

Oklahoma Modified Alternate Assessment Program

Grades 3–8 Assessments 2013 Technical Report

Submitted to The Oklahoma State Department of Education November 2013



Copyright

Developed and published under contract with Oklahoma State Department of Education by CTB/McGraw-Hill LLC, 20 Ryan Ranch Road, Monterey, California 93940-5703. Copyright © 2013 by Oklahoma State Department of Education. Portions copyright © CTB/McGraw-Hill LLC. All rights reserved. Only Oklahoma State educators and citizens may copy and/or download and print the document, located online at http://ok.gov/sde/accountability-state-testing-results. Any other use or reproduction of this document, in whole or in part, requires written approval of Oklahoma State Department of Education and the publisher.

Introduction

The Oklahoma School Testing Program (OSTP) is a statewide assessment program that was established to improve academic achievement for all Oklahoma students. It also meets the requirements of the No Child Left Behind Act (NCLB) introduced by the federal government in 2001. The OSTP includes grades 3–8 and high school End-of-Instruction (EOI) assessments, for which students who complete an area of instruction must also take the corresponding statewide standardized assessment.

The OSTP developed three types of tests to assess the three groups of students defined by NCLB: the Oklahoma Core Curriculum Tests (OCCT) for the general student population, the Oklahoma Modified Alternate Assessment Program (OMAAP) for students instructed on grade level but whose IEP designated the OMAAP as the appropriate assessment, and the Oklahoma Alternate Assessment Program (OAAP) for students with the most significant cognitive disabilities. All three tests cover students in grades 3–8 and high school.

The Oklahoma College, Career, and Citizen Ready Standards (OK C^3) academic content standards are the foundation for all three tests. The Curriculum Access Resource Guides (CARG) describe access points to the OK C^3 through scaffolding of skills. An alternate guide, the CARG-A, provides guidance for instruction and assessment of Oklahoma students with the most significant cognitive disabilities.

The Oklahoma Modified Alternate Assessment Program (OMAAP) Grades 3–8 tests are used to assess student proficiency in Reading, Mathematics, and Science (grades 5 and 8). The OMAAP is intended for a population of students for whom the general OCCT exams and the Oklahoma Alternate Assessment Program (OAAP) portfolio assessments are inappropriate. The OMAAP Grades 3–8 tests are based on modified blueprints and items from the corresponding OCCT Grades 3–8 tests.

In 2013, CTB/McGraw-Hill was contracted by the Oklahoma State Department of Education (SDE) to develop, administer, and maintain the OMAAP Grades 3–8 administration. This technical report focuses on the technical details of work accomplished for the spring test only.

Purpose

The purpose of this Technical Report is to provide objective information regarding technical aspects of the OMAAP Grades 3–8 assessments. This volume is intended to be one source of information to Oklahoma K–12 educational stakeholders (including test coordinators, educators, parents, and other interested citizens) about the development, implementation, scoring, and technical attributes of the OMAAP Grades 3–8 assessments. Other sources of information regarding the OMAAP Grades 3–8 tests include the administration manuals, interpretation manuals, student, teacher, and parent guides, implementation materials, and training materials.

The information provided here fulfills legal, professional, and scientific guidelines (AERA, APA, & NCME, 1999) for technical reports of large-scale educational assessments and is intended for use by qualified users within schools who use the OMAAP assessments and

interpret the results. Specifically, information was selected for inclusion in this report based on NCLB requirements and the following Standards for Educational and Psychological Testing:

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

This technical report documents the OMAAP Grades 3–8 test development methods, data analysis, and results for use by qualified users and technical experts. Section 1 provides an overview of the test design, test content, and content standards. Section 2 provides summary information about the test administration. Section 3 details the classical item analyses and reliability results, and Section 4 details the calibration, equating, scaling analyses, and results. Section 5 provides the results of the classification accuracy and classifications studies. Section 6 provides higher-level summaries of all the tests included in the OMAAP Grades 3–8 testing program.

Table of Contents

Copyright	i
Introduction	ii
Purpose	ii
Table of Contents	iv
List of Tables and Figures	v
Acronyms and Abbreviations	vi
Section 1	1
1.1 Overview of the OMAAP 3–8 Assessments	1
1.2 Content Standards	1
1.3 Blueprint	1
1.4 Universal Design and Modifications	3
1.5 Test Development and Content Validity	4
1.5.a. Item Development and Selection	4
1.5.b. Configuration of the Tests	5
Section 2	. 10
2.1 Packaging and Shipping	. 10
2.2 Materials Return	. 10
2.3 Materials Discrepancies Process	. 11
Section 3	. 12
3.1 Data Quality Check and Clean-Up	. 12
3.2 Classical Item Analyses	. 15
3.2.a. Test-Level Summaries of Classical Item Analyses	. 15
3.3 Procedures for Detecting Item Bias	. 16
3.3.a. Differential Item Functioning Results	. 16
3.4 Test Reliability	. 18
3.4.a. Overall Test Reliability	. 18
3.4.b. Test Reliability by Subgroup	. 19
Section 4	. 20
4.1 Item Response Theory (IRT) Models	. 20
4.2 Calibration and Equating	. 20
4.3 Scaling and Scoring Results	. 20
Section 5	. 27
5.1 Classification Consistency and Accuracy	. 27
Section 6	. 31
6.1 Descriptive Statistics	. 31
6.2 Performance Level Distribution	. 34
6.3 Conditional Standard Error of Measurement	. 35
6.4 Standard Error of Measurement	. 36
References	. 37
Appendix A	. 38
Appendix B: Scale Score Distributions	. 53

List of Tables and Figures

Table 1.1. Oklahoma OK C ³ Content Standards by Subject	2
Table 1.2. OCCT and OMAAP Grades 3–8 Item Count Comparison	3
Table 1.3. OMAAP Item Modification Rules	3
Table 1.4. Criteria for Aligning the Test with the $OK C^3$ Standards and Objectives	5
Table 1.5.a. Percentage of Items by DOK Level in Operational/Braille Forms	6
Table 1.5.b. Percentage of Items by DOK Level of Equivalent Forms	6
Table 1.6.a. Mathematics Blueprint (Number of Items by Standard)	7
Table 1.6.b. Number of Items of Operational/Braille Forms by Content Standard for	
Mathematics	7
Table 1.6.c. Number of Items of Equivalent Form by Content Standard for Mathematics	7
Table 1.6.d. Reading Blueprint (Number of Items by Standard)	7
Table 1.6.e. Number of Items of Operational/Braille Forms by Content Standard for Reading	8
Table 1.6.f. Number of Items of Equivalent Form by Content Standard for Reading	8
Table 1.6.g. Science Blueprint and Number of Items of Operational/Braille Forms Content	
Standards	8
Table 1.6.h. Science Blueprint and Number of Items of Equivalent Forms Content Standards	9
Table 3.1. Demographic Breakdowns by Gender, ELL, and Social Economic Status	. 12
Table 3.1.a. Demographic Breakdown by Ethnicity	. 13
Table 3.2. Secondary Statistical Key Check Criteria	. 14
Table 3.3. Number of Items Removed and Final Maximum Score Point Possible	. 14
Table 3.4. Test-Level Summaries of Classical Item Analyses	. 15
Table 3.5. DIF Flag Incidence by Subject and Grade	. 17
Table 3.6. Cronbach's Alpha by Subject and Grade	. 18
Table 3.7. Test Reliability by Subgroup	. 19
Table 3.7.a. Test Reliability by Subgroup (continued)	. 19
Table 4.1. OMAAP Scaling Constants, Scale Range, and Cut Scores by Subject and Grade	. 21
Table 4.2. Raw Score to Scale Score Conversion Table for Mathematics—Grades 3–5	. 22
Table 4.3. Raw Score to Scale Score Conversion Table for Mathematics—Grade 6–8	. 23
Table 4.4. Raw Score to Scale Score Conversion Table for Reading—Grades 3–5	. 24
Table 4.5. Raw Score to Scale Score Conversion Table for Reading—Grades 6–8	. 25
Table 4.6. Raw Score to Scale Score Conversion Table for Science—Grades 5 and 8	. 26
Table 5.1. Estimates of Accuracy and Consistency of Performance Classification	. 28
Table 5.2. Accuracy and Consistency Estimates by Cut Score. False Positive and False Negative	ve
Rates	.29
Table 5.2.a. Accuracy and Consistency Estimates by Cut Score. False Positive and False	>
Negative Rates (continued).	. 30
Table 6.1 Scale Score Descriptive Statistics—Overall	31
Table 6.2. Scale Score Descriptive Statistics by Gender	32
Table 6.3 Scale Score Descriptive Statistics by Race/Ethnicity	32
Table 6.3 a Scale Score Descriptive Statistics by Race/Ethnicity (continued)	33
Table 6.3 b. Scale Score Descriptive Statistics by Race/Ethnicity (continued)	33
Table 6.4. Scale Score Descriptive Statistics by Free/Reduced Lunch Status	34
Table 6.5 Percentage of Students by Performance Level	34
Table 6.6 Overall Estimates of SEM by Subject and Grade	36
rusie o.o. o volum Dominutes of SEM by Subject and Orade	. 50

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 $C^{\underline{3}}$ Oklahoma's Core curriculum, the College, Career and Citizen Ready CCSSO Council of Chief State School Officers **CE** Critical Element **CR** Constructed-Response CSEM Conditional Standard Error of Measurement **DIF Differential Item Functioning** DOK Depth of Knowledge DTC District Test Coordinator ELL English Language Learners **EOI End-of-Instruction EO** Equivalent **FN** False Negative **FP** False Positive **GR** Gridded-Response **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 Test OE Open-Ended OMAAP Oklahoma Modified Alternate Assessment Program $OK C^{\frac{3}{2}}$ Oklahoma College, Career, and Citizen Ready Standards **OP** Operational OSTP Oklahoma School Testing Program PASS Priority Academic Student Skills **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 TCC Test Characteristic Curve **TP** Test Proctor TPM Test Preparation Manual US DOE United States Department of Education WP Writing Prompt

Section 1 Overview of the Oklahoma Modified Alternate Assessment Program (OMAAP) Grades 3–8 Assessments

1.1 Overview of the OMAAP Grades 3-8 Assessments

The Oklahoma Modified Alternate Assessment Program for grades 3–8 evaluates student proficiency in Mathematics, Reading, and Science, and meets the state requirements for mandated, criterion-referenced tests. The OMAAP is specifically designed for students who would not benefit from taking the Oklahoma Alternate Assessment Program (OAAP) and/or the Oklahoma Core Curriculum Tests (OCCT).

The OMAAP Grades 3–8 tests assess student proficiency according to the *Oklahoma College*, *Career*, *and Citizen Ready Standards* (*OK* C^3), which are designed to ensure that students receive current and relevant learning experiences in order to become college, career, and citizenship ready. Certain subject exams are administered in specific grades. Mathematics and Reading OMAAP exams are administered in grades 3–8; however, the Science OMAAP exam is administered only in grades 5 and 8. The Oklahoma State Department of Education (SDE) and CTB/McGraw-Hill collaborated in the development and administration of the Spring 2013 OMAAP assessments. Scoring, equating, and scaling were performed by CTB/McGraw-Hill.

OMAAP Grades 3–8 are exclusively multiple-choice tests. In Spring 2013, each subject and grade had one operational form, which was also used as the base for the Braille test forms. Equivalent forms for each subject were selected from prior administrations' operational forms. The State Department of Education Office of Accountability and Assessments determined situations when an equivalent form was deemed appropriate, such as a student's illness during test administration or a security breach.

1.2 Content Standards

The OMAAP Grades 3–8 assessments were designed to measure the $OK C^3$ standards. Table 1.1 outlines the $OK C^3$ content standards by subject. Appendix A outlines the objectives of each content and/or process standard. The SDE's *Curriculum Access Resource Guides* (CARG) offers assistance to teachers by illustrating various methods of incorporating $OK C^3$ into classroom instruction through the appropriate development of skills.

1.3 Blueprint

The OMAAP blueprints have fewer items than the OCCT, but maintain similar proportions across standards. Guidelines dictate that in order for a standard to serve as a reporting category, it must have at least five items. The OMAAP adheres to this guideline, and test blueprints underwent thorough scrutiny before approval. First, committees of teachers and administrators proposed test blueprints for OMAAP. After the committees reviewed the $OK C^3$ content standards and the OCCT blueprints, they determined which proposed OMAAP Grades 3–8

blueprints were appropriate for the OMAAP student population. Final review of proposed test blueprints was conducted by the SDE before being submitted to the School Board of Education for approval.

OMAAP forms have 86–96% of operational items as the OCCT exams. Also, there are fewer or no field-test items in OMAAP tests. Table 1.2 shows a comparison of item counts across the Spring 2013 OCCT and the OMAAP forms.

	Mathematics				
Standard 1.	Algebraic Reasoning: Patterns and Relationships				
Standard 2.	Number Sense and Operation				
Standard 3.	Geometry				
Standard 4.	Measurement				
Standard 5.	Data Analysis				
	Reading				
Grade 3					
Standard 2.	Vocabulary				
Standard 4.	Comprehension/Critical Literacy				
Standard 5.	Literature				
Standard 6.	Research and Information				
Grades 4–8					
Standard 1.	Vocabulary				
Standard 3.	Comprehension/Critical Literacy				
Standard 4.	Literature				
Standard 5.	Research and Information				
	Science				
Process Stando	ards: Grades 5 and 8				
Standard 1.	Observe and Measure				
Standard 2.	Classify				
Standard 3.	Experiment				
Standard 4.	Interpret and Communicate				
Content Stande	ards: Grade 5				
Standard 1.	Properties of Matter and Energy				
Standard 2.	Organisms and Environments				
Standard 3.	Structures of the Earth and the Solar System				
Content Standards: Grade 8					
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				

Table 1.1. Oklahoma $OK C^3$ Content Standards by Subject

	OCCT Items		OMAA	P Items
Subject	OP	FT	OP	FT
Mathematics	50	10	43	0
Reading	50	10	43	0
Science	45	10	43	7

 Table 1.2. OCCT and OMAAP Grades 3–8 Item Count Comparison

*OP=Operational, FT=Field test.

