

**Technical Report  
Spring 2013 Test Administration**

**Washington, D.C.  
Comprehensive Assessment System  
(DC CAS)**

**November 27, 2013**



**CTB/McGraw-Hill Education  
Monterey, California 93940**

Developed and published under contract with District of Columbia Office of the State Superintendent of Education by CTB/McGraw-Hill LLC, 20 Ryan Ranch Road, Monterey, California 93940-5703. Copyright © 2013 by the District of Columbia Office of the State Superintendent of Education. Based on a template, copyright © CTB/McGraw-Hill LLC. All rights reserved. Only authorized customers may copy, download and/or print the document, located online at <http://osse.dc.gov>. Any other use or reproduction of this document, in whole or in part, requires written permission of the District of Columbia Office of the State Superintendent of Education and the publisher.

---

**Table of Contents**

**List of Tables..... 5**

**Section 1. Overview ..... 8**

**Section 2. Item and Test Development..... 9**

    Overview ..... 9

    Content Standards ..... 9

    Item Development ..... 9

    Test Development ..... 10

    Test Design ..... 10

**Section 3. Test Administration Guidelines and Requirements ..... 19**

    Overview ..... 19

    Guidelines and Requirements for Administering DC CAS ..... 20

    Materials Orders, Delivery, and Retrieval ..... 20

    Secure Inventory ..... 21

**Section 4. Student Participation ..... 22**

    Tests Administered ..... 22

    Participation in DC CAS ..... 22

    Definition of Valid Test Administration ..... 23

    Special Accommodation ..... 23

**Section 5. Scoring..... 30**

    Selection of Raters for Scoring ..... 30

    Recruitment ..... 30

    The Interview Process ..... 30

    Training Material Development ..... 30

    Preparation and Meeting Logistics for Rangefinding ..... 31

    Training and Qualifying Procedures ..... 31

    Breakdown of Scoring Teams ..... 32

    Monitoring the Scoring Process ..... 32

**Section 6. Methods..... 34**

    Classical Item Level Analyses ..... 34

    Item Bias Analyses ..... 35

    Calibration and Linking ..... 36

    Goodness of Fit ..... 36

    Year-to-Year Linking Procedures ..... 37

    Establishing Upper and Lower Bounds for the Grade Level Scales ..... 39

    Reliability Coefficients ..... 40

    Standard Errors of Measurement ..... 40

    Proficiency Level Analyses ..... 41

    Classification Consistency ..... 41

    Classification Accuracy ..... 41

**Section 7. Standard Setting..... 48**

    Reading and Mathematics Cut Score Reviews ..... 49

    Final, Approved DC CAS Cut Scores ..... 50

**Section 8. Evidence for Reliability and Validity ..... 52**

    Reliability ..... 52

    Validity ..... 53

    Item Level Evidence ..... 53

    Classical Item Statistics ..... 53

    Inter-Rater Reliability ..... 54

    Differential Item Function ..... 54

    Test and Strand Level Evidence ..... 55

    Operational Test Scores ..... 55

    Strand Level Scores ..... 55

Standard Errors of Measurement..... 56  
 Proficiency Level Evidence ..... 56  
 Correlational Evidence across Content Areas ..... 57  
**References ..... 99**  
**Appendix A: Checklist for DC Educator Review of DC CAS Items ..... 102**  
**Appendix B: DC CAS Composition Scoring Rubrics..... 104**  
**Appendix C: Operational and Field Test Item Adjusted *P* Values ..... 106**  
**Appendix D: Internal Consistency Reliability Coefficients for Examinee Subgroups ..... 143**  
**Appendix E: Classification Consistency and Accuracy Estimates for All Proficiency Levels for Examinee Subgroups..... 149**

**List of Tables**

Table 1. DC CAS 2013 Operational Test Form Blueprints: Reading..... 12

Table 2. DC CAS 2013 Operational Test Form Blueprints: Mathematics..... 14

Table 3. DC CAS 2013 Operational Test Form Blueprints: Science/Biology..... 17

Table 4. DC CAS 2013 Operational Test Form Blueprints: Composition..... 18

Table 5. Number and Percentage of Examinees with Valid Test Administrations on the 2013 DC CAS in Reading, Mathematics, Science/Biology, and Composition..... 24

Table 6. Number and Percentage of Students in Special Programs with Test Scores on the 2013 DC CAS in Reading, Mathematics, Science/Biology, and Composition..... 25

Table 7. Number and Percentage of Students Coded for ELL Access for Proficiency Levels 1–4 in Reading, Mathematics, Science/Biology, and Composition..... 26

Table 8. Number and Percentage of Students Receiving One or More English Language Learner Test Administration Accommodations in Reading, Mathematics, Science/Biology, and Composition..... 27

Table 9. Number and Percentage of Students Receiving One or More Special Education Test Administration Accommodations in Reading, Mathematics, Science/Biology, and Composition..... 28

Table 10. Number and Percentage of Students Receiving One or More Selected Special Education Test Administration Accommodations in Reading, Mathematics, Science/Biology, and Composition..... 29

Table 11. DC CAS 2013 Numbers of Operational Items Flagged for Poor Fit During Calibration..... 44

Table 12. Correlations Between the Item Parameters for the Reference Form and 2013 DC CAS Operational Test Form..... 45

Table 13. Scaling Constants Across Administrations, All Grades and Content Areas..... 46

Table 14. LOSS and HOSS for Relevant Grades in Reading, Mathematics, Science/Biology and Composition..... 47

Table 15. Final Cut Score Ranges..... 51

Table 16. DC CAS 2013 Classical Item Level Statistics..... 59

Table 17. DC CAS 2013 Operational Inter-Rater Agreement for CR Items: Reading..... 60

Table 18. DC CAS 2013 Operational Inter-Rater Agreement for CR Items: Mathematics..... 61

Table 19. DC CAS 2013 Operational Inter-Rater Agreement for CR Items: Science/Biology..... 62

Table 20. DC CAS 2013 Operational Inter-Rater Agreement for CR Items: Composition..... 63

Table 21. DC CAS 2013 Field Test Inter-Rater Agreement for CR Items: Reading..... 64

Table 22. DC CAS 2013 Field Test Inter-Rater Agreement for CR Items: Mathematics..... 65

Table 23. DC CAS 2013 Field Test Inter-Rater Agreement for CR Items: Science/Biology..... 67

Table 24. Numbers of Operational Items Flagged for DIF Using the Mantel-Haenszel Procedure: Reading..... 68

Table 25. Numbers of Operational Items Flagged for DIF Using the Mantel-Haenszel Procedure: Mathematics..... 70

Table 26. Numbers of Operational Items Flagged for DIF Using the Mantel-Haenszel Procedure: Science/Biology... 71

Table 27. Numbers of Operational Items Flagged for DIF Using the Mantel-Haenszel Procedure: Composition..... 72

Table 28. Numbers of Field Test Items Flagged for DIF Using the Mantel-Haenszel Procedure: Reading..... 73

Table 29. Numbers of Field Test Items Flagged for DIF Using the Mantel-Haenszel Procedure: Mathematics..... 75

Table 30. Numbers of Field Test Items Flagged for DIF Using the Mantel-Haenszel Procedure: Science/Biology..... 76

Table 31. Total Test Scale and Raw Score Means and Reliability Statistics..... 77

Table 32. Weighted *P* Value Mean, *P* Value Mean and Standard Deviation, and Coefficient Alpha Reliability for Reading Strand Scores..... 78

Table 33. Weighted <i>P</i> Value Mean, <i>P</i> Value Mean and Standard Deviation, and Coefficient Alpha Reliability for Mathematics Strand Scores .....	79
Table 34. Weighted <i>P</i> Value Mean, <i>P</i> Value Mean and Standard Deviation, and Coefficient Alpha Reliability for Science/Biology Strand Scores.....	80
Table 35. Raw score mean and STD, and correlations for Composition Strand Scores .....	81
Table 36. DC CAS 2013 Reading Strand Correlations by Grade .....	82
Table 37. DC CAS 2013 Mathematics Strand Correlations by Grade.....	83
Table 38. DC CAS 2013 Science/Biology Strand Correlations by Grade .....	85
Table 39. DC CAS 2013 Composition Rubric Score Correlations by Grade .....	86
Table 40. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Reading .....	87
Table 41. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Mathematics.....	89
Table 42. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Science/Biology .....	92
Table 43. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Composition.....	93
Table 44. DC CAS 2013 Percentages of Students at Each Performance Level.....	94
Table 45. Classification Consistency and Accuracy Rates by Grade and Cut Score: Reading.....	95
Table 46. Classification Consistency and Accuracy Rates by Grade and Cut Score: Mathematics .....	96
Table 47. Classification Consistency and Accuracy Rates by Grade and Cut Score: Science/Biology.....	97
Table 48. Classification Consistency and Accuracy Rates by Grade and Cut Score: Composition .....	97
Table 49. Correlations Between Reading, Mathematics, Science/Biology, and Composition Total Test Raw Scores, by Grade .....	98
Table C1. DC CAS 2013 Operational Form Item Adjusted <i>P</i> Values: Reading.....	106
Table C2. DC CAS 2013 Operational Form Item Adjusted <i>P</i> Values: Mathematics .....	115
Table C3. DC CAS 2013 Operational Form Item Adjusted <i>P</i> Values: Science/Biology .....	123
Table C4. DC CAS 2013 Operational Form Item Adjusted <i>P</i> Values: Composition .....	126
Table C5. DC CAS 2013 Field Test Form Item Adjusted <i>P</i> Values: Reading .....	127
Table C6. DC CAS 2013 Field Test Form Item Adjusted <i>P</i> Values: Mathematics.....	132
Table C7. DC CAS 2013 Field Test Form Item Adjusted <i>P</i> Values: Science/Biology .....	140
Table D1. Scale Score Mean and SD, and Internal Consistency Reliability Coefficients for Examinee Subgroups: Reading.....	143
Table D2. Scale Score Mean and SD, and Internal Consistency Reliability Coefficients for Examinee Subgroups: Mathematics .....	145
Table D3. Scale Score Mean and SD, and Internal Consistency Reliability Coefficients for Examinee Subgroups: Science/Biology.....	147
Table D4. Scale Score Mean and SD, and Internal Consistency Reliability Coefficients for Examinee Subgroups: Composition .....	148
Table E1. Classification Consistency and Accuracy Rates for All Cut Scores and Examinee Subgroups: Reading....	149
Table E2. Classification Consistency and Accuracy Rates for All Cut Scores and Examinee Subgroups: Mathematics .....	151

---

Table E3. Classification Consistency and Accuracy Rates for All Cut Scores and Examinee Subgroups:  
Science/Biology..... 153

Table E4. Classification Consistency and Accuracy Rates for All Cut Scores and Examinee Subgroups:  
Composition ..... 154

## Section 1. Overview

The primary purpose of the District of Columbia Comprehensive Assessment System (DC CAS) is to measure the mastery of Reading, Mathematics, Science, Biology, and Composition content standards of all District of Columbia (DC) public school students annually. The assessments provide the foundation for an accountability system that enables the District to determine whether students and schools are making adequate yearly progress on DC content standards as required by the No Child Left Behind (NCLB) Act. In addition, the assessments are used by district- and school-based instructional staff to focus their lessons on content standards and evaluate whether students and schools are achieving those standards. Parents use the results to monitor their children's educational progress and the effectiveness of their school and school district.

This document describes the operational DC CAS that was administered to students in the spring of 2013 to assess students' skills in Grades 2–10 Reading; Grades 2–8 and 10 Mathematics; Grades 5 and 8 Science and high school Biology; and Grades 4, 7, and 10 Composition. The DC CAS consists of multiple choice (MC) and constructed response (CR) items in Reading, Mathematics, and Science/Biology, and writing prompts for Composition. All items are administered under standardized conditions, where students are allowed accommodations when eligible.

Technical reports for assessment programs are the primary means for test developers and assessment program managers to communicate with test users (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2009, p. 67). The standards require technical reports to document, for example, rationales and recommended uses for tests (Standard 6.3) and technical characteristics, such as score reliability and validity of score interpretations (Standard 6.5). Because of the technical nature of developing, implementing, and validating achievement tests like the DC CAS, technical reports target audiences with some level of technical training and understanding. Furthermore, the evidence provided in this report is directly relevant to the *Standards and Assessments Peer Review Guidance*, Critical Elements (January 12, 2009; see <http://www.ed.gov/policy/elsec/guid/saaprguidance.pdf>).

This technical report is written to document procedures and results from developing, analyzing, and validating the 2013 DC CAS. It provides information relevant to an evaluation of the validity of intended interpretations and uses of results from the 2013 DC CAS tests. The design of the test administration, content development and forms construction, classical item analysis, differential item functioning (DIF), item response theory analyses (IRT), and proficiency level data are provided.

## Section 2. Item and Test Development

This section contains information relevant to the *Standards and Assessments Peer Review Guidance*, Critical Elements 4.1 and 5.1:

### 4.1

For each assessment, including all alternate assessments, has the State documented the issue of **validity** (in addition to the alignment of the assessment with the content standards), as described in the *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 1999), with respect to all of the following categories:

(d) Has the State ascertained that the scoring and reporting structures are consistent with the sub-domain structures of its academic content standards (i.e., are item interrelationships consistent with the framework from which the test arises)?

### 5.1

Has the State outlined a coherent approach to ensuring alignment between each of its assessments...based on grade-level achievement standards, and the academic content standards and academic achievement standards the assessment is designed to measure?

## Overview

A key piece of validity evidence is provided by the procedures used to develop the test's content and the alignment of items with the test blueprint and specifications. By setting forth a description of the events that took place in a test's development, we establish evidence of validity for the DC CAS based on test development procedures and test content.

Evidence of validity based on test content includes information about the item and test specifications. Test development involves creating a design framework from the statement of the achievement construct to be measured. Design elements include numbers and types of items and score points allocated to each content strand in each content area test.

## Content Standards

The DC CAS tests are aligned to either DC Content Standards, Common Core State Standards (CCSS), or to both for content areas transitioning to the common core. The standards serve as reporting categories. Since 2012, Reading content has been fully aligned to the CCSS in Reading. The total numbers of operational items included in the Science/Biology test design remained the same, although they were distributed differently under the new standard headings. In 2012, Mathematics content had been transitioning to the CCSS and remained partially aligned to the DC Content Standards, which served as the reporting categories. In 2013, the Mathematics content is fully aligned to the CCSS.

## Item Development

Each year, newly developed items are field tested in DC CAS in all grades and content areas. These items are developed by CTB and, prior to field testing, go through a rigorous content and psychometric review and approval process. CTB content and style editors, supervisors, and managers review all items and then provide items to participants in Content and Bias/Sensitivity Review workshops conducted in DC. The Office of the State Superintendent of Education (OSSE)

invited educators, members of the DC Public Schools (DCPS) administration, and community representatives to participate in workshops to review the items. A training session was provided by CTB, after which the participants reviewed all items for content and grade appropriateness, as well as for alignment to the content standards, then rated each item for acceptance, revision, or rejection. The reviewers used the criteria in the checklist in Appendix A to guide their rating decisions.

## **Test Development**

CTB's Research and Development teams, with the approval of the OSSE, developed test forms designed to measure student performance through both MC and CR item types. The total number of items and score points included in each reporting category serves as the test blueprint, details of which are provided in Tables 1–4.

The items that are available for selection in the DC CAS 2013 assessments originated from a pool of operational and formerly field tested items from the 2012 DC CAS administration, with a small number of items selected from older pools (excluding 2009). The majority of the Grade 2 Reading and Mathematics items and the Grade 9 Reading items originated from CTB-owned items in the *TerraNova*<sup>TM</sup> item pool.

DC CAS assessments are linked each year so that, from one form and year to the next, student scores remain comparable. The linking process requires a set of items to link or anchor one year to the next. These anchor items are a small subset of items that, proportionally, reflect the overall test blueprints. They are typically selected first, around which the remaining operational items are selected for each form.

The forms were assembled by CTB Development staff who attended to the blueprint requirements, as well as various other content and psychometric requirements, such as test length, score points, item types and statistical comparability. For example, all proposed selections for operational forms were compared to previous DC CAS test forms to ensure they remain parallel and comparable in terms of test difficulty and coverage of the DC CAS content standards, as specified in the 2013 test blueprints.

Once the forms were assembled, they went through an iterative review and approval process where they were reviewed and approved by CTB Research, and then by OSSE. The items were reviewed for content standards alignment and appropriateness by CTB Development and by OSSE.

## **Test Design**

The DC CAS tests are designed as operational tests with embedded field test items. In this way, newly developed items can be field tested in and amongst operational items. This is an advantage over separate field test designs that highlight the items that do not “count” towards students' scores and can decrease the motivation of their serious effort and response.

To maximize the number of items field tested while minimally impacting the testing time required, two forms were developed for all grades except for Reading and Composition. Each form included the same core set of operational items (which comprised the linking anchor subset) and a set of unique embedded field test items. The two forms were spiraled together and packaged to ensure near equal distribution of the forms in classrooms and so that field test data were based on randomly equivalent groups across all students in the District .

The Composition tests were operational tests. This design captures student performance on operational items. One writing prompt was administered in each of Grades 4, 7, and 10. The prompts were developed to reflect CCSS and scored three times based on the following traits/rubrics: Topic/Idea Development, Standard English Conventions, and Understanding Literary Text or Understanding Informational Text (depending upon the type of passage associated with each prompt: literary or informational). The rubrics used to score these items can be found in Appendix B. Note that the student reports, overall scale score, and proficiency level designation reflected the scores from all three rubrics.

### **Test Administration Design**

For Grades 4–8 and 10, both Reading and Mathematics items were included in the same test books. Reading items were in a stand-alone test book for Grade 9 because Mathematics was not tested at that grade level. For all grades, test books and answer booklets were color-coded. Students in Grades 2 and 3 used scannable test books in which they recorded their answers.

For Grades 3–10, each Reading and Mathematics test was divided into four sessions, for a total of eight sessions per grade level test. For Grade 2, each Reading and Mathematics test was divided into three sessions, for a total of six sessions. For all Mathematics grades, each session included both MC and CR items. For Reading grades 4, 6, 7 and 8, each session included both MC and CR items.

A similar configuration was used for the Science/Biology tests. Students responded to the test items in one of two test books. They recorded their answers in scannable answer documents. No manipulatives were provided. The Science/Biology tests were divided into three sessions, each with both MC and CR items.

Composition test books were provided for each prompt. The test books were scannable documents that included the following: directions to students, evaluation criteria, a writing prompt, three lined pages, and a biogrid. The prompts were administered within the established two-week testing window. Students were also issued two sheets of double-sided, lined draft paper, specially developed for the Composition test, for planning their writing.

Additional information regarding administration manuals and procedures is provided in Section 3.

**Table 1. DC CAS 2013 Operational Test Form Blueprints: Reading**

Grade	Content Strand		Operational					Anchor		Field Test	
			Number of MC Items/ Points	% of MC Points	Number of CR Items	Number of CR Points	% of CR Points	Total Number of Points	Number of Items	% of Points	Number of Items
2	1	Vocabulary Acquisition & Use	5	100.00	0	0	0.00	5	5	100.00	0
	3	Reading Informational Text	12	66.67	2	6	33.33	18	7	38.89	5
	4	Reading Literary Text	16	100.00	0	0	0.00	16	8	50.00	5
		Total	33	84.62	2	6	15.38	39	20	51.28	10
3	1	Vocabulary Acquisition & Use	9	100.00	0	0	0.00	9	4	44.44	1
	3	Reading Informational Text	16	72.73	2	6	27.27	22	9	40.91	4
	4	Reading Literary Text	20	86.96	1	3	13.04	23	9	39.13	5
		Total	45	83.33	3	9	16.67	54	22	40.74	10
4	1	Vocabulary Acquisition & Use	9	100.00	0	0	0.00	9	4	44.44	0
	3	Reading Informational Text	18	85.71	1	3	14.29	21	8	38.10	5
	4	Reading Literary Text	18	75.00	2	6	25.00	24	10	41.67	6
		Total	45	83.33	3	9	16.67	54	22	40.74	11
5	1	Vocabulary Acquisition & Use	8	100.00	0	0	0.00	8	4	50.00	1
	3	Reading Informational Text	15	71.43	2	6	28.57	21	7	33.33	5
	4	Reading Literary Text	22	88.00	1	3	12.00	25	11	44.00	4
		Total	45	83.33	3	9	16.67	54	22	40.74	10
6	1	Vocabulary Acquisition & Use	9	100.00	0	0	0.00	9	4	44.44	0
	3	Reading Informational Text	14	60.87	3	9	39.13	23	9	39.13	10
	4	Reading Literary Text	22	100.00	0	0	0.00	22	9	40.91	0
		Total	45	83.33	3	9	16.67	54	22	40.74	10

**Table 1. DC CAS 2013 Operational Test Form Blueprints: Reading (continued)**

Grade	Content Strand		Operational					Anchor		Field Test	
			Number of MC Items/ Points	% of MC Points	Number of CR Items	Number of CR Points	% of CR Points	Total Number of Points	Number of Items	% of Points	Number of Items
7	1	Vocabulary Acquisition & Use	7	100.00	0	0	0.00	7	3	42.86	0
	3	Reading Informational Text	19	76.00	2	6	24.00	25	10	40.00	5
	4	Reading Literary Text	19	86.36	1	3	13.64	22	9	40.91	5
		Total	45	83.33	3	9	16.67	54	22	40.74	10
8	1	Vocabulary Acquisition & Use	5	100.00	0	0	0.00	5	2	40.00	0
	3	Reading Informational Text	21	70.00	3	9	30.00	30	11	36.67	6
	4	Reading Literary Text	19	100.00	0	0	0.00	19	9	47.37	5
		Total	45	83.33	3	9	16.67	54	22	40.74	11
9	1	Vocabulary Acquisition & Use	7	100.00	0	0	0.00	7	4	57.14	1
	3	Reading Informational Text	20	90.91	1	2	9.09	22	10	45.45	10
	4	Reading Literary Text	18	75.00	2	6	25.00	24	8	33.33	0
		Total	45	84.91	3	8	15.09	53	22	41.51	11
10	1	Vocabulary Acquisition & Use	7	100.00	0	0	0.00	7	3	42.86	0
	3	Reading Informational Text	21	70.00	3	9	30.00	30	10	33.33	6
	4	Reading Literary Text	17	100.00	0	0	0.00	17	9	52.94	6
		Total	45	83.33	3	9	16.67	54	22	40.74	12

**Table 2. DC CAS 2013 Operational Test Form Blueprints: Mathematics**

Grade	Content Strand		Operational					Anchor		Field Test	
			Number of MC Items/ Points	% of MC Points	Number of CR Items	Number of CR Points	% of CR Points	Total Number of Points	Number of Items	% of Points	Number of Items
2	1	Operations & Algebraic Thinking	6	54.55	2	5	45.45	11	3	27.27	5
	2	Numbers & Operations Base Ten	17	100.00	0	0	0.00	17	7	41.18	10
	4	Measurement and Data	13	100.00	0	0	0.00	13	8	61.54	8
	11	Geometry	5	100.00	0	0	0.00	5	2	40.00	9
		Total	41	89.13	2	5	10.87	46	20	43.48	32
3	1	Operations & Algebraic Thinking	13	100.00	0	0	0.00	13	7	53.85	3
	2	Number & Operations Base Ten	5	100.00	0	0	0.00	5	1	20.00	7
	3	Number & Operations Fractions	10	76.92	1	3	23.08	13	2	15.38	6
	4	Measurement and Data	18	75.00	2	6	25.00	24	10	41.67	14
	11	Geometry	5	100.00	0	0	0.00	5	5	100.00	2
		Total	51	85.00	3	9	15.00	60	25	41.67	32
4	1	Operations & Algebraic Thinking	12	80.00	1	3	20.00	15	5	33.33	5
	2	Number & Operations Base Ten	8	100.00	0	0	0.00	8	4	50.00	8
	3	Number & Operations Fractions	14	100.00	0	0	0.00	14	5	35.71	9
	4	Measurement and Data	12	80.00	1	3	20.00	15	9	60.00	8
	11	Geometry	5	62.50	1	3	37.50	8	2	25.00	2
		Total	51	85.00	3	9	15.00	60	25	41.67	32

**Table 2. DC CAS 2013 Operational Test Form Blueprints: Mathematics (continued)**

Grade	Content Strand		Operational					Anchor		Field Test	
			Number of MC Items/ Points	% of MC Points	Number of CR Items	Number of CR Points	% of CR Points	Total Number of Points	Number of Items	% of Points	Number of Items
5	1	Operations & Algebraic Thinking	9	100.00	0	0	0.00	9	6	66.67	5
	2	Number & Operations Base Ten	9	100.00	0	0	0.00	9	5	55.56	7
	3	Number & Operations Fractions	15	100.00	0	0	0.00	15	3	20.00	10
	4	Measurement and Data	12	80.00	1	3	20.00	15	6	40.00	5
	11	Geometry	6	50.00	2	6	50.00	12	4	33.33	5
		Total	51	85.00	3	9	15.00	60	24	40.00	32
6	5	Ratios & Proportional Relationships	10	76.92	1	3	23.08	13	4	30.77	10
	6	The Number System	13	100.00	0	0	0.00	13	8	61.54	3
	7	Expressions & Equations	14	100.00	0	0	0.00	14	5	35.71	7
	11	Geometry	8	72.73	1	3	27.27	11	3	27.27	5
	12	Statistics & Probability	6	66.67	1	3	33.33	9	4	44.44	7
		Total	51	85.00	3	9	15.00	60	24	40.00	32
7	5	Ratios & Proportional Relationships	12	80.00	1	3	20.00	15	7	46.67	7
	6	The Number System	10	100.00	0	0	0.00	10	6	60.00	4
	7	Expressions & Equations	14	100.00	0	0	0.00	14	3	21.43	12
	11	Geometry	7	70.00	1	3	30.00	10	4	40.00	3
	12	Statistics & Probability	8	72.73	1	3	27.27	11	4	36.36	6
		Total	51	85.00	3	9	15.00	60	24	40.00	32

**Table 2. DC CAS 2013 Operational Test Form Blueprints: Mathematics (continued)**

Grade	Content Strand		Operational					Anchor		Field Test	
			Number of MC Items/ Points	% of MC Points	Number of CR Items	Number of CR Points	% of CR Points	Total Number of Points	Number of Items	% of Points	Number of Items
8	6	The Number System	9	100.00	0	0	0.00	9	4	44.44	1
	7	Expressions & Equations	10	76.92	1	3	23.08	13	7	53.85	7
	10	Functions	10	76.92	1	3	23.08	13	6	46.15	5
	11	Geometry	12	100.00	0	0	0.00	12	5	41.67	11
	12	Statistics & Probability	10	100.00	0	0	0.00	10	3	30.00	9
		Total		51	89.47	2	6	10.53	57	25	43.86
10	8	Number & Quantity	9	100.00	0	0	0.00	9	4	44.44	4
	9	Algebra	15	83.33	1	3	16.67	18	9	50.00	9
	10	Functions	5	62.50	1	3	37.50	8	3	37.50	6
	11	Geometry	16	84.21	1	3	15.79	19	6	31.58	10
	12	Statistics & Probability	6	100.00	0	0	0.00	6	2	33.33	3
		Total		51	85.00	3	9	15.00	60	24	40.00

*Note:* One grade 8 OP CR item was changed to a FT item.

**Table 3. DC CAS 2013 Operational Test Form Blueprints: Science/Biology**

Grade	Content Strand		Operational					Anchor		Field Test	
			Number of MC Items/Points	% of MC Points	Number of CR Items	Number of CR Points	% of CR Points	Total Number of Points	Number of Items	% of Points	Number of Items
5	1	Science and Technology	17	89.47	1	2	10.53	19	8	42.11	9
	2	Earth and Space Science	11	84.62	1	2	15.38	13	4	30.77	5
	3	Physical Science	8	100.00	0	0	0.00	8	5	62.50	7
	4	Life Science	11	84.62	1	2	15.38	13	6	46.15	7
		Total	47	88.68	3	6	11.32	53	23	43.40	28
8	1	Scientific Thinking and Inquiry	7	77.78	1	2	22.22	9	4	44.44	3
	2	Matter and Reactions	18	90.00	1	2	10.00	20	9	45.00	10
	3	Forces	7	77.78	1	2	22.22	9	3	33.33	5
	4	Energy and Waves	15	100.00	0	0	0.00	15	7	46.67	8
		Total	47	88.68	3	6	11.32	53	23	43.40	26
High School	1	Cell Biology & Biochemistry	15	88.24	1	2	11.76	17	6	35.29	13
	2	Genetics and Evolution	16	100.00	0	0	0.00	16	8	50.00	9
	3	Multicellular Organisms	7	63.64	2	4	36.36	11	4	36.36	2
	4	Ecosystems	9	100.00	0	0	0.00	9	4	44.44	3
		Total	47	88.68	3	6	11.32	53	22	41.51	27

*Note:* One High School FT item and two grade 8 FT items appear in both Form 1 and 2, but they are only counted once.

**Table 4. DC CAS 2013 Operational Test Form Blueprints: Composition**

Grade	Scoring Rubric	Number of CR Items	Number of CR Points	Contribution to Overall Scale Score	
				Number of Points	% of Points
4	Writing Topic Development	1	6	6	42.86
	Writing Language Conventions	1	4	4	28.57
	Understanding Literary Text	1	4	4	28.57
	Total	3	14	14	100.00
7	Writing Topic Development	1	6	6	42.86
	Writing Language Conventions	1	4	4	28.57
	Understanding Informational Text	1	4	4	28.57
	Total	3	14	14	100.00
10	Writing Topic Development	1	6	6	42.86
	Writing Language Conventions	1	4	4	28.57
	Understanding Literary Text	1	4	4	28.57
	Total	3	14	14	100.00

## Section 3. Test Administration Guidelines and Requirements

This section contains information relevant to the *Standards and Assessments Peer Review Guidance*, Critical Elements 4.3, 4.5, and 6.2:

### 4.3

Has the State ensured that its assessment system is fair and accessible to all students, including students with disabilities and students with limited English proficiency, with respect to each of the following issues:

- (a) Has the State ensured that the assessments provide an appropriate variety of accommodations for students with disabilities? *and*
- (b) Has the State ensured that the assessments provide an appropriate variety of linguistic accommodations for students with limited English proficiency?

### 4.5

Has the State established clear criteria for the administration, scoring, analysis, and reporting components of its assessment system, including all alternate assessments, and does the State have a system for monitoring and improving the on-going quality of its assessment system?

### 6.2

1. What guidelines does the State have in place for including all students with disabilities in the assessment system?

- (a) Has the State developed, disseminated information on, and promoted use of appropriate accommodations to increase the number of students with disabilities who are tested against academic achievement standards for the grade in which they are enrolled?
- (b) Has the State ensured that general and special education teachers and other appropriate staff know how to administer assessments, including making use of accommodations, for students with disabilities and students covered under Section 504.

## Overview

Administration of the DC CAS assessments each spring is managed by the OSSE, coordinated in each school by a Test Chairperson, and conducted by classroom teachers. Assessment office staff trained school Test Chairpersons on test administration guidelines and requirements using the 2013 *Test Chairperson's Manual*. Test Chairpersons, in turn, trained all Test Administrators and proctors. Test Administrators administered all DC CAS assessments according to requirements and steps in the *Test Directions*.

The *Test Chairperson's Manual* directs Test Chairpersons to follow the procedures for training Test Administrators and proctors on required procedures for administering each test and maintaining test security before, during, and after test administrations. It also provides information on available accommodations for students with disabilities and for English language learners.

The *Test Directions* covers similar topics and requirements. In addition, it provides instructions on scheduling test administrations, preparing students for the test administration, using

standardized testing procedures, and verbatim instructions for administering each test to students. It also provides information on available accommodations for students with disabilities and for English language learners.

## **Guidelines and Requirements for Administering DC CAS**

The *Test Chairperson's Manual* indicates that DC CAS administrations should be scheduled to ensure that all students have adequate time to respond to all test items under unhurried conditions. It also describes testing condition requirements to ensure that students can feel as comfortable as possible and are not distracted during administration. The manual requires each Test Chairperson to complete a Test Site Observation Report to ensure that adequate testing conditions can be provided. It also contains instructions on distributing test materials to Test Administrators, retrieving the materials, accounting for 100% of all secure materials, shipping the materials to CTB for processing, and maintaining security of the materials at all times and throughout the entire process.

