08	31021	0.00	0.06	2196	0.047
Caucasian	0.68	0.46	31021	0.64	0.48	2196	0.086
Male	0.49	0.50	31018	0.47	0.50	2195	0.048
After Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
Predicted Disruption	0.07	0.02	2179	0.07	0.02	2179	-0.005
Algebra Scale	744.67	49.41	2179	739.55	50.17	2179	0.103
2012 Math Scale	741.35	95.62	2179	737.50	96.85	2179	0.040
2012 Reading Scale	757.07	94.64	2179	755.75	89.15	2179	0.014
School Free-Lunch	0.45	0.14	2179	0.46	0.14	2179	-0.067
School Achievement	752.48	32.20	2179	750.52	30.73	2179	0.062
African American	0.13	0.34	2179	0.13	0.33	2179	0.012
American Indian	0.22	0.42	2179	0.22	0.41	2179	0.009
Asian	0.03	0.17	2179	0.02	0.15	2179	0.037
Hispanic	0.00	0.00	2179	0.00	0.00	2179	0.000
Pacific Islander	0.00	0.06	2179	0.00	0.06	2179	0.000
Caucasian	0.64	0.48	2179	0.64	0.48	2179	-0.006
Male	0.49	0.50	2179	0.47	0.50	2179	0.035

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A14. Mean Covariate Differences for 2013 Algebra (7th & 8th) – Cohort C

	Before Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
Predicted Disruption	0.07	0.03	8632	0.08	0.03	621	**-0.345**
Algebra Scale	775.49	47.84	9077	763.11	55.51	648	**0.240**
2012 Math Scale	810.83	65.01	8738	799.77	71.70	622	**0.162**
2013 Math Scale	802.69	69.26	9077	791.65	75.09	648	**0.153**
2013 Read Scale	806.86	67.91	9050	799.37	73.35	648	**0.106**
African American	0.09	0.29	9077	0.10	0.30	648	-0.023
American Indian	0.17	0.37	9077	0.19	0.40	648	-0.074
Asian	0.05	0.23	9077	0.04	0.21	648	0.042
Caucasian	0.75	0.43	9077	0.71	0.45	648	0.091
Hispanic	0.10	0.29	9077	0.07	0.26	648	0.082
Pacific Islander	0.00	0.07	9077	0.00	0.04	648	0.059
Free Lunch	0.33	0.47	9077	0.41	0.49	648	**-0.152**
Grade	7.89	0.32	9077	7.91	0.29	661	-0.084
School FRL Proportion	0.33	0.15	8991	0.35	0.16	647	-0.098
School Achievement	804.81	17.56	8991	799.35	22.83	647	**0.270**
Male	0.49	0.50	9076	0.46	0.50	648	0.057
	After Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
Predicted Disruption	0.08	0.03	621	0.08	0.03	621	0.001
Algebra Scale	771.51	48.81	621	763.50	55.45	621	**0.154**
2012 Math Scale	798.61	89.06	621	800.07	71.36	621	-0.018
2013 Math Scale	795.61	71.39	621	792.02	74.68	621	0.049
2013 Read Scale	804.61	70.84	621	800.23	73.21	621	0.061
African American	0.08	0.27	621	0.10	0.29	621	-0.057
American Indian	0.20	0.40	621	0.20	0.40	621	0.012
Asian	0.05	0.22	621	0.05	0.21	621	0.015
Caucasian	0.71	0.45	621	0.71	0.45	621	-0.007
Hispanic	0.07	0.25	621	0.07	0.26	621	-0.025
Pacific Islander	0.00	0.04	621	0.00	0.04	621	0.000
Free Lunch	0.41	0.49	621	0.40	0.49	621	0.007
Grade	7.92	0.27	621	7.91	0.28	621	0.023
School FRL Proportion	0.33	0.15	621	0.35	0.15	621	-0.099
School Achievement	801.66	20.94	621	799.72	22.11	621	0.090
Male	0.45	0.50	621	0.46	0.50	621	-0.019

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A15. Mean Covariate Differences for 2013 English Group 1 – Cohort D