1.4 Universal Design and Modifications

OMAAP item and test formats follow the *Universal Design* guidelines, ensuring tests are appropriate for students with various needs. Subject specific modifications have also been applied to increase test suitability. Table 1.3 lists the *Universal Design* and subject specific modifications.

Universal Design
Minimize the number of questions on the page (limit to 2)
Use a larger font size
Provide only three answer options instead of four
Highlight the main points in the question or passage by underlining and using boldface
Allow for the same accommodations as in the standard assessment
Avoid questions that require students to select the better/best answer
Eliminate answer choices that give students the option of making "no change" to the item
Mathematics
Allow for read-aloud and calculators
For lower grades, display numbers on all sides of figures for questions about perimeter
Unless required by standard, avoid items with negative and positive answer choices that use
the same number (e.g., -4 and $+4$)
Place any items with coordinate grids on one page
Be consistent with qualifiers in the stem and answer choices (e.g., use ml throughout or
milliliters throughout)
Avoid questions that use best or closest
Avoid complicated art
Reading
Display passages in a one-column format
Break passages into smaller portions, and place the questions that pertain to the smaller portion
Science
Reduce the amount of reading
Avoid complicated art
Simplify tables and charts by removing irrelevant rows or columns
Box formulas to make them stand out

1.5 Test Development and Content Validity

Development of a test relies upon test specifications to guide the construction process. Content validity is determined by specifications in content standards and test blueprints. Content standards address the knowledge and skills that are to be measured through the test, and test blueprints outline the number of items and item types to be included in each content area. The degree of content validity of a test is based on how closely it adheres to the specifications set forth. The closer the test is to meeting all specifications, the higher degree of content validity. This section describes the measures taken during the test construction process to ensure high content validity.

1.5.a. Item Development and Selection

The OMAAP test design requires items to be pulled from two sources: anchor items selected from previously-used OMAAP items and non-anchor items selected from previously-used OCCT Grades 3–8 items. The previously-used OMAAP items have been modified from OCCT items when they were first used on the OMAAP. Since anchor items are selected from previously-used OMAAP items, they can be used directly on the new test form. The non-anchor items are modified following the *Universal Design* guidelines.

Teacher committees identified the cognitive level each item measured through the Norman Webb's depth of knowledge (DOK) framework of each OCCT item. The OCCT items were then simplified through modification techniques that did not alter the content standards or cognitive level. Item review committees then examined the modified items for alignment to the $OK C^3$ and item appropriateness. The OMAAP item pool consists of only those items which passed the review.

Though newly modified OCCT items are eligible for inclusion in the operational form, they do not have a statistical history for their modified format. Item statistical quality is evaluated after the test is administered. This allows for occasional items with poor psychometric properties to be removed from scoring. Item quality control during test construction is facilitated by a plan of field testing items. Since Spring 2011, field-test items were included in the OMAAP Grades 3–8 tests. Field-test items serve to fill the gap between content standards and item banks. However, field-test items are limited to five or eight items per the one subject test form in each administration; this is reasonable for the OMAAP population size. Field testing gradually improves the OMAAP test construction process and test quality since these items have known psychometric properties, in contrast to the newly modified OCCT items which lack OMAAP population-based statistics.

CTB/McGraw-Hill used an embedded field-test design to incorporate newly developed field-test items into the tests. This method improves the likelihood that test-takers would treat every item as a scored item. Before a field-tested item is eligible for operational use, it is approved by the item review committee and data review committee.

Additional measures were taken in the test construction process. Test blueprints and *Universal Design* were applied in addition to five test construction guidelines focused on aligning the test with the $OK C^3$ standards and objectives. (See Table 1.4).

Туре	Guidelines
	The test is constructed so that there are at least six items $\frac{1}{2}$
	measuring each $OK C^{\circ}$ standard. The number of items is
1. Categorical Concurrence	based on estimating the number of items that could
	produce a reasonably reliable estimate of a student's
	mastery of the content measured.
	The test is constructed using items from a variety of Depth
2. Depth of Knowledge	of Knowledge levels that are consistent with the processes
Consistency	students need in order to demonstrate proficiency for each
	$OK C^3$ objective.
2 Danga of Knowledge	The test is constructed so that at least 75% of the
5. Kalige of Kliowledge	objectives for an $OK C^3$ standard have at least one
Correspondence	corresponding assessment item.
	The test is constructed according to the Test Blueprint
	which reflects the degree of representation given on the
	test to each $OK C^3$ standard and/or objective in terms of the
1 Delement of Demonstration	percent of total test items measuring each standard and the
4. Balance of Representation	number of test items measuring each standard and/or
	objective. The test construction shall yield a balance of
	representation with an index of 0.7 or higher of assessed
	objectives related to a standard.
	Each test item is constructed in such a way that the major
	cognitive demand comes directly from the targeted OKC^3
5. Source of Challenge	objective or concept being assessed, not from specialized
	knowledge or cultural background that the test-taker may
	bring to the testing situation.

Table 1.4. Criteria for Aligning the Test with the $OK C^3$ Standards and Objectives

1.5.b. Configuration of the Tests

The number of operational and field-test items for the Spring 2013 OMAAP Grades 3–8 assessment is presented in Table 1.2. Unlike the Spring 2012 test, the Spring 2013 tests for Mathematics and Reading are comprised of only operational items. No field-test items appear on the tests. The Science test is comprised of operational items and a set of field-test items. The field-test items are embedded in the operational test to attain reliable item statistics for future use.

In test assemblage, content experts followed two specifications: the test blueprint and DOK targets. Tables 1.5.a-1.5.b provide the targets and empirical percentages of items at each DOK level. Tables 1.6.a - 1.6.h shows the targets and empirical item counts of test blueprints of the OMAAP tests. Due to item bank limitations, some targets were short of the blueprint requirement.. This limitation is more noticeable in Reading tests that have passage-related

items. The Spring 2013 tests were constructed to have the maximum items possible—43 items for each subject and grade. After evaluating the item statistics, certain items were dropped from score reporting based on SDE's decision.

	DOK Level	1	2	3
	Target DOK	20-25	60-65	10-15
Subject	Grade			
	3	28	65	7
	4	26	67	7
M - 41-	5	33	58	9
Math	6	22	66	12
	7	19	72	9
	8	30	65	5
	3	21	70	9
	4	14	72	14
D.,	5	9	82	9
Reading	6	16	65	19
	7	9	82	9
	8	7	79	14
G	Target DOK	20-25	65-70	5-15
Science	5	23	65	12
Saianas	Target DOK	10-15	65-70	15-25
Science	8	19	67	14

 Table 1.5.a.
 Percentage of Items by DOK Level in Operational/Braille Forms

 Table 1.5.b.
 Percentage of Items by DOK Level of Equivalent Forms

	DOK Level	1	2	3
-	Target DOK	20-25	60-65	10-15
Subject	Grade			
	3	47	44	9
	4	35	60	5
M	5	28	67	5
Math	6	28	60	12
	7	19	71	10
	8	29	61	10
	3	23	65	12
	4	19	74	7
Deading	5	7	88	5
Reading	6	7	73	20
	7	7	73	20
	8	10	83	7
Saianaa	Target DOK	20-25	65-70	5-15
Science -	5	22	68	10
Sajanca	Target DOK	10-15	65-70	15-25
Science	8	26	72	2

	Grade					
Standard	3	4	5	6	7	8
Standard 1	6-7	6-7	10-11	10-11	12-13	13-14
Standard 2	15-16	14-15	12-13	12-13	8-9	8-9
Standard 3	6-7	7-8	6-7	6-7	6-7	7-8
Standard 4	7-8	7-8	6-7	6-7	7-8	6-7
Standard 5	6-7	6-7	6-7	6-7	6-7	6-7
Total	40-43	40-43	40-43	40-43	40-43	40-43

 Table 1.6.a.
 Mathematics
 Blueprint (Number of Items by Standard)

Table 1.6.b. Number of Items of Operational/Braille Forms by Content Standard for

 Mathematics

	Grade					
Standard	3	4	5	6	7	8
Standard 1	7	7	10	11	13	13
Standard 2	15	15	13	13	9	9
Standard 3	6	8	6	6	8	8
Standard 4	8	6	7	4	7	6
Standard 5	7	7	7	7	6	7
Total	43	43	43	41	43	43

Table 1.6.c. Number of Items of Equivalent Form by Content Standard for Mathematics

	Grade					
Standard	3	4	5	6	7	8
Standard 1	7	7	11	10	11	13
Standard 2	16	13	12	12	9	9
Standard 3	6	7	6	6	7	6
Standard 4	8	7	6	5	8	6
Standard 5	6	6	7	7	7	7
Total	43	40	42	40	42	41

Table 1.6.d. Reading Blueprint (Number of Items by Standard)

	Grade								
Standard*	3	4	5	6	7	8			
Standard (2)1	9-11	9-11	9-11	6-7	6-8	6-7			
Standard (4)2	18-20	17-19	15-17	15-17	15-17	16-18			
Standard (5)4	6-7	6-8	9-11	10-12	9-11	11-13			
Standard (6)5	6-7	6-7	6-7	6-7	6-7	6-7			
Total	40-43	40-43	40-43	40-43	40-43	40-43			

*Values in parentheses are grade 3 content standards. The other set of values are content standards for grades 4 through 8.

	Grade							
Standard*	3	4	5	6	7	8		
Standard (2)1	10	9	11	7	9	11		
Standard (4)2	20	18	17	20	18	17		
Standard (5)4	6	9	9	11	7	8		
Standard (6)5	7	7	6	5	9	7		
Total	43	43	43	43	43	43		

Table 1.6.e. Number of Items of Operational/Braille Forms by Content Standard for Reading

*Values in parentheses are grade 3 content standards. The other set of values are content standards for grades 4 through 8.

Table 1.6.f. Number of Items of Equivalent Form by Content Standard for Reading

	Grade							
Standard*	3	4	5	6	7	8		
Standard (2)1	9	8	9	7	8	7		
Standard (4)2	18	17	15	15	17	16		
Standard (5)4	8	9	11	13	9	11		
Standard (6)5	8	8	7	8	7	8		
Total	43	42	42	43	41	42		

*Values in parentheses are grade 3 content standards. The other set of values are content standards for grades 4 through 8.

Table 1.6.g. Science Blueprint and Number of Items of Operational/Braille Forms Content

 Standards

Standard		Grade 8		
Content	Target	Actual	Target	Actual
Standard 1	15-17	17	6-8	6
Standard 2	10-12	11	6-8	8
Standard 3	9-11	11	7-9	9
Standard 4			6-8	8
Standard 5			6-8	8
Total	37-40	39	36-39	39

		Grade 5		Grade 8
Process	Target	Actual	Target	Actual
Standard 1	8-10	9	6-8	8
Standard 2	8-10	10	6-8	9
Standard 3	9-11	10	13-15	14
Standard 4	12-14	14	11-13	12
Total	40-43	43	40-43	43

		Grade	2	
Standard		Grade 5		Grade 8
Content	Target	Actual	Target	Actual
Standard 1	15-17	14	6-8	7
Standard 2	10-12	11	6-8	8
Standard 3	9-11	8	7-9	9
Standard 4			6-8	6
Standard 5			6-8	7
Total	37-40	33	36-39	37
		Grade 5		Grade 8
Process	Target	Actual	Target	Actual
Standard 1	8-10	8	6-8	8
Standard 2	8-10	8	6-8	8
Standard 3	9-11	9	13-15	14
Standard 4	12-14	12	11-13	12
Total	40-43	37	40-43	42

Table 1.6.h. Science Blueprint and Number of Items of Equivalent Forms Content Standards

Section 2 Administration of the OMAAP Grades 3–8 Assessments

To ensure a valid and reliable assessment, the OMAAP Grades 3–8 assessments are first constructed in alignment with the *Oklahoma* C^3 *Standards* (now Oklahoma Academic Standards) by the Oklahoma State Department of Education in collaboration with CTB. 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 administering 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 OMAAP Grades 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 OMAAP Grades 3–8 assessments by ensuring fairness in administration of the test.

2.1 Packaging and Shipping

In order to provide secure and dependable services for the shipping of the OMAAP Grades 3–8 assessment materials, CTB'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 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 with the precode information needed to print student barcode labels, which are affixed on answer documents or consumable test books. The barcoding of all secure materials at the time of production allows for accurate tracking of these materials through the entire packing, delivery, and return process. Accurate tracking allows CTB to inventory all materials throughout the packaging and delivery process.

2.2 Materials Return

The Test Preparation Manual and Materials Return poster provides clear instructions on how to assemble, box, label, and return testing materials after test administration. CTB 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's Data Processing Facility.

2.3 Materials Discrepancies Process

The scanning process allows CTB to capture multiple-choice 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 Oklahoma Help Desk, and reported. The results are subsequently reported to the Oklahoma State Department of Education (SDE).

A pre-determined date is set by SDE and CTB 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 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 takes test security seriously and makes every effort to recover missing material.

Section 3

Classical Item Analysis and Results

Analyses presented in Sections 3, 4, and 5 focused on the Spring operational forms. This section used the student data for final score reporting. Invalid cases or second-time test-takers were excluded in these analyses.

3.1 Data Quality Check and Clean-Up

After all tests were scanned and scored, a data clean-up process was implemented to remove invalid cases, ineligible responses, absent students, and repeat test-takers. A statistical key check was also performed at this time. This 'cleaned' data was used for classical item analyses.

Exclusion Rules. Exclusionary rules were applied to form the final sample that was used for classical item analyses, calibration, and equating. Any student who had attempted at least five responses was included in the data analyses. However, students who: took the Braille form, were second-time test-takers, had invalidated codes, or attended a private school were not included in the equating and scaling processes. The demographic breakdown of the students in the item analysis and calibration sample is given in Table 3.1 and 3.1.a.