The *Test Chairperson's Manual* and *Test Directions* provide information on available test administration accommodations for students with disabilities and for English language learners. They specify approved accommodations that maintain standard testing conditions (e.g., reading only Mathematics, Science, or Health questions or Composition writing prompts to examinees) and identify accommodations that are considered modifications to the test that will result in invalidated test scores (e.g., assisted reading of Reading passages).

The *Test Chairperson's Manual* and *Test Directions* specify accommodations approved for students with disabilities in the following areas: timing/scheduling (e.g., providing breaks between prescribed timing sections of the tests), setting (e.g., individual and small group administrations), presentation (e.g., reading of [only] Mathematics, Science, or Health test questions or Composition writing prompts), and response accommodations (e.g., dictating responses). The *Test Chairperson's Manual* and *Test Directions* specify accommodations approved for English language learners; they are in the following areas: direct linguistic support—oral, direct linguistic support—written, and indirect linguistic support. Both manuals indicate that Test Administrators must record on the student's answer document all test administration accommodations that are provided.

CTB provides test administration sessions for school Test Chairpersons in the month prior to the test administration. School Test Chairpersons are required to conduct training sessions, and all school staff who will handle test materials must attend these sessions. School Test Chairpersons are explicitly required in the *Test Chairperson's Manual* to oversee the test administrations in their schools. They are required to ensure that test materials are available in adequate numbers and that school staff adhere to test security requirements, track materials by using security checklists, report breaches if they occur, document disruptions during testing, sign test materials in and out each day, account for 100% of secure test materials, and report missing or damaged materials immediately to CTB Customer Service and OSSE by completing the online Security Exceptions Survey.

## **Materials Orders, Delivery, and Retrieval**

Customer orders were managed in CTB's Online Enrollment System. Schools updated and validated their enrollments or indicated non-participation. CTB used the results for order fulfillment.

Prior to shipment of materials, barcodes were applied to the secure materials for the purpose of secure inventory tracking (a description of the Secure Inventory process is provided next in this section). Corresponding security checklists were also produced. Daily tracking reports were provided to OSSE for the purpose of monitoring the deliveries.

The appropriate district and school staff were previously trained to maintain security and monitor quantities of materials. Shortly after delivery, they unpacked and reviewed materials to ensure readiness for administration, as described in the previous section of this report, “Guidelines and Requirements for Administering DC CAS.” In the event that the materials received were not sufficient for administration, a short/add window functioned to permit CTB Customer Service to process requests for additional materials while maintaining a secure inventory.

After the test administration was complete, the materials were packaged for retrieval and picked up according to a verified schedule. Daily tracking reports also served for OSSE to monitor retrievals. When the materials were back in CTB’s custody, all books with security barcodes were accounted for as described in the following section of this report, “Secure Inventory.”

## Secure Inventory

To further support the full range of test security requirements for DC CAS, CTB has instituted a comprehensive Test Security/Test Inventory System. This system was created using industry best practices. Upon request, CTB further customized a security model to precisely match the needs of DC CAS security requirements. This security model for the DC CAS assessment maintains its own list of material deliverables and services, from assessment barcoding to inventory checking and shipment tracking, as described in the steps below.

1. Secure materials are barcoded at the printer, vertically banded, and inventoried. Barcode files are sent to CTB. Packing lists and test materials are sent to the schools.
2. Materials are distributed into the schools.
3. Following the test administration, school staff members separate secure and non-secure materials and package them for return to CTB following *Test Chairperson’s Manual* instructions.
4. The dedicated/secure carrier contacts the schools to schedule retrieval of their materials on a specified date.
5. Scorable secure documents are accounted for during answer document scanning, and nonscorable secure documents are scanned into an inventory return system. Materials sent to the wrong CTB facility are forwarded to the appropriate site, as needed.
6. Missing Materials Reports are sent to OSSE for resolution once scanning is completed. Given a list of shipped security barcodes minus the barcode numbers already received, the remaining list is considered to be missing inventory.
7. OSSE contacts schools and reports back to CTB on findings, including additional books that have been located, contaminated books that could not be returned to CTB, and damaged or destroyed books where no barcode was available for scanning.
8. CTB processes additional, received inventory and approved exceptions, and produces a final missing inventory report.

As of October 29, 2013, approximately 99.56% of secure materials were accounted for; 502 secure test books were missing for the 2013 administration, compared with 212 test books missing in 2012.

## Section 4. Student Participation

This section contains information relevant to *Standards and Assessments Peer Review Guidance*, Critical Elements 6.1 and 6.2:

### 6.1

1. Do the State's participation data indicate that all students in the tested grade levels or grade ranges are included in the assessment system (e.g., students with disabilities, students with limited English proficiency, economically disadvantaged students, race/ethnicity, migrant students, homeless students, etc.)?

2. Does the State report separately the number and percent of students with disabilities assessed on the regular assessment without accommodations, on the regular assessment with accommodations, on an alternate assessment against grade level standards, and, if applicable, on an alternate assessment against alternate achievement standards and/or on an alternate assessment against modified academic achievement standards?

### 6.2

1. What guidelines does the State have in place for including all students with disabilities in the assessment system?

(a) Has the State developed, disseminated information on, and promoted use of appropriate accommodations to increase the number of students with disabilities who are tested against academic achievement standards for the grade in which they are enrolled?

## Tests Administered

All public schools in the District of Columbia administered the DC CAS tests between April 22 and May 2, 2013.

The tests administered were:

- Reading, Grades 2–10
- Mathematics, Grades 2–8 and 10
- Composition, Grades 4, 7, and 10
- Science, Grades 5 and 8
- Biology, for those students in Grades 8–12 who were enrolled in a high school Biology course

## Participation in DC CAS

The DC CAS *Test Chairperson's Manual* states that all students enrolled in all public schools in DC must participate in DC CAS grade level test administrations, with one exception: A student with significant cognitive disabilities whose Individualized Education Program (IEP) indicates that the student meets OSSE's established criteria may participate in the DC CAS alternate assessment portfolio.

During the spring 2013 test administration, about 4,600 students were assessed in Reading and Mathematics at each tested grade. Science/Biology had about 4,300 and Composition had about 4,400 students at each tested grade. Only students with a valid test administration as required by the type of analysis, as defined below, are included in the reports.

## Definition of Valid Test Administration

In this technical report, two sets of rules are used to define a valid test administration. The first set of rules is for psychometric analyses included in this report (e.g., reliability, DIF, item parameter calibration, and linking). Answer documents are excluded when any of the following conditions are observed:

- Three or more of the first five items are invalidly marked or omitted.
- The operational test total raw score equals zero and the sum of the operational and field test item valid responses is less than 5.
- All operational and field test items are omitted.

The second set of valid test administration rules are for analyses summarizing test performance (e.g., overall numbers of examinees, descriptive statistics, and correlations of test scores). All students who have a valid test score, as defined in the DC CAS Spring 2013 Business Requirements, are included in these analyses. For the Reading, Mathematics, Science, and Biology assessments, the requirements document outlines a valid attempt on the test as:

- At least one item marked with a correct response OR
- At least 5 items validly marked in the content area

And for Composition, a valid attempt is defined as:

- A score of non-zero on all three parts of the item

Note: To maintain confidentiality of individual student results, this report does not show subgroup results for fewer than 25 students. The race/ethnicity subgroups Native Hawaiian/Pacific Islander and American Indian/Alaska Native contain fewer than 25 students per grade and are not shown in the following tables.

The total number and percentage of students with valid tests are provided at the total and subgroups of gender and race/ethnicity in Table 5. Participation rates for students in special populations, such as special education, Title 1, English Language Learners, and students with 504 plans is provided in Table 6. ELL students who participate in the DC CAS were classified by their schools into one of four language proficiency levels. These levels are related to levels of language instruction services and participation in school instruction, the number and percent of which are provided in Table 7.

## Special Accommodation

Students with disabilities and ELLs who participate in DC CAS grade level administrations may be provided approved test administration accommodations that are specified by special education IEP teams, Section 504 teams, or ELL teams. Test administration accommodations are categorized into one or more of four categories: timing/scheduling, setting, presentation, and response. For a student to receive an accommodation, the accommodation had to be in place during the school year and specified in the student's IEP or 504 plan. Within prescribed parameters, students in ELL programs received test administration accommodations in one or more of three categories: direct linguistic support—oral, direct linguistic support—written, and indirect linguistic support. The number and percent of the various accommodations documented are provided in Tables 6–10. For more information on these accommodations, please refer to the DC CAS *Test Chairperson's Manual*.

**Table 5. Number and Percentage of Examinees with Valid Test Administrations on the 2013 DC CAS in Reading, Mathematics, Science/Biology, and Composition**

Grade	Students with Test Scores	Males		Females		Asian		African American		Hispanic		White	
		N	%	N	%	N	%	N	%	N	%	N	%
<b>Reading</b>													
2	4,802	2,384	49.65	2,405	50.08	66	1.37	3,378	70.35	683	14.22	570	11.87
3	4,914	2,472	50.31	2,428	49.41	78	1.59	3,505	71.33	682	13.88	546	11.11
4	4,574	2,320	50.72	2,246	49.10	69	1.51	3,323	72.65	651	14.23	444	9.71
5	4,540	2,297	50.59	2,229	49.10	83	1.83	3,339	73.55	614	13.52	430	9.47
6	4,572	2,262	49.48	2,294	50.17	63	1.38	3,553	77.71	557	12.18	313	6.85
7	4,531	2,300	50.76	2,215	48.89	60	1.32	3,539	78.11	580	12.80	273	6.03
8	4,260	2,101	49.32	2,132	50.05	50	1.17	3,359	78.85	532	12.49	245	5.75
9	4,879	2,438	49.97	2,376	48.70	63	1.29	3,898	79.89	569	11.66	234	4.80
10	4,752	2,245	47.24	2,398	50.46	69	1.45	3,740	78.70	535	11.26	232	4.88
<b>Mathematics</b>													
2	4,840	2,404	49.67	2,420	50.00	70	1.45	3,390	70.04	700	14.46	575	11.88
3	4,947	2,487	50.27	2,446	49.44	83	1.68	3,511	70.97	698	14.11	550	11.12
4	4,620	2,345	50.76	2,266	49.05	75	1.62	3,335	72.19	671	14.52	449	9.72
5	4,572	2,313	50.59	2,243	49.06	83	1.82	3,344	73.14	633	13.85	436	9.54
6	4,598	2,276	49.50	2,306	50.15	65	1.41	3,557	77.36	575	12.51	313	6.81
7	4,563	2,320	50.84	2,227	48.81	63	1.38	3,544	77.67	596	13.06	279	6.11
8	4,302	2,119	49.26	2,153	50.05	52	1.21	3,356	78.01	569	13.23	247	5.74
10	4,729	2,236	47.28	2,384	50.41	69	1.46	3,721	78.68	530	11.21	233	4.93
<b>Science/Biology</b>													
5	4,523	2,295	50.74	2,215	48.97	81	1.79	3,306	73.09	626	13.84	432	9.55
8	4,135	2,030	49.09	2,074	50.16	50	1.21	3,209	77.61	554	13.40	239	5.78
High School	4,047	1,931	47.71	2,040	50.41	54	1.33	3,140	77.59	510	12.60	214	5.29
<b>Composition</b>													
4	4,470	2,246	50.25	2,210	49.44	69	1.54	3,232	72.30	643	14.38	438	9.80
7	4,403	2,222	50.47	2,154	48.92	60	1.36	3,415	77.56	572	12.99	265	6.02
10	4,396	2,022	46.00	2,236	50.86	70	1.59	3,377	76.82	519	11.81	228	5.19

**Table 6. Number and Percentage of Students in Special Programs with Test Scores on the 2013 DC CAS in Reading, Mathematics, Science/Biology, and Composition**

Grade	Students with Test Scores	Special Education		English Language Learner		Section 504		Title I Targeted		Home Schooling	
		N	%	N	%	N	%	N	%	N	%
<b>Reading and/or Mathematics</b>											
2	4,840	320	7	387	8	24	0	126	3	1	0
3	4,948	492	10	335	7	49	1	141	3	1	0
4	4,624	487	11	238	5	54	1	135	3	1	0
5	4,572	574	13	201	4	50	1	92	2	0	0
6	4,603	642	14	187	4	51	1	77	2	1	0
7	4,567	596	13	205	4	53	1	101	2	0	0
8	4,314	579	13	181	4	33	1	98	2	2	0
9	4,879	571	12	153	3	17	0	2	0	0	0
10	4,765	610	13	184	4	25	1	5	0	0	0
<b>Science/Biology</b>											
5	4,523	507	11	184	4	44	1	64	1	0	0
8	4,135	503	12	180	4	34	1	94	2	0	0
High School	4,047	453	11	116	3	7	0	3	0	1	0
<b>Composition</b>											
4	4,470	447	10	187	4	38	1	119	3	0	0
7	4,403	544	12	161	4	56	1	89	2	0	0
10	4,396	522	12	117	3	17	0	5	0	0	0

**Note:** Students who participated in more than one test administration are counted only once. Student subgroups are indicated in the Program Participation section on the biogrid on each student’s answer document.

**Table 7. Number and Percentage of Students Coded for ELL Access for Proficiency Levels 1–4 in Reading, Mathematics, Science/Biology, and Composition**

Grade	Students with Test Scores	Level 1		Level 2		Level 3		Level 4	
		N	%	N	%	N	%	N	%
<b>Reading and/or Mathematics</b>									
2	4,840	40	1	59	1	173	4	141	3
3	4,948	36	1	33	1	138	3	147	3
4	4,624	38	1	18	0	56	1	121	3
5	4,572	34	1	17	0	50	1	89	2
6	4,603	29	1	28	1	52	1	82	2
7	4,567	32	1	36	1	69	2	75	2
8	4,314	45	1	15	0	67	2	53	1
9	4,879	24	0	27	1	37	1	45	1
10	4,765	7	0	32	1	64	1	65	1
<b>Science/Biology</b>									
5	4,523	27	1	16	0	49	1	85	2
8	4,135	43	1	15	0	57	1	47	1
High School	4,047	32	1	26	1	25	1	35	1
<b>Composition</b>									
4	4,470	7	0	15	0	46	1	110	2
7	4,403	3	0	31	1	65	1	61	1
10	4,396	4	0	23	1	32	1	34	1

**Table 8. Number and Percentage of Students Receiving One or More English Language Learner Test Administration Accommodations in Reading, Mathematics, Science/Biology, and Composition**

Grade	Students with Test Scores	Direct Linguistic Support—Oral <sup>1</sup>		Direct Linguistic Support—Written		Indirect Linguistic Support		Other	
		N	%	N	%	N	%	N	%
<b>Reading and/or Mathematics</b>									
2	4,840	383	8	247	5	378	8	5	0
3	4,948	296	6	186	4	303	6	25	1
4	4,624	230	5	153	3	228	5	6	0
5	4,572	149	3	98	2	147	3	3	0
6	4,603	191	4	127	3	187	4	4	0
7	4,567	215	5	175	4	223	5	1	0
8	4,314	235	5	172	4	238	6	6	0
9	4,879	297	6	59	1	314	6	0	0
10	4,765	203	4	120	3	225	5	2	0
<b>Science/Biology</b>									
5	4,523	142	3	94	2	140	3	2	0
8	4,135	234	6	161	4	240	6	1	0
High School	4,047	245	6	142	4	258	6	0	0
<b>Composition</b>									
4	4,470	177	4	101	2	173	4	0	0
7	4,403	170	4	68	2	177	4	0	0
10	4,396	196	4	103	2	214	5	0	0

**Note:** Students who received more than one accommodation in a single content area test can be counted in multiple columns. Students who received accommodations in more than one content area test administration are counted only once. Students for whom the ELL bubble was not completed but who did receive these ELL test administration accommodations are counted here.

<sup>1</sup> The “Oral Reading of Test in English” accommodation is typically not permitted for the Reading test.

**Table 9. Number and Percentage of Students Receiving One or More Special Education Test Administration Accommodations in Reading, Mathematics, Science/Biology, and Composition**

Grade	Students with Test Scores	Timing/ Scheduling		Setting		Presentation <sup>1</sup>		Response		Other	
		N	%	N	%	N	%	N	%	N	%
<b>Reading/Mathematics</b>											
2	4,840	339	7	347	7	310	6	179	4	8	0
3	4,948	519	10	542	11	485	10	331	7	7	0
4	4,624	573	12	591	13	512	11	413	9	25	1
5	4,572	625	14	635	14	579	13	431	9	34	1
6	4,603	692	15	724	16	644	14	533	12	15	0
7	4,567	697	15	709	16	655	14	554	12	29	1
8	4,314	590	14	632	15	582	13	547	13	17	0
9	4,879	693	14	724	15	575	12	504	10	4	0
10	4,765	661	14	653	14	538	11	655	14	17	0
<b>Science/Biology</b>											
5	4,523	594	13	594	13	534	12	354	8	23	1
8	4,135	541	13	575	14	527	13	399	10	15	0
High School	4,047	474	12	481	12	380	9	410	10	2	0
<b>Composition</b>											
4	4,470	487	11	520	12	469	10	310	7	23	1
7	4,403	592	13	620	14	563	13	374	8	25	1
10	4,396	500	11	519	12	395	9	331	8	3	0

**Note:** Students who received more than one accommodation in a single content area test can be counted in multiple columns. Students who received accommodations in more than one content area test administration are counted only once. Students for whom the Special Education bubble was not completed and who did receive these Special Education test administration accommodations are counted here.

<sup>1</sup> The “Presentation” column contains 10 accommodations, two of which are not typically permitted for Reading assessments: “Reading Test Questions” and “Translation of Words or Phrases” are available for Mathematics, Science/Biology, and Composition only.

**Table 10. Number and Percentage of Students Receiving One or More Selected Special Education Test Administration Accommodations in Reading, Mathematics, Science/Biology, and Composition**

Grade	Students with Test Scores	Breaks		Small Group and Individual Administrations		Read or Translate Test Questions (MA, SC and WR only) <sup>1</sup>		Responses Dictated	
		N	%	N	%	N	%	N	%
<b>Reading and/or Mathematics</b>									
2	4,840	304	6	339	7	241	5	76	2
3	4,948	475	10	532	11	399	8	102	2
4	4,624	529	11	578	13	412	9	110	2
5	4,572	587	13	626	14	487	11	123	3
6	4,603	623	14	712	15	527	11	64	1
7	4,567	617	14	699	15	497	11	71	2
8	4,314	524	12	623	14	426	10	63	1
9	4,879	628	13	697	14	252	5	57	1
10	4,765	595	12	622	13	230	5	62	1
<b>Science/Biology</b>									
5	4,523	559	12	588	13	450	10	101	2
8	4,135	486	12	565	14	382	9	56	1
High School	4,047	427	11	464	11	214	5	56	1
<b>Composition</b>									
4	4,470	456	10	509	11	369	8	97	2
7	4,403	531	12	613	14	416	9	70	2
10	4,396	409	9	502	11	172	4	40	1

**Note:** Students who received more than one accommodation in a single content area test can be counted in multiple columns. Students who received accommodations in more than one content area test administration are counted only once. Accommodations are recorded by Test Administrators in the Accommodations section on the biogrid on each student’s answer document.

<sup>1</sup> The “Reading Test Questions” and “Translation of Words or Phrases” accommodations are typically not permitted for the Reading test.

## Section 5. Scoring

This section contains information relevant to *Standards and Assessments Peer Review Guidance*, Critical Element 4.5:

Has the State established clear criteria for the administration, scoring, analysis, and reporting components of its assessment system, including all alternate assessments, and does the State have a system for monitoring and improving the on-going quality of its assessment system?

MC items were scored by CTB using electronic scanning equipment. CR items were scored by human raters who were trained by CTB. Evidence of validity is provided by the procedures for hand-scoring described below.

### Selection of Raters for Scoring

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

#### *Recruitment*

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

#### *The Interview Process*

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

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

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

When Kelly Services Inc. felt applicants were qualified, the applicants were recommended for employment. All assignments were made according to availability and suitability. Before being hired, all employees were required to read, agree to, and sign a nondisclosure agreement outlining the CTB's business ethics and security procedures.

### Training Material Development

Scoring guides for the 2013 CR items were written by CTB's Development staff in conjunction with OSSE and, for Reading Grade 9 DCPS. In 2013, Composition's Understanding Literary

Text and Understanding Informational Text rubric was added to the scoring guides, and also underwent a rangefinding process in DC to identify anchor papers, which represent the exemplars at each score point.

Prior to actual scoring, CTB hand-scoring supervisors studied and internalized the scoring guides along with existing materials that were then used in training raters to hand-score the CR items for all content areas. This ensured consistency in scoring the same items across administrations (such as field test to operational), with the same anchor papers and training philosophy.

### **Preparation and Meeting Logistics for Rangefinding**

Rangefinding is the process of reviewing student responses to newly tested (field tested) items to identify anchor or exemplar papers at each score point. The anchor papers are concrete examples of particular score points and are delineated in the scoring guides used during training and scoring. All DC CAS CR items go through this process prior to operational scoring.

In preparation for rangefinding, CTB hand-scoring supervisors reviewed hundreds of student responses to identify a variety of papers for the reviews. These potential anchors were then assembled for review at rangefinding. During rangefinding, participants were placed in groups of three or more (plus the CTB hand-scoring supervisor/facilitator) to discuss a particular grade and content area, and were involved in discussion of all field test items for that grade. Rubrics were passed out and discussed so that all participants became familiar with the items and the criteria that raters would use to score the student responses after rangefinding. DC participants, along with their CTB facilitator, then reviewed packets containing approximately 35 to 50 responses per item and applied the rubrics and scoring criteria in order to choose appropriate anchor papers.

This process effectively sets the range of each score point for each item. At least one anchor paper for each score point was chosen for every item, and discussion within each group included insights, suggestions, and summary statements for future training on the item. These were recorded by the CTB facilitator. The chosen anchor papers and their final scores were also recorded by the CTB representative, and a DC participant provided sign-off that consensus on the scoring of the items was achieved.

### **Training and Qualifying Procedures**

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

All raters were trained and qualified in specific rater item blocks (RIBs), each of which consisted of a single item to be scored. Raters and team leaders were trained in the following steps:

- Reviewing the student answer booklet
- Reviewing rubrics
- Reviewing anchor papers and training papers and answering questions arising from the established scores
- Explaining scoring strategies, followed by a question-and-answer period
- Administering Qualifying Round 1
- Reviewing Qualifying Round 1 established scores, and answering questions arising from the scores
- Administering Qualifying Round 2 (if necessary)

- Explaining condition codes and sensitive paper procedures
- Explaining nonstandard response or computer-generated response (nsr/cgr) procedures
- Explaining unscannable image procedures

All raters were trained and qualified using the same procedures and criteria used for the team leaders, who had been trained previously. The CTB hand-scoring experts who supervised the training of the team leaders also supervised the training of the raters.

### **Breakdown of Scoring Teams**

Groups of CTB hand-scoring experts oversaw the training and scoring of the CR items for 2013 in Reading, Mathematics, Science/Biology, and Composition. Each expert was responsible for training and scoring all of the items in his or her content area. Teams of raters (numbers of which depended on the content and grade) trained on and scored all the operational items at their respective grades, and some cross-training was done across grades to ensure on-time completion.

Training and scoring of the operational CR items occurred May 16–29, 2013, for Reading, Mathematics, Science/Biology and Composition. Training and scoring of the field test CR items occurred July 16–22, 2013. Training consisted of a review of the rubrics, followed by an analysis of the anchor papers for each item. Raters then participated in qualifying rounds, which consisted of ten books of sample papers for the item in a given RIB. Raters were given two opportunities to achieve acceptable qualification ratings; those not meeting the minimum were dismissed.

### **Monitoring the Scoring Process**

After training was completed and live scoring began, a number of quality control measures were put in place to ensure that books were scored accurately and that raters' scoring ability remained consistent in scoring accuracy.

Throughout the course of hand-scoring, calibration sets of pre-scored papers (checksets/validity sets) were administered daily to each rater to monitor scoring accuracy and to maintain a consistent focus on the established rubrics and guidelines. Approximately 3% of books that the raters received were “checkset” papers rather than live books, where the checksets were “blind” or unknown to the rater. Raters whose checkset accuracy repeatedly dipped below the quality standards were flagged and retrained. In addition to the checkset process, CTB's hand-scoring protocol included the use of read-behinds (spot checks during live scoring). The read-behind was another valuable rater-reliability monitoring technique that allowed a team leader to review a rater's scored documents, providing feedback and counseling as appropriate. The CTB Data Monitoring staff also ran inter-rater reliability reports throughout live scoring to look for raters who were struggling and in need of retraining. Retraining involved a one-on-one discussion between the supervisor (or team leader) and the rater, who discussed the problem item(s) as well as the scoring guides and, if necessary, training papers. If the rater's accuracy on checkset scores did not meet the quality standards after this retraining, the rater was dismissed from the project.

Approximately 10% of all DC CAS tests were scored by a second rater to establish inter-rater reliability statistics for all CR items, results of which are provided in Section 8. This procedure is called a “double-blind read” because the second rater does not know the first rater's score.

***Scoring Security***

All raters had to sign nondisclosure forms indicating that they were not to disclose the items they were scoring. Security guards were on-site whenever employees were present in the building. All employees were issued identification badges and were required to wear them in plain view at all times. Visitors and employees who forgot their badges were issued visitors' badges and were required to wear them in plain view. All employees and visitors were subject to inspection of their personal effects.

## Section 6. Methods

This section contains information relevant to *Standards and Assessments Peer Review Guidance*, Critical Elements 4.4, 4.5, 4.6, and 5.6.

### 4.4

When different test forms or formats are used, the State must ensure that the meaning and interpretation of results are consistent.

(a) Has the State taken steps to ensure consistency of test forms over time?

### 4.5

Has the State established clear criteria for the administration, scoring, analysis, and reporting components of its assessment system, including all alternate assessments, and does the State have a system for monitoring and improving the on-going quality of its assessment system?

### 5.6

Assessment results must be expressed in terms of the achievement standards, not just scale scores or percentiles.

This section describes the methods used to analyze the item and test level data for the DC CAS. Results of the item and test level analyses described here are provided as evidence for reliability and validity in Section 8.

### Classical Item Level Analyses

Each operational test item was first reviewed in terms of classical raw score statistics. Each item's frequency distribution (number of students responding for each answer choice or score level) as well as each item's overall  $p$  value (proportion of students choosing the correct answer) and point-biserial or item-test correlation (the correlation between the score on the individual item and the sum of all item scores) were reviewed. Typically,  $p$  values should range between 0.30 and 0.90. Items with a  $p$  value less than 0.30 are considered more difficult since less than 30% of the students are getting the correct answer. Values larger than 0.90 indicate a fairly easy item, with more than 90% of students getting the correct answer. With newly tested content, the  $p$  value may dip lower than 0.30, at which point the item should be evaluated in light of the newness of content or students' opportunity to learn the content. Point-biserial correlations are usually in the range of 0.30 and above, although some items can be acceptable with a point-biserial correlation as low as 0.15. The point-biserial correlations of each item's distractors, or incorrect response options, were also computed for the MC items. Items with positive point-biserial correlations for distractors were reviewed for the possibility of more than one correct response option, or no correct response option.

It is also important to track the rate at which students do not respond to, or omit, items. Omitted items receive a zero score in scoring, but they are ignored in the adjusted  $p$  value computation. The rate of omission often provides some information about test times, or speededness, particularly if there is a high rate of items omitted at the end of a test session. It also provides an indication of items that may simply be unclear or illogically presented. When more than 5% of students omit an item, the item is reviewed by both CTB Research and Development and shared with OSSE.

## Item Bias Analyses

Differential item functioning (DIF) occurs when students with the same underlying ability, but from different demographic groups, have a different probability of obtaining a specific score on an item. Differential item functioning may be an indication that the item is biased in favor of or against a certain subgroup. To evaluate the potential bias, items are first reviewed from content perspectives. All items are screened in Content and Bias Review meetings comprised of DC educators to ensure that no obviously sensitive terms, phrases, scenarios, or illustrations that could influence examinee performance appear in the DC CAS items prior to field testing and selection for operational test forms.

For the DC CAS program, CTB uses Mantel-Haenszel statistics (Mantel & Haenszel, 1959) to evaluate DIF for both operational and field test items. The subgroups compared in the DIF analyses for the 2013 administration reflect conventional subgroupings, and were based on gender (male—reference and female—focal) and race/ethnicity (African American—reference, and Asian, Hispanic, and White—focal). As with all statistical tests, Mantel-Haenszel DIF statistics are subject to Type I and II errors. An item flagged for DIF may or may not provide an unfair advantage or disadvantage for one examinee subgroup compared with another. However, the flag does show when an item is more difficult for a particular focal subgroup of students than for the comparison or reference subgroup with equivalent total test scores OSSE and CTB screen all items that are flagged for DIF after each administration to identify items that may favor or disadvantage examinee subgroups.

For MC items, the Mantel-Haenszel ( $\chi^2_{MH}$ ) statistic flags items for potential DIF using the following criteria:

- B level DIF, where a “B” indicates DIF and has a value of the Mantel-Haenszel ( $\Delta_{MH}$ ) that is significantly different from zero (at the 0.05 level) and  $-1.5 \leq \Delta_{MH} \leq -1$  or  $1 \leq \Delta_{MH} \leq 1.5$ .
- C level DIF, where a “C” indicates DIF and has a value of the Mantel-Haenszel ( $\Delta_{MH}$ ) that is significantly different from zero (at the 0.05 level) and  $|\Delta_{MH}|$  exceeds 1.5.

Positive DIF values indicate items that favor the focal group, while negative values indicate items that disadvantage the focal group. In the tables, B and C indicate positive DIF values, and B- and C- indicate negative DIF values.

For CR items, an effect size (ES) statistic based on the Mantel  $\chi^2$  is used to flag items for potential DIF. ES is obtained by dividing the standardized mean difference (SMD) statistics by the standard deviation of the item. Items are flagged using the same rules that are used in National Assessment of Educational Progress (NAEP):

- B level, where the Mantel statistic is significant ( $p < 0.05$ ) and  $-0.25 \leq ES \leq -0.17$  or  $0.17 \leq ES \leq 0.25$
- C level, where the Mantel statistic is significant ( $p < 0.05$ ) and  $ES < -0.25$  or  $ES > 0.25$

C level flags indicate moderate to severe DIF. B level flags indicate moderate DIF. A level flags indicate negligible DIF. (A detailed description of these procedures can be found in Zwick, Donoghue, & Grima, 1993.)

## Calibration and Linking

Scaling and linking was accomplished using the PARDUX and SAS computer programs to implement the three-parameter logistic model (3PL) and the two-parameter partial-credit (2PPC) IRT models for item calibration and scaling, and the Stocking and Lord (1983) procedure was used for linking. These software programs were developed at CTB/McGraw-Hill Education to enable scaling and linking of complex assessment data.

In PARDUX, a marginal maximum likelihood procedure was used to simultaneously estimate the item parameters under the 3PL model (used for MC items) and the 2PPC model (used for CR items) (Bock & Aitkin, 1981; Thissen, 1982). These models were implemented using the microcomputer program PARDUX (CTB/McGraw-Hill Education, 2011). For setting the 2006 base scales for Reading and Mathematics, all scales were also calibrated in PARSCALE (Muraki & Bock, 1991) as verification of the PARDUX results.

Under the 3PL model, the probability that a student with trait or scale score  $\theta$  responds correctly to MC item  $j$  is as follows:

$$P_j(\theta) = c_j + (1 - c_j) / [1 + \exp(-1.7a_j(\theta - b_j))]. \quad (1)$$

In equation (1),  $a_j$  is the item discrimination,  $b_j$  is the item difficulty, and  $c_j$  is the probability of a correct response by a very low-scoring student. The 2PPC model holds that the probability that a student with trait or scale score  $\theta$  will respond in category  $k$  to partial-credit item  $j$  is given by

$$P_{jk}(\theta) = \exp(z_{jk}) / \sum_{i=1}^{m_j} \exp(z_{ji}), \quad (2)$$

where  $z_{jk} = (k - 1)f_j - \sum_{i=0}^{k-1} g_{ji}$ , and  $g_{j0} = 0$  for all  $j$ .