Before Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
Predicted Disruption	0.05	0.02	7991	0.06	0.03	427	**-0.377**
2013 English	761.18	71.68	35702	756.61	69.93	3113	0.065
2013 History Scale	706.90	98.52	7995	693.56	96.85	431	**0.137**
School Free-Lunch	0.46	0.09	35700	0.46	0.07	3112	0.060
African American	0.11	0.31	35702	0.10	0.30	3113	0.027
American Indian	0.20	0.40	35702	0.17	0.38	3113	0.078
Asian	0.03	0.17	35702	0.02	0.15	3113	0.044
Hispanic	0.00	0.00	35702	0.00	0.00	3113	0.000
Pacific Islander	0.01	0.09	35702	0.01	0.08	3113	0.037
Caucasian	0.67	0.47	35702	0.54	0.50	3113	**0.280**
Male	0.49	0.50	35696	0.49	0.50	3108	-0.007
After Matching							
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
Predicted Disruption	0.06	0.03	427	0.06	0.03	427	-0.025
2013 English	746.11	73.47	427	748.36	74.80	427	-0.030
2013 History Scale	692.31	95.73	427	694.80	96.44	427	-0.026
School Free-Lunch	0.46	0.09	427	0.45	0.06	427	0.087
African American	0.13	0.33	427	0.11	0.31	427	0.058
American Indian	0.21	0.41	427	0.23	0.42	427	-0.045
Asian	0.03	0.17	427	0.02	0.14	427	0.076
Hispanic	0.00	0.00	427	0.00	0.00	427	0.000
Pacific Islander	0.02	0.14	427	0.01	0.12	427	0.054
Caucasian	0.56	0.50	427	0.57	0.50	427	-0.009
Male	0.48	0.50	427	0.49	0.50	427	-0.033

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.

Table A16. Mean Covariate Differences for 2013 English Group 2 – Cohort D

	Before Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
Predicted Disruption	0.08	0.04	30130	0.12	0.10	2688	**-0.595**
2013 English Scale	761.18	71.68	35702	756.61	69.93	3113	0.065
2011 Reading Scale	704.12	171.71	30134	709.98	167.72	2692	-0.035
2011 Math Scale	721.04	108.29	30134	724.22	104.05	2692	-0.030
African American	0.11	0.31	35702	0.10	0.30	3113	0.027
American Indian	0.20	0.40	35702	0.17	0.38	3113	0.078
Asian	0.03	0.17	35702	0.02	0.15	3113	0.044
Caucasian	0.67	0.47	35702	0.54	0.50	3113	**0.280**
Hispanic	0.00	0.00	35702	0.00	0.00	3113	0.000
Pacific Islander	0.01	0.09	35702	0.01	0.08	3113	0.037
Free Lunch	0.47	0.50	35702	0.38	0.49	3113	**0.176**
Male	0.49	0.50	35696	0.49	0.50	3108	-0.007
	After Matching						
	Non-Disrupted			Disrupted			
	Mean	SD	*N*	Mean	SD	*N*	Cohen's *d*
Predicted Disruption	0.12	0.10	2688	0.12	0.10	2688	-0.001
2013 English Scale	765.24	67.92	2688	763.56	64.10	2688	0.025
2011 Reading Scale	708.85	165.13	2688	710.03	167.59	2688	-0.007
2011 Math Scale	726.26	97.85	2688	724.43	103.74	2688	0.018
African American	0.10	0.30	2688	0.09	0.29	2688	0.029
American Indian	0.18	0.39	2688	0.18	0.38	2688	0.004
Asian	0.02	0.14	2688	0.02	0.13	2688	0.014
Caucasian	0.53	0.50	2688	0.55	0.50	2688	-0.028
Hispanic	0.00	0.00	2688	0.00	0.00	2688	0.000
Pacific Islander	0.01	0.08	2688	0.00	0.07	2688	0.025
Free Lunch	0.39	0.49	2688	0.37	0.48	2688	0.044
Male	0.50	0.50	2688	0.49	0.50	2688	0.021

Note. Bolded values indicate Cohen's *d* greater than .10 or -.10.