		English					
					Language	Economically	
Subject	Grade	Total	Female	Male	Learner	Disadvantaged	
	3	2679	947	1725	340	2087	
	4	3076	1154	1917	380	2369	
	5	3615	1435	2173	532	2800	
Mamematics	6	3447	1264	2175	408	2581	
	7	3588	1315	2263	341	2649	
	8	3639	1392	2231	284	2600	
	3	3221	1106	2106	409	2476	
	4	3647	1289	2354	412	2787	
Deading	5	4053	1534	2513	585	3151	
Reading	6	3626	1273	2344	434	2746	
	7	3576	1253	2312	368	2654	
	8	3528	1276	2239	305	2600	
Science	5	3057	1163	1890	459	2378	
Science	8	2535	909	1620	219	1855	

Table 3.1. Demographic	Breakdowns by	Gender, ELL	and Social	Economic Status
i abie etit Demographie	Dicalitation in the o			Deononne Statas

		African	Native	•		Pacific		
Subject	Grade	American	American	Hispanic	Asian	Islander	White	Other
	3	399	467	401	13	4	1211	177
	4	418	517	449	14	7	1461	205
Mathamatica	5	528	604	561	11	4	1674	226
Mathematics	6	470	647	474	19	4	1584	241
	7	503	664	466	11	6	1706	222
	8	499	710	410	11	5	1783	205
	3	450	559	495	16	5	1478	209
	4	476	611	536	17	6	1751	246
Deading	5	580	684	654	14	5	1867	243
Reading	6	499	654	528	22	5	1648	261
	7	466	672	501	10	5	1687	224
	8	489	683	451	10	6	1692	184
Saianaa	5	426	520	488	9	5	1418	187
Science	8	344	500	302	6	3	1243	131

Table 3.1.a. Demographic Breakdown by Ethnicity

Statistical Key Check. To screen for potentially problematic items and to confirm multiplechoice items were accurately scored, a statistical item answer key check was conducted and items were flagged if:

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

Flagged operational items were submitted for answer key review by a CTB/McGraw-Hill content specialist. Any flagged items that were identified by content experts as mis-keyed would be corrected prior to further data analysis. Once the keys were verified, a secondary statistical key check and evaluation of items was conducted for the potential of removing items from scoring. There were no items identified as having a key issue for the 2013 tests.

Removal of operational items. Once the statistical key check was complete, all items were screened using the criteria defined in Table 3.2. This procedure identified items with poor statistics for potential removal from scoring. The CTB/McGraw-Hill research scientists and content specialists reviewed the flagged items and proposed suggestions to the SDE. The SDE then evaluated and decided any exclusion of the items.

Key Validation Item-Flagging Criteria							
If p value of keyed response < 0.35	Difficult item						
If p value of keyed response < 0.05 or > 0.95	Extreme item						
If p value of keyed response < p value of distracter	Possible mis-key						
If p value of distracter > 0.35	Possible second correct option						
If point-biserial of keyed response < 0.20	Poorly discriminating item						
If point-biserial of distracter > 0.05	Possible second correct option						
If point-biserial of keyed response < point-biserial of							
distracter	Possible mis-key						

 Table 3.2. Secondary Statistical Key Check Criteria

The SDE's decision on the 2013 tests is listed by subject in Table 3.3. These items were removed from operational scoring for the current operational and Braille forms and also removed from the OMAAP Grades 3–8 item bank. Table 3.3 also presents the final number of points possible on the 2013 OMAAP Grades 3–8 after the removal of the items.

Subject	Grada	Number of Items	Final Max. Score Point
Subject	Ulaue	Diopped	FOSSIDIE
	3	0	43
	4	0	43
Mathamatics	5	0	43
Mainematics	6	2	41
	7	0	43
	8	0	43
	3	0	43
	4	0	43
Deedine	5	0	43
Reading	6	0	43
	7	0	43
	8	0	43
Sajanaa	5	0	43
Science	8	0	43

 Table 3.3. Number of Items Removed and Final Maximum Score Point Possible

3.2 Classical Item Analyses

The following classical item analysis statistics were produced for operational test items:

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

3.2.a. Test-Level Summaries of Classical Item Analyses

The test-level raw score descriptive statistics for the calibration samples are shown in Table 3.4. Note that students whose tests were invalidated and those students taking the test for a second time were excluded. The test difficulty ranged between 0.47 and 0.60, median point biserials were above 0.30 in most cases, and omit rates were smaller than 3% for all subjects.

		Sample	Mean	Items	Mean	Median	Omit	Omit
Subject	Grade	Size	Raw Score	Points	P-value	r pb	Min	Max
	3	2679	24.8	43	0.58	0.42	0.00	1.38
	4	3076	25.9	43	0.60	0.39	0.03	0.39
Mathamatica	5	3615	21.3	43	0.50	0.37	0.00	0.41
Mathematics	6	3447	20.6	41	0.48	0.33	0.06	0.58
	7	3588	20.2	43	0.47	0.29	0.08	0.89
	8	3639	21.7	43	0.50	0.32	0.00	0.74
	3	3221	21.4	43	0.50	0.42	0.16	2.36
	4	3647	22.3	43	0.52	0.38	0.14	1.81
Daading	5	4053	23.5	43	0.55	0.38	0.25	2.20
Reading	6	3626	22.3	43	0.52	0.32	0.11	1.19
	7	3576	22.3	43	0.52	0.33	0.25	2.04
	8	3528	23.8	43	0.55	0.35	0.20	1.50
Saianaa	5	3057	25.1	43	0.58	0.35	0.00	0.72
Science	8	2535	26.0	43	0.60	0.36	0.04	1.26

 Table 3.4. Test-Level Summaries of Classical Item Analyses

* rpb= Point biserial; Item Points = Max. Possible Raw Score

3.3 Procedures for Detecting Item Bias

One of the goals of the OMAAP Grades 3–8 assessments is to assemble a set of items that provides a measure of a student's ability that is as fair and accurate as possible for all subgroups within the population. Differential item functioning (DIF) analysis refers to statistical procedures that assess whether items are differentially difficult for different groups of examinees. DIF procedures typically control for overall between-group differences on a criterion, usually total test scores. Between-group performance on each item is then compared within sets of examinees having the same total test scores. If the item is differentially more difficult for an identifiable subgroup when conditioned on ability, the item may be measuring something different from the intended construct. However, it is important to recognize that DIF-flagged items might be related to actual differences in relevant knowledge or skills or statistical Type I error. As a result, DIF statistics are used only to identify potential sources of item bias. Subsequent review by content experts and bias committees are required to determine the source and meaning of performance differences. For the OMAAP Grades 3-8 test DIF analyses, DIF statistics were estimated for race and gender. Items with statistically-significant differences in performance were flagged so that items could be carefully examined for possible biased or unfair content that was undetected in earlier fairness and bias content review meetings held prior to form construction.

CTB/McGraw-Hill used the Mantel-Haenszel (MH) approach for detecting DIF in items. The student group of interest is the *focal* group, and the group to which performance on the item is being compared is the *reference* group. The reference groups for these DIF analyses were White for race and male for gender. The focal groups for race were African American, Native American, and Hispanic students. The focal group for gender was female students.

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

- Classification C: $|\Delta| \ge 1.5$ and MH D-DIF chi-square < 0.05
- Classification B: $1 \le |\Delta| \le 1.5$ and MH D-DIF chi-square < 0.05
- Classification A: Otherwise

3.3.a. Differential Item Functioning Results

During field-test stage, items flagged for DIF were reviewed by expert content specialists from SDE and CTB/McGraw-Hill prior to inclusion as part of the operational scored set. The 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 race 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). Items that were found to exhibit bias as a result of the content of the item

would be removed from scoring and the item bank, excluding them from future use. This section presents DIF analysis results of the operational items to evaluate the performance of each item.

					African		Native			
			Fen	nale/	American/		American/		Hispanic/	
			Μ	ale	White		White		White	
Subject	Grade	Total Items	В	С	В	С	В	С	В	С
	3	43	1	0	2	0	0	0	3	0
	4	43	1	0	2	0	0	0	2	0
Mathamatica	5	43	2	0	0	0	0	0	0	0
Mathematics	6	41	1	0	1	0	0	0	0	0
	7	43	1	0	0	1	1	0	0	1
	8	43	2	0	4	0	0	0	1	0
	3	43	0	0	0	0	0	0	0	0
	4	43	0	0	1	0	0	0	1	0
Deading	5	43	0	0	0	0	0	0	1	0
Reading	6	43	0	1	0	0	0	0	1	0
	7	43	2	0	0	0	0	0	1	0
	8	43	2	0	1	0	0	0	0	0
<u> </u>	5	43	2	0	0	0	0	0	1	0
Science	8	43	2	0	0	1	0	0	1	0

Table 3.5. DIF Flag Incidence by Subject and Grade

*Classification C: $|\Delta| \ge 1.5$ and MH D-DIF chi-square < 0.05;

Classification B: $1 \le |\Delta| \le 1.5$ and MH D-DIF chi-square < 0.05

3.4 Test Reliability

3.4.a. Overall Test Reliability

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

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

Cronbach's Alpha of all operational items was estimated for each subject and grade. Table 3.6 shows that the reliability coefficients are above 0.75 in all cases. The values indicate that the OMAAP Grades 3–8 assessments had adequate internal consistency and that the tests produce relatively stable scores.

Subject	Grade	Number of Items	Alpha
	3	43	0.89
	4	43	0.87
Mathamatica	5	43	0.84
Mathematics	6	41	0.76
	7	43	0.75
	8	43	0.81
	3	43	0.87
	4	43	0.85
Deading	5	43	0.85
Reading	6	43	0.80
	7	43	0.81
	8	43	0.83
Soionaa	5	43	0.82
Science	8	43	0.84

 Table 3.6. Cronbach's Alpha by Subject and Grade

3.4.b. Test Reliability by Subgroup

Table 3.7 shows the reliability analysis results by reporting subgroups for the OMAAP assessments. This table illustrates the subject, subgroups, and Cronbach's Alpha reliability coefficients. In most cases, the reliability coefficients are well above the accepted lower limit of .70.

				African	Native			
Subject	Grade	Male	Female	American	American	Hispanic	Asian	White
	3	0.90	0.87	2.63	0.88	0.89	0.89	0.89
Mathematics	4	0.87	0.86	2.57	0.86	0.86	0.90	0.87
	5	0.85	0.84	2.49	0.84	0.84	0.93	0.85
Mamematics	6	0.76	0.76	2.15	0.77	0.73	0.85	0.77
	7	0.75	0.75	2.12	0.75	0.72	0.76	0.77
	8	0.81	0.81	2.36	0.82	0.78	0.69	0.81
	3	0.87	0.86	2.53	0.86	0.85	0.91	0.88
	4	0.85	0.85	2.46	0.85	0.82	0.85	0.86
Deading	5	0.85	0.84	2.50	0.85	0.84	0.81	0.85
Reading	6	0.79	0.80	2.26	0.78	0.75	0.84	0.82
	7	0.81	0.81	2.34	0.80	0.77	0.77	0.82
	8	0.83	0.83	2.44	0.84	0.81	0.87	0.83
Saianaa	5	0.83	0.81	2.42	0.81	0.80	0.92	0.83
Science	8	0.85	0.82	2.43	0.84	0.82	0.54	0.85

 Table 3.7. Test Reliability by Subgroup

Table 3.7.a. Test Reliability by Subgroup (continued)

		English	
		Language	Economically
Subject	Grade	Learner	Disadvantaged
	3	0.87	0.89
	4	0.87	0.87
Mathamatica	5	0.84	0.84
Mathematics	6	0.71	0.75
	7	0.73	0.74
	8	0.79	0.80
	3	0.83	0.86
	4	0.80	0.84
Daadina	5	0.83	0.84
Reading	6	0.73	0.79
	7	0.76	0.80
	8	0.81	0.83
Saianaa	5	0.80	0.82
Science	8	0.81	0.84

Section 4

Calibration, Equating, and Scaling

4.1 Item Response Theory (IRT) Models

Rasch Model. The Rasch model (Rasch, 1960) was used for calibrating all OMAAP Grades 3–8 items. In the Rasch model, the probability that a student with an ability level of θ responds correctly to item *i* is

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

where b_i is the item difficulty parameter.

4.2 Calibration and Equating

The Rasch model was used for all OMAAP Grades 3–8 assessments. Due to low student sample sizes, the traditional post-equating design could not be applied this year, pre-equating was used in this year's equating instead. All item parameters from previous (before Spring 2013) equating results were used for generating raw score to scale score relationships (scoring tables) directly.

4.3 Scaling and Scoring Results

The Lowest Obtainable Scale Score (LOSS), Highest Obtainable Scale Score (HOSS), and final scaling constants for each of the subjects are shown in Table 4.1. The scaling constants, M1 (multiplicative) and M2 (additive), place the true scores associated with each raw score point onto the reporting or operational scale using a straightforward linear transformation:

Scale Score =
$$(\hat{\tau} \times M1) + M2$$
 (3)

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

Also shown in Table 4.1 are the 3 cut scores for each subject and grade. The 3 cut scores divide scores into 4 performance levels: 1 - Unsatisfactory, 2 - Limited Knowledge, 3 - Satisfactory, and 4 - Advanced. The raw score to number-correct scale score conversion tables were generated using WINSTEPS Rasch model. The results are shown in Tables 4.2 through 4.6, along with the conditional standard error of measurement (*CSEM*; please see Section 6.3 for the computation of *CSEM*) associated with each of the scale score.