The summary output of the above equations is in two different metrics corresponding to the two item response models (3PL and 2PPC). The location and discrimination parameters for the MC items are in the traditional 3PL metric (labeled  $b$  and  $a$ , respectively). In the 2PPC model,  $f$  (alpha) and  $g$  (gamma) are analogous to  $b$  and  $a$ , where alpha is the discrimination parameter and gamma over alpha ( $g/f$ ) is the location where adjacent trace lines cross on the ability scale. Because of the different metrics used, the 3PL parameters  $b$  and  $a$  are not directly comparable to the 2PPC parameters  $f$  and  $g$ ; however, they can be converted to a common metric. The two metrics are related by  $b = g/f$  and  $a = f/1.7$  (CTB/McGraw-Hill Education, 2011). Application of this procedure locates both the MC and CR items on the same scale. Note that for the 2PPC model there are  $m_j - 1$  (where  $m_j$  is a score level  $j$ ), independent  $g$ 's, and one  $f$ , for a total of  $m_j$  independent parameters estimated for each item, while there is one  $a$  and one  $b$  per item in the 3PL model.

## Goodness of Fit

Goodness-of-fit statistics were computed for each item to examine how closely the item's data conform to the item response models. A procedure described by Yen (1981) was used to measure fit. In this procedure, students are rank ordered on the basis of their  $\hat{\theta}$  values and sorted into ten cells with 10% of the sample in each cell. Each item  $j$  in each decile  $i$  has a response from  $N_{ij}$

examinees. The fitted IRT models are used to calculate an expected proportion  $E_{ijk}$  of examinees who respond to item  $j$  in category  $k$ . The observed proportion  $O_{ijk}$  is also tabulated for each decile, and the approximate chi-square statistic:

$$Q_{1j} = \sum_{i=1}^{10} \sum_{k=1}^{m_j} \frac{N_{ij} (O_{ijk} - E_{ijk})^2}{E_{ijk}},$$

$Q_{1j}$  should be approximately chi-square distributed with degrees of freedom ( $DF$ ) equal to the number of “independent” cells,  $10(m_j - 1)$ , minus the number of estimated parameters. For the 3PL model,  $m_j = 2$ , so  $DF = 10(2-1) - 3 = 7$ . For the 2PPC model,  $DF = 10(m_j - 1) - m_j = 9m_j - 1$ . Since  $DF$  differs between MC and CR items and among CR items with different score levels  $m_j$ ,  $Q_{1j}$  is transformed, yielding the test statistic:

$$Z_j = \frac{Q_{1j} - DF}{\sqrt{2DF}}, \text{ where } j = \text{item } j.$$

This statistic is useful for flagging items that fit relatively poorly.  $Z_j$  is sensitive to sample size, and cut-off values for flagging an item based on  $Z_j$  have been developed and were used to identify items for the item review. The cut-off value is  $(N/1500 \times 4)$  for a given test, where  $N$  is the sample size.

Model-fit information is obtained from the  $Z$ -statistic. The  $Z$ -statistic is a transformation of the chi-square ( $Q_1$ ) statistic that takes into account differing numbers of score levels as well as sample size.

The  $Z$ -statistic is an index of the degree to which obtained proportions of students with each item score are close to the proportions that would be predicted by the estimated thetas and item parameters. These values are computed for ten intervals corresponding to deciles of the theta distribution (Burket, 1995). The  $Z$ -statistic is used to characterize item fit. The critical value of  $Z$  is different for each grade because it is dependent on sample size.

Evidence of the validity of the scalings is provided by model fit. If the IRT model fits the empirical item response distributions for the population we want to generalize to (i.e., DC students), then the claim that the scores are valid indicators of an underlying proficiency is strengthened. Fit statistics indicate the degree of difference between (a) expected probabilities of correct responses at each proficiency level and (b) observed probabilities examined when items are field tested and when they are used operationally. Table 11 indicates that only small numbers of operational items were flagged for poor fit to the IRT model. No items were removed from operational scaling and scoring due to poor fit.

## Year-to-Year Linking Procedures

Once the IRT scaling is accomplished, linking the scale across years enables comparability of scores from one year to the next and across all test forms in the same content area and grade. In 2007 through 2013, anchor item sets that link the current test forms to the previous year’s scale were used in a Stocking and Lord (1983) linking methodology.

The Stocking and Lord (1983) procedure, also called test characteristic curve (TCC) method, was used to place each grade on the horizontal scale that had been developed for each content area. It minimizes the mean squared difference between the two characteristic curves, one based on estimates from the previous calibration (reference form) and the other on transformed estimates from the current calibration (operational form). Let  $\hat{\psi}_q$  be the test characteristic curve at quadrature point  $\theta_q$  based on estimates from the previous year and  $\hat{\psi}_q^*$  be the test characteristic curve based on transformed estimates from the current year:

$$\hat{\psi}_q = \hat{\psi}(\theta_q) = \sum_{j=1}^n P_j(\theta_q; a_j, b_j, c_j),$$

$$\hat{\psi}_q^* = \hat{\psi}(\theta_q) = \sum_{j=1}^n P_j(\theta_q; \frac{a_j}{M_1}, M_1 b_j + M_2, c_j).$$

The TCC method determines the scaling constants (multiplicative -- M1 and additive -- M2) by minimizing the following quadratic loss function (F):

$$F = \frac{1}{N} \sum_{q=1}^N (\hat{\psi}_q - \hat{\psi}_q^*)^2$$

where  $N$  is total number of quadrature points.

All anchor items are MC items. Anchor items are rotated in and out of use each year, to the degree possible, to minimize item over-exposure. Anchor items are placed in approximately the same location or same third of the test form across administrations. Anchor item  $a$  and  $b$  parameters are calibrated separately for each administration. The number and representativeness of the anchor items relative to the overall test and blueprints is provided in Tables 1–4. The blueprint should be proportionally represented in the anchor sets.

Because Composition prompts are so few, the “items” or scores from each of the Writing rubrics were linked to the Reading scale by first matching students’ Reading and Composition item-level scored responses. The Reading operational MC items were treated as anchors and the Stocking and Lord common-item linking procedure was conducted.

Once calibrated, the anchor item set and linking results are carefully reviewed to ensure that it is performing very similarly across administrations. Standard CTB Research team quality checks are followed during calibration and linking analyses for all grades and content areas. Additional anchor item checks were conducted for items flagged in any of the following verifications, which were performed to ensure the quality and accuracy of the linking:

1. Correlation coefficients for the two sets of estimated IRT item parameters (reference and operational form) should be very high (0.90–1.00). Specifically, differential anchor item performance between the 2012 and 2013 administrations was evaluated by comparing the correlations between the reference and new form item difficulty ( $b$  parameter), discrimination ( $a$  parameter), and proportion correct ( $p$  value) values after linking. IRT guessing ( $c$ ) parameters typically fluctuate considerably, and were held to fixed values during linking, and were not considered in this evaluation. The correlations are shown in

Table 12 for the discrimination ( $a$ ) and difficulty ( $b$ ) parameters and are moderate to high, ranging from 0.71 to 0.96 for  $a$  parameters (0.85–0.97 in 2012) and from 0.89 to 0.99 for  $b$  parameters (0.96–0.99 in 2012). For several grades, the 2013 correlations are slightly lower than 2012, but these values are within the acceptable range that we typically see in large scale assessments. These correlations indicate that the items performed similarly in the two administrations and provide evidence that the linking results are reasonable and accurate.

2. Reference and linked anchor item parameters and TCCs should be closely aligned. The TCCs are reviewed after each linking cycle for each grade and content area. Further, statistical differences were evaluated with four difference statistics: root mean squared difference, mean absolute difference, maximum absolute difference, and the absolute value of the mean signed difference.
3. The scaling constants, or Stocking-Lord linear transformation parameters, should be fairly stable across administrations. There are two constants, a multiplicative constant ( $M1$ ) and an additive constant ( $M2$ ). Because PARDUX calibrations center the IRT scale close to the average proficiency of the test takers, the magnitude of the 2012–2013 differences in these scaling constants indicates the degree of differences in average difficulty of the reference and new test form administrations. The scaling constants from the 2013 administration along with constants across the 2007–current years of the DC CAS administration and scales are provided in Table 13.
4. The anchor  $p$  values for the estimated new form and the reference form should be similar and aligned on a regression line, showing the same direction and magnitude of change as do the scale scores. The correlations of the anchor item  $p$  values in Table 12 are highly correlated, ranging from 0.87 to 0.99 (0.96–0.99 in 2012) for all grades and content areas. The 2013 correlations are slightly lower than 2012 but they are still high ( $> 0.87$ ). This is an indication that the anchor items performed similarly in the examinee populations in 2012 and 2013.

Once the tests are linked, final parameter tables are developed into scoring tables, from which each student's scale score is derived. Examinee scale scores are estimated for DC CAS using number correct scoring.

### **Establishing Upper and Lower Bounds for the Grade Level Scales**

Upper and lower bound scale scores are called the lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS). A maximum likelihood procedure cannot produce scale score estimates for students with perfect scores or scores below the level expected from guessing. Also, while maximum likelihood estimates are available for students with extreme scores other than zero or perfect scores, occasionally these estimates have standard errors of measurement that are very large, and differences between these extreme values have very little meaning. Therefore, scores are established for these students based on a rational but necessarily non-maximum likelihood procedure.

For the DC CAS, LOSS and HOSS were set to be equal at the same grade for each content area. For example, the Grade 3 LOSS and HOSS are 300 and 399, (respectively) and the Grade 5 LOSS and HOSS are 500 and 599, respectively, for Reading, Mathematics, and Science. These values were established on the 2006 base scale for Reading and Mathematics,

the 2008 base scale for Science/Biology, the 2011 base scale for Reading Grade 9, the 2012 base scale for Reading and Mathematics Grade 2, and the 2012 base scale for Composition. These values remain constant from year to year. The LOSS and HOSS for all grades are provided in Table 14.

## Reliability Coefficients

Total test reliability statistics measure the level of consistency of performance over all test questions in a given form, the results of which imply how well the questions measure the content domain and could continue to do so over repeated administrations. Total test reliability coefficients (in this case measured by Cronbach's alpha [ $\alpha$ ; 1951]) may range from 0 to 1, where 1 refers to a perfectly reliable test. The DC CAS reliability data are based on DC students in the calibration sample with an average sample size of approximately 4,500 students per grade/content.

The total test reliabilities of the operational forms were evaluated first by Cronbach's  $\alpha$  index of internal consistency. The estimate for Cronbach's  $\alpha$  is:

$$\hat{\alpha} = \frac{k}{k-1} \left( 1 - \frac{\sum \hat{\sigma}_j^2}{\hat{\sigma}_x^2} \right), \quad (8.1)$$

where  $k$  is the number of items on the test form,  $\hat{\sigma}_j^2$  is the variance of item  $j$ , and  $\hat{\sigma}_x^2$  is the total test variance.

The stratified coefficient alpha is an internal consistency score reliability index. It measures the internal consistency of a test that contains both MC and CR items. The stratified alpha treats the MC and CR sections as separate subtests, estimates the reliability of the two subtests, and combines those estimates to estimate total test score internal consistency.

The Feldt-Raju index is a third index of internal consistency. It is also designed for mixed-format tests. Unlike the stratified alpha, the Feldt-Raju corrects the underestimation of Cronbach's alpha, which assumes that tests are parallel in classical test theory terms; mixed format tests are more appropriately assumed to be congeneric.

As a rule of thumb, reliability coefficients for test scores that are equal to or greater than 0.80 are considered acceptable for tests of moderate lengths. All of the reliability indices calculated provide evidence that these tests are performing as expected and that they support inferences about what students know and can do in relation to the content knowledge and skills that the tests target.

## Standard Errors of Measurement

Whereas reliability coefficients indicate the degree of consistency in test scores, the standard error of measurement (SEM) indicates the degree of unreliability in test scores. The standard error is an estimate of the standard deviation of observed scores to expect if an examinee were retested under unchanged conditions. Conditional standard deviations of observed scores can be found for each score level. The conditional estimate of measurement error increases as the number of items that coincide with examinees' levels of performance decreases. Generally, there are few students with extreme scores; these score levels are measured less accurately than

moderate scores. If all of the items are very difficult or very easy for examinees, the error of measurement will be larger than when the items' difficulties are distributed across the ability levels of the students being tested.

In addition to classic internal consistency reliability coefficients, the SEM based on IRT is also provided as reliability evidence for the DC CAS scores. The IRT SEM provides conditional standard errors that are specific to each scale score. These standard errors were estimated as a function of the scale scores using IRT. Accuracy of measurement is especially important when applied to individual scores. The IRT-based SEM indicates the expected standard deviation of observed scores if an examinee at a specific level of ability were tested repeatedly under unchanged conditions.

## **Proficiency Level Analyses**

One of the cornerstones of the NCLB Act (US DOE, 2002) is the measurement of Adequate Yearly Progress (AYP) for states with respect to the percentage of students at or above the academic performance standards established by states. Because of a heavy emphasis on moving all students to or above the "Proficient" category by the year 2014, the consistency and accuracy of the classification of students into these proficiency categories is of particular interest. The statistical quality of cut scores that define the proficiency levels in which students are placed per their performance serves as additional validity evidence. Details about the standard setting workshops and Bookmark Standard Setting Procedure used to set the cut scores are given in the DC CAS Cut Score Setting Technical Report (CTB/McGraw-Hill, 2012). It may be useful to note that the Bookmark procedure (Mitzel, Lewis, Patz, & Green, 2001) is a well-documented and highly regarded procedure that has been demonstrated by independent research to produce reasonable cut scores on tests across the country.

It is also important to review the specific scale score SEM for each cut score. Comparison of these SEMs to the SEMs associated with other DC CAS scale scores for each test should almost always be among the lowest, meaning that the DC CAS tests tend to measure most accurately near the cut score. This is a desirable quality when cut scores are used to classify examinees.

## **Classification Consistency**

Not only is it important that the amount of measurement error around the cut score be minimal; also important is the expected consistency with which students would be classified into performance levels if given the test over repeated occasions. Classification consistency, or decision consistency, is defined as the extent to which the classifications of examinees agree on the basis of two independent administrations of a test or administration of two parallel test forms. However, it is practically infeasible to obtain data from repeated administrations of a test because of cost, time, and students' recall and growth of the first administration. Therefore, a common practice is to estimate decision consistency from one administration of a test.

## **Classification Accuracy**

Classification accuracy, or decision accuracy, is defined as the extent to which the actual classifications of test-takers based on observed test scores agree with classifications that would be made on the basis of their true scores (Livingston & Lewis, 1995). It is common practice to estimate decision accuracy using a psychometric model to estimate true scores that correspond to observed scores as the basis for estimating classification accuracy. In other words, classification

*consistency* refers to the agreement between two observed scores, while classification *accuracy* refers to the agreement between the observed score and the estimated true score.

A straightforward classification consistency estimation can be expressed in terms of a contingency table representing the probability of a particular classification outcome under specific scenarios. For example, the table below is a contingency table of (H+1) rows × (H+1) columns, where H is the number of cut scores, such that two cut scores yield a 3×3 contingency table.

**Example of Contingency Table with Two Cut Scores**

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

Hambleton and Novick (1973) proposed  $P$  as a measure of classification consistency, where  $P$  is defined as the sum of the diagonal values of the contingency table (shaded above):

$$P = P_{11} + P_{22} + P_{33}.$$

To account for statistical chance agreement, Swaminathan, Hambleton, & Algina (1974) suggested using Cohen’s kappa (1960):

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

where  $P_c$  is the chance probability of a consistent classification under two completely independent assignments. This probability,  $P_c$ , is the sum of the probabilities obtained by multiplying the marginal probability of the first administration and the corresponding marginal probability of the second administration:

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

Kolen and Kim (2005) suggested a method for estimating consistency and accuracy that involves the generation of item responses using item parameters based on the IRT model (see also Kim, Choi, Um, & Kim, 2006, as well as Kim, Barton, & Kim, 2008). Two sets of item responses are generated using a set of item parameters and an examinee’s ability distribution from a single test administration.

CTB used the KKCLASS program (Kim, 2007) to calculate these statistics on the 2013 DC CAS results. The KKCLASS program implements an IRT-based procedure that is consistent with DC CAS IRT scaling and scoring. The procedure is described below.

Step 1: Obtain item parameters and ability distribution weights ( $\hat{g}(\theta_q)$ ) at each quadrature point  $q$  from a single test.

Step 2: Compute two raw scores at each quadrature point. At a given quadrature point

$\theta_q$ , simulate two sets of item responses using the item parameters from a test form, assuming that the same test form was administered twice to an examinee with the true ability  $\theta_q$ .

Step 3: Construct a classification matrix at each quadrature point. Determine the joint event for the cells in the contingency table using the raw scores obtained from Step 2.

Step 4: Repeat Steps 2 and 3  $R$  times and get average values over  $R$  replications.

Step 5: Multiply distribution weight ( $\hat{g}(\theta_q)$ ) by average values in Step 4 for each quadrature point, and sum across all quadrature points. From this final contingency table, classification consistency indices, such as consistency agreement and kappa, can be computed.

Step 6: Because examinees' abilities are estimated at each quadrature point, this quadrature point can be considered the true score. Therefore, classification accuracy is computed using both examinees' estimated abilities (observed scores) and quadrature point (true score).

**Table 11. DC CAS 2013 Numbers of Operational Items Flagged for Poor Fit During Calibration**

<b>Content</b>	<b>Grade</b>	<b>Flagged for Poor Fit</b>
<b>Reading</b>	2	1
	3	0
	4	1
	5	1
	6	0
	7	1
	8	0
	9	1
	10	0
<b>Mathematics</b>	2	1
	3	1
	4	1
	5	2
	6	2
	7	0
	8	0
	10	3
<b>Science/Biology</b>	5	1
	8	1
	High School	2
<b>Composition</b>	4	2
	7	1
	10	3

**Table 12. Correlations Between the Item Parameters for the Reference Form and 2013 DC CAS Operational Test Form**

<b>Grade</b>	<b>Discrimination (a)</b>	<b>Difficulty (b)</b>	<b>P Value Correlation</b>
<b>Reading</b>			
2	0.91	0.97	0.98
3	0.72	0.97	0.97
4	0.89	0.97	0.99
5	0.91	0.95	0.95
6	0.92	0.96	0.97
7	0.82	0.98	0.99
8	0.88	0.99	0.99
9	0.85	0.90	0.87
10	0.93	0.96	0.97
<b>Mathematics</b>			
2	0.87	0.98	0.98
3	0.91	0.97	0.97
4	0.86	0.97	0.97
5	0.72	0.96	0.96
6	0.83	0.95	0.96
7	0.90	0.98	0.96
8	0.71	0.95	0.94
10	0.73	0.89	0.88
<b>Science/Biology</b>			
5	0.94	0.97	0.98
8	0.92	0.96	0.98
High School	0.96	0.96	0.95

**Table 13. Scaling Constants Across Administrations, All Grades and Content Areas**

Grade	2007		2008		2009		2010		2011		2012		2013	
	Mult	Add												
<b>Reading</b>														
2	N/A	N/A	11.96	245.18										
3	10.40	352.60	10.70	354.00	10.70	353.10	14.30	349.60	13.60	350.40	13.09	349.66	13.31	351.27
4	11.80	451.20	11.70	453.30	12.40	453.40	13.40	451.60	13.50	451.10	12.22	453.49	11.97	454.99
5	11.40	552.20	11.30	554.90	11.40	553.70	12.40	553.20	12.20	554.20	12.23	555.64	10.95	555.85
6	10.80	652.10	10.40	652.90	10.40	653.00	11.40	651.60	11.20	652.70	11.39	652.04	12.17	651.49
7	10.40	751.30	10.40	752.70	10.20	754.70	11.50	754.30	11.60	754.30	11.55	755.52	11.05	757.47
8	11.10	851.80	10.40	853.80	11.10	853.50	12.30	854.60	12.00	856.90	11.75	855.29	11.38	856.60
9	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	13.50	950.00	13.85	948.32	13.27	951.81
10	11.30	954.50	10.90	953.40	10.70	954.10	12.10	952.10	13.00	955.60	12.60	954.87	13.16	954.91
<b>Mathematics</b>														
2	N/A	N/A	11.07	253.82										
3	14.50	353.90	16.20	354.00	17.30	357.00	16.70	352.40	17.30	353.80	15.87	353.15	16.15	355.66
4	14.10	452.10	13.20	456.40	14.10	457.90	13.80	455.10	14.10	455.10	13.68	457.78	14.09	459.90
5	14.10	552.20	14.80	555.90	15.10	556.40	14.20	556.80	14.70	557.00	14.62	558.84	14.13	560.31
6	13.40	647.30	14.50	649.80	14.30	650.10	14.30	650.30	14.20	652.40	14.36	652.35	15.91	653.87
7	13.70	746.90	13.40	750.00	14.60	751.00	15.10	751.20	14.50	753.60	15.02	754.39	14.90	754.49
8	13.00	844.80	12.50	847.40	12.90	848.50	13.50	848.60	13.00	851.60	13.08	851.59	12.97	855.54
10	15.50	945.20	16.90	945.40	16.70	947.30	16.50	944.30	16.20	948.00	16.25	948.32	16.72	947.84
<b>Science/Biology</b>														
5	N/A	N/A	8.00	550.00	8.70	549.90	9.10	549.20	9.20	549.40	8.91	550.22	8.99	550.33
8	N/A	N/A	8.00	850.00	8.90	851.00	9.40	851.90	9.40	852.90	9.52	852.86	8.70	853.34
High School	N/A	N/A	8.00	950.00	7.70	946.60	7.50	949.50	7.80	949.90	8.36	950.88	8.65	949.62
<b>Composition</b>														
4	N/A	N/A	12.78	452.38	12.19	454.07								
7	N/A	N/A	11.34	754.68	10.84	756.66								
10	N/A	N/A	13.11	952.72	13.15	953.82								

**Table 14. LOSS and HOSS for Relevant Grades in Reading, Mathematics, Science/Biology and Composition**

<b>Grade</b>	<b>LOSS</b>	<b>HOSS</b>
2	200	299
3	300	399
4	400	499
5	500	599
6	600	699
7	700	799
8	800	899
9	900	999
10	900	999

## Section 7. Standard Setting

This section contains information relevant to the *Standards and Assessments Peer Review Guidance*, Critical Elements 2.1, 2.2, and 2.3:

### 2.1

Has the State formally approved/adopted challenging academic achievement standards in Reading/Language Arts and Mathematics for each of Grades 3 through 8 and for the 10-12 grade range? These standards were to be completed by school year 2005–2006.

### 2.2

Has the State formally approved/adopted academic achievement descriptors in Science for each of the grade spans 3-5, 6-9, and 10-12 as required by school year 2005–06?

### 2.3

1. Do these academic achievement standards (including modified and alternate academic achievement standards, if applicable) include for each content area--

(a) At least three levels of achievement, including two levels of high achievement (proficient and advanced) that determine how well students are mastering a State’s academic content standards and a third level of achievement (basic) to provide information about the progress of lower-achieving students toward mastering the proficient and advanced levels of achievement; *and*

(b) Descriptions of the competencies associated with each achievement level; *and*

(c) Assessment scores (“cut scores”) that differentiate among the achievement levels and a rationale and procedure used to determine each achievement level?

The DC CAS cut scores associated with each of the four proficiency levels (*Below Basic, Basic, Proficient, Advanced*) for each grade and content area were set using content, statistical, and policy-based standard setting processes. The content and related statistics were reviewed through standard setting workshops conducted with DC teachers and the resulting cut score recommendations were provided to the DC Technical Advisory Council and OSSE for review and approvals.

Standard settings were conducted to determine cut scores in 2006 for Reading and Mathematics Grades 3–8 and 10; in 2008 for Science Grades 5, 8, and High School Biology; in 2011 for Reading Grade 9; in 2012 for Reading and Mathematics Grade 2 and Composition Grades 4, 7, and 10. Additional standard setting activities were conducted for Reading in 2012 and for Mathematics in 2013, but they did not result in changes to the cut scores. These standard settings are summarized in this section.

The Bookmark Standard Setting Procedure (BSSP; Lewis, Mitzel, & Green, 1996; Lewis, Mitzel, Mercado, & Schultz, 2012) was implemented to establish performance standards for DC CAS Reading, Mathematics, and Science/Biology. The Reading and Mathematics Grades 3–8 and 10 cut scores were established in 2006 and the Science/Biology cut scores were established in 2008. Cut scores for Grade 2 Reading and Mathematics were established in 2012. In 2012 and 2013, DC educators reviewed the cut scores for the Reading and Mathematics assessments, respectively; however, the cut scores for the Reading and Mathematics tests were not adjusted from the values established in 2006.

The 2006 standard setting workshops for Mathematics and Reading lasted four-and-a-half days, with the morning of the first day devoted to orientation and bookmark training, two-and-a-half days to standard setting, and one-and-a-half days to description writing. Participants recommended three cut scores at the Basic, Proficient, and Advanced levels, which would separate students into four performance levels: Below Basic, Basic, Proficient, and Advanced. Participants engaged in training, discussion, and three rounds of bookmark placements. The table leaders reviewed the participant-recommended cut scores and associated impact data and suggested changes to promote cross-grade articulation. Impact data are the percentages of students who are classified in each performance level based on the recommended cut scores.

The Judgmental Policy Capturing procedure (Jaeger, 1995) was implemented to set standards for the Composition assessments in Grades 4, 7, and 10 in 2012. Judgmental Policy Capturing is a rubric-centered, content-based method that has been used in recent years to establish performance standards on unscaled assessments, (see Jaeger, 1995; Perie, 2007; Roeber, 2002). During the one-and-a-half-day procedure, DC educators were trained to examine the DC CAS scoring rubrics and to consider the knowledge and skills associated with the attainment of each successive score level. Two separate rubrics were used to score the Composition tests: students received 0–6 points for Topic Development, and 0–4 points for Standard English Conventions. A third rubric for Composition, Literary Analysis (0–4 points), was also used to score students' responses; however, scores from this rubric did not contribute to students' total scores in 2012. Participants studied these scoring rubrics, the DC CAS content standards, and performance level descriptions and discussed their expectations of the knowledge and skills students must have in order to associate a score level with a performance level for each writing prompt.

The cut score recommendations from the committees for all content areas and grades were reviewed by the DC CAS Technical Advisory Committee and DCPS in 2006 and 2012 and by OSSE in 2008 and 2012. Certain cut scores were adjusted each time to achieve articulated standards and impact data. The DC Board of Education approved these cut scores in 2006 and 2008, and the OSSE approved the cut scores for Composition and for Grade 2 Reading and Mathematics in 2012.

## **Reading and Mathematics Cut Score Reviews**

In 2012 and 2013, the OSSE sponsored cut score reviews for Reading Grades 3–10 and for Mathematics Grades 3–8 and 10, respectively. The purpose of these cut score reviews was to determine whether the cut scores for the assessments should be adjusted to reflect changes to the tests and content standards.

For Reading Grades 3–10, the Bookmark Standard Setting Procedure (BSSP; Lewis, Mitzel, & Green, 1996; Lewis, Mitzel, Mercado, & Schulz, 2012) was implemented to review the cut scores. The workshop lasted one-and-a-half days.

Participants engaged in training, discussion, and two rounds of bookmark placements. Participants reviewed the existing cut scores and recommended adjustments for content-based reasons. Participants also adjusted the performance level descriptions (PLDs) to improve their clarity and alignment with the tested content.

For Mathematics Grades 3–8 and 10, a two-phased standard setting process was conducted using a content sorting methodology modeled after the Yes/No modified Angoff procedure (Plake & Cizek, 2012), followed by a Briefing Book workshop (Hartel & Miles, 2012). In the first phase,

participants studied the test items and made content-based recommendations regarding the knowledge, skills, and abilities that could be mastered by students in each performance level. These participants then recommended adjustments to the PLDs to improve their clarity and alignment with the tested content. In the second phase, a smaller committee reviewed the cut scores recommended by the first committee in light of students' actual performance on the Mathematics tests. The second committee then reviewed the consistency of the cut scores across grades and recommended adjustments to the cut scores to improve that consistency.

### **Final, Approved DC CAS Cut Scores**

The cut score recommendations from the 2012 and 2013 cut score review committees were reviewed by staff from the OSSE in those respective years. In addition, the OSSE reviewed the impact data associated with the recommended cut scores: impact data are the percentages of students classified in each performance level, based on the cut scores. The Reading cut scores from the 2012 committee, as recommended by District of Columbia educators, were reviewed by the Technical Advisory Committee and approved by the OSSE in August 2012.

After the 2013 cut score review, OSSE determined that the cut scores established in 2006 for the tests in Reading Grades 3–10 and Mathematics Grades 3–8 and 10 were still appropriate for DC CAS and that no change to the cut scores was warranted. Accordingly, those existing cut scores remained in place and were applied to the tests administered in Spring 2013. (Note that the Reading cut scores recommended during the 2012 cut score review were not implemented in Spring 2012; accordingly, the cut scores for Reading Grades 3–10 were never adjusted operationally from those established in 2006.) For transparency, OSSE released information about the cut scores recommended during the 2012 and 2013 cut score reviews, including the potential impact data associated with those recommended cut scores, in Summer 2013.

The standard setting technical reports summarize procedures and results of the 2012 and 2013 cut score reviews. The reports include round-by-round synopses, agendas, training materials, and the recommended cut scores. See District of Columbia Comprehensive Assessment System (DC CAS) Standard Setting Technical Report 2012 (CTB/McGraw-Hill, 2012) and the District of Columbia Comprehensive Assessment System (DC CAS) Cut Score Review Technical Report 2013 (CTB/McGraw-Hill, 2013).

Table 15 shows the final, approved cut scores for the DC CAS program.

**Table 15. Final Cut Score Ranges**

<b>Reading</b>				
<b>Grade</b>	<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>
2	200 – 231	232 – 245	246 – 263	264 – 299
3	300 – 338	339 – 353	354 – 372	373 – 399
4	400 – 438	439 – 454	455 – 471	472 – 499
5	500 – 539	540 – 555	556 – 572	573 – 599
6	600 – 639	640 – 654	655 – 671	672 – 699
7	700 – 738	739 – 755	756 – 767	768 – 799
8	800 – 839	840 – 855	856 – 869	870 – 899
9	900 – 930	931 – 949	950 – 959	960 – 999
10	900 – 939	940 – 955	956 – 969	970 – 999
<b>Mathematics</b>				
<b>Grade</b>	<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>
2	200 – 243	244 – 254	255 – 267	268 – 299
3	300 – 339	340 – 359	360 – 375	376 – 399
4	400 – 442	443 – 457	458 – 473	474 – 499
5	500 – 542	543 – 559	560 – 574	575 – 599
6	600 – 635	636 – 653	654 – 667	668 – 699
7	700 – 735	736 – 751	752 – 769	770 – 799
8	800 – 835	836 – 849	850 – 867	868 – 899
10	900 – 932	933 – 950	951 – 970	971 – 999
<b>Science/Biology</b>				
<b>Grade</b>	<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>
5	500 – 540	541 – 552	553 – 563	564 – 599
8	800 – 848	849 – 855	856 – 867	868 – 899
High School	900 – 945	946 – 951	952 – 965	966 – 999
<b>Composition</b>				
<b>Grade</b>	<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>
4	400 – 443	444 – 455	456 – 469	470 – 499
7	700 – 743	744 – 755	756 – 766	767 – 799
10	900 – 942	943 – 954	955 – 966	967 – 999

## Section 8. Evidence for Reliability and Validity

This section contains information relevant to *Standards and Assessments Peer Review Guidance*, Critical Elements 4.1, 4.2, and 4.3:

### 4.1

For each assessment, including all alternate assessments, has the State documented the issue of **validity** (in addition to the alignment of the assessment with the content standards), as described in the *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 1999), with respect to all of the following categories:

- (a) Has the State specified the purposes of the assessments, delineating the types of uses and decisions most appropriate to each?
- (c) Has the State ascertained that the scoring and reporting structures are consistent with the sub-domain structures of its academic content standards (i.e., are item interrelationships consistent with the framework from which the test arises)?
- (e) Has the State ascertained that test and item scores are related to outside variables as intended (e.g., scores are correlated strongly with relevant measures of academic achievement and are weakly correlated, if at all, with irrelevant characteristics, such as demographics)?

### 4.2

For each assessment, including all alternate assessments, has the State considered the issue of **reliability**, as described in the *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 1999), with respect to all of the following categories:

- (a) Has the State determined the reliability of the scores it reports, based on data for its own student population and each reported subpopulation? *and*
- (b) Has the State quantified and reported within the technical documentation for its assessments the conditional standard error of measurement and student classification that are consistent at each cut score specified in its academic achievement standards? *and*
- (c) Has the State reported evidence of generalizability for all relevant sources, such as variability of groups, internal consistency of item responses, variability among schools, consistency from form to form of the test, and inter-rater consistency in scoring?