		Scaling								
		Cons	stants	Scale	Range		Cut Score			
						Limited				
Subject	Grade	Slope	Mean	LOSS	HOSS	Knowledge	Satisfactory	Advanced		
	3	19.94	244.59	100	350	233	250	277		
	4	20.09	247.68	100	350	238	250	277		
Mathamatica	5	20.31	243.66	100	350	240	250	271		
Mathematics	6	19.36	245.86	100	350	237	250	272		
	7	20.53	240.05	100	350	232	250	265		
	8	20.15	244.33	100	350	235	250	271		
	3	20.89	247.50	100	350	238	250	269		
	4	20.75	247.42	100	350	237	250	266		
Deading	5	20.73	248.66	100	350	231	250	269		
Reading	6	21.77	242.02	100	350	229	250	261		
	7	21.82	242.05	100	350	229	250	271		
	8	21.31	241.48	100	350	236	250	276		
Science	5	20.92	245.91	100	350	238	250	277		
Science	8	20.13	251.39	100	350	241	250	288		

Table 4.1. OMAAP Scaling Constants, Scale Range, and Cut Scores by Subject and Grade

*LOSS = Lowest Obtainable Scale Score; HOSS = Highest Obtainable Scale Score

		3			4			5	
Raw	Scale	Perf.		Scale	Perf.		Scale	Perf.	
Score	Score	Level	CSEM	Score	Level	CSEM	Score	Level	CSEM
0	148	1	37	147	1	37	152	1	37
1	172	1	20	172	1	20	177	1	21
2	187	1	15	186	1	15	192	1	15
3	196	1	12	195	1	12	201	1	12
4	202	1	11	202	1	11	208	1	11
5	207	1	10	207	1	10	213	1	10
6	212	1	9	211	1	9	217	1	9
7	216	1	9	215	1	9	221	1	9
8	219	1	8	219	1	8	225	1	8
9	223	1	8	222	1	8	228	1	8
10	226	1	8	225	1	8	231	1	8
11	228	1	7	228	1	7	234	1	7
12	231	1	7	230	1	7	236	1	7
13	234	2	7	233	1	7	239	1	7
14	236	2	7	235	1	7	241	2	7
15	238	2	7	237	1	7	244	2	7
16	241	2	7	240	2	7	246	2	7
17	243	2	7	242	2	7	248	2	7
18	245	2	7	244	2	7	250	3	7
19	247	2	7	246	2	6	252	3	7
20	249	2	7	248	2	6	255	3	7
21	252	3	7	250	3	6	257	3	6
22	254	3	7	252	3	6	259	3	6
23	256	3	7	254	3	6	261	3	7
24	258	3	7	256	3	6	263	3	7
25	260	3	7	258	3	7	265	3	7
26	262	3	7	261	3	7	267	3	7
27	265	3	7	263	3	7	269	3	7
28	267	3	7	265	3	7	271	4	7
29	269	3	7	267	3	7	274	4	7
30	272	3	7	270	3	7	276	4	7
31	274	3	7	272	3	7	279	4	7
32	277	4	7	275	3	7	281	4	7
33	280	4	8	278	4	8	284	4	8
34	283	4	8	280	4	8	287	4	8
35	286	4	8	284	4	8	290	4	8
36	290	4	9	287	4	9	293	4	9
37	294	4	9	291	4	9	297	4	9
38	298	4	10	295	4	10	302	4	10
39	303	4	11	301	4	11	307	4	11
40	310	4	12	307	4	12	313	4	12
41	319	4	15	316	4	15	322	4	15
42	333	4	20	331	4	20	337	4	21

 Table 4.2. Raw Score to Scale Score Conversion Table for Mathematics—Grades 3–5

43	350	4	30	350	4	32	350	4	28
Table 4	.3. Raw Sc	ore to Sca	ale Score	Conversion	n Table fo	or Mathen	natics—Gra	ade 6–8	
		6			7			8	
Raw	Scale	Perf.		Scale	Perf.		Scale	Perf.	
Score	Score	Level	CSEM	Score	Level	CSEM	Score	Level	CSEM
0	157	1	36	146	1	38	151	1	37
1	181	1	20	172	1	21	176	1	20
2	195	1	14	187	1	15	190	1	15
3	204	1	12	196	1	12	199	1	12
4	211	1	10	202	1	11	205	1	11
5	216	1	10	208	1	10	211	1	10
6	220	1	9	212	1	9	215	1	9
7	224	1	8	216	1	9	219	1	8
8	227	1	8	220	1	8	222	1	8
9	230	1	8	223	1	8	225	1	8
10	233	1	7	226	1	8	228	1	7
11	236	1	7	229	1	7	231	1	7
12	238	2	7	231	1	7	233	1	7
13	241	2	7	234	2	7	236	2	7
14	243	2	7	236	2	7	238	2	7
15	245	2	7	238	2	7	240	2	7
16	248	2	6	241	2	7	242	2	7
17	250	3	6	243	2	7	244	2	6
18	252	3	6	245	2	7	246	2	6
19	254	3	6	247	2	7	249	2	6
20	256	3	6	249	2	7	251	3	6
21	258	3	6	251	3	7	253	3	6
22	260	3	6	253	3	7	255	3	6
23	262	3	6	255	3	7	257	3	6
24	264	3	6	258	3	7	259	3	6
25	266	3	6	260	3	7	261	3	6
26	268	3	6	262	3	7	263	3	6
27	271	3	7	264	3	7	265	3	7
28	273	4	7	266	4	7	267	3	7
29	275	4	7	268	4	7	269	3	7
30	278	4	7	271	4	7	271	4	7
31	280	4	7	273	4	7	274	4	7
32	283	4	7	276	4	7	276	4	7
33	286	4	8	279	4	8	279	4	7
34	289	4	8	282	4	8	282	4	8
35	293	4	9	285	4	8	285	4	8
36	297	4	9	288	4	9	288	4	8
37	302	4	10	292	4	9	292	4	9
38	309	4	12	296	4	10	297	4	10
39	317	4	14	302	4	11	302	4	11
40	331	4	20	308	4	12	308	4	12
41	350	4	31	317	4	15	317	4	15
42				332	4	21	332	4	20
43	•	•		350	4	32	350	4	32

Copyright © 2013 by Oklahoma State Department of Education.

14,510		3		0011101010	4		8 01440	5	
Raw	Scale	Perf.		Scale	Perf.		Scale	Perf.	
Score	Score	Level	CSEM	Score	Level	CSEM	Score	Level	CSEM
0	150	1	38	155	1	38	147	1	38
1	176	1	21	180	1	21	173	1	21
2	191	1	15	195	1	15	188	1	15
3	200	1	13	204	1	12	197	1	12
4	207	1	11	210	1	11	203	1	11
5	212	1	10	216	1	10	209	1	10
6	216	1	9	220	1	9	213	1	9
7	220	1	9	224	1	9	217	1	9
8	224	1	8	227	1	8	220	1	8
9	227	1	8	231	1	8	223	1	8
10	230	1	8	234	1	8	226	1	8
11	232	1	7	236	1	7	229	1	7
12	235	1	7	239	2	7	232	2	7
13	237	1	7	241	2	7	234	2	7
14	240	2	7	244	2	7	236	2	7
15	242	2	7	246	2	7	239	2	7
16	244	2	7	248	2	7	241	2	7
17	246	2	7	250	3	7	243	2	7
18	248	2	7	252	3	7	245	2	7
19	250	3	7	254	3	7	247	2	7
20	252	3	6	256	3	6	249	2	6
21	254	3	6	258	3	6	251	3	6
22	256	3	6	260	3	6	253	3	6
23	258	3	6	262	3	6	255	3	6
24	260	3	7	264	3	7	257	3	7
25	262	3	7	266	4	7	259	3	7
26	264	3	7	269	4	7	261	3	7
27	267	3	7	271	4	7	264	3	7
28	269	4	7	273	4	7	266	3	7
29	271	4	7	275	4	7	268	3	7
30	273	4	7	277	4	7	270	4	7
31	276	4	7	280	4	7	273	4	7
32	278	4	7	282	4	7	275	4	7
33	281	4	8	285	4	8	278	4	8
34	284	4	8	288	4	8	281	4	8
35	287	4	8	291	4	8	284	4	8
36	290	4	9	295	4	9	287	4	9
37	294	4	9	299	4	9	291	4	9
38	299	4	10	303	4	10	296	4	10
39	304	4	11	308	4	11	301	4	11
40	311	4	13	315	4	13	308	4	12
41	320	4	15	324	4	15	317	4	15
42	335	4	21	339	4	21	332	4	21
43	350	4	30	350	4	27	350	4	32

Table 4.4. Raw Score to Scale Score Conversion Table for Reading—Grades 3–5

		6			7			8	
Raw	Scale	Perf.		Scale	Perf.		Scale	Perf.	
Score	Score	Level	CSEM	Score	Level	CSEM	Score	Level	CSEM
0	143	1	40	141	1	40	145	1	39
1	170	1	22	168	1	22	171	1	22
2	186	1	16	184	1	16	186	1	16
3	196	1	13	193	1	13	195	1	13
4	203	1	12	200	1	12	202	1	11
5	208	1	11	206	l	11	208	1	10
6	213	1	10	210	1	10	212	1	10
/	217	1	9	214	1	9	216	1	9
8	221	1	9	218	1	9	220	1	9
9	224	1	8	221	1	ð	225	1	8
10	227	1	8 0	224	1	8	220	1	8
11	230	2	0	227	1	0	229	1	0 7
12	235	$\frac{2}{2}$	0 7	230	$\frac{2}{2}$	0 7	231	1	7
13	230	2	7	233	$\frac{2}{2}$	7	234	1	7
14	238	$\frac{2}{2}$	7	233	$\frac{2}{2}$	7	230	$\frac{2}{2}$	7
15	241	$\frac{2}{2}$	7	237	$\frac{2}{2}$	7	239	2	7
17	245	$\frac{2}{2}$	7	240 242	$\frac{2}{2}$	7	241	$\frac{2}{2}$	7
18	243 247	$\frac{2}{2}$	7	242	$\frac{2}{2}$	7	245	$\frac{2}{2}$	7
19	250	3	7	244	$\frac{2}{2}$, 7	243	$\frac{2}{2}$	7
20	252	3	7	249	$\frac{2}{2}$, 7	250	3	7
21	254	3	, 7	251	3	7	252	3	7
22	256	3	7	253	3	7	254	3	7
23	258	3	7	255	3	7	256	3	7
24	260	3	7	257	3	7	258	3	7
25	262	4	7	259	3	7	260	3	7
26	265	4	7	262	3	7	262	3	7
27	267	4	7	264	3	7	264	3	7
28	269	4	7	266	3	7	267	3	7
29	272	4	7	269	3	7	269	3	7
30	274	4	7	271	4	7	271	3	7
31	277	4	8	274	4	8	274	3	7
32	279	4	8	276	4	8	277	4	8
33	282	4	8	279	4	8	279	4	8
34	285	4	8	282	4	8	282	4	8
35	289	4	9	286	4	9	286	4	8
36	292	4	9	289	4	9	289	4	9
37	296	4	10	293	4	10	293	4	9
38	301	4	10	298	4	11	298	4	10
39	306	4	12	304	4	12	303	4	11
40	313	4	13	310	4	13	310	4	13
41	323	4	16	320	4	16	319	4	16
42	339	4	22	336	4	22	334	4	22
43	350	4	28	350	4	30	350	4	31

 Table 4.5. Raw Score to Scale Score Conversion Table for Reading—Grades 6–8

		5			8	
Raw	Scale	Perf.		Scale	Perf.	
Score	Score	Level	CSEM	Score	Level	CSEM
0	144	1	38	155	1	37
1	170	1	21	180	1	20
2	185	1	15	195	1	15
3	194	1	13	204	1	12
4	201	1	11	210	1	11
5	206	1	10	215	1	10
6	211	1	9	220	1	9
7	215	1	9	224	1	9
8	219	1	8	227	1	8
9	222	1	8	230	1	8
10	225	1	8	233	1	8
11	228	1	8	236	1	7
12	230	1	7	239	1	7
13	233	1	7	241	2	7
14	235	1	7	243	2	7
15	238	2	7	246	2	7
16	240	2	7	248	2	7
17	242	2	7	250	3	7
18	245	2	7	252	3	7
19	247	2	7	254	3	6
20	249	2	7	256	3	6
21	251	3	7	258	3	6
22	253	3	7	261	3	6
23	255	3	7	263	3	6
24	257	3	7	265	3	6
25	260	3	7	267	3	6
26	262	3	7	269	3	7
27	264	3	7	271	3	7
28	266	3	7	273	3	7
29	269	3	7	275	3	7
30	271	3	7	278	3	7
31	274	3	7	280	3	7
32	276	3	8	283	3	7
33	279	4	8	285	3	7
34	282	4	8	288	4	8
35	285	4	8	291	4	8
36	289	4	9	295	4	9
37	293	4	9	299	4	9
38	297	4	10	303	4	10
39	303	4	11	308	4	11
40	309	4	13	315	4	12
41	319	4	15	323	4	15
42	334	4	21	338	4	20
43	350	4	31	350	4	27
			<i>.</i> .		•	

 Table 4.6. Raw Score to Scale Score Conversion Table for Science—Grades 5 and 8

Section 5

Classification Consistency and Accuracy Studies

5.1 Classification Consistency and Accuracy

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

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

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

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

To evaluate decisions at specific cut scores, the joint distribution of all the performance levels are collapsed into dichotomized distributions around that specific cut score (for example collapsing Unsatisfactory and Limited Knowledge and then Satisfactory and Advanced to assess decisions at the Satisfactory cut score). The accuracy index at cut score is computed as the sum of the proportions of correct classifications around this selected cut score. The consistency at a specific cut score is obtained in a similar way, but by dichotomizing the distributions at the cut score performance level and between all other performance levels combined. Table 5.1 presents the overall accuracy and consistency indices for the OMAAP Grades 3–8 assessments.

				False	False	
Subject	Grade	Accuracy	Consistency	Positive	Negative	Kappa
	3	0.77	0.69	0.11	0.12	0.54
	4	0.76	0.67	0.11	0.13	0.51
Mathamatica	5	0.72	0.62	0.14	0.15	0.46
Mainematics	6	0.72	0.61	0.13	0.15	0.36
	7	0.68	0.57	0.16	0.16	0.36
	8	0.72	0.62	0.13	0.14	0.43
	3	0.72	0.63	0.14	0.14	0.49
	4	0.73	0.64	0.12	0.15	0.47
Deading	5	0.74	0.64	0.13	0.14	0.48
Reading	6	0.7	0.61	0.14	0.16	0.42
	7	0.73	0.63	0.13	0.14	0.44
	8	0.75	0.65	0.12	0.13	0.46
Caianaa	5	0.75	0.65	0.12	0.13	0.45
Science	8	0.83	0.76	0.08	0.09	0.49

Table 5.1. Estimates of Accuracy and Consistency of Performance Classification

As shown in Table 5.1, the overall accuracy indices range between 68 and 83 percent and overall consistency ranges between 57 and 76 percent. Kappa coefficients range from 36 and 54 percent. The rate of false positives range from 8 to 16 percent. The false negative rates range from 9 to 16 percent.

Table 5.2 provides the accuracy, consistency, false positive, and false negative rates by cut-score. The data in these tables reveal that the level of agreement for both accuracy and consistency is above 70 percent in all cases, with most cases above 80 percent. In general, the high rates of accuracy and consistency support the cut decisions made using these assessments. The false positive and false negative rates are low compared to Table 5.1.

The importance of the dichotomous categorization is particularly notable when they map onto pass/fail decisions for the assessments. For the OMAAP Grades 3–8 tests, the U+L/S+A is the important dichotomization because it directly translates to the pass/fail decision point. Similar to other dichotomization distinctions, there are three main scenarios at this cut point: 1) observed performance is accurately reflective of the true ability level (i.e., the examinee passed and should

have passed); 2) the true ability level is below the standard, but the observed test score is above the standard (i.e., a false positive); and 3) the true ability level is above the standard, but the observed test score is below the standard (i.e., a false negative). In examining Table 5.2/5.2.a, Math Grade 3, for example, 91 percent of students are correctly classified as pass or fail based on their performance (scenario 1), 4 percent passed, but their true ability is below the standard (scenario 2), and 5 percent failed although their true ability is above the standard (scenario 3). Overall, the accuracy rates for accurate classification are above 80% for all tests – students are appropriately (more than 80% of the time) categorized into pass/fail classifications based on their true ability using their observed score (raw score) as their classification score.

		A	Accurac	у	Co	onsisten	су
		U	U+L	U+L+S	U	U+L	U+L+S
		/	/	/	/	/	/
Subject	Grade	L+S+A	S+A	А	L+S+A	S+A	А
	3	0.95	0.91	0.91	0.93	0.87	0.88
	4	0.94	0.90	0.91	0.92	0.87	0.87
Mathamatics	5	0.92	0.88	0.90	0.89	0.84	0.86
Mainematics	6	0.95	0.86	0.91	0.92	0.80	0.87
	7	0.93	0.83	0.92	0.90	0.77	0.88
	8	0.94	0.86	0.92	0.91	0.81	0.89
	3	0.92	0.89	0.90	0.89	0.84	0.87
	4	0.95	0.90	0.88	0.93	0.86	0.83
Daadina	5	0.95	0.88	0.90	0.94	0.84	0.85
Reading	6	0.97	0.87	0.85	0.95	0.82	0.80
	7	0.95	0.86	0.91	0.93	0.81	0.88
	8	0.94	0.88	0.92	0.92	0.84	0.89
Saianaa	5	0.94	0.89	0.91	0.92	0.85	0.87
Science	8	0.97	0.93	0.92	0.95	0.90	0.89

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

* U =Unsatisfactory; L = Limited Knowledge; S = Satisfactory; and A = Advanced.
| | | U/L+S+A | | U+L | ∠/S+A | U+L+S/A | |
|-------------|-------|----------|----------|----------|----------|----------|----------|
| | | False | False | False | False | False | False |
| Subject | Grade | Positive | Negative | Positive | Negative | Positive | Negative |
| | 3 | 0.02 | 0.03 | 0.04 | 0.05 | 0.05 | 0.04 |
| | 4 | 0.02 | 0.04 | 0.04 | 0.05 | 0.06 | 0.04 |
| Mathamatica | 5 | 0.03 | 0.05 | 0.05 | 0.06 | 0.06 | 0.04 |
| Mathematics | 6 | 0.01 | 0.04 | 0.06 | 0.09 | 0.06 | 0.03 |
| | 7 | 0.02 | 0.05 | 0.08 | 0.08 | 0.06 | 0.02 |
| | 8 | 0.02 | 0.04 | 0.06 | 0.07 | 0.05 | 0.03 |
| | 3 | 0.03 | 0.05 | 0.05 | 0.06 | 0.06 | 0.04 |
| | 4 | 0.02 | 0.03 | 0.04 | 0.06 | 0.07 | 0.06 |
| Deedlar | 5 | 0.01 | 0.03 | 0.05 | 0.06 | 0.06 | 0.04 |
| Reading | 6 | 0.01 | 0.02 | 0.06 | 0.08 | 0.08 | 0.07 |
| | 7 | 0.01 | 0.03 | 0.06 | 0.07 | 0.06 | 0.03 |
| | 8 | 0.02 | 0.04 | 0.05 | 0.07 | 0.05 | 0.03 |
| Saianaa | 5 | 0.02 | 0.04 | 0.05 | 0.06 | 0.06 | 0.03 |
| Science | 8 | 0.01 | 0.02 | 0.02 | 0.04 | 0.05 | 0.03 |

Table 5.2.a. Accuracy and Consistency Estimates by Cut Score, False Positive and False Negative Rates (continued)

* U =Unsatisfactory; L = Limited Knowledge; S = Satisfactory; and A = Advanced.

Section 6

Summary Statistics

6.1 Descriptive Statistics

Descriptive statistics were calculated on the student data in the final score reporting.

Subject	Crada	Total						
Subject	Grade	Ν	Mean	SD	Med.			
	3	2679	261.8	22.5	260			
	4	3076	262.5	20.4	261			
Mathamatica	5	3615	257.6	17.9	255			
Mathematics	6	3447	257.3	13.8	256			
	7	3588	249.5	13.8	247			
	8	3639	254.3	15.4	253			
	3	3221	255.9	19.6	250			
	4	3647	261.6	18.5	260			
Dooding	5	4053	257.1	18.3	255			
Reading	6	3626	257.2	16.3	254			
	7	3576	254.0	17.1	253			
	8	3528	258.3	17.6	258			
Science	5	3057	261.0	17.9	260			
	8	2535	270.2	18.0	269			

 Table 6.1. Scale Score Descriptive Statistics—Overall

*SD = standard deviation; Med. = median

		Female			Male				
Subject	Grade	Ν	Mean	SD	Med.	Ν	Mean	SD	Med.
	3	947	259.4	20.3	258	1725	263.1	23.5	260
	4	1154	261.1	19.7	261	1917	263.3	20.8	263
Mathamatica	5	1435	256.6	17.3	255	2173	258.3	18.3	255
Mathematics	6	1264	258.2	13.7	256	2175	256.8	13.8	256
	7	1315	249.1	13.9	247	2263	249.8	13.8	247
	8	1392	254.7	15.4	253	2231	254.1	15.4	253
	3	1106	256.9	19.3	252	2106	255.3	19.8	250
	4	1289	262.8	18.3	260	2354	261.0	18.6	258
Deading	5	1534	258.8	17.7	257	2513	256.1	18.6	253
Reading	6	1273	259.0	16.6	256	2344	256.2	16.1	254
	7	1253	256.6	17.2	255	2312	252.5	16.8	251
	8	1276	260.5	17.7	260	2239	257.1	17.4	256
Saianaa	5	1163	260.5	16.9	260	1890	261.4	18.4	260
Science	8	909	268.8	17.1	267	1620	271.0	18.4	271

Table 6.2. Scale Score Descriptive Statistics by Gender

*SD = standard deviation; Med. = median

Table 6.3. Scale Score Descriptive Statistics by Race/Ethnicity

		African American		Native American					
Subject	Grade	Ν	Mean	SD	Med.	Ν	Mean	SD	Med.
	3	399	254.3	19.6	252	470	260.7	21.4	258
	4	419	255.7	18.1	254	517	263.4	20.3	263
Mathamatica	5	528	253.7	16.1	250	606	257.3	18.3	255
Mathematics	6	473	252.8	11.9	252	649	257.1	13.9	256
	7	503	245.2	11.5	245	664	249.9	13.9	247
	8	501	250.5	14.0	251	710	254.3	16.3	255
	3	451	251.7	17.8	246	561	255.4	18.8	252
	4	477	256.8	16.6	254	611	261.5	18.1	260
Deading	5	581	253.3	17.2	251	684	257.3	18.6	255
Reading	6	499	253.3	14.4	252	656	256.7	15.5	254
	7	469	249.4	15.5	246	673	254.0	16.6	253
	8	489	252.6	15.9	252	687	260.1	18.2	260
Science	5	426	256.7	16.9	255	520	262.0	17.3	262
Science	8	344	261.2	14.5	261	500	272.0	17.9	271

*SD = standard deviation; Med. = median

		Hispanic			Asian				
Subject	Grade	Ν	Mean	SD	Med.	Ν	Mean	SD	Med.
	3	402	261.4	23.4	258	13	262.0	21.8	260
	4	449	262.3	19.8	261	14	273.2	21.2	276.5
Mathematics	5	562	258.1	17.8	255	11	261.4	25.5	255
	6	475	256.5	13.0	256	19	257.2	16.7	254
	7	467	248.6	13.4	247	11	246.6	13.9	243
	8	413	252.8	14.2	251	11	264.7	11.4	265
	3	495	254.1	18.2	250	16	260.6	22.8	250
	4	536	258.8	16.6	256	18	262.8	17.5	263
Deading	5	654	255.9	17.6	253	14	259.4	14.9	263.5
Reading	6	529	254.8	14.7	254	22	258.0	18.1	254
	7	501	251.2	15.1	249	10	255.6	16.0	252
	8	452	256.3	17.1	254	10	255.6	18.7	258
Science	5	488	257.9	16.5	255	9	258.8	25.6	249
Science	8	303	268.1	16.4	267	6	251.0	9.8	250

Table 6.3.a. Scale Score Descriptive Statistics by Race/Ethnicity (continued)

*SD = standard deviation; Med. = median

 Table 6.3.b.
 Scale Score Descriptive Statistics by Race/Ethnicity (continued)

		White			Other				
Subject	Grade	Ν	Mean	SD	Med.	Ν	Mean	SD	Med.
	3	1212	264.9	23.1	265	179	260.7	21.4	258
	4	1464	263.9	20.9	263	206	263.6	20.9	263
Mathamatica	5	1675	258.9	18.2	257	229	256.6	17.0	252
Mathematics	6	1586	259.1	14.2	258	241	256.6	13.3	256
	7	1712	250.8	14.4	249	225	250.7	13.1	249
	8	1785	255.6	15.3	255	214	255.1	16.7	254
	3	1479	258.2	20.5	254	214	253.6	19.6	248
	4	1751	263.7	19.3	262	247	263.4	19.8	262
Deading	5	1869	258.8	18.7	257	246	256.0	18.4	252
Reading	6	1651	259.4	17.2	258	264	256.9	17.1	254
	7	1688	255.9	17.8	255	230	256.4	17.7	254
	8	1697	259.7	17.6	260	187	258.9	17.4	258
Saianaa	5	1418	263.1	18.5	262	191	261.6	17.4	260
Science	8	1245	272.4	18.3	271	134	271.8	19.0	270

*SD = standard deviation; Med. = median

		Free/	Free/Reduced Lunch = Yes		Free/Reduced Lunch = No				
Subject	Grade	Ν	Mean	SD	Med.	Ν	Mean	SD	Med.
	3	2089	260.6	22.0	258	590	266.0	23.6	265
	4	2371	261.8	20.4	261	705	264.8	20.4	265
Mathamatica	5	2803	256.9	17.4	255	812	260.2	19.2	257
Mathematics	6	2584	256.4	13.6	256	863	260.0	13.9	258
	7	2651	248.8	13.6	247	937	251.6	14.4	251
	8	2603	253.4	14.9	253	1036	256.8	16.4	257
	3	2479	254.3	18.7	250	742	260.9	21.4	258
	4	2788	260.3	17.9	258	859	266.1	19.8	264
Deading	5	3153	256.0	17.9	253	900	261.0	19.3	259
Reading	6	2748	256.0	16.1	254	878	260.9	16.6	258
	7	2660	252.7	16.5	251	916	258.0	18.2	257
	8	2610	257.3	17.6	256	918	261.1	17.2	260
Saianaa	5	2380	260.4	17.4	260	677	263.3	19.2	262
Science	8	1856	269.3	17.8	269	679	272.7	18.4	273

Table 6.4. Scale Score Descriptive Statistics by Free/Reduced Lunch Status

*SD = standard deviation; Med. = median

6.2 Performance Level Distribution

The distributions of students in the four performance levels are presented in Table 6.5. (Please see Appendix B for distributions by scale score).

Table 0.5. Tereentage of Students by Terrormanee Lever								
			Limited					
Subject	Grade	Ν	Unsatisfactory	Knowledge	Satisfactory	Advanced		
	3	2679	7.2%	26.7%	41.8%	24.1%		
	4	3076	11.3%	16.9%	48.6%	23.0%		
Mathematics	5	3615	14.5%	22.1%	41.1%	22.1%		
	6	3447	4.9%	22.5%	57.8%	14.6%		
	7	3588	7.0%	50.8%	28.8%	13.2%		
	8	3639	8.8%	32.0%	43.7%	15.2%		
	3	3221	17.4%	28.1%	28.1%	26.1%		
	4	3647	6.4%	21.0%	34.5%	37.8%		
Deading	5	4053	4.2%	35.3%	35.8%	24.5%		
Reading	6	3626	1.3%	31.9%	31.3%	35.3%		
	7	3576	4.1%	39.7%	37.9%	18.1%		
	8	3528	8.1%	22.9%	50.9%	17.8%		
Science	5	3057	6.6%	21.7%	54.7%	16.8%		
Science	8	2535	2.4%	9.4%	70.8%	17.2%		

 Table 6.5. Percentage of Students by Performance Level

6.3 Conditional Standard Error of Measurement

The Rasch model standard error (SE) for ability estimate ($\hat{\beta}$) is as follows (Andrich & Luo, 2004):

$$\sigma_{\hat{\beta}=} \frac{1}{\sqrt{\sum_{i=1}^{L} p_{vi}(1-p_{vi})}} , \qquad (4)$$

where

v = subscript for a person,

i = subscript for an item,

L =length of the test,

 $\hat{\beta}$ = ability estimate, and

 p_{vi} = the probability that a person answers an item correctly and defined as follows:

$$p_{\nu i} = \frac{e^{\beta_{\nu} - \delta_i}}{1 + e^{\beta_{\nu} - \delta_i}} , \qquad (5)$$

where β_{v} is person's ability and δ_{i} is item's difficulty.

A confidence band can be found for use in interpreting the ability estimate. For example, an approximate 68% confidence interval for $\hat{\beta}$ is given by $\hat{\beta} \pm SE$. Because different ability estimates ($\hat{\beta}$) have different SE, Rasch SE is generally referred to as the conditional standard error of measurement (CSEM) to differentiate from the standard error of measurement of the classical measurement model. The *CSEMs* by subject are reported in Tables 4.3 through 4.7.

6.4 Standard Error of Measurement

From the classical measurement theory aspect, the observed score (raw score) has two components; true score and error. A student's true score is the hypothetical average score that would result if the student took the test repeatedly under similar conditions. The error is the difference between true score and observed score. Among the three scores, only the observed score is known; the true score and error are derived from theory.