### 4.3

Has the State ensured that its assessment system is fair and accessible to all students, including students with disabilities and students with limited English proficiency, with respect to each of the following issues:

- (c) Has the State taken steps to ensure fairness in the development of the assessments?

## Reliability

Reliability refers to the degree to which students' scores are free from multiple occasion effects and provides a measure of consistency. In other words, reliability helps to describe how consistent students' performances would be if given the assessment over multiple occasions. The degree of score reliability that is required for an interpretation of an individual student's test score must be

carefully considered. Individual score reliability is estimated using internal consistency coefficients that are computed on all student responses in each grade and content area of the DC CAS. They are computed using the operational items administered to all students in a grade and content area.

## Validity

The collection of reliability sources of evidence is a necessary precursor to establishing evidence of validity. How the scores are ultimately used is a key component to validity evidence, such that the trustworthiness of the scores is well established. As noted in the introduction, test validation is an ongoing process of gathering evidence from many sources to evaluate the trustworthiness of the desired score interpretation or use. This evidence is provided throughout this technical report specific to procedures and processes that support the integrity of the content of the test, test development, blueprints, alignment, scoring and rater reliability, psychometric analyses (item analyses, scaling, linking, and comparative analyses across administrations), and student-level performance results.

## Item Level Evidence

### *Classical Item Statistics*

DC CAS operational and field test items are all reviewed for statistical accuracy and quality. Table 16 summarizes item level classical statistics for operational and field test items. Items that were suppressed from final scoring are noted in these tables. In all cases where items were suppressed, the blueprints remained comparable. For MC items, proportion correct (adjusted  $p$  values) is reported. For CR items, the adjusted  $p$  value is calculated as the mean score across all students divided by the maximum number of score points possible. Note omitted responses are ignored in the adjusted  $p$  value computation and all  $p$  values are adjusted in this report hereafter. On average, the collection of operational items on a test ranged from moderately difficult ( $p$  value mean of 0.39 for Mathematics Grade 10 and Biology) to moderately easy ( $p$  value mean of 0.72 for Grade 2 Mathematics). Tables in Appendix C display the item difficulty for each item at each grade. With respect to field test items, a test ranged from difficult ( $p$  value mean of 0.31 for Mathematics Grade 10) to moderately easy ( $p$  value mean of 0.65 for Reading Grade 9). The weighted  $p$  value means (weighted by CR scores) are also provided in Table 16. The weighted  $p$  value means can be slightly lower than non-weighted means because CR item  $p$  values are lower than most MC item  $p$  values (see Appendix C).

The point-biserial correlation is one measure of the correlation between each item and the overall test. The point-biserial correlations for each content area and grade for operational and field test items are shown in Table 16. The operational test form correlations range from 0.40 to 0.46 (Reading); from 0.35 to 0.44 (Mathematics); from 0.28 to 0.35 (Science/Biology); and from 0.64 to 0.70 (Composition). Field test form correlations range from 0.27 to 0.45 (Reading); from 0.23 to 0.44 (Mathematics); and from 0.22 to 0.29 (Science/Biology).

Table 16 also displays the mean item omit rates calculated across students for each grade and content area. CTB flags items when more than 5% of students omit an item. Flagged items are reviewed to ensure that they are appropriate for examinees in the tested grade. In addition, omitted items near the end of the test are reviewed as not reached items to ensure the administration conditions, such as testing time and accurate printing and scanning. Overall, the omit rates are low. The largest mean percentage omit rate is 3.20% in Mathematics Grade 10. All of the not reached rates are less than 1% except for Reading Grade 9 (1.42%), Reading Grade 10 (1.33%),

Mathematics Grade 10 (1.35%), Composition Grade 7 ( 1.37%), and Composition Grade 10 (1.97%), indicating that the students were provided with ample time to complete the DC CAS tests.

Both content specialists and research staff carefully reviewed items flagged by the above classic item statistics for possible answer key errors. There was no any key issue in 2013 operational items.

### ***Inter-Rater Reliability***

The DC CAS CR questions require a response composed by the examinee, usually in the form of one or more sentences, where the ideas expressed are scored as correct, partially correct, or incorrect. Since the ideas rather than the specific written expressions are scored, the response cannot be scored by applying a clerical key. Raters use judgment to determine whether the ideas expressed match those described in a scoring guide. In other words, raters interpret what the student has written. In order to minimize the difference in interpretations that raters make, raters are required to have certain hiring qualifications and on-site training using examples of responses that match and do not match the desired answers. Even so, the match between a student's response and the scoring guide description of a correct response is a matter of degree.

The inter-rater reliabilities of CR items rely heavily on the solid and consistent training of the hand-scorers, as was described in Section 5. Statistical data are presented in terms of the kappa and intraclass correlations as ways to measure the consistency (reliability) of the scores. Tables 17–20 provide the relevant operational inter-rater reliability statistics. In general, the values are within acceptable limits. The lowest operational statistics fall on one item in Reading that presents an intraclass correlation statistic of 0.79 and a kappa statistic of 0.57. Intraclass correlations for all operational Reading items range from 0.79 to 0.94 with a mean of 0.87; from 0.90 to 0.99 with a mean of 0.96 in Mathematics; from 0.87 to 0.96 with a mean of 0.92 in Science/Biology; and from 0.85 to 0.88 with a mean of 0.87 in Composition. Kappa statistics range from 0.57 to 0.88 with a mean of 0.74 in Reading; from 0.79 to 0.97 with a mean of 0.91 in Mathematics; from 0.73 to 0.91 with a mean of 0.83 in Science/Biology; and from 0.70 to 0.77 with a mean of 0.74 in Composition. These operational values are within acceptable limits.

Tables 21–23 provide the relevant field test inter-rater reliability statistics. In general, the values are within acceptable limits. The lowest field test statistics fall on one item in Mathematics that presents an intraclass correlation statistic of 0.82 and a kappa statistic of 0.64. Intraclass correlations for all field test Reading items range from 0.83 to 0.94 with a mean of 0.90; from 0.82 to 1.00 with a mean of 0.95 in Mathematics; and from 0.83 to 0.99 with a mean of 0.92 in Science/Biology. Kappa statistics range from 0.65 to 0.87 with a mean of 0.80 in Reading; from 0.64 to 0.99 with a mean of 0.90 in Mathematics; and from 0.65 to 0.97 with a mean of 0.83 in Science/Biology. These field test values are within acceptable limits.

### ***Differential Item Function***

Differential item function (DIF) analyses were conducted for all grades and content areas for gender and race/ethnicity. DIF analyses were conducted with at least 400 cases for reference groups and 200 cases for focal groups to provide data adequate for Mantel-Haenszel DIF analysis procedures, which require subdividing each comparison group based on total test raw scores.

Tables 24–27 summarize the 2013 DIF analysis results for operational items, and Tables 28–30 for field items. Positive flags indicate DIF that favors the focal group. Statistics with fewer than 200

focal group examinees and 400 reference group examinees are not calculated for these analyses to provide appropriate subgroup comparisons. Recall that A corresponds to no DIF, B to moderate DIF, and C to considerable DIF. Modest numbers of MC and CR items were flagged for DIF at levels B and C. The majority of items flagged for DIF were in race/ethnicity comparisons; many of those were positive values that indicated DIF that favored the focal group (e.g., Hispanic and White students).

Overall, the number of operational items flagged for DIF was moderate. For example, the total of 155 Reading item flags for DIF represents 12.3% of the 1,257 flagging opportunities in Reading; the total of 141 item flags in Mathematics for DIF represents 11.2% of the 1,257 flagging opportunities in Mathematics; the total of 34 Science/Biology item flags for DIF represents 7.6% of the 448 flagging opportunities in Science/Biology; and the total of 4 Composition item flags for DIF represents 14.8% of the 27 flagging opportunities in Composition.

The number of field test items flagged for DIF was moderate. For example, the total of 38 Reading item flags for DIF represents 13.3% of the 285 flagging opportunities in Reading; the total of 77 item flags in Mathematics for DIF represents 12% of the 643 flagging opportunities in Mathematics; and the total of 9 Science/Biology item flags for DIF represents 4.7% of the 193 flagging opportunities in Science/Biology. There were no field test items for Composition.

## **Test and Strand Level Evidence**

### ***Operational Test Scores***

Operational test level raw score and scale score means and standard deviations for the District are provided in Table 31, along with the test level reliability coefficients, including Cronbach's alpha, stratified coefficient alpha, and Feldt-Raju. The scale score and raw score means and standard deviations are consistent across grades within content area. The reliabilities all show high levels of internal consistency, with reliabilities all greater than 0.85. Subgroup performance and total test reliabilities are provided in Appendix D.

Similarly, the content strand item number, weighted  $p$  value means,  $p$  value means, standard deviations, and reliabilities are provided for each grade and content area in Tables 32–35. Teachers and educational decision makers frequently want diagnostic information that can be used to inform instructional strategies within a content area and to help identify student strengths and weaknesses. This information can be derived from student scores on subsets of test questions called content strands (e.g., Informational Text, Number Sense, etc.).

### ***Strand Level Scores***

The raw score means and standard deviations highlight strands in which students show better or lesser mean performance, and the variability of that performance given the spread represented by the standard deviations. The weighted  $p$  value means are a better indicator of the strand level difficulty, however, given they are not swayed by the number of items in a given strand, as the mean raw score is. Therefore, a review of the weighted  $p$  value means in each strand highlights the strands that tend to be the more or less difficult for students. Specifically, the strands that tend to be the most difficult in each content area are Reading Informational Text in Reading (most grades), Statistics and Probability in Mathematics (grade 10), and in Science—Science and Technology (Grade 5), Forces (Grade 8), and Cell Biology and Biochemistry (High School). In Composition, we look to the mean raw scores, noting that Writing Topic Development represents 6 points and other strands represent a single rubric of 4 points each. So the mean raw score for

Writing Topic Development looks higher than other strands. The mean raw scores for Writing Language Conventions is slightly more than that for Understanding Literary or Informational Text in all grades.

In strands where there are very few items, reliabilities are lower, as would be expected. The degree of reliability that is required to interpret these strand scores, as for any test score, must therefore be carefully considered. These coefficients are computed on all valid student responses in each grade and content area for each content strand. The internal reliability estimates for these strand scores, which include as few as 5 items and as many as 24, range between 0.19 and 0.88.

As an additional measure of internal consistency, correlations have been produced between strands within each grade and content area. These are provided in Tables 36–39. A review of the correlations shows fairly strong relationships amongst strands within content area. Specifically, in Table 36, the DC CAS 2013 Reading strand and total test correlations for all grades are presented. The Reading strand correlations are moderate to high for all grades.

Table 37 displays the correlations for the DC CAS 2013 Mathematics strand and total test correlations by grade. The correlations are mostly moderate to high. The strand with the lowest correlation between other Mathematics strands is Geometry in Grades 2 through 5, Statistics & Probability in Grades 6, 7 and 10, and The Number System in Grade 8. Geometry and Statistics & Probability also tend to have the lowest correlations with the Mathematics total raw score at most grades. This is due in part to the smaller number of items used to measure Geometry and Statistics & Probability in relation to the rest of the content strands.

In Table 38, the DC CAS 2013 Science/Biology strand and total test correlations for all grades are presented. The correlations are moderate to high range from 0.75 to 0.92.

The DC CAS 2013 rubric score and total Composition test correlations for all grades are presented in Table 39. The correlations between the Content Strand scores are moderate, suggesting that each rubric assesses somewhat different composing skills, as intended. The correlations between the rubric scores and total Composition scores are high, as expected.

### ***Standard Errors of Measurement***

Standard errors of measurement (SEMs) indicate the degree of unreliability in the test scores, and conditional SEMs specific to each scale score provide further evidence. Tables 40–43 list the number correct to scale score values, along with their associated IRT SEM values. It is most important to review these at the cut scores that differentiated students by proficient level. The cut score SEMs range from 3 to 7 (Reading), 3 to 14 (Mathematics), 2 to 5 (Science/Biology) and 5 to 8 (Composition). The lowest SEMs are typically at the “Proficient” cut in most grades and content areas.

### **Proficiency Level Evidence**

Student performance relative to their score is classified into one of four proficiency levels: Below Basic, Basic, Proficient, and Advanced. The categorizations are important for accountability purposes, as well as for teacher, students, and parents to understand the content meaning of the associated scale scores. The percentage of students in each category, referred to as “impact data,” is provided in Table 44. The “overall pass rate” represents the combined impact data of the two upper levels, Proficient and Advanced, and is often the sum percentage referenced in accountability measures.

Tables 45-48 display the classification consistency and accuracy results for each cut score and across all cut scores for the 2013 DC CAS in Reading, Mathematics, Science/Biology, and Composition. (The same information is provided for each subgroup in tables in Appendix D). These statistics provide indication of the reliability of the proficiency cut scores, which designate the categories within which student performance would be classified over multiple administrations of the same assessment. The classification consistency statistics can be interpreted like the correlations, where the closer to 1.00 the statistics, the stronger the reliability. As with other measures of reliability, the statistics are impacted by the number of data points or, in this case, items and score points. Step 2 of the classification consistency calculations rests on the total raw scores. For that reason, the reliabilities for Composition are likely to be lower than the assessments in other content areas with higher possible total raw scores/points. What can be seen from the results described, however, is that Composition remains comparable to the other content areas, even with fewer points.

The classification consistency in all grades in Reading, Mathematics, and Science/Biology range from 0.65 to 0.98, and is comparable to that in 2012, which ranged between 0.65 and 0.82. The classification consistency ranged from 0.59 to 0.89 in Composition. The kappa values, which indicate classification consistency beyond chance consistency, represent moderate to substantial consistency levels (Landis & Koch, 1997). The kappa coefficients in Reading, Mathematics, and Science/Biology coefficients range between 0.45 and 0.82, which is comparable with the 2012 results (0.48 to 0.77). Kappa coefficients in Composition this year range from 0.45 to 0.71.

The classification accuracy results range from 0.69 to 0.98 in Reading, Mathematics, and Science/Biology. The results are comparable with those in 2012, which ranged between 0.73 and 0.85. In Composition, classification accuracies range from 0.69 to 0.93. These results suggest that the 2013 DC CAS assessments in all content areas classify examinees into DC CAS proficiency levels based on observed test scores with reasonably strong accuracy.

The false positive rates are estimates of the percentages of examinees that are classified into a proficiency level higher than their true proficiency level. The false negative rates are estimates of the percentages of examinees that are classified into a proficiency level lower than their true proficiency level. These are reasonably low false positive and negative rates in absolute terms. It is a policy question as to how much higher or lower false positive rates should be relative to false negative rates. A review of the tables though, shows these rates quite low, ranging from 0.01 to 0.18 in Composition and from 0.00 to 0.23 in Reading, Mathematics, and Science/Biology.

The magnitude of classification consistency and accuracy measures is influenced by key features of the test design, including the number of items and number of cut scores, score reliability and associated standard errors of measurement, and the locations of the cut scores in relation to the examinee proficiency frequency distributions. The classification consistency and accuracy results observed for 2013 suggest that consistent and accurate performance level classifications are being made for students based on the DC CAS assessments.

### **Correlational Evidence across Content Areas**

Using all scored data, the correlations across the Reading, Mathematics, Science/Biology, and Composition raw scores were calculated as a way of examining evidence of the validity of inferences about student achievement based on relationships between content area tests. This evidence is referred to as evidence of convergent and discriminant validity. The correlations

between Reading, Mathematics, Science/Biology, and Composition total raw scores appear in Table 49.

Correlations are somewhat higher in the elementary grades than in the middle and high school grades. Correlations between Reading and Mathematics are 0.68 and higher; correlations of Reading and Mathematics scores with Science/Biology scores are 0.56 and higher; correlations with the Composition total scores are in the range of 0.47 to 0.69. Composition correlations are relatively lower because Composition scores range from 3 to 14, which restricts variability and covariance. These results are consistent with typical content area correlations for educational achievement tests in these content areas.

These correlations are moderately high. They indicate that approximately 25%–50% of the variability in performance on these separate content area tests can be accounted for by skills and proficiency shared across the content areas (i.e., disregarding measurement error). This observation suggests that approximately one half to three quarters of the performance on each content area assessment can be explained by knowledge, skills, and proficiency that are unique to each content area (i.e., disregard measurement error).

**Table 16. DC CAS 2013 Classical Item Level Statistics**

Grade	Operational						Field Test					
	Number of Items	Weighted P Value Mean	P Value Mean	Mean Item-Total Correlation	Mean Omit Rate	Mean Not Reached Rate	Number of Items	Weighted P Value Mean	P Value Mean	Mean Item-Total Correlation	Mean Omit Rate	Mean Not Reached Rate
<b>Reading</b>												
2	35	0.63	0.64	0.41	1.74	0.35	10	0.45	0.45	0.38	1.82	0.50
3	48	0.59	0.62	0.46	1.40	0.24	10	0.51	0.56	0.45	0.93	0.23
4	48	0.61	0.63	0.43	0.56	0.23	11	0.40	0.42	0.34	1.27	0.08
5	48	0.62	0.64	0.41	0.50	0.16	10	0.54	0.55	0.42	0.81	0.20
6	48	0.60	0.63	0.43	0.85	0.18	10	0.51	0.51	0.33	0.64	0.12
7	48	0.58	0.60	0.40	0.81	0.38	10	0.37	0.39	0.27	1.01	0.20
8	48	0.61	0.63	0.40	1.14	0.45	11	0.55	0.53	0.36	1.14	0.53
9	48	0.56	0.58	0.42	2.44	1.42	11	0.65	0.65	0.45	2.34	2.00
10	48	0.59	0.61	0.42	2.20	1.33	12	0.54	0.54	0.39	1.42	0.92
<b>Mathematics</b>												
2	43	0.70	0.72	0.43	0.61	0.12	32	0.61	0.64	0.44	1.28	0.22
3	54	0.62	0.62	0.44	1.39	0.14	32	0.46	0.51	0.43	1.72	0.23
4	54	0.58	0.60	0.44	0.57	0.20	32	0.47	0.51	0.42	1.23	0.31
5	54	0.58	0.59	0.44	0.50	0.15	32	0.43	0.45	0.40	0.76	0.22
6	54	0.51	0.54	0.42	0.55	0.19	32	0.45	0.47	0.37	0.97	0.25
7	54	0.46	0.48	0.39	0.83	0.36	32	0.33	0.37	0.32	1.17	0.39
8	53	0.50	0.51	0.37	0.92	0.45	33	0.39	0.43	0.35	1.37	0.43
10	53	0.37	0.39	0.35	3.20	1.35	32	0.30	0.31	0.23	4.72	1.48
<b>Science/Biology</b>												
5	50	0.48	0.48	0.35	0.55	0.15	28	0.38	0.40	0.29	0.71	0.09
8	50	0.40	0.41	0.33	1.07	0.40	26	0.38	0.40	0.29	1.74	0.31
High School	50	0.39	0.39	0.28	1.93	0.63	27	0.29	0.32	0.22	3.31	0.47
<b>Composition</b>												
4	3	0.46	0.47	0.64*	0.86	0.86	N/A	N/A	N/A	N/A	N/A	N/A
7	3	0.58	0.59	0.68*	1.37	1.37	N/A	N/A	N/A	N/A	N/A	N/A
10	3	0.59	0.60	0.70*	1.97	1.97	N/A	N/A	N/A	N/A	N/A	N/A

\*Item-total correlations for Composition include the Reading items along with which the Composition prompts were scaled.

**Table 17. DC CAS 2013 Operational Inter-Rater Agreement for CR Items: Reading**

Grade	Item No.	Score Points	N	Perfect Proportion	Adjacent Proportion	Discrepant Proportion	Intraclass Correlation	Weighted Kappa	Mean Raw Score
2	25	0-3	489	0.85	0.05	0.10	0.91	0.83	1.40
	36	0-3	489	0.69	0.26	0.05	0.92	0.84	1.43
3	6	0-3	471	0.66	0.31	0.03	0.84	0.67	0.82
	32	0-3	471	0.77	0.18	0.05	0.92	0.84	1.21
	38	0-3	471	0.73	0.24	0.02	0.94	0.88	1.18
4	5	0-3	455	0.66	0.30	0.04	0.88	0.77	1.20
	38	0-3	455	0.63	0.34	0.04	0.90	0.80	1.60
	59	0-3	455	0.67	0.28	0.04	0.90	0.80	1.12
5	9	0-3	479	0.81	0.18	0.01	0.94	0.88	0.92
	19	0-3	479	0.76	0.23	0.01	0.92	0.85	1.50
	38	0-3	479	0.55	0.36	0.09	0.85	0.69	1.31
6	5	0-3	501	0.57	0.36	0.07	0.85	0.71	1.37
	24	0-3	501	0.64	0.32	0.04	0.88	0.76	0.94
	53	0-3	501	0.66	0.33	0.02	0.88	0.77	1.29
7	13	0-3	462	0.54	0.42	0.04	0.80	0.61	0.95
	42	0-3	462	0.65	0.31	0.04	0.89	0.77	1.26
	54	0-3	462	0.56	0.36	0.08	0.82	0.63	1.22
8	11	0-3	418	0.51	0.42	0.06	0.82	0.63	1.06
	40	0-3	418	0.69	0.27	0.04	0.89	0.77	1.16
	58	0-3	418	0.54	0.38	0.08	0.82	0.64	1.40
9	14	0-2	527	0.63	0.31	0.05	0.79	0.57	0.83
	18	0-3	527	0.63	0.32	0.04	0.86	0.72	0.79
	32	0-3	527	0.59	0.35	0.06	0.86	0.72	1.08
10	4	0-3	452	0.68	0.28	0.04	0.90	0.80	1.02
	30	0-3	452	0.61	0.36	0.03	0.87	0.73	1.33
	60	0-3	452	0.62	0.33	0.05	0.84	0.69	1.05

**Table 18. DC CAS 2013 Operational Inter-Rater Agreement for CR Items: Mathematics**

Grade	Item No.	Score Points	N	Perfect Proportion	Adjacent Proportion	Discrepant Proportion	Intraclass Correlation	Weighted Kappa	Mean Raw Score
2	6	0-2	489	0.88	0.11	0.01	0.94	0.88	0.68
	25	0-3	489	0.87	0.12	0.01	0.96	0.93	1.32
3	6	0-3	471	0.92	0.04	0.04	0.98	0.95	2.12
	25	0-3	471	0.90	0.09	0.01	0.98	0.96	1.21
	60	0-3	471	0.93	0.04	0.02	0.98	0.96	2.21
4	6	0-3	455	0.88	0.10	0.02	0.97	0.94	1.33
	25	0-3	455	0.83	0.14	0.03	0.94	0.87	1.09
	60	0-3	455	0.93	0.07	0.00	0.99	0.97	1.02
5	6	0-3	479	0.82	0.18	0.00	0.96	0.91	1.46
	25	0-3	479	0.86	0.13	0.01	0.97	0.94	2.23
	60	0-3	478	0.87	0.12	0.01	0.97	0.93	0.99
6	24	0-3	501	0.88	0.11	0.00	0.97	0.93	0.71
	39	0-3	501	0.77	0.19	0.04	0.93	0.87	1.11
	59	0-3	501	0.91	0.09	0.00	0.98	0.95	0.59
7	6	0-3	462	0.88	0.11	0.00	0.95	0.89	0.59
	25	0-3	462	0.90	0.10	0.00	0.98	0.95	0.83
	60	0-3	462	0.65	0.30	0.05	0.90	0.79	1.49
8	24	0-3	418	0.85	0.14	0.01	0.96	0.92	0.95
	60	0-3	418	0.76	0.22	0.02	0.93	0.86	0.87
10	6	0-3	452	0.92	0.07	0.01	0.91	0.81	0.17
	25	0-3	452	0.91	0.09	0.00	0.97	0.93	0.38
	60	0-3	452	0.93	0.06	0.02	0.95	0.90	0.40

**Table 19. DC CAS 2013 Operational Inter-Rater Agreement for CR Items: Science/Biology**

<b>Grade</b>	<b>Item No.</b>	<b>Score Points</b>	<b>N</b>	<b>Perfect Proportion</b>	<b>Adjacent Proportion</b>	<b>Discrepant Proportion</b>	<b>Intraclass Correlation</b>	<b>Weighted Kappa</b>	<b>Mean Raw Score</b>
5	16	0-2	452	0.76	0.21	0.03	0.89	0.77	1.03
	36	0-2	452	0.92	0.07	0.01	0.94	0.88	1.06
	55	0-2	452	0.95	0.05	0.00	0.96	0.91	0.35
8	16	0-2	433	0.79	0.17	0.04	0.89	0.78	0.65
	29	0-2	433	0.84	0.15	0.02	0.94	0.87	0.71
	51	0-2	433	0.87	0.13	0.00	0.87	0.73	0.33
High School	25	0-2	411	0.92	0.06	0.02	0.94	0.88	0.27
	53	0-2	411	0.80	0.17	0.02	0.92	0.85	1.30
	54	0-2	411	0.88	0.11	0.01	0.90	0.80	0.31

**Table 20. DC CAS 2013 Operational Inter-Rater Agreement for CR Items: Composition**

<b>Grade</b>	<b>Item No.</b>	<b>Score Points</b>	<b>N</b>	<b>Perfect Proportion</b>	<b>Adjacent Proportion</b>	<b>Discrepant Proportion</b>	<b>Intraclass Correlation</b>	<b>Weighted Kappa</b>	<b>Mean Raw Score</b>
4	1 A	1-6	441	0.54	0.41	0.05	0.88	0.75	2.30
	1 B	1-4	441	0.67	0.32	0.02	0.85	0.70	2.01
	1 C	1-4	441	0.65	0.33	0.02	0.85	0.70	1.93
7	1 A	1-6	414	0.49	0.42	0.09	0.87	0.74	2.83
	1 B	1-4	414	0.62	0.36	0.03	0.87	0.74	2.47
	1 C	1-4	414	0.60	0.36	0.04	0.87	0.74	2.44
10	1 A	1-6	455	0.50	0.43	0.07	0.88	0.77	2.75
	1 B	1-4	455	0.59	0.38	0.03	0.87	0.75	2.56
	1 C	1-4	455	0.58	0.37	0.04	0.88	0.76	2.55

**Table 21. DC CAS 2013 Field Test Inter-Rater Agreement for CR Items: Reading**

<b>Grade</b>	<b>Item No.</b>	<b>Score Points</b>	<b>N</b>	<b>Perfect Proportion</b>	<b>Adjacent Proportion</b>	<b>Discrepant Proportion</b>	<b>Intraclass Correlation</b>	<b>Weighted Kappa</b>	<b>Mean Raw Score</b>
3	54	0-3	496	0.77	0.20	0.03	0.93	0.86	0.76
4	34	0-2	439	0.86	0.13	0.02	0.88	0.77	0.30
5	29	0-3	429	0.79	0.17	0.04	0.94	0.87	1.27
6	39	0-3	500	0.67	0.28	0.04	0.91	0.83	1.52
7	17	0-3	443	0.70	0.25	0.06	0.89	0.77	1.20
	18	0-3	443	0.73	0.22	0.05	0.83	0.65	0.56
8	21	0-3	420	0.85	0.12	0.03	0.93	0.86	1.73

**Table 22. DC CAS 2013 Field Test Inter-Rater Agreement for CR Items: Mathematics**

Grade	Item No.	Score Points	N	Perfect Proportion	Adjacent Proportion	Discrepant Proportion	Intraclass Correlation	Weighted Kappa	Mean Raw Score
2	15	0-3	246	0.86	0.13	0.01	0.97	0.93	0.84
	15	0-3	208	0.97	0.02	0.00	1.00	0.99	1.69
	55	0-3	246	0.90	0.06	0.04	0.97	0.93	1.83
	55	0-3	208	0.94	0.02	0.04	0.99	0.98	1.20
3	32	0-3	250	0.95	0.05	0.00	0.99	0.98	1.14
	32	0-3	246	0.94	0.04	0.01	0.99	0.98	0.77
	49	0-3	250	0.92	0.06	0.02	0.93	0.86	0.23
	49	0-3	246	0.89	0.08	0.03	0.98	0.96	0.78
4	32	0-3	245	0.80	0.16	0.04	0.94	0.88	1.27
	32	0-3	194	0.88	0.11	0.01	0.95	0.90	0.58
	49	0-3	245	0.69	0.26	0.05	0.82	0.64	0.55
	49	0-3	194	0.91	0.06	0.03	0.97	0.93	1.42
5	32	0-3	205	0.95	0.03	0.02	0.98	0.95	0.80
	32	0-3	224	0.82	0.17	0.02	0.96	0.91	1.54
	49	0-3	205	0.85	0.14	0.00	0.93	0.86	0.79
	49	0-3	224	0.88	0.09	0.03	0.95	0.90	0.96
6	15	0-3	256	0.85	0.14	0.01	0.96	0.91	1.66
	15	0-3	244	0.89	0.11	0.00	0.96	0.91	0.95
	49	0-3	256	0.94	0.05	0.00	0.99	0.97	0.66
	49	0-3	244	0.96	0.04	0.00	0.99	0.97	0.53
7	32	0-3	232	0.92	0.08	0.00	0.94	0.88	0.23
	32	0-3	211	0.82	0.15	0.03	0.93	0.86	0.72
	49	0-3	232	0.97	0.02	0.01	0.96	0.92	0.07
	49	0-3	211	0.81	0.18	0.01	0.88	0.76	0.57

**Table 22. DC CAS 2013 Field Test Inter-Rater Agreement for CR Items: Mathematics (continued)**

<b>Grade</b>	<b>Item No.</b>	<b>Score Points</b>	<b>N</b>	<b>Perfect Proportion</b>	<b>Adjacent Proportion</b>	<b>Discrepant Proportion</b>	<b>Intraclass Correlation</b>	<b>Weighted Kappa</b>	<b>Mean Raw Score</b>
8	6	0-3	420	0.74	0.24	0.02	0.90	0.80	1.02
	15	0-3	215	0.83	0.11	0.07	0.94	0.88	0.71
	15	0-3	205	0.81	0.18	0.01	0.96	0.92	1.07
	49	0-3	215	0.99	0.01	0.00	0.99	0.98	0.19
	49	0-3	205	0.90	0.08	0.02	0.96	0.91	0.58
10	15	0-3	240	0.95	0.03	0.03	0.98	0.95	0.17
	15	0-3	246	0.89	0.04	0.07	0.96	0.91	1.02
	49	0-3	240	0.87	0.09	0.04	0.95	0.89	0.83
	49	0-3	246	0.82	0.12	0.06	0.87	0.73	0.37

**Table 23. DC CAS 2013 Field Test Inter-Rater Agreement for CR Items: Science/Biology**

<b>Grade</b>	<b>Item No.</b>	<b>Score Points</b>	<b>N</b>	<b>Perfect Proportion</b>	<b>Adjacent Proportion</b>	<b>Discrepant Proportion</b>	<b>Intraclass Correlation</b>	<b>Weighted Kappa</b>	<b>Mean Raw Score</b>
5	17	0-2	247	0.88	0.09	0.03	0.93	0.85	0.80
	17	0-2	221	0.78	0.19	0.02	0.89	0.77	0.70
	39	0-2	247	0.91	0.07	0.02	0.93	0.86	0.39
	39	0-2	221	0.92	0.08	0.00	0.92	0.85	0.22
8	17	0-2	239	0.95	0.05	0.01	0.92	0.83	0.15
	17	0-2	213	0.94	0.04	0.01	0.98	0.96	0.49
	41	0-2	239	0.96	0.03	0.01	0.99	0.97	0.81
	41	0-2	213	0.76	0.23	0.02	0.85	0.69	0.50
High School	17	0-2	219	0.88	0.11	0.01	0.88	0.75	0.24
	17	0-2	208	0.91	0.07	0.02	0.92	0.84	0.23
	41	0-2	219	0.88	0.11	0.01	0.83	0.65	0.19
	41	0-2	208	0.96	0.02	0.02	0.97	0.92	0.08

**Table 24. Numbers of Operational Items Flagged for DIF Using the Mantel-Haenszel Procedure: Reading**

Reference Group	Focal Group	A	B	B-	C	C-
Grade 2 (total 35 items)						
Male	Female	35	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	34	0	0	1	0
	White	29	0	0	6	0
Grade 3 (total 48 items)						
Male	Female	48	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	46	0	2	0	0
	White	35	6	0	7	0
Grade 4 (total 48 items)						
Male	Female	44	2	2	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	47	1	0	0	0
	White	36	3	0	9	0
Grade 5 (total 48 items)						
Male	Female	46	2	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	46	0	1	0	1
	White	28	5	0	15	0
Grade 6 (total 48 items)						
Male	Female	47	1	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	46	2	0	0	0
	White	35	6	0	7	0
Grade 7 (total 48 items)						
Male	Female	46	2	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	46	1	0	0	1
	White	36	3	0	9	0
Grade 8 (total 48 items)						
Male	Female	45	2	0	0	1
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	44	1	3	0	0
	White	32	2	3	11	0

N/A= not applicable because case count requirements for the reference (400) and focal (200) groups were not met.

See Table 5 for the numbers of examinees in each grade and subgroup.

**Table 24. Numbers of Operational Items Flagged for DIF Using the Mantel-Haenszel Procedure: Reading (continued)**

Reference Group	Focal Group	A	B	B-	C	C-
Grade 9 (total 48 items)						
Male	Female	44	1	2	1	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	44	0	2	0	2
	White	37	6	3	2	0
Grade 10 (total 48 items)						
Male	Female	46	1	1	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	45	1	2	0	0
	White	35	3	1	9	0

N/A= not applicable because case count requirements for the reference (400) and focal (200) groups were not met.

See Table 5 for the numbers of examinees in each grade and subgroup.

**Table 25. Numbers of Operational Items Flagged for DIF Using the Mantel-Haenszel Procedure: Mathematics**

Reference Group	Focal Group	A	B	B-	C	C-
Grade 2 (total 43 items)						
Male	Female	42	0	1	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	41	1	1	0	0
	White	30	6	0	5	2
Grade 3 (total 54 items)						
Male	Female	54	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	53	1	0	0	0
	White	34	9	3	5	3
Grade 4 (total 54 items)						
Male	Female	54	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	52	0	2	0	0
	White	36	6	2	7	3
Grade 5 (total 54 items)						
Male	Female	53	1	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	50	3	1	0	0
	White	36	3	6	8	1
Grade 6 (total 54 items)						
Male	Female	53	1	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	51	3	0	0	0
	White	37	6	2	7	2
Grade 7 (total 54 items)						
Male	Female	52	0	0	2	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	53	1	0	0	0
	White	41	5	2	4	2
Grade 8 (total 53 items)						
Male	Female	53	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	51	0	2	0	0
	White	40	5	1	4	3
Grade 10 (total 53 items)						
Male	Female	53	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	53	0	0	0	0
	White	44	5	1	2	1

N/A= not applicable because case count requirements for the reference (400) and focal (200) groups were not met.

See Table 5 for the numbers of examinees in each grade and subgroup.

**Table 26. Numbers of Operational Items Flagged for DIF Using the Mantel-Haenszel Procedure: Science/Biology**

Reference Group	Focal Group	A	B	B-	C	C-
Grade 5 (total 50 items)						
Male	Female	49	0	1	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	50	0	0	0	0
	White	35	9	2	3	1
Grade 8 (total 50 items)						
Male	Female	49	1	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	50	0	0	0	0
	White	41	2	0	6	1
High School (total 50 items)						
Male	Female	50	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	50	0	0	0	0
	White	40	4	0	3	1

N/A= not applicable because case count requirements for the reference (400) and focal (200) groups were not met.

See Table 5 for the numbers of examinees in each grade and subgroup. In High School Biology two items did not meet the focal group N count criterion, therefore DIF statistics were not calculated for these items.

**Table 27. Numbers of Operational Items Flagged for DIF Using the Mantel-Haenszel Procedure: Composition**

Reference Group	Focal Group	A	B	B-	C	C-
Grade 4 (total 3 items)						
Male	Female	3	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	3	0	0	0	0
	White	3	0	0	0	0
Grade 7 (total 3 items)						
Male	Female	1	2	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	3	0	0	0	0
	White	2	0	0	1	0
Grade 10 (total 3 items)						
Male	Female	2	1	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	3	0	0	0	0
	White	3	0	0	0	0

N/A= not applicable because case count requirements for the reference (400) and focal (200) groups were not met.

See Table 5 for the numbers of examinees in each grade and subgroup.

**Table 28. Numbers of Field Test Items Flagged for DIF Using the Mantel-Haenszel Procedure: Reading**

Reference Group	Focal Group	A	B	B-	C	C-
Grade 2 (total 10 items)						
Male	Female	10	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	10	0	0	0	0
	White	5	4	0	1	0
Grade 3 (total 10 items)						
Male	Female	10	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	10	0	0	0	0
	White	7	2	1	0	0
Grade 4 (total 11 items)						
Male	Female	11	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	11	0	0	0	0
	White	6	2	0	3	0
Grade 5 (total 10 items)						
Male	Female	10	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	9	1	0	0	0
	White	6	2	0	2	0
Grade 6 (total 10 items)						
Male	Female	8	1	1	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	10	0	0	0	0
	White	7	1	0	2	0
Grade 7 (total 10 items)						
Male	Female	10	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	10	0	0	0	0
	White	6	1	0	3	0
Grade 8 (total 11 items)						
Male	Female	11	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	11	0	0	0	0
	White	6	3	0	2	0

N/A= not applicable because case count requirements for the reference (400) and focal (200) groups were not met.

See Table 5 for the numbers of examinees in each grade and subgroup.

**Table 28. Numbers of Field Test Items Flagged for DIF Using the Mantel-Haenszel Procedure: Reading (continued)**

Reference Group	Focal Group	A	B	B-	C	C-
Grade 9 (total 11 items)						
Male	Female	11	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	10	1	0	0	0
	White	10	1	0	0	0
Grade 10 (total 12 items)						
Male	Female	12	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	12	0	0	0	0
	White	8	2	0	2	0

N/A= not applicable because case count requirements for the reference (400) and focal (200) groups were not met.

See Table 5 for the numbers of examinees in each grade and subgroup.

**Table 29. Numbers of Field Test Items Flagged for DIF Using the Mantel-Haenszel Procedure: Mathematics**

Reference Group	Focal Group	A	B	B-	C	C-
Grade 2 (total 32 items)						
Male	Female	28	2	2	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	29	1	2	0	0
	White	20	5	0	7	0
Grade 3 (total 32 items)						
Male	Female	30	0	1	0	1
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	29	3	0	0	0
	White	18	2	2	7	3
Grade 4 (total 32 items)						
Male	Female	32	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	27	4	1	0	0
	White	20	2	3	3	4
Grade 5 (total 32 items)						
Male	Female	32	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	31	0	1	0	0
	White	19	4	1	5	3
Grade 6 (total 32 items)						
Male	Female	31	0	1	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	31	1	0	0	0
	White	N/A	N/A	N/A	N/A	N/A
Grade 7 (total 32 items)						
Male	Female	30	0	2	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	31	0	1	0	0
	White	N/A	N/A	N/A	N/A	N/A
Grade 8 (total 33 items)						
Male	Female	33	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	30	3	0	0	
	White	1	0	0	0	0
Grade 10 (total 32 items)						
Male	Female	32	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	32	0	0	0	0
	White	N/A	N/A	N/A	N/A	N/A

N/A= not applicable because case count requirements for the reference (400) and focal (200) groups were not met.

See Table 5 for the numbers of examinees in each grade and subgroup. In Math Grade 8 thirty-two items did not meet the focal group N count criterion, therefore DIF statistics were not calculated for these items.

**Table 30. Numbers of Field Test Items Flagged for DIF Using the Mantel-Haenszel Procedure: Science/Biology**

Reference Group	Focal Group	A	B	B-	C	C-
Grade 5 (total 28 items)						
Male	Female	28	0	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	28	0	0	0	0
	White	23	1	2	2	0
Grade 8 (total 26 items)						
Male	Female	24	2	0	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	25	1	0	0	0
	White	2	0	0	0	0
High School (total 27 items)						
Male	Female	26	0	1	0	0
African American	Asian	N/A	N/A	N/A	N/A	N/A
	Hispanic	27	0	0	0	0
	White	1	0	0	0	0

N/A= not applicable because case count requirements for the reference (400) and focal (200) groups were not met.

See Table 5 for the numbers of examinees in each grade and subgroup. In Science Grade 8 and High School Biology, twenty-four items and twenty-six items respectively, did not meet the focal group N count criterion, therefore DIF statistics were not calculated for these items.

**Table 31. Total Test Scale and Raw Score Means and Reliability Statistics**

Grade	Students with Test Scores	Number of Items	Alpha	Stratified Alpha	Feldt-Raju	Scale Score		Raw Score	
						Mean	SD	Mean	SD
<b>Reading</b>									
2	4,692	35	0.88	0.89	0.89	243.23	14.84	23.84	8.05
3	4,851	48	0.93	0.94	0.94	349.96	16.07	31.56	11.94
4	4,467	48	0.92	0.92	0.92	453.84	14.47	32.64	11.13
5	4,532	48	0.91	0.92	0.92	554.99	13.32	33.04	10.79
6	4,560	48	0.92	0.92	0.92	650.64	14.35	32.25	11.20
7	4,520	48	0.91	0.91	0.91	755.82	13.75	30.99	10.45
8	4,236	48	0.91	0.92	0.91	855.51	14.03	32.39	10.52
9	4,779	48	0.92	0.93	0.93	948.36	16.86	29.06	11.47
10	4,684	48	0.92	0.93	0.93	952.47	15.89	31.21	11.66
<b>Mathematics</b>									
2	4,754	43	0.91	0.91	0.92	253.45	13.84	32.00	9.10
3	4,876	54	0.93	0.93	0.93	354.75	18.27	36.66	12.76
4	4,551	54	0.93	0.94	0.94	458.34	16.95	34.73	12.75
5	4,554	54	0.93	0.94	0.94	558.61	16.95	34.45	12.86
6	4,591	54	0.93	0.93	0.93	652.35	18.02	30.68	12.58
7	4,545	54	0.92	0.92	0.92	752.90	18.67	27.62	11.86
8	4,280	53	0.91	0.91	0.91	853.73	16.51	28.15	11.00
10	4,669	53	0.90	0.91	0.91	945.09	21.78	20.90	10.61
<b>Science/Biology</b>									
5	4,520	50	0.89	0.89	0.89	549.26	11.91	25.19	9.68
8	4,106	50	0.88	0.88	0.89	850.05	17.53	21.12	9.51
High School	4,031	50	0.85	0.85	0.85	946.47	16.38	20.08	8.37
<b>Composition*</b>									
4	4,400	3	0.91	0.87	0.93	453.44	17.18	6.37	2.55
7	4,387	3	0.94	0.92	0.95	756.81	15.51	7.99	2.87
10	4,217	3	0.93	0.88	0.94	954.31	18.60	7.97	3.10

\* Stratified Alpha is based on two items.

**Table 32. Weighted *P* Value Mean, *P* Value Mean and Standard Deviation, and Coefficient Alpha Reliability for Reading Strand Scores**

Grade	Content Strand		Number of Items	Weighted <i>P</i> Value Mean	<i>P</i> Value Mean	Standard Deviation	Reliability
2	1	Vocabulary Acquisition & Use	5	0.70	0.70	0.14	0.60
	3	Reading Informational Text	14	0.56	0.58	0.12	0.76
	4	Reading Literary Text	16	0.69	0.69	0.19	0.76
	Total Number of Items on DC CAS		35				
3	1	Vocabulary Acquisition & Use	9	0.67	0.67	0.11	0.73
	3	Reading Informational Text	18	0.49	0.52	0.16	0.80
	4	Reading Literary Text	21	0.67	0.69	0.12	0.88
	Total Number of Items on DC CAS		48				
4	1	Vocabulary Acquisition & Use	9	0.68	0.68	0.14	0.71
	3	Reading Informational Text	19	0.59	0.61	0.17	0.80
	4	Reading Literary Text	20	0.60	0.63	0.11	0.83
	Total Number of Items on DC CAS		48				
5	1	Vocabulary Acquisition & Use	8	0.64	0.64	0.13	0.68
	3	Reading Informational Text	17	0.60	0.65	0.17	0.78
	4	Reading Literary Text	23	0.62	0.63	0.15	0.84
	Total Number of Items on DC CAS		48				
6	1	Vocabulary Acquisition & Use	9	0.64	0.64	0.15	0.73
	3	Reading Informational Text	17	0.56	0.61	0.16	0.80
	4	Reading Literary Text	22	0.63	0.63	0.15	0.84
	Total Number of Items on DC CAS		48				
7	1	Vocabulary Acquisition & Use	7	0.63	0.63	0.22	0.59
	3	Reading Informational Text	21	0.56	0.60	0.16	0.83
	4	Reading Literary Text	20	0.58	0.60	0.15	0.79
	Total Number of Items on DC CAS		48				
8	1	Vocabulary Acquisition & Use	5	0.65	0.65	0.22	0.54
	3	Reading Informational Text	24	0.56	0.60	0.17	0.83
	4	Reading Literary Text	19	0.67	0.67	0.15	0.81
	Total Number of Items on DC CAS		48				
9	1	Vocabulary Acquisition & Use	7	0.53	0.53	0.14	0.56
	3	Reading Informational Text	21	0.56	0.57	0.12	0.82
	4	Reading Literary Text	20	0.57	0.62	0.13	0.86
	Total Number of Items on DC CAS		48				
10	1	Vocabulary Acquisition & Use	7	0.67	0.67	0.17	0.61
	3	Reading Informational Text	24	0.58	0.61	0.10	0.87
	4	Reading Literary Text	17	0.59	0.59	0.12	0.81
	Total Number of Items on DC CAS		48				

**Note:** All *p* values are adjusted and weighted *p* value means are weighted by CR item scores if there is any CR item in the strand.

**Table 33. Weighted *P* Value Mean, *P* Value Mean and Standard Deviation, and Coefficient Alpha Reliability for Mathematics Strand Scores**

Grade	Content Strand		Number of Items	Weighted <i>P</i> Value Mean	<i>P</i> Value Mean	Standard Deviation	Reliability
2	1	Operations & Algebraic Thinking	8	0.54	0.59	0.14	0.69
	2	Numbers & Operations Base Ten	17	0.79	0.79	0.12	0.83
	4	Measurement and Data	13	0.65	0.65	0.14	0.80
	11	Geometry	5	0.87	0.87	0.09	0.38
	Total Number of Items on DC CAS		43				
3	1	Operations & Algebraic Thinking	13	0.69	0.69	0.14	0.82
	2	Number & Operations Base Ten	5	0.73	0.73	0.13	0.60
	3	Number & Operations Fractions	11	0.56	0.58	0.25	0.73
	4	Measurement and Data	20	0.59	0.56	0.21	0.79
	11	Geometry	5	0.66	0.66	0.08	0.80
	Total Number of Items on DC CAS		54				
4	1	Operations & Algebraic Thinking	13	0.50	0.51	0.15	0.76
	2	Number & Operations Base Ten	8	0.75	0.75	0.09	0.74
	3	Number & Operations Fractions	14	0.66	0.66	0.13	0.79
	4	Measurement and Data	13	0.50	0.52	0.21	0.76
	11	Geometry	6	0.58	0.65	0.18	0.65
	Total Number of Items on DC CAS		54				
5	1	Operations & Algebraic Thinking	9	0.67	0.67	0.16	0.72
	2	Number & Operations Base Ten	9	0.69	0.69	0.11	0.75
	3	Number & Operations Fractions	15	0.52	0.52	0.15	0.77
	4	Measurement and Data	13	0.47	0.49	0.21	0.81
	11	Geometry	8	0.64	0.66	0.12	0.70
	Total Number of Items on DC CAS		54				
6	5	Ratios & Proportional Relationships	11	0.47	0.52	0.22	0.71
	6	The Number System	13	0.63	0.63	0.17	0.80
	7	Expressions & Equations	14	0.53	0.53	0.16	0.77
	11	Geometry	9	0.45	0.50	0.22	0.67
	12	Statistics & Probability	7	0.47	0.49	0.09	0.67
	Total Number of Items on DC CAS		54				
7	5	Ratios & Proportional Relationships	13	0.47	0.51	0.16	0.75
	6	The Number System	10	0.52	0.52	0.15	0.69
	7	Expressions & Equations	14	0.41	0.41	0.12	0.73
	11	Geometry	8	0.41	0.44	0.18	0.62
	12	Statistics & Probability	9	0.54	0.55	0.22	0.61
	Total Number of Items on DC CAS		54				

**Table 33. Weighted *P* Value Mean, *P* Value Mean and Standard Deviation, and Coefficient Alpha Reliability for Mathematics Strand Scores (continued)**

Grade	Content Strand		Number of Items	Weighted <i>P</i> Value Mean	<i>P</i> Value Mean	Standard Deviation	Reliability
8	6	The Number System	9	0.43	0.43	0.12	0.60
	7	Expressions & Equations	11	0.51	0.55	0.21	0.74
	10	Functions	11	0.48	0.52	0.15	0.76
	11	Geometry	12	0.52	0.52	0.17	0.63
	12	Statistics & Probability	10	0.54	0.54	0.20	0.54
	Total Number of Items on DC CAS			53			
10	8	Number & Quantity	9	0.33	0.33	0.10	0.51
	9	Algebra	15	0.41	0.44	0.11	0.81
	10	Functions	6	0.33	0.37	0.17	0.68
	11	Geometry	17	0.38	0.41	0.14	0.71
	12	Statistics & Probability	6	0.31	0.31	0.17	0.19
	Total Number of Items on DC CAS			53			

*Note:* All *p* values are adjusted and weighted *p* value means are weighted by CR item scores if there is any CR item in the strand.

**Table 34. Weighted *P* Value Mean, *P* Value Mean and Standard Deviation, and Coefficient Alpha Reliability for Science/Biology Strand Scores**

Grade	Content Strand		Number of Items	Weighted <i>P</i> Value Mean	<i>P</i> Value Mean	Standard Deviation	Reliability
5	1	Science and Technology	18	0.45	0.44	0.17	0.73
	2	Earth and Space Science	12	0.51	0.54	0.19	0.71
	3	Physical Science	8	0.45	0.45	0.13	0.53
	4	Life Science	12	0.51	0.51	0.07	0.68
	Total Number of Items on DC CAS			50			
8	1	Scientific Thinking and Inquiry	8	0.45	0.46	0.10	0.59
	2	Matter and Reactions	19	0.39	0.40	0.09	0.75
	3	Forces	8	0.38	0.40	0.13	0.59
	4	Energy and Waves	15	0.40	0.40	0.15	0.62
	Total Number of Items on DC CAS			50			
High School	1	Cell Biology & Biochemistry	16	0.37	0.38	0.10	0.65
	2	Genetics and Evolution	16	0.39	0.39	0.11	0.63
	3	Multicellular Organisms	9	0.42	0.41	0.17	0.56
	4	Ecosystems	9	0.38	0.38	0.10	0.46
	Total Number of Items on DC CAS			50			

*Note:* All *p* values are adjusted and weighted *p* value means are weighted by CR item scores if there is any CR item in the strand.

**Table 35. Raw score mean and STD, and correlations for Composition Strand Scores**

<b>Grade</b>		<b>Content Strand</b>	<b>Number of Items</b>	<b>Students with Test Scores</b>	<b>Mean Raw Score</b>	<b>STD of Raw Score</b>	<b>Correlation Between Strand 2</b>	<b>Correlation Between Strand 3</b>
4	1	Writing Topic Development	1	4,400	2.35	1.12	0.76	0.88
	2	Writing Language Conventions	1	4,400	2.05	0.78	--	0.78
	3	Understanding Literary Text	1	4,400	1.97	0.83	--	--
7	1	Writing Topic Development	1	4,387	2.91	1.22	0.83	0.89
	2	Writing Language Conventions	1	4,387	2.57	0.87	--	0.85
	3	Understanding Informational Text	1	4,387	2.51	0.93	--	--
10	1	Writing Topic Development	1	4,217	2.81	1.25	0.81	0.89
	2	Writing Language Conventions	1	4,217	2.59	0.97	--	0.79
	3	Understanding Literary Text	1	4,217	2.58	1.06	--	--

**Table 36. DC CAS 2013 Reading Strand Correlations by Grade**

<b>Grade</b>	<b>Content Strand</b>	<b>Acquisition &amp; Use</b>	<b>Informational Text</b>	<b>Literary Text</b>	<b>Total Reading</b>
2	Acquisition & Use	--	0.67	0.67	0.79
	Informational Text		--	0.74	0.94
	Literary Text			--	0.91
	Total Raw Score				--
3	Acquisition & Use	--	0.74	0.80	0.87
	Informational Text		--	0.81	0.93
	Literary Text			--	0.96
	Total Raw Score				--
4	Acquisition & Use	--	0.75	0.77	0.86
	Informational Text		--	0.80	0.93
	Literary Text			--	0.95
	Total Raw Score				--
5	Acquisition & Use	--	0.71	0.71	0.83
	Informational Text		--	0.79	0.93
	Literary Text			--	0.95
	Total Raw Score				--
6	Acquisition & Use	--	0.74	0.76	0.87
	Informational Text		--	0.79	0.94
	Literary Text			--	0.94
	Total Raw Score				--
7	Acquisition & Use	--	0.71	0.70	0.81
	Informational Text		--	0.81	0.95
	Literary Text			--	0.94
	Total Raw Score				--
8	Acquisition & Use	--	0.67	0.69	0.77
	Informational Text		--	0.79	0.96
	Literary Text			--	0.92
	Total Raw Score				--
9	Acquisition & Use	--	0.70	0.70	0.80
	Informational Text		--	0.82	0.94
	Literary Text			--	0.96
	Total Raw Score				--
10	Acquisition & Use	--	0.73	0.71	0.82
	Informational Text		--	0.81	0.97
	Literary Text			--	0.92
	Total Raw Score				--

**Table 37. DC CAS 2013 Mathematics Strand Correlations by Grade**

Grade	Content Strand	Operations & Algebraic Thinking	Numbers & Operations Base Ten	Measurement & Data	Geometry	Total Mathematics	
2	Operations & Algebraic Thinking	--	0.71	0.71	0.42	0.87	
	Numbers & Operations Base Ten		--	0.76	0.49	0.92	
	Measurement & Data			--	0.45	0.91	
	Geometry				--	0.58	
	Total Raw Score					--	
Grade	Content Strand	Operations & Algebraic Thinking	Numbers & Operations Base Ten	Numbers & Operations Fractions	Measurement & Data	Geometry	Total Mathematics
3	Operations & Algebraic Thinking	--	0.67	0.71	0.74	0.58	0.88
	Numbers & Operations Base Ten		--	0.60	0.64	0.48	0.76
	Numbers & Operations Fractions			--	0.74	0.55	0.87
	Measurement & Data				--	0.58	0.93
	Geometry					--	0.71
	Total Raw Score						--
4	Operations & Algebraic Thinking	--	0.68	0.75	0.77	0.63	0.90
	Numbers & Operations Base Ten		--	0.70	0.69	0.60	0.82
	Numbers & Operations Fractions			--	0.77	0.68	0.91
	Measurement & Data				--	0.64	0.91
	Geometry					--	0.79
Total Raw Score						--	
5	Operations & Algebraic Thinking	--	0.67	0.66	0.72	0.66	0.84
	Numbers & Operations Base Ten		--	0.66	0.68	0.64	0.83
	Numbers & Operations Fractions			--	0.74	0.64	0.88
	Measurement & Data				--	0.68	0.91
	Geometry					--	0.84
Total Raw Score						--	

**Table 37. DC CAS 2013 Mathematics Strand Correlations by Grade (continued)**

<b>Grade</b>	<b>Content Strand</b>	<b>Ratio &amp; Proportional Relationships</b>	<b>The Number System</b>	<b>Expressions &amp; Equations</b>	<b>Geometry</b>	<b>Statistics &amp; Probability</b>	<b>Total Mathematics</b>
6	Ratio & Proportional Relationships	--	0.69	0.70	0.67	0.66	0.86
	The Number System		--	0.73	0.68	0.66	0.88
	Expressions and Equations			--	0.69	0.68	0.89
	Geometry				--	0.64	0.84
	Statistics & Probability					--	0.83
	Total Raw Score						--
7	Ratio & Proportional Relationships	--	0.68	0.73	0.70	0.66	0.89
	The Number System		--	0.68	0.63	0.62	0.83
	Expressions and Equations			--	0.69	0.64	0.89
	Geometry				--	0.62	0.84
	Statistics & Probability					--	0.82
	Total Raw Score						--
<b>Grade</b>	<b>Content Strand</b>	<b>The Number System</b>	<b>Expressions &amp; Equations</b>	<b>Functions</b>	<b>Geometry</b>	<b>Statistics &amp; Probability</b>	<b>Total Mathematics</b>
8	The Number System	--	0.59	0.58	0.53	0.46	0.74
	Expressions and Equations		--	0.75	0.69	0.61	0.89
	Functions			--	0.69	0.63	0.90
	Geometry				--	0.58	0.84
	Statistics & Probability					--	0.77
	Total Raw Score						--
<b>Grade</b>	<b>Content Strand</b>	<b>Number &amp; Quantity</b>	<b>Algebra</b>	<b>Functions</b>	<b>Geometry</b>	<b>Statistics &amp; Probability</b>	<b>Total Mathematics</b>
10	Number & Quantity	--	0.61	0.58	0.58	0.31	0.75
	Algebra		--	0.77	0.72	0.36	0.92
	Functions			--	0.69	0.35	0.86
	Geometry				--	0.38	0.89
	Statistics & Probability					--	0.49
	Total Raw Score						--

**Table 38. DC CAS 2013 Science/Biology Strand Correlations by Grade**

<b>Grade</b>	<b>Content Strand</b>	<b>Science and Technology</b>	<b>Earth and Space Science</b>	<b>Physical Science</b>	<b>Life Science</b>	<b>Total Science</b>
5	Science and Technology	--	0.70	0.62	0.68	0.90
	Earth and Space Science		--	0.62	0.68	0.87
	Physical Science			--	0.59	0.78
	Life Science				--	0.87
	Total Raw Score					--
<b>Grade</b>	<b>Content Strand</b>	<b>Scientific Thinking and Inquiry</b>	<b>Matter and Reactions</b>	<b>Forces</b>	<b>Energy and Waves</b>	<b>Total Science</b>
8	Scientific Thinking and Inquiry	--	0.66	0.61	0.60	0.82
	Matter & Reactions		--	0.66	0.67	0.92
	Forces			--	0.61	0.81
	Energy and Waves				--	0.85
	Total Raw Score					--
<b>Grade</b>	<b>Content Strand</b>	<b>Cell Biology and Biochemistry</b>	<b>Genetics and Evolution</b>	<b>Multicellular Organisms</b>	<b>Ecosystems</b>	<b>Total Biology</b>
High School	Cell Biology and Biochemistry	--	0.61	0.60	0.55	0.87
	Genetics & Evolution		--	0.56	0.52	0.84
	Multicellular Organisms			--	0.54	0.81
	Ecosystems				--	0.75
	Total Raw Score					--

**Table 39. DC CAS 2013 Composition Rubric Score Correlations by Grade**

<b>Grade</b>	<b>Content Strand</b>	<b>Writing Topic Development</b>	<b>Writing Language Conventions</b>	<b>Understanding Literary Text</b>	<b>Total Composition</b>
4	Writing Topic Development	--	0.73	0.87	0.95
	Writing Language Conventions		--	0.75	0.87
	Understanding Literary Text			--	0.94
	Total Raw Score				--
<b>Grade</b>	<b>Content Strand</b>	<b>Writing Topic Development</b>	<b>Writing Language Conventions</b>	<b>Understanding Informational Text</b>	<b>Total Composition</b>
7	Writing Topic Development	--	0.81	0.89	0.96
	Writing Language Conventions		--	0.83	0.92
	Understanding Informational Text			--	0.95
	Total Raw Score				--
<b>Grade</b>	<b>Content Strand</b>	<b>Writing Topic Development</b>	<b>Writing Language Conventions</b>	<b>Understanding Literary Text</b>	<b>Total Composition</b>
10	Writing Topic Development	--	0.77	0.87	0.96
	Writing Language Conventions		--	0.75	0.89
	Understanding Literary Text			--	0.94
	Total Raw Score				--

**Table 40. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Reading**

Raw Score	Grade 2		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 9		Grade 10	
	Scale Score	SEM																
0	200	28	300	36	400	37	500	40	600	35	700	41	800	39	900	35	900	37
1	200	28	300	36	400	37	500	40	600	35	700	41	800	39	900	35	900	37
2	200	28	300	36	400	37	500	40	600	35	700	41	800	39	900	35	900	37
3	200	28	300	36	400	37	500	40	600	35	700	41	800	39	900	35	900	37
4	200	28	300	36	400	37	500	40	600	35	700	41	800	39	900	35	900	37
5	200	28	300	36	400	37	500	40	600	35	700	41	800	39	900	35	900	37
6	200	28	300	36	400	37	500	40	600	35	700	41	800	39	900	35	900	37
7	200	28	300	36	400	37	500	40	600	35	700	41	800	39	900	35	900	37
8	202	26	300	36	400	37	500	40	600	35	700	41	800	39	900	35	900	37
9	213	15	300	36	408	29	500	40	602	33	700	41	806	33	900	35	905	32
10	218	10	313	23	420	18	512	27	617	19	721	20	821	18	907	28	917	19
11	221	8	321	15	425	12	523	17	622	13	728	13	826	13	917	18	923	13
12	224	7	326	10	429	9	528	12	626	9	732	10	830	9	923	13	927	10
13	227	6	329	8	431	7	531	9	629	7	735	8	833	7	926	10	930	8
14	229	5	331	6	433	6	534	7	631	6	737	6	835	6	929	8	933	7
15	230	5	333	6	435	5	536	6	632	5	739*	6	836	5	932*	7	934	6
16	232*	5	334	5	436	5	538	5	634	5	740	5	838	5	934	6	936	5
17	234	5	336	4	438	4	539	5	635	4	742	5	839	4	936	6	938	5
18	235	4	337	4	439*	4	541*	4	636	4	743	4	841*	4	937	5	939	4
19	237	4	338	4	440	4	542	4	638	4	744	4	842	4	939	5	940*	4
20	238	4	339*	4	441	4	543	4	639	4	745	4	843	4	940	5	942	4
21	240	4	341	4	442	4	544	4	640*	3	747	4	844	4	942	4	943	4
22	241	4	342	3	443	3	545	4	641	3	748	4	845	4	943	4	944	4
23	242	4	343	3	444	3	546	3	642	3	749	3	846	4	944	4	945	4
24	244	4	343	3	445	3	547	3	643	3	750	3	847	3	945	4	946	4
25	245	4	344	3	446	3	548	3	643	3	751	3	848	3	946	4	947	3
26	246*	4	345	3	447	3	549	3	644	3	752	3	849	3	947	4	948	3
27	248	4	346	3	448	3	550	3	645	3	752	3	850	3	948	3	949	3

\*Proficiency Level Scale Score cuts (Basic, Proficient, Advanced).

**Table 40. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Reading (continued)**

Raw Score	Grade 2		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 9		Grade 10	
	Scale Score	SEM																
28	249	4	347	3	449	3	551	3	646	3	753	3	851	3	949	3	950	3
29	251	4	348	3	450	3	552	3	647	3	754	3	852	3	950*	3	951	3
30	252	4	349	3	451	3	552	3	648	3	755	3	853	3	951	3	952	3
31	254	4	350	3	452	3	553	3	649	3	756*	3	854	3	952	3	953	3
32	256	4	351	3	453	3	554	3	650	3	757	3	855	3	953	3	954	3
33	257	4	351	3	454	3	555	3	651	3	758	3	856*	3	954	3	954	3
34	260	5	352	3	455*	3	556*	3	652	3	759	3	857	3	955	3	955	3
35	262	5	353	3	456	3	557	3	653	3	760	3	858	3	956	3	956*	3
36	265*	6	354*	3	457	3	558	3	654	3	761	3	859	3	957	3	957	3
37	270	8	355	3	458	3	559	3	655*	3	762	3	860	3	958	3	958	3
38	278	12	356	3	459	3	559	3	656	3	763	3	862	3	959	3	960	3
39	299	30	357	3	460	3	560	3	657	3	764	3	863	3	960*	3	961	3
40	.	.	358	3	461	4	561	3	658	4	766	3	864	3	961	3	962	4
41	.	.	360	3	462	4	563	3	660	4	767	3	865	3	962	3	963	4
42	.	.	361	4	464	4	564	3	661	4	768*	4	866	4	964	4	964	4
43	.	.	362	4	465	4	565	4	662	4	769	4	868	4	965	4	966	4
44	.	.	364	4	466	4	566	4	664	4	771	4	869	4	966	4	967	4
45	.	.	365	4	468	4	568	4	665	4	772	4	870*	4	968	4	969	4
46	.	.	367	4	470	5	569	4	667	4	774	4	872	4	969	4	970*	4
47	.	.	369	5	472*	5	571	4	669	5	776	5	874	4	971	5	972	5
48	.	.	371	5	474	5	573*	5	671	5	778	5	876	5	974	5	974	5
49	.	.	374*	6	477	6	575	5	674*	6	781	6	878	5	976	6	977	5
50	.	.	378	6	480	7	578	6	677	6	784	6	881	6	980	7	980	6
51	.	.	382	8	484	8	582	7	681	8	788	8	885	7	985	9	983	7
52	.	.	390	10	491	11	587	9	687	10	794	10	891	10	994	13	989	9
53	.	.	399	15	499	15	597	15	698	15	799	12	899	14	999	17	998	13
54	.	.	399	15	499	15	599	16	699	16	799	12	899	14	.	.	999	13

\*Proficiency Level Scale Score cuts (Basic, Proficient, Advanced).