The standard error of measurement (*SEM*), as an overall test-level measure of error, is the average of all errors associated with student scores. Instead of using errors of student scores, the classical SEM is derived using test reliability:

$$SEM = SD\sqrt{(1-r)} \tag{6}$$

where,

SEM = test Standard Error of Measurement of classical theory,

SD = standard deviation of raw score, and

r = test reliability, Cronbach's Alpha in this case.

The equation indicates that test reliability and *SEM* are in reverse relation; while test reliability increases, the SEM decreases. Table 6.6 presents the overall estimates of *SEM* for each of the content areas.

Subject	Grade	SEM
	3	2.83
	4	2.88
	5	2.98
Mathematics	6	2.98
	7	3.04
	8	3.03
	3	3.02
	4	3.03
Deeller	5	3.00
Reading	6	3.07
	7	3.06
	8	3.01
а :	5	2.98
Science	8	2.93

Table 6.6. Overall Estimates of SEM by Subject and Grade

*SEM = Standard Error of Measurement

References

- American Educational Research Association (AERA), American Psychological Association (APA), & the National Council on Measurement in Education (NCME) (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Andrich, A., & Luo, G. (2004). Modern measurement and analysis in social science. Murdoch University, Perth, Western Australia.
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum.
- Holland, P. W., & Thayer, D.T. (1988). *Differential item performance and the Mantel-Haenszel* procedure. (ETS RR-86-31). Princeton, NJ: Educational Testing Service.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: methods and practices (2nd ed.)*. New York: Springer.
- Kraemer, H. C. (1982). Kappa coefficient. Encyclopedia of Statistical Sciences. Wiley.
- Linacre, J.M. (2009). *Winsteps*® (*Version 3.61*) [Computer Software]. Beaverton, Oregon: Winsteps.com.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, *32*, 179–197.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.
- 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.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

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

Ideal	
Number of	Actual
Items for	Number of
Alignment to	Items on
OKC^3	2013 Test
6–7	
1–3	
1–3	
1–3	2
15-16	15
7–8	8
7–8	7
6–7	6
1–3	1
1–3	1
1–3	4
7–8	8
2–4	3
1–3	3
1–3	2
6–7	7
2–4	4
2–4	3
40–43	43
	Ideal Number of Items for Alignment to <i>OKC</i> ³ 6–7 1–3 1–3 1–3 1–3 7–8 6–7 1–3 1–3 1–3 1–3 1–3 1–3 1–3 1–3

OMAAP Test Blueprint and Actual Item Counts: Grade 3 Mathematics

OKC ³ Standard and Objective	Ideal Number of Items for Alignment to <i>OKC</i> ³	Actual Number of Items on 2013 Test
Algebraic Reasoning: Patterns and Relationships	6–7	7
Algebra Patterns (1.1)	1–3	3
Equations (1.2)	1–3	2
Number Properties (1.3)	1–3	2
Number Sense and Operation	14–15	15
Number Sense (2.1)	6–7	8
Number Operations (2.2)	7–8	7
Geometry	7–8	8
Lines (3.1)	1–2	2
Angles (3.2)	1–2	3
Polygons (3.3)	1–2	1
Transformations (3.4)	1–2	2
Measurement	7–8	6
Measurement (4.1)	2–4	3
Time and Temperature (4.2)	1–3	1
Money (4.3)	1–3	2
Data Analysis	6–7	7
Data Analysis (5.1)	1–3	5
Probability (5.2)	1–3	1
Central Tendency (5.3)	1–3	1
Total Test	40–43	43

OMAAP Test Blueprint and Actual Item Counts: Grade 4 Mathematics

OKC ³ Standard and Objective	Ideal Number of Items for Alignment to <i>OKC³</i>	Actual Number of Items on 2013 Test
Algebraic Reasoning: Patterns and Relationships	10–11	10
Algebra Patterns (1.1)	3–5	5
Equations (1.2)	2–4	3
Number Properties (1.3)	2–4	2
Number Sense and Operation	12–13	13
Number Sense (2.1)	5–7	7
Number Operations (2.2)	5–7	6
Geometry	6–7	6
Circles and Polygons (3.1)	3–4	2
Angles (3.2)	2–4	4
Measurement	6–7	7
Measurement (4.1)	3–4	5
Money (4.2)	2–3	2
Data Analysis	6–7	7
Data Analysis (5.1)	1–3	2
Probability (5.2)	1–3	5
Central Tendency (5.3)	1–3	0
Total Test	40–43	43

offin in rost Dideprint and rietaal from County. Grade 5 maintennatios
--

	Ideal	
	Number of	Actual
	Items for	Number of
	Alignment to	Items on
OKC ³ Standard and Objective	OKC^3	2013 Test
Algebraic Reasoning: Patterns and Relationships	10–11	11
Algebra Patterns (1.1)	2–3	3
Expressions and Equations (1.2)	2–3	3
Number Properties (1.3)	2–3	3
Solving Equations (1.4)	2–3	2
Number Sense and Operation	12–13	13
Number Sense (2.1)	3–5	4
Number Operations (2.2)	7–9	9
Geometry	6–7	6
Three Dimensional Figures (3.1)	1–3	1
Congruent and Similar Figures (3.2)	1–3	3
Coordinate Geometry (3.3)	1–3	2
Measurement	6–7	4
Circles (4.1)	3–4	1
Conversions (4.2)	2–3	3
Data Analysis	6–7	7
Data Analysis (5.1)	1–3	3
Probability (5.2)	1–3	2
Central Tendency (5.3)	1–3	2
Total Test	40–43	41

OMAAP Test Blueprint and Actual Item Counts: Grade 6 Mathematics

*Two operational items aligned to objective 4.1 was suppressed.

	Ideal	
	Number of	Actual
	Items for	Number of
	Alignment to	Items on
OKC ^o Standard and Objective	OKC	2013 Test
Algebraic Reasoning: Patterns and Relationships	12–13	13
Linear Relationships (1.1)	3–5	2
Solving Equations (1.2)	3–5	6
Solving and Graphing Inequalities (1.3)	3–5	5
Number Sense and Operation	8–9	9
Number Sense (2.1)	4–5	3
Number Operations (2.2)	4–5	6
Geometry	6–7	8
Classifying Figures (3.1)	2–3	1
Lines and Angles (3.2)	2–3	3
Transformations (3.3)	2–3	4
Measurement	7–8	7
Perimeter and Area (4.1)	3–4	4
Circles (4.2)	1–3	1
Composite Figures (4.3)	1–3	2
Data Analysis	6–7	6
Data Analysis (5.1)	1–3	0
Probability (5.2)	1–3	3
Central Tendency (5.3)	1–3	3
Total Test	40–43	43

OMAAP Test Blueprint and Actual Item Counts: Grade 7 Mathematics

<i>OKC</i> ³ Standard and Objective	Ideal Number of Items for Alignment to <i>OKC</i> ³	Actual Number of Items on 2013 Test
Algebraic Reasoning: Patterns and Relationships	13–14	13
Equations (1.1)	8–9	8
Inequalities (1.2)	4–5	5
Number Sense and Operation	8–9	9
Number Sense (2.1)	2–3	3
Number Operations (2.2)	5–6	6
Geometry	7–8	8
Three Dimensional Figures (3.1)	4–5	4
Pythagorean Theorem (3.2)	2–3	4
Measurement	6–7	6
Surface Area and Volume (4.1)	1–3	3
Ratio and Proportions (4.2)	1–3	2
Composite Figures (4.3)	1–3	1
Data Analysis	6–7	7
Data Analysis (5.1)	2–4	3
Central Tendency (5.3)	2–4	4
Total Test	40–43	43

OKC ³ Standard and Objective	Ideal Number of Items for Alignment to <i>OKC</i> ³	Actual Number of Items on 2013 Test
Vocabulary	9–11	10
Words in Context (2.1)	2–4	3
Affixes, Roots, and Stems (2.2)	2–4	2
Synonyms, Antonyms, and Homonyms (2.3)	2–4	4
Using Resource Materials (2.4)	1–3	1
Comprehension/Critical Literacy	18–24	20
Literal Understanding (4.1)	4–6	5
Inferences and Interpretation (4.2)	4–6	6
Summary and Generalization (4.3)	4–6	5
Analysis and Evaluation (4.4)	2–4	4
Literature	6–7	6
Literary Elements (5.2) & Figurative Language/Sound Devices (5.3)	6–7	6
Research and Information	6–7	7
Accessing Information (6.1)	6–7	7
Total Test	40–43	43

OMAAP Test Blueprint and Actual Item Counts: Grade 3 Reading

OKC ³ Standard and Objective	Ideal Number of Items for Alignment to <i>OKC</i> ³	Actual Number of Items on 2013 Test
Vocabulary	9–11	9
Words in Context (1.1)	2–4	4
Affixes, Roots, and Stems (1.2)	2–4	0
Synonyms, Antonyms, Homonyms/Homophones (1.3)	2–4	5
Comprehension/Critical Literacy	17–19	18
Literal Understanding (3.1)	3–5	4
Inferences and Interpretation (3.2)	3–5	6
Summary and Generalization (3.3)	3–5	4
Analysis and Evaluation (3.4)	3–5	4
Literature	6–8	9
Literary Elements (4.2)	2–4	8
Figurative Language/Sound Devices (4.3)	2–4	1
Research and Information	6–7	7
Accessing Information (5.1)	6–7	7
Total Test	40-43	43

OMAAP Test Blueprint and Actual Item Counts: Grade 4 Reading

<i>OKC</i> ³ Standard and Objective	Ideal Number of Items for Alignment to <i>OKC</i> ³	Actual Number of Items on 2013 Test
Vocabulary	9–11	11
Words in Context (1.1)	2–4	4
Affixes, Roots, and Stems (1.2)	2–4	5
Synonyms, Antonyms, and Homonyms/Homophones (1.3)	2–4	2
Comprehension/Critical Literacy	15–17	17
Literal Understanding (3.1)	3–5	3
Inferences and Interpretation (3.2)	3–5	3
Summary and Generalization (3.3)	3–5	7
Analysis and Evaluation (3.4)	3–5	4
Literature	9–11	9
Literary Genre (4.1)	2–4	3
Literary Elements (4.2)	2–4	3
Figurative Language/Sound Devices (4.3)	2–4	3
Research and Information	6–7	6
Accessing Information (5.1)	2-4	3
Interpreting Information (5.2)	2-4	3
Total Test	40–43	43

OMAAP Test Blueprint and Actual Item Counts: Grade 5 Reading

OKC ³ Standard and Objective	Ideal Number of Items for Alignment to <i>OKC</i> ³	Actual Number of Items on 2013 Test
Vocabulary	6–7	7
Words in Context (1.1)	4–5	3
Word Origins (1.2)	2–3	4
Comprehension/Critical Literacy	15–17	20
Literal Understanding (3.1)	4–5	7
Inferences and Interpretation (3.2)	3–4	4
Summary and Generalization (3.3)	3–4	5
Analysis and Evaluation (3.4)	3–4	4
Literature	10–12	11
Literary Genres (4.1)	3	2
Literary Elements (4.2)	3–4	7
Figurative Language/Sound Devices (4.3)	3–4	2
Research and Information	6–7	5
Accessing Information (5.1)	3–5	3
Interpreting Information (5.2)	2–4	2
Total Test	40-43	43

OMAAP Test Blueprint and Actual Item Counts: Grade 6 Reading

OKC ³ Standard and Objective	Ideal Number of Items for Alignment to <i>OKC</i> ³	Actual Number of Items on 2013 Test
Vocabulary	6–8	9
Words in Context (1.1)	2–3	5
Word Origins (1.2)	1–2	2
Idioms and Comparisons (1.3)	2–3	2
Comprehension/Critical Literacy	15–17	18
Literal Understanding (3.1)	3–4	4
Inferences and Interpretation (3.2)	4–6	5
Summary and Generalization (3.3)	4–6	5
Analysis and Evaluation (3.4)	3–4	4
Literature	9–11	7
Literary Genres (4.1)	3–4	4
Literary Elements (4.2)	3–4	2
Figurative Language and Sound Devices (4.3)	2–3	1
Research and Information	6–7	9
Accessing Information (5.1)	3–5	5
Interpreting Information (5.2)	2–4	4
Total Test	40-43	43

OMAAP Test Blueprint and Actual Item Counts: Grade 7 Reading

OKC³ Standard and Objective	Ideal Number of Items for Alignment to <i>OKC</i> ³	Actual Number of Items on 2013 Test
Vocabulary	6–7	12
Words in Context (1.1)	2–3	6
Word Origins (1.2)	0–1	0
Idioms and Comparisons (1.3)	2–3	6
Comprehension/Critical Literacy	16-18	17
Literal Understanding (3.1)	3–4	5
Inferences and Interpretation (3.2)	4–5	3
Summary and Generalization (3.3)	4–5	2
Analysis and Evaluation (3.4)	4–5	7
Literature	11–13	8
Literary Genre (4.1)	3–4	3
Literary Elements (4.2)	5–6	4
Figurative Language and Sound Devices (4.3)	3–4	1
Research and Information	6–7	6
Accessing Information (5.1)	3–4	4
Interpreting Information (5.2)	2–4	2
Total Test	40-43	43

OMAAP Test Blueprint and Actual Item Counts: Grade 8 Readin

OKC^3 Standard and Objective	Ideal Number of Items for Alignment	Actual Number of Items on 2013 Test	Number of Items Field- Tested in 2013
Process Standards	to oke	2013 1050	2013
Observe and Measure	8–10	9	1
SI Metric (P1.1)	3–5	5	0
Similar/different characteristics (P1.2)	3–5	4	1
Classify	8–10	10	3
Observable properties (P2.1)	3–5	5	1
Serial order (P2.2)	3–5	5	2
Experiment	9–11	10	1
Experimental design (P3.2)	5–7	6	1
Hazards/practice safety (P3.4)	3–5	4	0
Interpret and Communicate	12–14	14	2
Data tables/line/bar/trend and circle graphs (P4.2)	4–6	4	0
Prediction based on data (P4.3)	3–5	4	1
Explanations based on data (P4.4)	3–5	6	1
Total Test	40–43	43	7
Content Standards			
Properties of Matter and Energy	15–17	17	3
Matter has physical properties (C1.1)	4–6	5	0
Physical properties can be measured (C1.2)	4–6	6	0
Energy can be transferred (C1.3)	4–6	6	0
Potential/Kinetic energy (C1.4)	0	0	3
Organisms and Environments	10–12	11	1
Dependence upon community (C2.1)	4–6	5	1
Individual organism and species survival (C2.2)	4–6	6	0
Structures of the Earth and the Solar System	9–11	11	3
Properties of soils (C3.1)	0	0	3
Weather patterns (C3.2)	4–6	6	0
Earth as a planet (C3.3)	4–6	5	0
Total Test	37–40	39	7