**Table 41. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Mathematics**

Raw Score	Grade 2		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
	Scale Score	SEM														
0	200	34	300	26	400	41	500	42	600	36	700	45	800	44	900	49
1	200	34	300	26	400	41	500	42	600	36	700	45	800	44	900	49
2	200	34	300	26	400	41	500	42	600	36	700	45	800	44	900	49
3	200	34	300	26	400	41	500	42	600	36	700	45	800	44	900	49
4	200	34	300	26	400	41	500	42	600	36	700	45	800	44	900	49
5	200	34	300	26	400	41	500	42	600	36	700	45	800	44	900	49
6	200	34	300	26	400	41	500	42	600	36	700	45	800	44	900	49
7	200	34	300	26	400	41	500	42	600	36	700	45	800	44	900	49
8	200	34	300	26	400	41	500	42	600	36	700	45	800	44	900	49
9	207	26	300	26	400	41	500	42	600	36	700	45	800	44	900	49
10	217	17	300	26	400	41	500	42	605	30	700	45	800	44	900	49
11	222	12	306	21	412	29	512	30	616	20	707	38	806	39	916	33
12	225	9	311	15	422	19	523	19	622	14	722	24	824	20	929	20
13	228	7	316	13	427	13	529	14	626	11	729	17	830	14	935*	14
14	230	6	319	11	431	10	532	10	629	9	733	12	834	11	939	11
15	232	5	322	10	434	8	535	8	632	8	737*	10	837*	9	942	9
16	234	5	325	9	437	7	537	7	634	7	739	8	840	7	945	7
17	235	4	327	8	439	6	539	6	636*	6	742	7	842	6	947	6
18	237	4	330	8	440	6	541	6	638	6	744	7	844	6	949	6
19	238	4	332	7	442	5	543*	5	640	5	746	6	845	5	950	5
20	239	4	334	7	444*	5	544	5	641	5	747	5	847	5	952*	5
21	240	4	335	6	445	5	545	5	643	5	749	5	848	5	953	5
22	242	3	337	6	446	4	547	4	644	5	750	5	849	4	955	4
23	243	3	339	6	448	4	548	4	646	5	751	4	851*	4	956	4
24	244*	3	340*	5	449	4	549	4	647	4	753*	4	852	4	957	4
25	245	3	341	5	450	4	550	4	648	4	754	4	853	4	958	4
26	246	3	343	5	451	4	551	4	649	4	755	4	854	4	959	4
27	247	3	344	5	452	4	552	4	651	4	756	4	855	4	960	4

\*Proficiency Level Scale Score cuts (Basic, Proficient, Advanced).

**Table 41. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Mathematics (continued)**

Raw Score	Grade 2		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
	Scale Score	SEM														
28	248	3	345	4	453	3	553	4	652	4	757	4	856	3	961	3
29	249	3	346	4	454	3	554	4	653	4	758	4	857	3	962	3
30	250	3	348	4	455	3	555	4	654*	4	759	3	858	3	963	3
31	251	3	349	4	456	3	556	4	655	4	760	3	859	3	964	3
32	252	3	350	4	457	3	557	4	656	4	761	3	860	3	965	3
33	253	3	351	4	458*	3	558	4	657	4	762	3	861	3	966	3
34	254	3	352	4	459	3	559	3	658	4	763	3	862	3	967	3
35	255*	3	353	4	459	3	561*	3	659	3	764	3	863	3	968	3
36	256	3	354	4	460	3	562	3	660	3	765	3	864	3	969	3
37	258	3	355	4	461	3	563	3	661	3	766	3	865	3	970	3
38	259	3	356	4	462	3	564	3	662	3	767	3	866	3	971*	3
39	261	4	357	4	463	3	564	3	663	3	768	3	867	3	971*	3
40	262	4	358	4	464	3	565	3	664	3	769	3	868*	3	972	3
41	264	4	360*	4	465	3	566	3	665	3	770*	3	869	3	973	3
42	267	5	361	4	466	3	567	3	666	3	771	3	870	3	974	3
43	270*	6	362	4	467	3	568	3	667	3	772	3	871	3	975	3
44	274	8	363	4	468	3	569	3	668*	3	773	3	872	3	976	3
45	284	15	364	4	469	3	570	3	669	4	774	3	874	3	977	3
46	299	30	366	4	470	3	571	3	671	4	775	3	875	3	978	3
47	.	.	367	4	472	3	572	3	672	4	776	3	876	3	980	3
48	.	.	368	4	473	4	574	3	673	4	777	4	877	4	981	4
49	.	.	370	4	474*	4	575*	3	674	4	779	4	879	4	982	4
50	.	.	371	4	476	4	576	3	676	4	780	4	881	4	984	4
51	.	.	373	4	477	4	577	3	677	4	782	4	882	4	985	4
52	.	.	374	4	479	4	578	3	679	4	783	4	884	4	987	4
53	.	.	376*	5	480	4	580	4	681	5	785	4	887	5	989	5
54	.	.	378	5	482	4	581	4	683	5	787	5	890	6	992	5
55	.	.	380	5	485	4	583	4	686	5	790	5	894	7	994	6

\*Proficiency Level Scale Score cuts (Basic, Proficient, Advanced).

**Table 41. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Mathematics (continued)**

Raw Score	Grade 2		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 10	
	Scale Score	SEM														
56	.	.	383	6	487	5	585	5	689	5	792	6	899	9	998	7
57	.	.	386	6	491	6	588	6	692	6	796	7	899	9	999	7
58	.	.	391	8	496	8	592	7	697	9	799	8	.	.	999	7
59	.	.	398	12	499	10	599	10	699	10	799	8	.	.	999	7
60	.	.	399	12	499	10	599	10	699	10	799	8	.	.	.	.

\*Proficiency Level Scale Score cuts (Basic, Proficient, Advanced).

**Table 42: DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Science/Biology**

Raw Score	Grade 5		Grade 8		High School	
	Scale Score	SEM	Scale Score	SEM	Scale Score	SEM
0	500	43	800	53	900	50
1	500	43	800	53	900	50
2	500	43	800	53	900	50
3	500	43	800	53	900	50
4	500	43	800	53	900	50
5	500	43	800	53	900	50
6	500	43	800	53	900	50
7	500	43	800	53	900	50
8	500	43	800	53	900	50
9	500	43	800	53	900	50
10	516	27	800	53	900	50
11	527	16	837	16	930	20
12	532	11	842	11	937	13
13	535	8	845	8	941	10
14	538	7	847	6	943	7
15	540	6	849*	5	945	6
16	541*	5	851	4	947*	5
17	543	5	852	4	949	4
18	544	4	853	3	950	4
19	545	4	854	3	951	4
20	546	4	855	3	952*	3
21	548	3	856*	3	953	3
22	549	3	856*	3	954	3
23	549	3	857	3	955	3
24	550	3	858	2	956	3
25	551	3	859	2	956	2
26	552	3	859	2	957	2
27	553*	3	860	2	958	2
28	554	3	861	2	958	2
29	554	3	861	2	959	2
30	555	2	862	2	960	2
31	556	2	862	2	960	2
32	557	2	863	2	961	2
33	558	2	864	2	961	2
34	558	2	864	2	962	2
35	559	2	865	2	963	2
36	560	2	866	2	963	2
37	561	2	866	2	964	2
38	561	2	867	2	964	2
39	562	2	868*	2	965	2
40	563	2	868*	2	966*	2
41	564*	2	869	2	967	2
42	565	2	870	2	967	2
43	566	3	871	2	968	2
44	567	3	872	2	969	2

\*Proficiency Level Scale Score cuts (Basic, Proficient, Advanced).

**Table 42. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Science/Biology (continued)**

Raw Score	Grade 5		Grade 8		High School	
	Scale Score	SEM	Scale Score	SEM	Scale Score	SEM
45	568	3	873	2	970	2
46	569	3	874	3	971	2
47	570	3	875	3	972	3
48	572	3	877	3	973	3
49	574	4	878	4	975	3
50	576	4	881	5	977	4
51	580	6	885	6	981	6
52	586	9	891	8	987	8
53	599	21	899	13	999	17

\*Proficiency Level Scale Score cuts (Basic, Proficient, Advanced).

**Table 43. DC CAS 2013 Number Correct to Scale Score Conversions with Associated Standard Errors of Measurement (SEM): Composition**

Raw Score	Grade 4		Grade 7		Grade 10	
	Scale Score	SEM	Scale Score	SEM	Scale Score	SEM
0	400	15	700	25	900	19
1	410	10	719	8	914	9
2	420	9	727	6	922	7
3	429	9	733	6	928	7
4	438	8	738	6	934	7
5	446*	8	743	5	939	7
6	453	7	747*	5	944*	6
7	459*	7	752	5	949	6
8	465	7	756*	6	953	6
9	470*	7	761	6	958*	7
10	475	7	766	6	964	7
11	481	7	772*	6	970*	8
12	487	7	778	7	978	9
13	495	9	786	9	990	12
14	499	11	799	15	999	15

\*Proficiency Level Scale Score cuts (Basic, Proficient, Advanced).

**Table 44. DC CAS 2013 Percentages of Students at Each Performance Level**

Content	Grade	Spring 2012 Impact Data					Spring 2013 Impact Data				
		N	Percent of Students at Each Performance Level				N	Percent of Students at Each Performance Level			
			Below Basic	Basic	Proficient	Advanced		Below Basic	Basic	Proficient	Advanced
Reading	2	4,491	22.24	33.44	36.12	8.19	4,802	19.24	35.53	39.53	5.71
	3	4,754	21.64	38.16	36.52	3.68	4,914	20.57	36.37	37.71	5.35
	4	4,589	15.86	35.93	41.73	6.47	4,574	12.79	35.64	42.30	9.27
	5	4,744	14.42	38.26	38.85	8.47	4,540	11.32	35.57	46.21	6.89
	6	4,545	17.43	42.22	36.13	4.22	4,572	19.42	39.48	36.59	4.51
	7	4,301	11.04	39.64	35.97	13.35	4,531	7.92	37.36	37.19	17.52
	8	4,359	14.18	38.15	37.62	10.05	4,260	11.60	33.85	42.49	12.07
	9	4,164	16.55	41.50	22.79	19.16	4,879	13.59	35.64	26.54	24.23
	10	4,272	18.38	39.79	32.44	9.39	4,752	18.75	37.90	31.69	11.66
Mathematics	2	4,514	20.36	32.19	36.89	10.57	4,840	19.96	34.28	35.25	10.52
	3	4,781	21.31	42.25	27.25	9.18	4,947	18.42	39.28	30.50	11.81
	4	4,603	15.75	33.67	37.98	12.60	4,620	15.58	26.99	41.32	16.10
	5	4,759	17.08	34.33	36.25	12.33	4,572	13.34	36.70	32.79	17.17
	6	4,567	16.33	35.84	32.95	14.87	4,598	15.44	32.99	32.95	18.62
	7	4,325	12.79	29.36	43.56	14.29	4,563	13.89	29.24	39.84	17.03
	8	4,381	14.13	29.35	45.04	11.48	4,302	10.60	25.06	47.09	17.25
	10	4,245	22.00	36.35	34.65	7.00	4,729	21.76	35.08	33.69	9.47
Science/Biology <sup>1</sup>	5	4,707	18.89	42.91	30.74	7.46	4,523	15.85	43.91	31.46	8.78
	8	4,263	35.05	25.17	34.37	5.42	4,135	28.30	28.13	37.03	6.55
	10	3,715	28.34	26.59	41.48	3.58	4,047	34.20	22.44	40.45	2.92
Composition	4	4,470	26.60	31.99	23.76	17.65	4,470	24.81	30.47	24.23	20.49
	7	4,146	18.07	28.87	32.34	20.72	4,403	14.85	25.66	44.56	14.92
	10	3,511	28.60	25.43	23.84	22.13	4,396	16.56	34.76	25.86	22.82

**Note:** Total percentages for a grade may not sum to 100 due to rounding.

<sup>1</sup> Biology is administered to students in Grades 8–12, the grade in which they elect to take the Biology course.

**Table 45. Classification Consistency and Accuracy Rates by Grade and Cut Score: Reading**

Grade	Reading Classification Consistency and Accuracy		Basic	Proficient	Advanced	All Cuts
2	Classification Consistency	Consistency	0.91	0.87	0.95	0.72
		Kappa	0.70	0.73	0.61	0.60
	Classification Accuracy	Accuracy	0.94	0.91	0.96	0.80
		False Positive Errors	0.03	0.04	0.01	0.09
		False Negative Errors	0.03	0.05	0.03	0.11
3	Classification Consistency	Consistency	0.93	0.90	0.95	0.78
		Kappa	0.78	0.81	0.60	0.69
	Classification Accuracy	Accuracy	0.95	0.93	0.97	0.84
		False Positive Errors	0.03	0.02	0.01	0.06
		False Negative Errors	0.02	0.05	0.02	0.09
4	Classification Consistency	Consistency	0.94	0.88	0.92	0.75
		Kappa	0.74	0.77	0.60	0.63
	Classification Accuracy	Accuracy	0.96	0.91	0.94	0.81
		False Positive Errors	0.02	0.02	0.01	0.05
		False Negative Errors	0.02	0.07	0.05	0.14
5	Classification Consistency	Consistency	0.95	0.89	0.94	0.77
		Kappa	0.73	0.77	0.58	0.64
	Classification Accuracy	Accuracy	0.96	0.91	0.95	0.83
		False Positive Errors	0.02	0.02	0.01	0.06
		False Negative Errors	0.02	0.06	0.04	0.12
6	Classification Consistency	Consistency	0.93	0.89	0.95	0.77
		Kappa	0.77	0.78	0.54	0.66
	Classification Accuracy	Accuracy	0.95	0.92	0.96	0.84
		False Positive Errors	0.02	0.04	0.01	0.07
		False Negative Errors	0.03	0.04	0.02	0.10
7	Classification Consistency	Consistency	0.95	0.89	0.91	0.74
		Kappa	0.67	0.77	0.70	0.63
	Classification Accuracy	Accuracy	0.96	0.92	0.94	0.82
		False Positive Errors	0.02	0.04	0.02	0.08
		False Negative Errors	0.02	0.04	0.04	0.11
8	Classification Consistency	Consistency	0.94	0.88	0.92	0.75
		Kappa	0.72	0.76	0.66	0.63
	Classification Accuracy	Accuracy	0.96	0.92	0.94	0.82
		False Positive Errors	0.02	0.04	0.02	0.08
		False Negative Errors	0.02	0.04	0.03	0.10
9	Classification Consistency	Consistency	0.92	0.89	0.91	0.73
		Kappa	0.66	0.79	0.78	0.63
	Classification Accuracy	Accuracy	0.94	0.92	0.93	0.80
		False Positive Errors	0.03	0.04	0.02	0.08
		False Negative Errors	0.03	0.04	0.05	0.12
10	Classification Consistency	Consistency	0.92	0.90	0.93	0.75
		Kappa	0.74	0.79	0.71	0.65
	Classification Accuracy	Accuracy	0.95	0.93	0.95	0.83
		False Positive Errors	0.03	0.04	0.01	0.08
		False Negative Errors	0.03	0.03	0.03	0.09

**Table 46. Classification Consistency and Accuracy Rates by Grade and Cut Score:  
Mathematics**

Grade	Mathematics Classification Consistency and Accuracy		Basic	Proficient	Advanced	All Cuts
2	Classification Consistency	Consistency	0.93	0.89	0.91	0.73
		Kappa	0.77	0.77	0.61	0.62
	Classification Accuracy	Accuracy	0.94	0.92	0.94	0.80
		False Positive Errors	0.02	0.03	0.02	0.07
		False Negative Errors	0.04	0.05	0.04	0.13
3	Classification Consistency	Consistency	0.92	0.91	0.94	0.77
		Kappa	0.73	0.81	0.73	0.67
	Classification Accuracy	Accuracy	0.94	0.93	0.96	0.83
		False Positive Errors	0.03	0.02	0.01	0.06
		False Negative Errors	0.03	0.04	0.03	0.11
4	Classification Consistency	Consistency	0.93	0.91	0.93	0.77
		Kappa	0.73	0.82	0.75	0.68
	Classification Accuracy	Accuracy	0.95	0.94	0.95	0.84
		False Positive Errors	0.02	0.02	0.01	0.06
		False Negative Errors	0.03	0.04	0.04	0.11
5	Classification Consistency	Consistency	0.93	0.90	0.94	0.77
		Kappa	0.70	0.81	0.78	0.68
	Classification Accuracy	Accuracy	0.95	0.93	0.95	0.83
		False Positive Errors	0.02	0.04	0.01	0.07
		False Negative Errors	0.03	0.03	0.04	0.09
6	Classification Consistency	Consistency	0.92	0.91	0.93	0.75
		Kappa	0.69	0.81	0.77	0.66
	Classification Accuracy	Accuracy	0.94	0.93	0.95	0.82
		False Positive Errors	0.03	0.02	0.02	0.07
		False Negative Errors	0.03	0.05	0.02	0.10
7	Classification Consistency	Consistency	0.89	0.88	0.94	0.72
		Kappa	0.55	0.76	0.80	0.60
	Classification Accuracy	Accuracy	0.92	0.92	0.96	0.80
		False Positive Errors	0.04	0.04	0.01	0.09
		False Negative Errors	0.04	0.05	0.03	0.11
8	Classification Consistency	Consistency	0.91	0.88	0.93	0.71
		Kappa	0.54	0.73	0.75	0.58
	Classification Accuracy	Accuracy	0.93	0.91	0.95	0.79
		False Positive Errors	0.04	0.05	0.02	0.11
		False Negative Errors	0.03	0.04	0.03	0.10
10	Classification Consistency	Consistency	0.81	0.87	0.96	0.65
		Kappa	0.45	0.73	0.80	0.50
	Classification Accuracy	Accuracy	0.86	0.91	0.97	0.74
		False Positive Errors	0.07	0.05	0.01	0.12
		False Negative Errors	0.08	0.05	0.02	0.14

**Table 47. Classification Consistency and Accuracy Rates by Grade and Cut Score:  
Science/Biology**

Grade	Science/Biology Classification Consistency and Accuracy		Basic	Proficient	Advanced	All Cuts
5	Classification Consistency	Consistency	0.87	0.88	0.95	0.70
		Kappa	0.57	0.75	0.72	0.57
	Classification Accuracy	Accuracy	0.91	0.92	0.96	0.79
		False Positive Errors	0.05	0.04	0.01	0.10
		False Negative Errors	0.04	0.05	0.03	0.12
8	Classification Consistency	Consistency	0.82	0.91	0.97	0.71
		Kappa	0.64	0.81	0.82	0.58
	Classification Accuracy	Accuracy	0.80	0.92	0.96	0.69
		False Positive Errors	0.17	0.06	0.00	0.23
		False Negative Errors	0.03	0.01	0.04	0.08
High School	Classification Consistency	Consistency	0.81	0.84	0.98	0.68
		Kappa	0.59	0.67	0.68	0.51
	Classification Accuracy	Accuracy	0.86	0.89	0.98	0.74
		False Positive Errors	0.08	0.04	0.00	0.12
		False Negative Errors	0.06	0.07	0.01	0.13

**Table 48. Classification Consistency and Accuracy Rates by Grade and Cut Score:  
Composition**

Grade	Composition Classification Consistency and Accuracy		Basic	Proficient	Advanced	All Cuts
4	Classification Consistency	Consistency	0.89	0.78	0.87	0.59
		Kappa	0.71	0.56	0.62	0.45
	Classification Accuracy	Accuracy	0.93	0.84	0.91	0.69
		False Positive Errors	0.04	0.07	0.04	0.13
		False Negative Errors	0.03	0.10	0.05	0.18
7	Classification Consistency	Consistency	0.88	0.86	0.87	0.64
		Kappa	0.64	0.70	0.59	0.50
	Classification Accuracy	Accuracy	0.92	0.91	0.91	0.74
		False Positive Errors	0.07	0.04	0.02	0.13
		False Negative Errors	0.01	0.05	0.07	0.14
10	Classification Consistency	Consistency	0.88	0.82	0.87	0.60
		Kappa	0.62	0.63	0.66	0.46
	Classification Accuracy	Accuracy	0.91	0.87	0.91	0.71
		False Positive Errors	0.06	0.07	0.03	0.16
		False Negative Errors	0.03	0.05	0.06	0.13

**Table 49. Correlations Between Reading, Mathematics, Science/Biology, and Composition Total Test Raw Scores, by Grade**

<b>Grade</b>	<b>Mathematics</b>	<b>Science/Biology*</b>	<b>Composition</b>
<b>Reading</b>			
Grade 2	0.75	--	--
Grade 3	0.79	--	--
Grade 4	0.79	--	0.63
Grade 5	0.76	0.79	--
Grade 6	0.78	--	--
Grade 7	0.77	--	0.64
Grade 8	0.76	0.73	--
Grade 9	--	0.73	--
Grade 10	0.68	0.65	0.69
<b>Mathematics</b>			
Grade 4	--	--	0.57
Grade 5	--	0.78	--
Grade 7	--	--	0.58
Grade 8	--	0.76	--
Grade 10	--	0.56	0.57
<b>Science/Biology</b>			
Grade 10	--	--	0.47

*Note:* "--" = not applicable.

\*In Biology all grades were used in the analyses but only Grades 9 and 10 can be used for the correlations since the other grades are not in common with other content areas.

## References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2009). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, *46*, 443–459.
- Burket, G. R. (2000). ITEMWIN [Computer program]. Unpublished.
- CTB/McGraw-Hill. (2011). *District of Columbia Comprehensive Assessment System (DC CAS) grade 9 reading bookmark standard setting technical report 2011*. Monterey, CA: Author.
- CTB/McGraw-Hill. (2011). *District of Columbia Comprehensive Assessment System (DC CAS) test chairperson's manual: Reading and mathematics, composition, science, and biology*. Monterey, CA: Author.
- CTB/McGraw-Hill. (2011). *District of Columbia Comprehensive Assessment System (DC CAS) test directions: Reading and mathematics (grades 4–8 and 10), composition (grades 4, 7, and 10), science (grades 5 and 8), and biology*. Monterey, CA: Author.
- CTB/McGraw-Hill. (2012). *District of Columbia Comprehensive Assessment System (DC CAS) Standard Setting Technical Report for Grades 3–10 Reading, Grade 2 Reading and Mathematics, and Grades 4, 7, and 10 Composition*. Monterey, CA: Author.
- CTB/McGraw-Hill. (2011). *District of Columbia Public Schools (DCPS) grade 9 reading technical report 2011*. Monterey, CA: Author.
- CTB/McGraw-Hill Education. (2011). PARDUX (Version 1.66) [Software]. Monterey, CA: CTB/McGraw Hill Education.
- CTB/McGraw-Hill Education. (2013). *District of Columbia Comprehensive Assessment System (DC CAS) Cutscore Review Technical Report 2013*. Monterey, CA: Author.
- CTB/McGraw-Hill Education. (2013). *District of Columbia Comprehensive Assessment System (DC CAS) test chairperson's manual: Reading and mathematics, composition, science, biology, and health and physical education*. Monterey, CA: Author.
- CTB/McGraw-Hill Education. (2013). *District of Columbia Comprehensive Assessment System (DC CAS) test directions: Reading and mathematics (grade 2)*. Monterey, CA: Author.
- CTB/McGraw-Hill Education. (2013). *District of Columbia Comprehensive Assessment System (DC CAS) test directions: Reading and mathematics (grade 3)*. Monterey, CA: Author.
- CTB/McGraw-Hill Education. (2013). *District of Columbia Comprehensive Assessment System (DC CAS) test directions: Reading and mathematics (grades 4–8 and 10), composition (grades 4, 7, and 10), science (grades 5, 8, and biology), and health (grades 5, 8, and high school)*. Monterey, CA: Author.
- Haertel, E.H., Beimers, J., & Miles, J. (2012). The briefing book method. In G.J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives* (2nd Ed., pp. 283–300). Mahwah, NJ: Lawrence Erlbaum.
- Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, *10*, 159–170.
- Jaeger, R. M. (1995). Setting standards for complex performances: An iterative, judgmental

- policy-capturing strategy. *Educational Measurement: Issues and Practice*, 14(4), 16–20.
- Kim, D., Barton, K., & Kim, X. (2008). *Estimating Classification Consistency and Classification Accuracy With Pattern Scoring*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Kim, D., Choi, S., Um, K., & Kim, J. (2006). *A comparison of methods for estimating classification consistency*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.
- Kolen, M. J., & Kim, D. (2005). Personal correspondence.
- Landis, J. R., & Koch, G. G. (1997). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996, June). Standard setting: A bookmark approach. In D. R. Green (Chair), *IRT-based standard setting procedures utilizing behavioral anchoring*. Symposium presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment. Phoenix, AZ.
- Linn, R. L., & Harnisch, D. L. (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurement*, 18(2), 109–118.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–197.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.
- Muraki, E., & Bock, R. D. (1991). *PARSCALE: Parameter Scaling of Rating Data* [Computer program]. Chicago, IL: Scientific Software, Inc.
- No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat.1425 (2002).
- Perie, M. (2007). *Setting alternate achievement standards*. Dover, NH: National Center for the Improvement of Educational Assessment. Retrieved January 11, 2008, from [http://www.nciea.org/publications/CCSSO\\_MAP07.pdf](http://www.nciea.org/publications/CCSSO_MAP07.pdf)
- Plake, B.S., & Cizek, G.J. (2012). Variations on a theme: The modified Angoff, extended Angoff, and yes/no Angoff standard setting methods. In G.J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives* (2nd Ed., pp. 181–200). Mahwah, NJ: Lawrence Erlbaum.
- Roeber, E. (2002). *Setting standards on alternate assessments (Synthesis Report 42)*. Minneapolis, MN: National Center on Educational Outcomes. Retrieved January 11, 2008, from <http://cehd.umn.edu/NCEO/OnlinePubs/Synthesis42.html>
- Standards and Assessments Peer Review Guidance*. (2009). Retrieved December 7, 2010, from <http://www.ed.gov/policy/elsec/guid/saaprguidance.pdf>
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201–210.
- Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of Criterion-Referenced Tests: A Decision-Theoretic Formulation, *Journal of Educational Measurement*, 11(4), 263–267.

- Thissen, D. (1982). Marginal maximum-likelihood estimation for the one-parameter logistic model. *Psychometrika*, *47*, 175–186.
- U.S. Department of Education. (2009). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Retrieved December 7, 2010, from <http://www.ed.gov/policy/elsec/guid/saaprguidance.pdf>
- Yen, W.M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, *5*, 245–262.
- Zwick, R., Donoghue, J.R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, *30*, 233–251.

## Appendix A: Checklist for DC Educator Review of DC CAS Items

### A. Checklist for the Content Reviewer

#### For All Items:

##### *Check to ensure that the content of each item:*

- is targeted to assess only one strand or skill
- deals with material that is important in testing the targeted strand or skill
- uses grade-appropriate content and thinking skills
- is presented at a reading level suitable for the grade level being tested
- is accurate and documented against reliable, up-to-date sources

#### For Multiple Choice Items:

##### *Check to ensure that the content of each item:*

- has a stem that facilitates answering the question or completing the statement without looking at the answer choices
- has a stem that does not present clues to the correct answer choice
- has answer choices that are plausible and attractive to the student who has not mastered the Strand or skill
- is conceptually, grammatically, and syntactically consistent—between the stem and answer choices, and among the answer choices
- has mutually exclusive distractors
- has one and only one correct answer choice

#### For Constructed Response Items:

##### *Check to ensure that the content of each item:*

- is written so that a student possessing the knowledge or skill being tested can construct a response that is scorable with the specified rubric or scoring tool; that is, the range of possible correct responses must be wide enough to allow for diversity of responses, but narrow enough so that students who do not clearly show their grasp of the Strand or skill being assessed cannot obtain the maximum score
- is presented without clues to the correct response
- has precise and unambiguous directions for the desired response
- is free of extraneous words or expressions
- is appropriate for the question being asked and the intended response (For example, the item does not ask students to draw pictures of abstract ideas.)
- is conceptually, grammatically, and syntactically consistent

## **B. Checklist for the Sensitivity Reviewer**

To have confidence in test results, it is important to ensure that students are given a reasonable chance to do their best on the test. Test items must be accessible to a diverse student population with respect to gender, race, ethnicity, geographic region, socioeconomic status, and other factors.