OMAAP Test Blueprint and Actual Item Counts: Grade 5 Science

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

Ideal Actual Number of Number of Number of **Items Field-Items for** Items on Tested in Alignment 2013 Test 2013 OKC³Standard and Objective to OKC^3 **Process Standards Observe and Measure** 6-8 8 2 Qualitative/quantitative observations/changes (P1.1) 3–5 3 0 SI (metrics) units/appropriate tools (P1.2 and P1.3) 3–5 5 1 9 Classify 6-8 2 Classification system (P2.1) 3-5 4 2 Properties ordered (P2.2) 3-5 5 0 13-15 14 2 Experiment Experimental design (P3.2) 4-6 5 1 5 0 Identify variables (P3.3) 4-6 3–5 4 Hazards/practice safety (P3.6) 1 **Interpret and Communicate** 11-13 12 2 Data tables/line/bar/trend and circle graphs (P4.2) 6–8 7 1 Explanations/prediction (P4.3) 4–6 5 1 7 43 **Total Test** 40-43 **Content Standards Properties and Chemical Changes in Matter** 6 1 6-8 4 0 Chemical reactions (C1.1) 2-42 - 42 Conservation of matter (C1.2) 1 **Motion and Forces** 6-8 8 0 Motion of an object (C2.1) 2 - 45 0 3 Object subjected to a force (C2.2) 2-40 7_9 9 **Diversity and Adaptations of Organisms** 0 Classification (C3.1) 2-46 0 2-43 Internal and external structures (C3.2) 0 8 Structures/Forces of the Earth/Solar System 6-8 1 Landforms result from constructive and destructive forces 2 - 44 0 (C4.1) Rock cycle (C4.2) 2 - 44 0 Global weather patterns (C4.3) 0 0 1 **Earth's History** 6-8 8 2 Catastrophic events (C5.1) 2-43 2 2-4Fossil evidence (C5.2) 5 0 **Total Test** 36-39 39 4

OMAAP Test Blueprint and Actual Item Counts: Grade 8 Science

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

Appendix B: Scale Score Distributions

Tables below show the Spring 2013 operational form score distribution. These analyses are based on the final student data file that is used for reporting.

Raw	Scale	Ene cur en er	Doncomt	Cumulative	Cumulative
Score	Score	Frequency	Percent	Frequency	Percent
4	202	1	0.04	1	0.04
5	207	1	0.04	2	0.07
6	212	2	0.07	4	0.15
7	216	8	0.30	12	0.45
8	219	13	0.49	25	0.93
9	223	25	0.93	50	1.87
10	226	28	1.05	78	2.91
11	228	47	1.75	125	4.67
12	231	69	2.58	194	7.24
13	234	72	2.69	266	9.93
14	236	90	3.36	356	13.29
15	238	92	3.43	448	16.72
16	241	81	3.02	529	19.75
17	243	106	3.96	635	23.70
18	245	87	3.25	722	26.95
19	247	88	3.28	810	30.24
20	249	101	3.77	911	34.01
21	252	98	3.66	1009	37.66
22	254	109	4.07	1118	41.73
23	256	96	3.58	1214	45.32
24	258	107	3.99	1321	49.31
25	260	114	4.26	1435	53.56
26	262	81	3.02	1516	56.59
27	265	99	3.70	1615	60.28
28	267	92	3.43	1707	63.72
29	269	96	3.58	1803	67.30
30	272	113	4.22	1916	71.52
31	274	115	4.29	2031	75.81
32	277	75	2.80	2106	78.61
33	280	73	2.72	2179	81.34
34	283	84	3.14	2263	84.47
35	286	85	3.17	2348	87.64
36	290	62	2.31	2410	89.96
37	294	73	2.72	2483	92.68
38	298	46	1.72	2529	94.40
39	303	58	2.16	2587	96.57
40	310	37	1.38	2624	97.95
41	319	27	1.01	2651	98.95
42	333	20	0.75	2671	99.70
43	350	8	0.30	2679	100.00

MATH Grade 03 Scale Score Distribution



MATH Grade 03 Scale Score Distribution

Raw	Scale	Б	D (Cumulative	Cumulative
Score	Score	F requency	rercent	Frequency	Percent
5	207	1	0.03	1	0.03
6	211	2	0.07	3	0.10
7	215	3	0.10	6	0.20
8	219	8	0.26	14	0.46
9	222	14	0.46	28	0.91
10	225	24	0.78	52	1.69
11	228	20	0.65	72	2.34
12	230	41	1.33	113	3.67
13	233	65	2.11	178	5.79
14	235	72	2.34	250	8.13
15	237	100	3.25	350	11.38
16	240	91	2.96	441	14.34
17	242	104	3.38	545	17.72
18	244	98	3.19	643	20.90
19	246	119	3.87	762	24.77
20	248	109	3.54	871	28.32
21	250	94	3.06	965	31.37
22	252	115	3.74	1080	35.11
23	254	89	2.89	1169	38.00
24	256	134	4.36	1303	42.36
25	258	140	4.55	1443	46.91
26	261	137	4.45	1580	51.37
27	263	132	4.29	1712	55.66
28	265	135	4.39	1847	60.05
29	267	142	4.62	1989	64.66
30	270	134	4.36	2123	69.02
31	272	116	3.77	2239	72.79
32	275	129	4.19	2368	76.98
33	278	109	3.54	2477	80.53
34	280	90	2.93	2567	83.45
35	284	96	3.12	2663	86.57
36	287	93	3.02	2756	89.60
37	291	89	2.89	2845	92.49
38	295	70	2.28	2915	94.77
39	301	53	1.72	2968	96.49
40	307	57	1.85	3025	98.34
41	316	25	0.81	3050	99.15
42	331	25	0.81	3075	99.97
43	350	1	0.03	3076	100.00

MATH Grade 04 Scale Score Distribution



MATH Grade 04 Scale Score Distribution

Raw Scale				Cumulative	Cumulative
Score	Score	Frequency	Percent	Frequency	Percent
4	208	1	0.03	1	0.03
5	213	2	0.06	3	0.08
6	217	6	0.17	9	0.25
7	221	18	0.50	27	0.75
8	225	18	0.50	45	1.24
9	228	49	1.36	94	2.60
10	231	60	1.66	154	4.26
11	234	86	2.38	240	6.64
12	236	124	3.43	364	10.07
13	239	163	4.51	527	14.58
14	241	173	4.79	700	19.36
15	244	231	6.39	931	25.75
16	246	203	5.62	1134	31.37
17	248	194	5.37	1328	36.74
18	250	203	5.62	1531	42.35
19	252	191	5.28	1722	47.63
20	255	144	3.98	1866	51.62
21	257	156	4.32	2022	55.93
22	259	149	4.12	2171	60.06
23	261	143	3.96	2314	64.01
24	263	127	3.51	2441	67.52
25	265	140	3.87	2581	71.40
26	267	116	3.21	2697	74.61
27	269	117	3.24	2814	77.84
28	271	116	3.21	2930	81.05
29	274	100	2.77	3030	83.82
30	276	89	2.46	3119	86.28
31	279	79	2.19	3198	88.46
32	281	66	1.83	3264	90.29
33	284	62	1.72	3326	92.01
34	287	70	1.94	3396	93.94
35	290	62	1.72	3458	95.66
36	293	45	1.24	3503	96.90
37	297	28	0.77	3531	97.68
38	302	31	0.86	3562	98.53
39	307	20	0.55	3582	99.09
40	313	19	0.53	3601	99.61
41	322	9	0.25	3610	99.86
42	337	5	0.14	3615	100.00

MATH Grade 05 Scale Score Distribution



MATH Grade 05 Scale Score Distribution

Raw	Scale	_		Cumulative	Cumulative
Score	Score	Frequency	Percent	Frequency	Percent
1	181	1	0.03	1	0.03
5	216	2	0.06	3	0.09
6	220	4	0.12	7	0.20
7	224	7	0.20	14	0.41
8	227	12	0.35	26	0.75
9	230	22	0.64	48	1.39
10	233	40	1.16	88	2.55
11	236	81	2.35	169	4.90
12	238	100	2.90	269	7.80
13	241	127	3.68	396	11.49
14	243	162	4.70	558	16.19
15	245	189	5.48	747	21.67
16	248	200	5.80	947	27.47
17	250	223	6.47	1170	33.94
18	252	211	6.12	1381	40.06
19	254	234	6.79	1615	46.85
20	256	196	5.69	1811	52.54
21	258	223	6.47	2034	59.01
22	260	199	5.77	2233	64.78
23	262	179	5.19	2412	69.97
24	264	158	4.58	2570	74.56
25	266	140	4.06	2710	78.62
26	268	116	3.37	2826	81.98
27	271	116	3.37	2942	85.35
28	273	104	3.02	3046	88.37
29	275	96	2.79	3142	91.15
30	278	69	2.00	3211	93.15
31	280	70	2.03	3281	95.18
32	283	45	1.31	3326	96.49
33	286	40	1.16	3366	97.65
34	289	24	0.70	3390	98.35
35	293	22	0.64	3412	98.98
36	297	15	0.44	3427	99.42
37	302	10	0.29	3437	99.71
38	309	5	0.15	3442	99.85
39	317	3	0.09	3445	99.94
40	331	2	0.06	3447	100.00

MATH Grade 06 Scale Score Distribution

Copyright © 2013 by Oklahoma State Department of Education.



MATH Grade 06 Scale Score Distribution

Raw	Scale	-		Cumulative	Cumulative
Score	Score	Frequency	Percent	Frequency	Percent
2	187	1	0.03	1	0.03
6	212	2	0.06	3	0.08
7	216	8	0.22	11	0.31
8	220	11	0.31	22	0.61
9	223	23	0.64	45	1.25
10	226	33	0.92	78	2.17
11	229	75	2.09	153	4.26
12	231	100	2.79	253	7.05
13	234	145	4.04	398	11.09
14	236	208	5.80	606	16.89
15	238	254	7.08	860	23.97
16	241	235	6.55	1095	30.52
17	243	271	7.55	1366	38.07
18	245	244	6.80	1610	44.87
19	247	236	6.58	1846	51.45
20	249	231	6.44	2077	57.89
21	251	221	6.16	2298	64.05
22	253	198	5.52	2496	69.57
23	255	153	4.26	2649	73.83
24	258	143	3.99	2792	77.81
25	260	106	2.95	2898	80.77
26	262	106	2.95	3004	83.72
27	264	109	3.04	3113	86.76
28	266	99	2.76	3212	89.52
29	268	74	2.06	3286	91.58
30	271	68	1.90	3354	93.48
31	273	51	1.42	3405	94.90
32	276	36	1.00	3441	95.90
33	279	37	1.03	3478	96.93
34	282	29	0.81	3507	97.74
35	285	27	0.75	3534	98.49
36	288	22	0.61	3556	99.11
37	292	12	0.33	3568	99.44
38	296	12	0.33	3580	99.78
39	302	2	0.06	3582	99.83
40	308	3	0.08	3585	99.92
42	332	3	0.08	3588	100.00

MATH Grade 07 Scale Score Distribution

Copyright © 2013 by Oklahoma State Department of Education.



MATH Grade 07 Scale Score Distribution

Raw	Scale	F	D	Cumulative	Cumulative
Score	Score	Frequency	Percent	Frequency	Percent
4	205	1	0.03	1	0.03
5	211	2	0.05	3	0.08
6	215	3	0.08	6	0.16
7	219	10	0.27	16	0.44
8	222	15	0.41	31	0.85
9	225	32	0.88	63	1.73
10	228	45	1.24	108	2.97
11	231	90	2.47	198	5.44
12	233	125	3.44	323	8.88
13	236	121	3.33	444	12.20
14	238	155	4.26	599	16.46
15	240	160	4.40	759	20.86
16	242	171	4.70	930	25.56
17	244	187	5.14	1117	30.70
18	246	186	5.11	1303	35.81
19	249	187	5.14	1490	40.95
20	251	183	5.03	1673	45.97
21	253	187	5.14	1860	51.11
22	255	188	5.17	2048	56.28
23	257	190	5.22	2238	61.50
24	259	168	4.62	2406	66.12
25	261	153	4.20	2559	70.32
26	263	148	4.07	2707	74.39
27	265	132	3.63	2839	78.02
28	267	121	3.33	2960	81.34
29	269	123	3.38	3083	84.72
30	271	121	3.33	3204	88.05
31	274	88	2.42	3292	90.46
32	276	87	2.39	3379	92.86
33	279	70	1.92	3449	94.78
34	282	56	1.54	3505	96.32
35	285	45	1.24	3550	97.55
36	288	26	0.71	3576	98.27
37	292	22	0.60	3598	98.87
38	297	19	0.52	3617	99.40
39	302	8	0.22	3625	99.62
40	308	11	0.30	3636	99.92
41	317	2	0.05	3638	99.97
42	332	1	0.03	3639	100.00

MATH Grade 08 Scale Score Distribution



MATH Grade 08 Scale Score Distribution

Dow	Scolo		o seule s	Cumulative	Cumulativa
Naw Score	Scale	Frequency	Percent	Eroquonov	Domoont
2	101	1	0.02	requency	
2	191	1	0.03	1	0.03
3	200	1	0.03	2	0.06
4	207	1	0.03	3	0.09
5	212	2	0.06	5	0.16
6	216	3	0.09	8	0.25
7	220	13	0.40	21	0.65
8	224	21	0.65	42	1.30
9	227	38	1.18	80	2.48
10	230	72	2.24	152	4.72
11	232	110	3.42	262	8.13
12	235	142	4.41	404	12.54
13	237	159	4.94	563	17.48
14	240	198	6.15	761	23.63
15	242	179	5.56	940	29.18
16	244	214	6.64	1154	35.83
17	246	175	5.43	1329	41.26
18	248	142	4.41	1471	45.67
19	250	140	4.35	1611	50.02
20	252	116	3.60	1727	53.62
21	254	125	3.88	1852	57.50
22	256	91	2.83	1943	60.32
23	258	91	2.83	2034	63.15
24	260	96	2.98	2130	66.13
25	262	88	2.73	2218	68.86
26	264	80	2.48	2298	71.34
27	267	81	2.51	2379	73.86
28	269	76	2.36	2455	76.22
29	271	81	2.51	2536	78.73
30	273	78	2.42	2614	81.15
31	276	95	2.95	2709	84.10
32	278	68	2.11	2777	86.22
33	281	66	2.05	2843	88.26
34	284	103	3.20	2946	91.46
35	287	67	2.08	3013	93.54
36	290	65	2.02	3078	95.56
37	294	39	1.21	3117	96.77
38	299	34	1.06	3151	97.83
39	304	29	0.90	3180	98.73
40	311	19	0.59	3199	99.32
41	320	12	0.37	3211	99.69
42	335	9	0.28	3220	99.97
43	350	1	0.03	3221	100.00
	000	•	0.00	0221	100.00

READING Grade 03 Scale Score Distribution


READING Grade 03 Scale Score Distribution

Dow	Seele			Cumulativa	Cumulativa
Kaw	Scale	Frequency	Percent	Encourante	Democrat
Score	105	1	0.02	Frequency	Percent
2	195	1	0.03	1	0.03
3	204	1	0.03	2	0.05
4	210	1	0.03	3	0.08
5	216	2	0.05	5	0.14
6	220	2	0.05	1	0.19
7	224	8	0.22	15	0.41
8	227	22	0.60	37	1.01
9	231	38	1.04	75	2.06
10	234	65	1.78	140	3.84
11	236	96	2.63	236	6.47
12	239	105	2.88	341	9.35
13	241	143	3.92	484	13.27
14	244	171	4.69	655	17.96
15	246	182	4.99	837	22.95
16	248	168	4.61	1005	27.56
17	250	189	5.18	1194	32.74
18	252	180	4.94	1374	37.67
19	254	155	4.25	1529	41.92
20	256	152	4.17	1681	46.09
21	258	136	3.73	1817	49.82
22	260	151	4.14	1968	53.96
23	262	154	4.22	2122	58.18
24	264	144	3.95	2266	62.13
25	266	143	3.92	2409	66.05
26	269	120	3.29	2529	69.34
27	271	107	2.93	2636	72.28
28	273	128	3.51	2764	75.79
29	275	114	3.13	2878	78.91
30	277	111	3.04	2989	81.96
31	280	98	2.69	3087	84.64
32	282	91	2.50	3178	87.14
33	285	114	3.13	3292	90.27
34	288	81	2.22	3373	92.49
35	291	69	1.89	3442	94.38
36	295	51	1.40	3493	95.78
37	299	51	1.40	3544	97.18
38	303	28	0.77	3572	97.94
39	308	27	0.74	3599	98.68
40	315	25	0.69	3624	99.37
41	324	14	0.38	3638	99.75
42	339	8	0.22	3646	99.97
43	350	1	0.03	3647	100.00

READING Grade 04 Scale Score Distribution



READING Grade 04 Scale Score Distribution

D			o beare b		
Kaw	Scale	Frequency	Percent	Cumulative	Cumulative
Score	Score			Frequency	Percent
1	173	<u> </u>	0.02	l	0.02
3	197	1	0.02	2	0.05
4	203	2	0.05	4	0.10
5	209	2	0.05	6	0.15
6	213	3	0.07	9	0.22
7	217	6	0.15	15	0.37
8	220	15	0.37	30	0.74
9	223	31	0.76	61	1.51
10	226	28	0.69	89	2.20
11	229	83	2.05	172	4.24
12	232	89	2.20	261	6.44
13	234	132	3.26	393	9.70
14	236	130	3.21	523	12.90
15	239	146	3.60	669	16.51
16	241	184	4.54	853	21.05
17	243	182	4.49	1035	25.54
18	245	195	4.81	1230	30.35
19	247	205	5.06	1435	35.41
20	249	170	4.19	1605	39.60
21	251	191	4.71	1796	44.31
22	253	178	4.39	1974	48.70
23	255	183	4.52	2157	53.22
24	257	159	3.92	2316	57.14
25	259	148	3.65	2464	60.79
26	261	145	3.58	2609	64.37
27	264	150	3.70	2759	68.07
28	266	166	4.10	2925	72.17
29	268	133	3.28	3058	75.45
30	270	126	3.11	3184	78.56
31	273	133	3.28	3317	81.84
32	275	122	3.01	3439	84.85
33	278	108	2.66	3547	87.52
34	281	108	2.66	3655	90.18
35	284	96	2.37	3751	92.55
36	287	87	2.15	3838	94.70
37	291	71	1.75	3909	96.45
38	296	50	1.23	3959	97.68
39	301	49	1.21	4008	98.89
40	308	25	0.62	4033	99.51
41	317	17	0.42	4050	99.93
42	332	1	0.02	4051	99.95
43	350	2	0.05	4053	100.00

READING Grade 05 Scale Score Distribution



READING Grade 05 Scale Score Distribution

Darr	Caala			Cumulative	Cumulative
Kaw Score	Scale	Frequency	Percent	Encouonov	Domoont
	212	2	0.06	rrequency	
0	213	2	0.00	2	0.08
/	217	3	0.08	5	0.14
8	221	6	0.17	11	0.30
9	224	15	0.41	26	0.72
10	227	22	0.61	48	1.32
11	230	57	1.57	105	2.90
12	233	70	1.93	175	4.83
13	236	111	3.06	286	7.89
14	238	131	3.61	417	11.50
15	241	156	4.30	573	15.80
16	243	199	5.49	772	21.29
17	245	216	5.96	988	27.25
18	247	217	5.98	1205	33.23
19	250	223	6.15	1428	39.38
20	252	217	5.98	1645	45.37
21	254	187	5.16	1832	50.52
22	256	176	4.85	2008	55.38
23	258	177	4.88	2185	60.26
24	260	158	4.36	2343	64.62
25	262	154	4.25	2497	68.86
26	265	141	3.89	2638	72.75
27	267	131	3.61	2769	76.37
28	269	123	3.39	2892	79.76
29	272	101	2.79	2993	82.54
30	274	114	3.14	3107	85.69
31	277	100	2.76	3207	88.44
32	279	88	2.43	3295	90.87
33	282	83	2.29	3378	93.16
34	285	64	1.77	3442	94.93
35	289	51	1.41	3493	96.33
36	292	40	1.10	3533	97.44
37	296	43	1.19	3576	98.62
38	301	22	0.61	3598	99.23
39	306	12	0.33	3610	99.56
40	313	11	0.30	3621	99.86
41	323	2	0.06	3623	99.92
42	339	3	0.08	3626	100.00

READING Grade 06 Scale Score Distribution

Copyright © 2013 by Oklahoma State Department of Education.



READING Grade 06 Scale Score Distribution

Dow	Seele		, beare b	Cumulative	Cumulative
Score	Scale	Frequency	Percent	Frequency	Percent
5	206	3	0.08	3	0.08
6	210	2	0.06	5	0.14
7	214	7	0.20	12	0.34
8	218	18	0.50	30	0.84
9	221	25	0.70	55	1.54
10	224	30	0.84	85	2.38
11	227	65	1.82	150	4.19
12	230	85	2.38	235	6.57
13	233	99	2.77	334	9.34
14	235	148	4.14	482	13.48
15	237	178	4.98	660	18.46
16	240	191	5.34	851	23.80
17	242	184	5.15	1035	28.94
18	244	183	5.12	1218	34.06
19	246	187	5.23	1405	39.29
20	249	165	4.61	1570	43.90
21	251	177	4.95	1747	48.85
22	253	170	4.75	1917	53.61
23	255	192	5.37	2109	58.98
24	257	150	4.19	2259	63.17
25	259	166	4.64	2425	67.81
26	262	138	3.86	2563	71.67
27	264	145	4.05	2708	75.73
28	266	105	2.94	2813	78.66
29	269	113	3.16	2926	81.82
30	271	117	3.27	3043	85.10
31	274	98	2.74	3141	87.84
32	276	104	2.91	3245	90.74
33	279	86	2.40	3331	93.15
34	282	72	2.01	3403	95.16
35	286	43	1.20	3446	96.36
36	289	35	0.98	3481	97.34
37	293	38	1.06	3519	98.41
38	298	25	0.70	3544	99.11
39	304	17	0.48	3561	99.58
40	310	8	0.22	3569	99.80
41	320	6	0.17	3575	99.97
43	350	1	0.03	3576	100.00

READING Grade 07 Scale Score Distribution



READING Grade 07 Scale Score Distribution

Darr	Casla			Cumulativa	Cumulativa
Kaw Seoro	Scale	Frequency	Percent	Encourage	Domont
1	171	1	0.02	rrequency	
1	1/1	1	0.03	2	0.05
<u> </u>	200	2	0.05		0.00
5	208	2	0.00	4	0.11
0	212	3	0.09	/	0.20
7	216	4	0.11	11	0.31
8	220	9	0.26	20	0.57
9	223	17	0.48	37	1.05
10	226	44	1.25	81	2.30
11	229	40	1.13	121	3.43
12	231	67	1.90	188	5.33
13	234	101	2.86	289	8.19
14	236	117	3.32	406	11.51
15	239	133	3.77	539	15.28
16	241	121	3.43	660	18.71
17	243	138	3.91	798	22.62
18	245	143	4.05	941	26.67
19	247	157	4.45	1098	31.12
20	250	166	4.71	1264	35.83
21	252	148	4.20	1412	40.02
22	254	170	4.82	1582	44.84
23	256	162	4.59	1744	49.43
24	258	152	4.31	1896	53.74
25	260	162	4.59	2058	58.33
26	262	147	4.17	2205	62.50
27	264	173	4.90	2378	67.40
28	267	143	4.05	2521	71.46
29	269	135	3.83	2656	75.28
30	271	120	3.40	2776	78.68
31	274	121	3.43	2897	82.11
32	277	124	3.51	3021	85.63
33	279	117	3.32	3138	88.95
34	282	121	3.43	3259	92.38
35	286	82	2.32	3341	94.70
36	289	60	1.70	3401	96.40
37	293	44	1.25	3445	97.65
38	298	36	1.02	3481	98.67
39	303	26	0.74	3507	99.40
40	310	14	0.40	3521	99.80
41	319	4	0.11	3525	99.91
42	334	3	0.09	3528	100.00

READING Grade 08 Scale Score Distribution



READING Grade 08 Scale Score Distribution

Pow	Scolo			Cumulative	Cumulative
Score	Score	Frequency	Percent	Frequency	Percent
3	194	1	0.03	1	0.03
5	206	1	0.03	2	0.05
6	200	1	0.03	3	0.10
7	211	1	0.03	4	0.13
8	213	2	0.07	6	0.15
0	21)	10	0.33	16	0.52
10	225	10	0.55	33	1.08
10	223	28	0.92	61	2.00
11	220	40	1.31	101	3 30
12	230	40	1.51	144	4 71
13	235	59	1.11	203	6.64
15	233	79	2.58	203	9.22
15	230	97	2.50	379	12.40
10	240	105	3.17	184	15.83
17	242	103	3.45	484 597	19.53
10	243	115	3.70 4.45	733	23.08
20	247	130	4.43	867	23.98
20	249	142	4.50	1009	33.01
21	251	142	4.05	1145	37.46
22	255	150	5.23	1305	42.69
23	255	145	J.23	1450	42.09
25	260	145	4.74	1587	51 91
25	262	146	4.40	1733	56.69
20	262	162	5 30	1895	61.99
27	264	148	4 84	2043	66.83
20	269	131	4 29	2045	71.12
30	20)	129	4 22	2303	75 34
31	274	136	4.45	2439	79.78
32	276	103	3.37	2542	83.15
33	279	105	3.43	2647	86.59
34	282	91	2.98	2738	89.56
35	285	88	2.88	2826	92.44
36	289	59	1.93	2885	94.37
37	293	50	1.64	2935	96.01
38	297	42	1.37	2977	97.38
39	303	34	1.11	3011	98.50
40	309	27	0.88	3038	99.38
41	319	12	0.39	3050	99.77
42	334	6	0.20	3056	99.97
43	350	1	0.03	3057	100.00

SCIENCE Grade 05 Scale Score Distribution



SCIENCE Grade 05 Scale Score Distribution

	Seilli	CE Graac o	o seare s		
Raw Score	Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
8	227	2	0.08	2	0.08
9	230	9	0.36	11	0.43
10	233	6	0.24	17	0.67
11	236	14	0.55	31	1.22
12	239	31	1.22	62	2.45
13	241	50	1.97	112	4.42
14	243	39	1.54	151	5.96
15	246	67	2.64	218	8.60
16	248	83	3.27	301	11.87
17	250	75	2.96	376	14.83
18	252	104	4.10	480	18.93
19	254	77	3.04	557	21.97
20	256	111	4.38	668	26.35
21	258	107	4.22	775	30.57
22	261	105	4.14	880	34.71
23	263	86	3.39	966	38.11
24	265	110	4.34	1076	42.45
25	267	114	4.50	1190	46.94
26	269	122	4.81	1312	51.76
27	271	104	4.10	1416	55.86
28	273	115	4.54	1531	60.39
29	275	127	5.01	1658	65.40
30	278	109	4.30	1767	69.70
31	280	118	4.65	1885	74.36
32	283	113	4.46	1998	78.82
33	285	100	3.94	2098	82.76
34	288	72	2.84	2170	85.60
35	291	92	3.63	2262	89.23
36	295	76	3.00	2338	92.23
37	299	54	2.13	2392	94.36
38	303	51	2.01	2443	96.37
39	308	46	1.81	2489	98.19
40	315	25	0.99	2514	99.17
41	323	18	0.71	2532	99.88
42	338	2	0.08	2534	99.96
43	350	1	0.04	2535	100.00

SCIENCE Grade 08 Scale Score Distribution



SCIENCE Grade 08 Scale Score Distribution