*Check to ensure that the content of each item is free of explicit references to or descriptions of:*

- events involving extreme sadness or adversity
- acts of physical or psychological violence
- alcohol or drug abuse
- vulgar language
- sex

*Check to ensure that if any religious, political, social, or philosophical issues are addressed:*

- more than one point of view is expressed
- beliefs or biases do not interfere with factual accuracy
- contemporary issues that have already been proven to be controversial are absent
- stereotypic descriptions of beliefs or customs are absent

*Test items must:*

- be free of offensive, disturbing, or inappropriate language or content
- be free of stereotyping based on:
  - gender
  - race
  - ethnicity
  - religion
  - socioeconomic status
  - age
  - regional or geographic area
  - disability
  - occupation
- demonstrate sensitivity to historical representation of groups
- be free of differential familiarity for any group based on:
  - language
  - socioeconomic status
  - regional or geographic area
  - prior knowledge or experiences unrelated to the subject matter being tested

## Appendix B: DC CAS Composition Scoring Rubrics

### Topic/Idea Development

Score	Description
6	<ul style="list-style-type: none"> <li>• Rich topic/idea development</li> <li>• Careful and/or subtle organization</li> <li>• Effective/rich use of language</li> </ul>
5	<ul style="list-style-type: none"> <li>• Full topic/idea development</li> <li>• Logical organization</li> <li>• Strong details</li> <li>• Appropriate use of language</li> </ul>
4	<ul style="list-style-type: none"> <li>• Moderate topic/idea development and organization</li> <li>• Adequate, relevant details</li> <li>• Some variety in language</li> </ul>
3	<ul style="list-style-type: none"> <li>• Rudimentary topic/idea development and/or organization</li> <li>• Basic supporting ideas</li> <li>• Simplistic language</li> </ul>
2	<ul style="list-style-type: none"> <li>• Limited or weak topic/idea development, organization, and/or details</li> <li>• Limited awareness of audience and/or task</li> </ul>
1	<ul style="list-style-type: none"> <li>• Limited topic/idea development, organization, and/or details</li> <li>• Little or no awareness of audience and/or task</li> </ul>

### Standard English Conventions

Score	Description
4	<ul style="list-style-type: none"> <li>• Control of sentence structure, grammar and usage, and mechanics (length and complexity of essay provide opportunity for student to show control of standard English conventions)</li> </ul>
3	<ul style="list-style-type: none"> <li>• Errors do not interfere with communication and/or</li> <li>• Few errors relative to length of essay or complexity of sentence structure, grammar and usage, and mechanics</li> </ul>
2	<ul style="list-style-type: none"> <li>• Errors interfere somewhat with communication and/or</li> <li>• Too many errors relative to length of the essay or complexity of sentence structure, grammar and usage, and mechanics</li> </ul>
1	<ul style="list-style-type: none"> <li>• Errors seriously interfere with communication AND</li> <li>• Little control of sentence structure, grammar and usage, and mechanics</li> </ul>

### Understanding Literary or Informational Text

Score	Description
4	<p>The response demonstrates an understanding of the complexities of the text.</p> <ul style="list-style-type: none"> <li>• Fully addresses the demands of the question or prompt</li> <li>• Effectively uses explicitly stated text as well as inferences drawn from the text to support an answer or claim</li> </ul>
3	<p>The response demonstrates an understanding of the text.</p> <ul style="list-style-type: none"> <li>• Addresses the demands of the question or prompt</li> <li>• Uses some explicitly stated text and/or some inferences drawn from the text to support an answer or claim</li> </ul>
2	<p>The response is incomplete or oversimplified and demonstrates a partial or literal understanding of the text.</p> <ul style="list-style-type: none"> <li>• Attempts to answer the question or address the prompt</li> <li>• Uses explicitly stated text that demonstrates some understanding</li> </ul>
1	<p>The response shows evidence of a minimal understanding of the text.</p> <ul style="list-style-type: none"> <li>• Shows evidence that some meaning has been derived from the text to answer the question</li> <li>• Has minimal textual evidence</li> </ul>

Note: The Composition prompt will also be aligned to a Common Core Reading standard. Responses will demonstrate degrees of mastery of that reading standard. Reading standards that the composition prompts will align to may include:

- Grade 4: CC.4.R.I.1, CC.4.R.L.2, and CC.4.R.L.4 (see Reading tested standards)
- Grade 7: CC.7.R.I.1, CC.7.R.I.8, CC.7.R.L.1, and CC.7.R.L.2 (see Reading tested standards)
- Grade 10: CC.9-10.R.I.1, CC.9-10.R.I.2, CC.10.R.I.3, CC.9-10.R.L.2, and CC.9-10.R.L.6 (see Reading tested standards)

### Appendix C: Operational and Field Test Item Adjusted *P* Values

**Table C1. DC CAS 2013 Operational Form Item Adjusted *P* Values: Reading**

Reading Grade 2							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,690	1	0.95	25	3,771	1	0.69
2	4,656	1	0.91	26	4,663	1	0.50
3	4,553	1	0.86	27	4,647	1	0.63
4	4,672	1	0.40	28	4,641	1	0.42
5	4,627	1	0.43	29	4,653	1	0.62
6	4,645	1	0.59	30	4,629	1	0.51
7	4,619	1	0.69	31	4,361	3	0.50
8	4,604	1	0.61	32	4,582	1	0.82
9	4,645	1	0.55	33	4,571	1	0.54
10	4,592	1	0.50	34	4,567	1	0.36
11	4,573	1	0.64	35	4,531	1	0.71
12	4,673	1	0.89				
13	4,652	1	0.67				
14	4,532	1	0.67				
15	4,660	1	0.57				
16	4,644	1	0.66				
17	4,585	1	0.66				
18	4,647	1	0.65				
19	4,591	1	0.76				
20	4,488	3	0.46				
21	4,575	1	0.84				
22	4,541	1	0.49				
23	4,623	1	0.92				
24	4,634	1	0.88				

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C1. DC CAS 2013 Operational Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 3							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,844	1	0.75	25	4,823	1	0.70
2	4,845	1	0.65	26	4,824	1	0.61
3	4,832	1	0.65	27	4,670	3	0.42
4	4,832	1	0.38	28	4,793	1	0.61
5	4,814	1	0.30	29	4,765	1	0.56
6	4,699	3	0.30	30	4,790	1	0.50
7	4,817	1	0.83	31	4,775	1	0.55
8	4,748	1	0.75	32	4,768	1	0.80
9	4,792	1	0.65	33	4,645	3	0.43
10	4,780	1	0.70	34	4,780	1	0.84
11	4,675	1	0.79	35	4,776	1	0.65
12	4,780	1	0.51	36	4,798	1	0.76
13	4,765	1	0.71	37	4,786	1	0.65
14	4,772	1	0.48	38	4,788	1	0.69
15	4,755	1	0.79	39	4,820	1	0.85
16	4,839	1	0.56	40	4,807	1	0.75
17	4,829	1	0.60	41	4,816	1	0.54
18	4,827	1	0.48	42	4,802	1	0.74
19	4,796	1	0.71	43	4,410	1	0.73
20	4,823	1	0.90	44	4,828	1	0.73
21	4,814	1	0.65	45	4,762	1	0.45
22	4,835	1	0.55	46	4,763	1	0.27
23	4,829	1	0.64	47	4,743	1	0.40
24	4,807	1	0.70	48	4,739	1	0.53

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C1. DC CAS 2013 Operational Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 4							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,462	1	0.63	25	4,449	1	0.67
2	4,464	1	0.55	26	4,437	1	0.79
3	4,460	1	0.65	27	4,364	3	0.54
4	4,445	1	0.59	28	4,443	1	0.45
5	4,293	3	0.42	29	4,447	1	0.74
6	4,453	1	0.77	30	4,447	1	0.69
7	4,450	1	0.52	31	4,442	1	0.81
8	4,446	1	0.32	32	4,444	1	0.75
9	4,445	1	0.58	33	4,438	1	0.51
10	4,449	1	0.51	34	4,445	1	0.57
11	4,447	1	0.80	35	4,436	1	0.71
12	4,459	1	0.71	36	4,441	1	0.51
13	4,459	1	0.80	37	4,436	1	0.52
14	4,457	1	0.82	38	4,415	1	0.77
15	4,455	1	0.78	39	4,439	1	0.61
16	4,456	1	0.47	40	4,442	1	0.82
17	4,450	1	0.41	41	4,439	1	0.82
18	4,453	1	0.87	42	4,436	1	0.61
19	4,453	1	0.46	43	4,439	1	0.60
20	4,456	1	0.45	44	4,438	1	0.78
21	4,453	1	0.66	45	4,435	1	0.59
22	4,456	1	0.79	46	4,432	1	0.72
23	4,454	1	0.49	47	4,415	1	0.50
24	4,449	1	0.73	48	4,331	3	0.38

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C1. DC CAS 2013 Operational Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 5							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,531	1	0.80	25	4,517	1	0.62
2	4,527	1	0.38	26	4,520	1	0.92
3	4,530	1	0.57	27	4,519	1	0.76
4	4,530	1	0.73	28	4,518	1	0.62
5	4,527	1	0.72	29	4,507	1	0.73
6	4,526	1	0.51	30	4,513	1	0.90
7	4,527	1	0.85	31	4,494	1	0.64
8	4,509	1	0.76	32	4,418	3	0.47
9	4,421	3	0.33	33	4,513	1	0.64
10	4,507	1	0.84	34	4,512	1	0.51
11	4,507	1	0.82	35	4,509	1	0.55
12	4,500	1	0.69	36	4,504	1	0.64
13	4,489	1	0.58	37	4,509	1	0.66
14	4,526	1	0.76	38	4,509	1	0.22
15	4,526	1	0.67	39	4,502	1	0.47
16	4,522	1	0.75	40	4,497	1	0.48
17	4,522	1	0.71	41	4,511	1	0.43
18	4,483	1	0.52	42	4,512	1	0.46
19	4,436	3	0.50	43	4,510	1	0.79
20	4,519	1	0.66	44	4,510	1	0.50
21	4,521	1	0.66	45	4,510	1	0.58
22	4,519	1	0.61	46	4,509	1	0.77
23	4,516	1	0.77	47	4,505	1	0.59
24	4,518	1	0.77	48	4,492	1	0.79

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C1. DC CAS 2013 Operational Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 6							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,552	1	0.61	25	4,504	1	0.30
2	4,558	1	0.82	26	4,507	1	0.59
3	4,554	1	0.75	27	4,505	1	0.68
4	4,544	1	0.61	28	4,505	1	0.67
5	4,495	3	0.49	29	4,504	1	0.61
6	4,546	1	0.60	30	4,541	1	0.89
7	4,546	1	0.75	31	4,541	1	0.84
8	4,541	1	0.74	32	4,536	1	0.55
9	4,540	1	0.68	33	4,539	1	0.60
10	4,541	1	0.38	34	4,540	1	0.84
11	4,541	1	0.44	35	4,539	1	0.79
12	4,477	1	0.84	36	4,536	1	0.40
13	4,477	1	0.59	37	4,530	1	0.40
14	4,477	1	0.63	38	4,536	1	0.59
15	4,475	1	0.76	39	4,535	1	0.65
16	4,466	1	0.26	40	4,536	1	0.71
17	4,466	1	0.80	41	4,533	1	0.64
18	4,548	1	0.69	42	4,492	1	0.70
19	4,537	1	0.55	43	4,457	3	0.45
20	4,448	3	0.34	44	4,528	1	0.68
21	4,546	1	0.75	45	4,527	1	0.64
22	4,541	1	0.68	46	4,527	1	0.52
23	4,501	1	0.67	47	4,525	1	0.53
24	4,498	1	0.54	48	4,520	1	0.77

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C1. DC CAS 2013 Operational Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 7							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,516	1	0.87	25	4,494	1	0.54
2	4,519	1	0.86	26	4,490	1	0.38
3	4,519	1	0.67	27	4,492	1	0.60
4	4,518	1	0.61	28	4,491	1	0.34
5	4,519	1	0.33	29	4,493	1	0.21
6	4,517	1	0.70	30	4,483	1	0.72
7	4,515	1	0.44	31	4,467	1	0.57
8	4,499	1	0.55	32	4,404	3	0.43
9	4,504	1	0.70	33	4,448	1	0.81
10	4,502	1	0.49	34	4,481	1	0.67
11	4,497	1	0.74	35	4,482	1	0.58
12	4,482	1	0.41	36	4,484	1	0.72
13	4,370	3	0.32	37	4,481	1	0.76
14	4,498	1	0.60	38	4,480	1	0.76
15	4,497	1	0.41	39	4,481	1	0.70
16	4,497	1	0.85	40	4,479	1	0.51
17	4,500	1	0.90	41	4,477	1	0.38
18	4,497	1	0.64	42	4,479	1	0.47
19	4,497	1	0.79	43	4,459	1	0.73
20	4,489	1	0.60	44	4,347	3	0.39
21	4,482	1	0.59	45	4,457	1	0.61
22	4,491	1	0.66	46	4,456	1	0.49
23	4,493	1	0.73	47	4,457	1	0.71
24	4,486	1	0.68	48	4,452	1	0.75

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C1. DC CAS 2013 Operational Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 8							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,232	1	0.74	25	4,210	1	0.82
2	4,236	1	0.75	26	4,207	1	0.47
3	4,233	1	0.70	27	4,206	1	0.52
4	4,235	1	0.82	28	4,206	1	0.54
5	4,228	1	0.62	29	4,193	1	0.73
6	4,225	1	0.75	30	4,194	1	0.50
7	4,221	1	0.48	31	4,191	1	0.35
8	4,226	1	0.47	32	4,194	1	0.72
9	4,222	1	0.42	33	4,188	1	0.86
10	4,179	1	0.63	34	4,127	3	0.40
11	4,110	3	0.34	35	4,195	1	0.39
12	4,122	1	0.78	36	4,192	1	0.74
13	4,116	1	0.66	37	4,192	1	0.80
14	4,111	1	0.63	38	4,190	1	0.40
15	4,112	1	0.62	39	4,188	1	0.86
16	4,198	1	0.69	40	4,189	1	0.84
17	4,197	1	0.40	41	4,190	1	0.88
18	4,197	1	0.52	42	4,187	1	0.53
19	4,197	1	0.68	43	4,191	1	0.84
20	4,187	1	0.30	44	4,191	1	0.75
21	4,187	1	0.78	45	4,181	1	0.72
22	4,207	1	0.79	46	4,169	1	0.63
23	4,203	1	0.64	47	4,062	3	0.48
24	4,207	1	0.83	48	4,128	1	0.57

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C1. DC CAS 2013 Operational Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 9							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,772	1	0.54	25	4,641	1	0.60
2	4,767	1	0.75	26	4,634	1	0.56
3	4,774	1	0.42	27	4,628	1	0.56
4	4,772	1	0.58	28	4,309	3	0.41
5	4,768	1	0.60	29	4,637	1	0.74
6	4,772	1	0.79	30	4,635	1	0.59
7	4,766	1	0.62	31	4,632	1	0.58
8	4,767	1	0.70	32	4,633	1	0.43
9	4,772	1	0.78	33	4,635	1	0.53
10	4,746	1	0.59	34	4,629	1	0.36
11	4,750	1	0.84	35	4,628	1	0.67
12	4,747	1	0.77	36	4,571	1	0.65
13	4,735	1	0.63	37	4,598	1	0.63
14	4,432	2	0.45	38	4,588	1	0.56
15	4,735	1	0.85	39	4,589	1	0.61
16	4,730	1	0.77	40	4,587	1	0.39
17	4,720	1	0.68	41	4,619	1	0.59
18	4,401	3	0.30	42	4,632	1	0.48
19	4,713	1	0.70	43	4,625	1	0.54
20	4,706	1	0.48	44	4,624	1	0.52
21	4,710	1	0.61	45	4,611	1	0.43
22	4,714	1	0.64	46	4,618	1	0.59
23	4,703	1	0.53	47	4,619	1	0.53
24	4,699	1	0.48	48	4,618	1	0.32

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C1. DC CAS 2013 Operational Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 10							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,672	1	0.71	25	4,595	1	0.69
2	4,677	1	0.67	26	4,599	1	0.82
3	4,676	1	0.77	27	4,596	1	0.79
4	4,388	3	0.41	28	4,593	1	0.57
5	4,672	1	0.66	29	4,595	1	0.72
6	4,667	1	0.63	30	4,571	1	0.63
7	4,670	1	0.66	31	4,581	1	0.49
8	4,669	1	0.56	32	4,582	1	0.46
9	4,615	1	0.63	33	4,582	1	0.57
10	4,617	1	0.47	34	4,570	1	0.54
11	4,614	1	0.39	35	4,572	1	0.36
12	4,610	1	0.59	36	4,571	1	0.56
13	4,600	1	0.43	37	4,557	1	0.73
14	4,595	1	0.60	38	4,553	1	0.68
15	4,589	1	0.41	39	4,548	1	0.81
16	4,637	1	0.64	40	4,553	1	0.75
17	4,641	1	0.68	41	4,546	1	0.60
18	4,642	1	0.68	42	4,545	1	0.65
19	4,640	1	0.61	43	4,545	1	0.47
20	4,640	1	0.85	44	4,548	1	0.58
21	4,630	1	0.61	45	4,536	1	0.68
22	4,632	1	0.66	46	4,539	1	0.67
23	4,622	1	0.66	47	4,519	1	0.66
24	4,282	3	0.50	48	4,121	3	0.42

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C2. DC CAS 2013 Operational Form Item Adjusted *P* Values: Mathematics**

Mathematics Grade 2							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,731	1	0.74	28	4,710	1	0.48
2	4,739	1	0.93	29	4,731	1	0.73
3	4,735	1	0.64	30	4,728	1	0.73
4	4,718	1	0.59	31	4,715	1	0.77
5	4,737	1	0.86	32	4,720	1	0.59
6	4,724	2	0.35	33	4,721	1	0.84
7	4,738	1	0.82	34	4,722	1	0.41
8	4,733	1	0.83	35	4,700	1	0.90
9	4,718	1	0.73	36	4,715	1	0.62
10	4,736	1	0.98	37	4,712	1	0.89
11	4,731	1	0.88	38	4,711	1	0.46
12	4,725	1	0.73	39	4,715	1	0.92
13	4,733	1	0.86	40	4,716	1	0.91
14	4,720	1	0.57	41	4,718	1	0.56
15	4,744	1	0.60	42	4,713	1	0.93
16	4,728	1	0.81	43	4,709	1	0.65
17	4,740	1	0.72				
18	4,725	1	0.75				
19	4,721	1	0.74				
20	4,678	3	0.44				
21	4,726	1	0.81				
22	4,733	1	0.83				
23	4,722	1	0.68				
24	4,709	1	0.68				
25	4,708	1	0.77				
26	4,663	1	0.74				
27	4,713	1	0.56				

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C2. DC CAS 2013 Operational Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 3							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,867	1	0.49	28	4,852	1	0.85
2	4,869	1	0.37	29	4,840	1	0.68
3	4,868	1	0.50	30	4,790	1	0.67
4	4,855	1	0.94	31	4,806	1	0.26
5	4,827	1	0.55	32	4,844	1	0.53
6	4,617	3	0.72	33	4,819	1	0.84
7	4,852	1	0.76	34	4,805	1	0.57
8	4,820	1	0.30	35	4,826	1	0.69
9	4,819	1	0.64	36	4,831	1	0.90
10	4,812	1	0.54	37	4,797	1	0.79
11	4,849	1	0.78	38	4,758	1	0.66
12	4,845	1	0.90	39	4,834	1	0.70
13	4,832	1	0.73	40	4,825	1	0.37
14	4,844	1	0.27	41	4,841	1	0.69
15	4,860	1	0.97	42	4,833	1	0.81
16	4,843	1	0.65	43	4,832	1	0.43
17	4,855	1	0.72	44	4,836	1	0.28
18	4,841	1	0.32	45	4,817	1	0.43
19	4,805	1	0.63	46	4,808	1	0.29
20	4,784	1	0.30	47	4,812	1	0.45
21	4,752	3	0.44	48	4,374	3	0.83
22	4,830	1	0.48	49	4,815	1	0.62
23	4,802	1	0.54	50	4,841	1	0.96
24	4,770	1	0.69	51	4,830	1	0.63
25	4,636	1	0.73	52	4,763	1	0.66
26	4,830	1	0.81	53	4,766	1	0.72
27	4,823	1	0.56	54	4,729	1	0.81

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C2. DC CAS 2013 Operational Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 4							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,547	1	0.70	28	4,532	1	0.41
2	4,540	1	0.15	29	4,531	1	0.56
3	4,539	1	0.66	30	4,531	1	0.56
4	4,549	1	0.59	31	4,527	1	0.80
5	4,494	1	0.48	32	4,529	1	0.53
6	4,494	3	0.45	33	4,532	1	0.91
7	4,536	1	0.41	34	4,530	1	0.24
8	4,538	1	0.76	35	4,527	1	0.49
9	4,519	1	0.70	36	4,531	1	0.60
10	4,535	1	0.76	37	4,528	1	0.86
11	4,538	1	0.73	38	4,526	1	0.72
12	4,538	1	0.75	39	4,515	1	0.78
13	4,535	1	0.65	40	4,512	1	0.85
14	4,523	1	0.66	41	4,526	1	0.86
15	4,538	1	0.75	42	4,522	1	0.44
16	4,538	1	0.72	43	4,516	1	0.64
17	4,534	1	0.36	44	4,514	1	0.31
18	4,524	1	0.35	45	4,522	1	0.70
19	4,525	1	0.55	46	4,519	1	0.72
20	4,470	1	0.54	47	4,462	1	0.54
21	4,494	3	0.37	48	4,472	3	0.38
22	4,536	1	0.53	49	4,521	1	0.48
23	4,532	1	0.84	50	4,523	1	0.87
24	4,533	1	0.32	51	4,521	1	0.78
25	4,528	1	0.54	52	4,524	1	0.53
26	4,525	1	0.50	53	4,520	1	0.61
27	4,505	1	0.86	54	4,524	1	0.65

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C2. DC CAS 2013 Operational Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 5							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,552	1	0.55	28	4,542	1	0.50
2	4,550	1	0.83	29	4,537	1	0.60
3	4,542	1	0.30	30	4,543	1	0.59
4	4,550	1	0.68	31	4,539	1	0.81
5	4,513	1	0.73	32	4,534	1	0.32
6	4,392	3	0.48	33	4,539	1	0.67
7	4,551	1	0.68	34	4,533	1	0.65
8	4,549	1	0.61	35	4,535	1	0.82
9	4,549	1	0.36	36	4,538	1	0.40
10	4,544	1	0.67	37	4,530	1	0.32
11	4,548	1	0.65	38	4,537	1	0.63
12	4,545	1	0.42	39	4,536	1	0.80
13	4,537	1	0.76	40	4,502	1	0.70
14	4,538	1	0.31	41	4,528	1	0.43
15	4,543	1	0.76	42	4,531	1	0.73
16	4,545	1	0.77	43	4,529	1	0.55
17	4,547	1	0.74	44	4,533	1	0.21
18	4,545	1	0.85	45	4,532	1	0.84
19	4,531	1	0.39	46	4,528	1	0.40
20	4,530	1	0.51	47	4,490	1	0.90
21	4,503	3	0.72	48	4,460	3	0.33
22	4,539	1	0.61	49	4,527	1	0.76
23	4,543	1	0.35	50	4,528	1	0.40
24	4,541	1	0.59	51	4,526	1	0.68
25	4,536	1	0.26	52	4,525	1	0.50
26	4,531	1	0.64	53	4,524	1	0.59
27	4,515	1	0.72	54	4,522	1	0.51

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C2. DC CAS 2013 Operational Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 6							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,588	1	0.43	28	4,556	1	0.40
2	4,590	1	0.80	29	4,568	1	0.48
3	4,572	1	0.42	30	4,550	1	0.51
4	4,579	1	0.28	31	4,508	3	0.37
5	4,584	1	0.67	32	4,573	1	0.53
6	4,585	1	0.86	33	4,567	1	0.44
7	4,586	1	0.54	34	4,570	1	0.75
8	4,583	1	0.74	35	4,566	1	0.64
9	4,586	1	0.81	36	4,570	1	0.65
10	4,585	1	0.47	37	4,570	1	0.81
11	4,573	1	0.17	38	4,565	1	0.76
12	4,578	1	0.45	39	4,554	1	0.29
13	4,583	1	0.72	40	4,555	1	0.58
14	4,552	1	0.27	41	4,561	1	0.35
15	4,567	1	0.51	42	4,563	1	0.76
16	4,574	1	0.43	43	4,562	1	0.35
17	4,566	1	0.43	44	4,554	1	0.52
18	4,572	1	0.48	45	4,557	1	0.72
19	4,563	1	0.60	46	4,541	1	0.50
20	4,492	3	0.25	47	4,481	3	0.21
21	4,577	1	0.83	48	4,566	1	0.74
22	4,576	1	0.65	49	4,566	1	0.68
23	4,571	1	0.55	50	4,565	1	0.71
24	4,573	1	0.85	51	4,558	1	0.56
25	4,576	1	0.49	52	4,566	1	0.25
26	4,572	1	0.44	53	4,561	1	0.54
27	4,572	1	0.49	54	4,564	1	0.45

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C2. DC CAS 2013 Operational Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 7							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,535	1	0.35	28	4,518	1	0.65
2	4,542	1	0.86	29	4,506	1	0.56
3	4,540	1	0.62	30	4,512	1	0.42
4	4,519	1	0.41	31	4,509	1	0.34
5	4,517	1	0.33	32	4,511	1	0.63
6	4,425	3	0.20	33	4,506	1	0.25
7	4,530	1	0.28	34	4,516	1	0.50
8	4,527	1	0.41	35	4,512	1	0.41
9	4,530	1	0.47	36	4,513	1	0.68
10	4,527	1	0.35	37	4,511	1	0.45
11	4,532	1	0.30	38	4,508	1	0.51
12	4,524	1	0.26	39	4,507	1	0.51
13	4,519	1	0.33	40	4,491	1	0.41
14	4,521	1	0.54	41	4,500	1	0.53
15	4,525	1	0.63	42	4,491	1	0.36
16	4,511	1	0.23	43	4,493	1	0.35
17	4,520	1	0.60	44	4,496	1	0.49
18	4,523	1	0.76	45	4,496	1	0.68
19	4,521	1	0.60	46	4,497	1	0.52
20	4,500	1	0.33	47	4,472	1	0.62
21	4,426	3	0.27	48	4,390	3	0.50
22	4,525	1	0.54	49	4,494	1	0.21
23	4,523	1	0.51	50	4,494	1	0.49
24	4,517	1	0.55	51	4,493	1	0.65
25	4,516	1	0.69	52	4,494	1	0.30
26	4,520	1	0.76	53	4,496	1	0.30
27	4,509	1	0.83	54	4,500	1	0.62

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C2. DC CAS 2013 Operational Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 8							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,278	1	0.72	28	4,245	1	0.76
2	4,272	1	0.61	29	4,241	1	0.53
3	4,205	1	0.32	30	4,237	1	0.39
4	4,271	1	0.67	31	4,244	1	0.58
5	4,258	1	0.53	32	4,240	1	0.45
6	4,273	1	0.58	33	4,231	1	0.41
7	4,273	1	0.27	34	4,244	1	0.47
8	4,260	1	0.29	35	4,245	1	0.68
9	4,261	1	0.36	36	4,241	1	0.81
10	4,269	1	0.66	37	4,242	1	0.60
11	4,259	1	0.43	38	4,232	1	0.50
12	4,267	1	0.30	39	4,231	1	0.49
13	4,257	1	0.61	40	4,224	1	0.55
14	4,260	1	0.91	41	4,227	1	0.55
15	4,255	1	0.69	42	4,229	1	0.31
16	4,249	1	0.37	43	4,228	1	0.29
17	4,241	1	0.44	44	4,224	1	0.24
18	4,236	1	0.23	45	4,219	1	0.72
19	4,147	3	0.33	46	4,217	1	0.61
20	4,259	1	0.68	47	4,097	3	0.30
21	4,245	1	0.42	48	4,233	1	0.63
22	4,260	1	0.80	49	4,234	1	0.59
23	4,260	1	0.44	50	4,232	1	0.63
24	4,257	1	0.85	51	4,231	1	0.57
25	4,249	1	0.30	52	4,229	1	0.30
26	4,256	1	0.60	53	4,233	1	0.36
27	4,224	1	0.44				

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C2. DC CAS 2013 Operational Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 10							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,658	1	0.40	28	4,539	1	0.25
2	4,606	1	0.36	29	4,547	1	0.55
3	4,648	1	0.34	30	4,559	1	0.39
4	4,621	1	0.41	31	4,561	1	0.63
5	4,588	1	0.49	32	4,555	1	0.49
6	3,299	3	0.11	33	4,552	1	0.42
7	4,647	1	0.45	34	4,552	1	0.16
8	4,577	1	0.41	35	4,545	1	0.55
9	4,639	1	0.30	36	4,561	1	0.29
10	4,646	1	0.49	37	4,551	1	0.28
11	4,636	1	0.10	38	4,548	1	0.40
12	4,646	1	0.53	39	4,553	1	0.41
13	4,638	1	0.28	40	4,532	1	0.50
14	4,625	1	0.39	41	4,486	1	0.35
15	4,590	1	0.31	42	4,505	1	0.30
16	4,627	1	0.63	43	4,510	1	0.34
17	4,589	1	0.46	44	4,524	1	0.51
18	4,624	1	0.32	45	4,513	1	0.51
19	4,621	1	0.65	46	4,495	1	0.39
20	4,562	1	0.44	47	4,475	1	0.42
21	3,691	3	0.19	48	3,827	3	0.20
22	4,617	1	0.55	49	4,508	1	0.32
23	4,626	1	0.60	50	4,514	1	0.57
24	4,600	1	0.12	51	4,499	1	0.35
25	4,608	1	0.40	52	4,504	1	0.34
26	4,616	1	0.37	53	4,523	1	0.42
27	4,610	1	0.18	54	suppressed*	1	.

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

\*Items deemed statically unacceptable were suppressed.

**Table C3. DC CAS 2013 Operational Form Item Adjusted *P* Values: Science/Biology**

Science Grade 5							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,517	1	0.68	26	4,448	1	0.70
2	4,517	1	0.55	27	4,382	2	0.54
3	4,519	1	0.57	28	4,505	1	0.62
4	4,512	1	0.38	29	4,500	1	0.47
5	4,515	1	0.79	30	4,506	1	0.30
6	4,519	1	0.65	31	4,504	1	0.35
7	4,515	1	0.47	32	4,503	1	0.46
8	4,515	1	0.62	33	4,504	1	0.56
9	4,515	1	0.39	34	4,499	1	0.61
10	4,516	1	0.26	35	4,500	1	0.53
11	4,511	1	0.50	36	4,502	1	0.61
12	4,445	1	0.40	37	4,500	1	0.26
13	4,416	2	0.54	38	4,501	1	0.25
14	4,502	1	0.70	39	4,497	1	0.54
15	4,503	1	0.54	40	4,497	1	0.39
16	4,494	1	0.37	41	4,490	1	0.30
17	4,487	1	0.31	42	4,463	1	0.49
18	4,513	1	0.29	43	4,362	2	0.18
19	4,516	1	0.73	44	4,503	1	0.33
20	4,506	1	0.31	45	4,495	1	0.52
21	4,508	1	0.53	46	4,495	1	0.75
22	4,516	1	0.39	47	4,491	1	0.54
23	4,511	1	0.65	48	4,490	1	0.28
24	4,512	1	0.36	49	4,489	1	0.38
25	4,491	1	0.46	50	4,474	1	0.70

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C3. DC CAS 2013 Operational Form Item Adjusted *P* Values: Science/Biology (continued)**

Science Grade 8							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,099	1	0.63	26	4,077	1	0.59
2	4,102	1	0.54	27	4,075	1	0.26
3	4,104	1	0.64	28	4,059	1	0.29
4	4,099	1	0.52	29	4,060	1	0.35
5	4,100	1	0.45	30	4,075	1	0.45
6	4,091	1	0.26	31	4,073	1	0.63
7	4,098	1	0.51	32	4,072	1	0.44
8	4,099	1	0.49	33	4,070	1	0.41
9	4,100	1	0.49	34	4,061	1	0.47
10	4,100	1	0.45	35	4,066	1	0.32
11	4,084	1	0.43	36	4,063	1	0.39
12	4,049	1	0.38	37	4,056	1	0.41
13	3,826	2	0.33	38	4,032	1	0.28
14	4,089	1	0.34	39	3,807	2	0.17
15	4,091	1	0.38	40	4,073	1	0.42
16	4,088	1	0.47	41	4,073	1	0.50
17	4,081	1	0.18	42	4,069	1	0.36
18	4,080	1	0.12	43	4,066	1	0.40
19	4,081	1	0.49	44	4,073	1	0.28
20	4,074	1	0.26	45	4,068	1	0.32
21	4,075	1	0.45	46	4,072	1	0.39
22	4,072	1	0.48	47	4,068	1	0.35
23	3,826	2	0.39	48	4,067	1	0.40
24	4,080	1	0.59	49	4,068	1	0.55
25	4,077	1	0.45	50	4,060	1	0.34

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C3. DC CAS 2013 Operational Form Item Adjusted *P* Values: Science/Biology (continued)**

High School Biology							
Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Operational Item Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,029	1	0.45	26	4,004	1	0.36
2	4,028	1	0.40	27	3,999	1	0.31
3	4,016	1	0.31	28	3,993	1	0.40
4	4,024	1	0.33	29	3,999	1	0.61
5	4,022	1	0.47	30	4,002	1	0.42
6	4,022	1	0.26	31	3,998	1	0.37
7	4,014	1	0.33	32	3,996	1	0.42
8	4,018	1	0.29	33	3,998	1	0.40
9	4,020	1	0.36	34	3,977	1	0.29
10	4,019	1	0.57	35	3,981	1	0.52
11	4,016	1	0.46	36	3,979	1	0.37
12	4,010	1	0.46	37	3,974	1	0.52
13	3,995	1	0.39	38	3,973	1	0.35
14	4,010	1	0.35	39	3,970	1	0.38
15	4,007	1	0.35	40	3,962	1	0.40
16	4,004	1	0.54	41	3,545	2	0.75
17	3,999	1	0.41	42	3,297	2	0.18
18	3,976	1	0.23	43	3,971	1	0.61
19	2,985	2	0.15	44	3,973	1	0.35
20	4,007	1	0.37	45	3,957	1	0.47
21	4,008	1	0.67	46	3,963	1	0.35
22	4,000	1	0.25	47	3,970	1	0.35
23	4,002	1	0.34	48	3,960	1	0.23
24	3,990	1	0.45	49	3,963	1	0.35
25	4,000	1	0.27	50	3,962	1	0.28

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C4. DC CAS 2013 Operational Form Item Adjusted *P* Values: Composition**

<b>Composition Grade 4</b>			
<b>Operational Item Sequence Number</b>	<b>N</b>	<b>Max Points</b>	<b>Adjusted <i>P</i> Value</b>
49	4,321	6	0.40
50	4,321	4	0.52
51	4,321	4	0.50
<b>Composition Grade 7</b>			
<b>Operational Item Sequence Number</b>	<b>N</b>	<b>Max Points</b>	<b>Adjusted <i>P</i> Value</b>
49	4,325	6	0.49
50	4,325	4	0.65
51	4,325	4	0.64
<b>Composition Grade 10</b>			
<b>Operational Item Sequence Number</b>	<b>N</b>	<b>Max Points</b>	<b>Adjusted <i>P</i> Value</b>
49	4,099	6	0.48
50	4,099	4	0.66
51	4,099	4	0.66

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C5. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Reading**

Reading Grade 2				Reading Grade 3			
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,719	1	0.69	1	4,872	1	0.52
2	4,688	1	0.26	2	4,865	1	0.66
3	4,690	1	0.53	3	4,851	1	0.85
4	4,689	1	0.65	4	4,856	1	0.65
5	4,676	1	0.45	5	4,854	1	0.49
6	4,683	1	0.56	6	4,873	1	0.60
7	4,656	1	0.22	7	4,871	1	0.62
8	4,634	1	0.44	8	4,854	1	0.55
9	4,657	1	0.30	9	4,868	1	0.36
10	4,609	1	0.44	10	4,787	3	0.26

**Note:** The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C5. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 4				Reading Grade 5			
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,458	1	0.42	1	4,521	1	0.53
2	4,459	1	0.48	2	4,518	1	0.42
3	4,459	1	0.38	3	4,516	1	0.70
4	4,451	1	0.33	4	4,511	1	0.70
5	4,441	1	0.45	5	4,477	1	0.77
6	4,491	1	0.54	6	4,363	3	0.47
7	4,489	1	0.29	7	4,511	1	0.56
8	4,484	1	0.51	8	4,509	1	0.49
9	4,478	1	0.67	9	4,508	1	0.52
10	4,430	1	0.45	10	4,507	1	0.37
11	4,288	2	0.15				

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C5. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 6				Reading Grade 7			
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,544	1	0.34	1	4,495	1	0.54
2	4,549	1	0.66	2	4,498	1	0.55
3	4,547	1	0.48	3	4,488	1	0.41
4	4,549	1	0.55	4	4,440	3	0.41
5	4,541	1	0.25	5	4,308	3	0.20
6	4,549	1	0.74	6	4,491	1	0.50
7	4,547	1	0.45	7	4,490	1	0.44
8	4,539	1	0.51	8	4,490	1	0.23
9	4,493	1	0.66	9	4,491	1	0.40
10	4,422	3	0.51	10	4,491	1	0.22

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C5. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 8				Reading Grade 9			
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,209	1	0.56	1	4,723	1	0.73
2	4,213	1	0.54	2	4,724	1	0.78
3	4,204	1	0.43	3	4,721	1	0.72
4	4,206	1	0.40	4	4,718	1	0.70
5	4,173	1	0.53	5	4,638	1	0.43
6	4,103	3	0.61	6	4,636	1	0.45
7	4,192	1	0.63	7	4,635	1	0.66
8	4,191	1	0.48	8	4,636	1	0.51
9	4,187	1	0.51	9	4,637	1	0.71
10	4,191	1	0.67	10	4,634	1	0.74
11	4,188	1	0.53	11	4,631	1	0.71

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C5. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Reading (continued)**

Reading Grade 10			
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,652	1	0.31
2	4,652	1	0.56
3	4,651	1	0.62
4	4,649	1	0.49
5	4,647	1	0.55
6	4,649	1	0.59
7	4,582	1	0.65
8	4,582	1	0.39
9	4,585	1	0.49
10	4,587	1	0.72
11	4,585	1	0.48
12	4,581	1	0.57

**Note:** The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C6. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Mathematics**

Mathematics Grade 2							
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	2,498	3	0.31	17	2,277	3	0.59
2	2,496	1	0.48	18	2,277	1	0.68
3	2,502	1	0.40	19	2,270	1	0.26
4	2,509	1	0.89	20	2,279	1	0.59
5	2,501	1	0.47	21	2,243	1	0.40
6	2,519	1	0.39	22	2,281	1	0.64
7	2,514	1	0.87	23	2,269	1	0.78
8	2,517	1	0.74	24	2,278	1	0.84
9	2,516	1	0.59	25	2,267	1	0.85
10	2,496	1	0.67	26	2,271	1	0.77
11	2,499	1	0.68	27	2,265	1	0.66
12	2,487	3	0.63	28	2,031	3	0.47
13	2,497	1	0.73	29	2,262	1	0.70
14	2,513	1	0.89	30	2,262	1	0.87
15	2,510	1	0.55	31	2,274	1	0.58
16	2,507	1	0.61	32	2,243	1	0.82

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C6. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 3								
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value		Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	2,482	1	0.62		17	2,395	1	0.49
2	2,440	1	0.71		18	2,406	1	0.59
3	2,478	1	0.91		19	2,386	1	0.64
4	2,475	1	0.47		20	2,396	1	0.21
5	2,487	3	0.37		21	2,343	3	0.26
6	2,487	1	0.59		22	2,399	1	0.81
7	2,475	1	0.58		23	2,385	1	0.62
8	2,483	1	0.73		24	2,402	1	0.82
9	2,473	3	0.08		25	2,076	3	0.33
10	2,478	1	0.72		26	2,403	1	0.58
11	2,442	1	0.53		27	2,378	1	0.53
12	2,483	1	0.48		28	2,402	1	0.46
13	2,473	1	0.46		29	2,401	1	0.52
14	2,448	1	0.50		30	2,331	1	0.43
15	2,464	1	0.40		31	2,397	1	0.39
16	2,478	1	0.28		32	2,404	1	0.20

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C6. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 4								
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value		Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	2,313	1	0.36		17	2,256	1	0.38
2	2,313	1	0.46		18	2,253	1	0.70
3	2,304	1	0.55		19	2,240	1	0.69
4	2,293	1	0.44		20	2,247	1	0.77
5	2,241	3	0.39		21	2,215	3	0.20
6	2,302	1	0.43		22	2,247	1	0.59
7	2,302	1	0.81		23	2,237	1	0.35
8	2,290	1	0.55		24	2,236	1	0.84
9	2,254	3	0.17		25	2,222	3	0.48
10	2,300	1	0.69		26	2,242	1	0.69
11	2,298	1	0.76		27	2,240	1	0.69
12	2,293	1	0.61		28	2,231	1	0.41
13	2,311	1	0.47		29	2,248	1	0.66
14	2,305	1	0.71		30	2,236	1	0.37
15	2,296	1	0.39		31	2,242	1	0.27
16	2,297	1	0.25		32	2,227	1	0.25

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C6. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 5								
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value		Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	2,309	1	0.39		17	2,232	1	0.34
2	2,305	1	0.59		18	2,231	1	0.51
3	2,308	1	0.49		19	2,225	1	0.45
4	2,304	1	0.34		20	2,224	1	0.39
5	2,261	3	0.29		21	2,198	3	0.51
6	2,303	1	0.54		22	2,225	1	0.60
7	2,300	1	0.42		23	2,223	1	0.51
8	2,292	1	0.49		24	2,223	1	0.33
9	2,281	3	0.27		25	2,175	3	0.35
10	2,310	1	0.38		26	2,227	1	0.60
11	2,305	1	0.44		27	2,222	1	0.33
12	2,306	1	0.82		28	2,221	1	0.52
13	2,301	1	0.62		29	2,220	1	0.38
14	2,302	1	0.46		30	2,220	1	0.48
15	2,299	1	0.45		31	2,219	1	0.34
16	2,294	1	0.51		32	2,214	1	0.35

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C6. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 6								
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value		Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	2,276	3	0.56		17	2,247	3	0.34
2	2,296	1	0.50		18	2,261	1	0.64
3	2,298	1	0.51		19	2,260	1	0.40
4	2,293	1	0.58		20	2,259	1	0.28
5	2,308	1	0.78		21	2,263	1	0.41
6	2,308	1	0.17		22	2,262	1	0.21
7	2,308	1	0.37		23	2,262	1	0.59
8	2,304	1	0.75		24	2,258	1	0.65
9	2,218	3	0.22		25	2,204	3	0.22
10	2,287	1	0.56		26	2,251	1	0.77
11	2,285	1	0.32		27	2,250	1	0.38
12	2,281	1	0.38		28	2,247	1	0.58
13	2,300	1	0.23		29	2,261	1	0.50
14	2,301	1	0.65		30	2,260	1	0.30
15	2,300	1	0.63		31	2,258	1	0.71
16	2,299	1	0.57		32	2,256	1	0.43

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C6. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 7								
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value		Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	2,301	1	0.53		17	2,218	1	0.18
2	2,304	1	0.60		18	2,218	1	0.52
3	2,300	1	0.35		19	2,218	1	0.54
4	2,300	1	0.18		20	2,215	1	0.43
5	2,252	3	0.10		21	2,150	3	0.27
6	2,291	1	0.60		22	2,219	1	0.39
7	2,293	1	0.29		23	2,214	1	0.21
8	2,289	1	0.49		24	2,216	1	0.69
9	2,204	3	0.03		25	2,159	3	0.23
10	2,293	1	0.27		26	2,212	1	0.19
11	2,292	1	0.23		27	2,212	1	0.32
12	2,293	1	0.30		28	2,210	1	0.67
13	2,289	1	0.40		29	2,206	1	0.65
14	2,292	1	0.40		30	2,205	1	0.31
15	2,288	1	0.38		31	2,205	1	0.33
16	2,286	1	0.36		32	2,207	1	0.43

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C6. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 8								
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value		Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	4,114	3	0.35		18	2,003	3	0.36
2	2,119	3	0.26		19	2,076	1	0.52
3	2,175	1	0.17		20	2,077	1	0.34
4	2,181	1	0.47		21	2,074	1	0.30
5	2,181	1	0.34		22	2,080	1	0.54
6	2,175	1	0.39		23	2,082	1	0.53
7	2,178	1	0.40		24	2,082	1	0.51
8	2,177	1	0.65		25	2,081	1	0.49
9	2,175	1	0.78		26	1,966	3	0.18
10	2,067	3	0.09		27	2,065	1	0.52
11	2,173	1	0.52		28	2,068	1	0.60
12	2,173	1	0.49		29	2,068	1	0.72
13	2,170	1	0.45		30	2,059	1	0.48
14	2,172	1	0.32		31	2,058	1	0.76
15	2,169	1	0.17		32	2,058	1	0.31
16	2,175	1	0.47		33	2,056	1	0.44
17	2,171	1	0.22					

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C6. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Mathematics (continued)**

Mathematics Grade 10								
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value		Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	1,781	3	0.04		17	1,920	3	0.43
2	2,323	1	0.21		18	2,286	1	0.27
3	2,321	1	0.34		19	2,293	1	0.44
4	2,315	1	0.23		20	2,285	1	0.12
5	2,327	1	0.37		21	2,284	1	0.21
6	2,328	1	0.48		22	2,280	1	0.34
7	2,328	1	0.30		23	2,283	1	0.50
8	2,329	1	0.24		24	2,285	1	0.19
9	1,926	3	0.34		25	1,689	3	0.19
10	2,286	1	0.33		26	2,233	1	0.52
11	2,282	1	0.09		27	2,231	1	0.32
12	2,285	1	0.37		28	2,225	1	0.26
13	2,279	1	0.57		29	2,229	1	0.40
14	2,282	1	0.18		30	2,224	1	0.30
15	2,285	1	0.45		31	2,230	1	0.33
16	2,285	1	0.19		32	2,233	1	0.26

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C7. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Science/Biology**

Science Grade 5							
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value	Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	2,278	1	0.47	15	2,236	1	0.92
2	2,282	1	0.65	16	2,235	1	0.43
3	2,281	1	0.15	17	2,234	1	0.56
4	2,198	2	0.37	18	2,170	2	0.36
5	2,272	1	0.45	19	2,231	1	0.50
6	2,273	1	0.46	20	2,231	1	0.16
7	2,277	1	0.38	21	2,232	1	0.27
8	2,277	1	0.45	22	2,236	1	0.46
9	2,272	1	0.34	23	2,234	1	0.40
10	2,172	2	0.19	24	2,134	2	0.12
11	2,277	1	0.31	25	2,231	1	0.30
12	2,275	1	0.24	26	2,231	1	0.45
13	2,271	1	0.33	27	2,222	1	0.23
14	2,272	1	0.56	28	2,222	1	0.57

*Note:* The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C7. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Science/Biology (continued)**

Science Grade 8								
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value		Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	2,087	1	0.16		15	2,016	1	0.49
2	2,083	1	0.26		16	2,012	1	0.51
3	4,099	1	0.43		17	1,827	2	0.23
4	1,973	2	0.10		18	2,009	1	0.50
5	4,091	1	0.54		19	2,007	1	0.46
6	2,084	1	0.53		20	2,006	1	0.46
7	2,071	1	0.33		21	2,000	1	0.50
8	2,071	1	0.59		22	1,993	1	0.21
9	2,065	1	0.62		23	1,990	1	0.21
10	2,062	1	0.42		24	1,878	2	0.26
11	2,052	1	0.33		25	2,001	1	0.27
12	1,904	2	0.48		26	2,000	1	0.67
13	2,070	1	0.26					
14	2,070	1	0.47					

**Note:** The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

**Table C7. DC CAS 2013 Field Test Form Item Adjusted *P* Values: Science/Biology (continued)**

High School Biology								
Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value		Field Test Sequence Number	N	Max Points	Adjusted <i>P</i> Value
1	2,036	1	0.41		15	1,989	1	0.61
2	2,031	1	0.37		16	1,982	1	0.37
3	2,034	1	0.19		17	1,829	2	0.12
4	1,667	2	0.16		18	1,985	1	0.57
5	2,027	1	0.15		19	1,986	1	0.23
6	2,029	1	0.48		20	1,980	1	0.20
7	2,019	1	0.18		21	1,983	1	0.46
8	4,004	1	0.25		22	1,982	1	0.37
9	2,023	1	0.59		23	1,974	1	0.52
10	2,008	1	0.29		24	1,969	1	0.30
11	2,004	1	0.40		25	1,465	2	0.06
12	1,647	2	0.10		26	1,955	1	0.27
13	2,008	1	0.26		27	1,960	1	0.44
14	2,011	1	0.29					

**Note:** The adjusted *p* value for an item only includes responses for examinees with valid responses to that item.

## Appendix D: Internal Consistency Reliability Coefficients for Examinee Subgroups

(See Section 8. Evidence for Reliability and Validity, *Internal Consistency Reliability*, Table 31)

**Table D1. Scale Score Mean and SD, and Internal Consistency Reliability Coefficients for Examinee Subgroups: Reading**

Grade	Subgroup	Students with Test Scores	Number of Items	Alpha	Stratified Alpha	Feldt-Raju	Scale Score	
							Mean	SD
<b>Reading</b>								
2	All Examinees	4,692	35	0.88	0.89	0.89	243.23	14.84
	Male	2,363	35	0.88	0.89	0.89	241.39	15.12
	Female	2,316	35	0.88	0.89	0.89	245.12	14.32
	Asian	66	35	0.85	0.86	0.86	252.20	10.68
	African American	3,278	35	0.86	0.87	0.87	240.47	13.65
	Hispanic	675	35	0.87	0.88	0.88	242.06	13.92
	White	568	35	0.84	0.85	0.85	258.18	12.93
3	All Examinees	4,851	48	0.93	0.94	0.94	349.96	16.07
	Male	2,467	48	0.93	0.94	0.94	348.02	16.52
	Female	2,371	48	0.93	0.93	0.93	352.00	15.27
	Asian	78	48	0.91	0.92	0.92	359.86	12.87
	African American	3,445	48	0.92	0.92	0.92	346.83	15.09
	Hispanic	681	48	0.92	0.92	0.93	349.45	14.28
	White	546	48	0.88	0.89	0.89	367.02	12.31
4	All Examinees	4,467	48	0.92	0.92	0.92	453.84	14.47
	Male	2,282	48	0.92	0.92	0.92	452.42	14.86
	Female	2,177	48	0.91	0.92	0.92	455.34	13.90
	Asian	69	48	0.92	0.92	0.92	460.61	12.28
	African American	3,227	48	0.91	0.91	0.91	451.37	13.59
	Hispanic	640	48	0.90	0.91	0.91	453.72	13.79
	White	444	48	0.86	0.87	0.86	468.93	11.23
5	All Examinees	4,532	48	0.91	0.92	0.92	554.99	13.32
	Male	2,290	48	0.92	0.92	0.92	552.86	14.44
	Female	2,228	48	0.90	0.90	0.90	557.19	11.69
	Asian	83	48	0.92	0.92	0.92	561.28	11.97
	African American	3,331	48	0.90	0.91	0.91	552.71	12.85
	Hispanic	614	48	0.89	0.90	0.90	556.23	11.83
	White	430	48	0.85	0.86	0.86	568.51	10.04

**Table D1. Scale Score Mean and SD, and Internal Consistency Reliability Coefficients for Examinee Subgroups: Reading (continued)**

Grade	Subgroup	Students with Test Scores	Number of Items	Alpha	Stratified Alpha	Feldt-Raju	Scale Score	
							Mean	SD
<b>Reading</b>								
6	All Examinees	4,560	48	0.92	0.92	0.92	650.64	14.35
	Male	2,253	48	0.92	0.93	0.93	648.17	14.83
	Female	2,291	48	0.91	0.91	0.91	653.10	13.40
	Asian	63	48	0.90	0.90	0.90	659.56	11.99
	African American	3,542	48	0.91	0.92	0.92	648.83	13.65
	Hispanic	556	48	0.91	0.91	0.91	650.84	13.47
	White	313	48	0.88	0.89	0.89	667.25	11.84
7	All Examinees	4,520	48	0.91	0.91	0.91	755.82	13.75
	Male	2,294	48	0.91	0.91	0.91	753.65	14.27
	Female	2,210	48	0.90	0.91	0.91	758.07	12.85
	Asian	60	48	0.90	0.91	0.91	766.77	11.36
	African American	3,528	48	0.90	0.90	0.90	754.17	13.35
	Hispanic	580	48	0.90	0.90	0.90	757.09	12.94
	White	273	48	0.90	0.91	0.91	769.55	11.69
8	All Examinees	4,236	48	0.91	0.92	0.91	855.51	14.03
	Male	2,086	48	0.91	0.92	0.92	852.98	14.68
	Female	2,123	48	0.90	0.91	0.90	858.02	12.90
	Asian	50	48	0.89	0.89	0.89	863.02	13.72
	African American	3,338	48	0.90	0.91	0.91	854.03	13.50
	Hispanic	531	48	0.91	0.91	0.91	856.96	13.55
	White	243	48	0.91	0.92	0.92	869.63	13.79
9	All Examinees	4,779	48	0.92	0.93	0.93	948.36	16.86
	Male	2,379	48	0.92	0.93	0.93	945.71	17.38
	Female	2,343	48	0.92	0.92	0.92	951.41	15.60
	Asian	62	48	0.92	0.93	0.93	960.13	15.75
	African American	3,816	48	0.91	0.91	0.92	947.27	15.88
	Hispanic	558	48	0.92	0.93	0.93	947.82	17.45
	White	234	48	0.93	0.94	0.93	966.54	15.31
10	All Examinees	4,684	48	0.92	0.93	0.93	952.47	15.89
	Male	2,208	48	0.92	0.93	0.93	950.18	15.81
	Female	2,370	48	0.92	0.93	0.93	955.07	15.35
	Asian	69	48	0.90	0.90	0.90	961.77	12.62
	African American	3,687	48	0.92	0.92	0.92	951.06	15.18
	Hispanic	524	48	0.92	0.93	0.93	954.96	14.84
	White	231	48	0.92	0.93	0.92	970.88	13.34

**Table D2. Scale Score Mean and SD, and Internal Consistency Reliability Coefficients for Examinee Subgroups: Mathematics**

Grade	Subgroup	Students with Test Scores	Number of Items	Alpha	Stratified Alpha	Feldt-Raju	Scale Score	
							Mean	SD
<b>Mathematics</b>								
2	All Examinees	4,754	43	0.91	0.91	0.92	253.45	13.84
	Male	2,393	43	0.92	0.92	0.92	253.04	13.80
	Female	2,346	43	0.91	0.91	0.91	253.90	13.89
	Asian	70	43	0.86	0.87	0.88	265.19	14.92
	African American	3,308	43	0.90	0.90	0.91	250.58	12.59
	Hispanic	698	43	0.90	0.91	0.91	253.56	12.73
	White	575	43	0.85	0.86	0.86	266.93	12.48
3	All Examinees	4,876	54	0.93	0.93	0.93	354.75	18.27
	Male	2,478	54	0.93	0.94	0.94	354.31	18.96
	Female	2,384	54	0.92	0.93	0.93	355.27	17.46
	Asian	83	54	0.88	0.89	0.88	372.47	12.66
	African American	3,443	54	0.91	0.92	0.92	350.51	16.91
	Hispanic	696	54	0.91	0.92	0.92	356.74	15.95
	White	550	54	0.89	0.90	0.90	374.28	13.96
4	All Examinees	4,551	54	0.93	0.94	0.94	458.34	16.95
	Male	2,325	54	0.93	0.94	0.94	457.82	17.34
	Female	2,217	54	0.93	0.93	0.94	458.90	16.52
	Asian	75	54	0.90	0.90	0.91	471.45	11.39
	African American	3,270	54	0.92	0.93	0.93	455.20	16.32
	Hispanic	667	54	0.92	0.93	0.93	459.86	14.94
	White	449	54	0.91	0.91	0.91	474.56	13.31
5	All Examinees	4,554	54	0.93	0.94	0.94	558.61	16.95
	Male	2,301	54	0.93	0.94	0.94	557.27	17.91
	Female	2,237	54	0.93	0.93	0.93	560.10	15.65
	Asian	83	54	0.93	0.94	0.94	572.25	13.72
	African American	3,327	54	0.92	0.92	0.92	555.42	16.25
	Hispanic	632	54	0.92	0.92	0.92	561.08	14.97
	White	436	54	0.91	0.91	0.91	575.51	11.89

**Table D2. Scale Score Mean and SD, and Internal Consistency Reliability Coefficients for Examinee Subgroups: Mathematics (continued)**

Grade	Subgroup	Students with Test Scores	Number of Items	Alpha	Stratified Alpha	Feldt-Raju	Scale Score	
							Mean	SD
<b>Mathematics</b>								
6	All Examinees	4,591	54	0.93	0.93	0.93	652.35	18.02
	Male	2,272	54	0.93	0.93	0.93	650.64	18.51
	Female	2,303	54	0.92	0.93	0.93	654.09	17.34
	Asian	65	54	0.91	0.91	0.92	667.55	14.01
	African American	3,551	54	0.92	0.92	0.92	649.78	17.44
	Hispanic	574	54	0.91	0.92	0.92	654.90	15.76
	White	313	54	0.91	0.92	0.92	671.63	13.61
7	All Examinees	4,545	54	0.92	0.92	0.92	752.90	18.67
	Male	2,308	54	0.92	0.92	0.92	751.67	19.28
	Female	2,221	54	0.92	0.92	0.92	754.19	17.94
	Asian	63	54	0.93	0.93	0.93	771.48	13.86
	African American	3,527	54	0.90	0.90	0.90	750.58	17.94
	Hispanic	596	54	0.91	0.91	0.91	754.38	17.71
	White	278	54	0.93	0.93	0.94	771.95	16.32
8	All Examinees	4,280	53	0.91	0.91	0.91	853.73	16.51
	Male	2,107	53	0.91	0.91	0.91	852.39	17.28
	Female	2,143	53	0.90	0.91	0.91	855.13	15.58
	Asian	52	53	0.91	0.92	0.92	867.92	13.44
	African American	3,337	53	0.89	0.90	0.90	852.30	15.99
	Hispanic	569	53	0.90	0.91	0.91	853.72	16.22
	White	244	53	0.92	0.93	0.93	868.63	14.96
10	All Examinees	4,669	53	0.90	0.91	0.91	945.09	21.78
	Male	2,207	53	0.90	0.91	0.91	943.91	22.05
	Female	2,359	53	0.90	0.91	0.91	946.55	21.41
	Asian	69	53	0.92	0.93	0.93	964.77	16.45
	African American	3,672	53	0.88	0.88	0.88	943.16	21.07
	Hispanic	525	53	0.90	0.90	0.90	947.11	21.40
	White	233	53	0.93	0.94	0.94	966.90	18.89

**Table D3. Scale Score Mean and SD, and Internal Consistency Reliability Coefficients for Examinee Subgroups: Science/Biology**

Grade	Subgroup	Students with Test Scores	Number of Items	Alpha	Stratified Alpha	Feldt-Raju	Scale Score	
							Mean	SD
<b>Science/Biology</b>								
5	All Examinees	4,520	50	0.89	0.89	0.89	549.26	11.91
	Male	2,294	50	0.90	0.90	0.90	548.63	12.48
	Female	2,213	50	0.89	0.89	0.89	549.94	11.21
	Asian	81	50	0.91	0.91	0.92	555.85	10.45
	African American	3,303	50	0.85	0.85	0.85	547.00	11.48
	Hispanic	626	50	0.85	0.85	0.85	550.10	9.94
	White	432	50	0.85	0.85	0.85	563.04	7.08
8	All Examinees	4,106	50	0.88	0.88	0.89	850.05	17.53
	Male	2,015	50	0.89	0.89	0.89	849.55	18.12
	Female	2,061	50	0.87	0.88	0.88	850.67	16.86
	Asian	50	50	0.90	0.91	0.91	861.14	8.14
	African American	3,185	50	0.85	0.85	0.85	848.83	17.28
	Hispanic	551	50	0.87	0.88	0.88	849.21	18.29
	White	238	50	0.93	0.93	0.93	864.47	13.14
High School	All Examinees	4,031	50	0.85	0.85	0.85	946.47	16.38
	Male	1,923	50	0.85	0.86	0.86	945.47	17.09
	Female	2,033	50	0.84	0.84	0.84	947.70	15.41
	Asian	54	50	0.86	0.87	0.87	957.20	7.39
	African American	3,127	50	0.81	0.81	0.81	945.44	16.20
	Hispanic	508	50	0.82	0.82	0.82	946.89	15.79
	White	214	50	0.91	0.91	0.92	958.93	14.03

**Table D4. Scale Score Mean and SD, and Internal Consistency Reliability Coefficients for Examinee Subgroups: Composition**

Grade	Subgroup	Students with Test Scores	Number of Items	Alpha	Stratified Alpha*	Feldt-Raju	Scale Score	
							Mean	SD
<b>Composition</b>								
4	All Examinees	4,400	3	0.91	0.87	0.93	453.44	17.18
	Male	2,245	3	0.91	0.88	0.93	450.73	17.38
	Female	2,148	3	0.91	0.86	0.92	456.28	16.51
	Asian	69	3	0.88	0.81	0.92	461.58	15.26
	African American	3,177	3	0.91	0.86	0.92	451.10	16.64
	Hispanic	635	3	0.90	0.86	0.92	454.91	16.30
	White	438	3	0.91	0.88	0.93	465.39	15.77
7	All Examinees	4,387	3	0.94	0.92	0.95	756.81	15.51
	Male	2,227	3	0.94	0.92	0.95	753.63	14.97
	Female	2,150	3	0.93	0.90	0.94	760.16	15.33
	Asian	60	3	0.91	0.90	0.94	770.30	14.76
	African American	3,421	3	0.94	0.91	0.95	755.00	15.01
	Hispanic	572	3	0.93	0.91	0.94	758.58	13.93
	White	265	3	0.92	0.92	0.94	771.24	15.53
10	All Examinees	4,217	3	0.93	0.88	0.94	954.31	18.60
	Male	1,972	3	0.93	0.88	0.94	950.50	18.09
	Female	2,149	3	0.92	0.87	0.93	958.24	18.13
	Asian	70	3	0.89	0.81	0.91	966.41	16.89
	African American	3,251	3	0.93	0.87	0.93	952.58	17.78
	Hispanic	499	3	0.92	0.87	0.93	957.75	18.01
	White	226	3	0.92	0.91	0.94	971.27	18.69

*Note:* \*Stratified Alpha is based on two items.

## Appendix E: Classification Consistency and Accuracy Estimates for All Proficiency Levels for Examinee Subgroups

**Table E1. Classification Consistency and Accuracy Rates for All Cut Scores and Examinee Subgroups: Reading**

Grade/Subgroup	Classification Consistency		Classification Accuracy		
	Consistency	Kappa	Accuracy	False Positive Errors	False Negative Errors
<b>Grade 2</b>					
Males	0.72	0.60	0.80	0.10	0.11
Females	0.72	0.59	0.80	0.10	0.11
Asian	0.73	0.55	0.80	0.09	0.11
African American	0.72	0.58	0.79	0.10	0.11
Hispanic	0.72	0.59	0.80	0.09	0.10
White	0.74	0.55	0.80	0.09	0.10
<b>Grade 3</b>					
Males	0.78	0.68	0.85	0.07	0.08
Females	0.77	0.66	0.84	0.08	0.08
Asian	0.77	0.60	0.85	0.06	0.09
African American	0.78	0.67	0.85	0.08	0.08
Hispanic	0.77	0.65	0.84	0.08	0.08
White	0.76	0.57	0.83	0.08	0.09
<b>Grade 4</b>					
Males	0.75	0.64	0.83	0.08	0.10
Females	0.75	0.63	0.83	0.08	0.10
Asian	0.75	0.60	0.82	0.08	0.10
African American	0.75	0.63	0.83	0.08	0.09
Hispanic	0.76	0.64	0.83	0.07	0.10
White	0.74	0.53	0.81	0.09	0.10
<b>Grade 5</b>					
Males	0.77	0.66	0.84	0.07	0.09
Females	0.77	0.63	0.83	0.08	0.09
Asian	0.79	0.66	0.85	0.06	0.09
African American	0.77	0.64	0.83	0.07	0.09
Hispanic	0.78	0.65	0.85	0.08	0.08
White	0.75	0.52	0.82	0.09	0.09
<b>Grade 6</b>					
Males	0.77	0.67	0.84	0.07	0.10
Females	0.76	0.64	0.83	0.07	0.10
Asian	0.74	0.60	0.82	0.08	0.10
African American	0.77	0.65	0.83	0.07	0.10
Hispanic	0.76	0.64	0.83	0.08	0.09
White	0.75	0.56	0.83	0.08	0.09

**Table E1. Classification Consistency and Accuracy Rates for All Cut Scores and Examinee Subgroups: Reading (continued)**

Grade/Subgroup	Classification Consistency		Classification Accuracy		
	Consistency	Kappa	Accuracy	False Positive Errors	False Negative Errors
<b>Grade 7</b>					
Males	0.74	0.62	0.81	0.09	0.10
Females	0.74	0.63	0.82	0.09	0.09
Asian	0.80	0.66	0.86	0.07	0.08
African American	0.74	0.61	0.81	0.09	0.09
Hispanic	0.73	0.61	0.81	0.10	0.10
White	0.82	0.63	0.87	0.08	0.05
<b>Grade 8</b>					
Males	0.74	0.63	0.82	0.09	0.09
Females	0.74	0.61	0.82	0.09	0.09
Asian	0.73	0.59	0.80	0.13	0.07
African American	0.74	0.62	0.82	0.09	0.09
Hispanic	0.74	0.63	0.82	0.09	0.09
White	0.78	0.62	0.84	0.09	0.07
<b>Grade 9</b>					
Males	0.73	0.63	0.81	0.10	0.10
Females	0.74	0.63	0.81	0.10	0.09
Asian	0.80	0.68	0.85	0.09	0.06
African American	0.72	0.62	0.80	0.10	0.10
Hispanic	0.72	0.62	0.80	0.11	0.09
White	0.88	0.71	0.91	0.05	0.04
<b>Grade 10</b>					
Males	0.74	0.64	0.82	0.09	0.10
Females	0.74	0.63	0.81	0.09	0.09
Asian	0.73	0.60	0.81	0.07	0.12
African American	0.74	0.63	0.82	0.09	0.10
Hispanic	0.73	0.62	0.80	0.10	0.09
White	0.78	0.61	0.84	0.09	0.08

**Table E2. Classification Consistency and Accuracy Rates for All Cut Scores and Examinee Subgroups: Mathematics**

Grade/Subgroup	Classification Consistency		Classification Accuracy		
	Consistency	Kappa	Accuracy	False Positive Errors	False Negative Errors
<b>Grade 2</b>					
Males	0.74	0.64	0.81	0.09	0.10
Females	0.74	0.64	0.81	0.08	0.11
Asian	0.75	0.62	0.81	0.10	0.09
African American	0.74	0.63	0.81	0.08	0.11
Hispanic	0.73	0.62	0.80	0.09	0.11
White	0.75	0.59	0.82	0.10	0.09
<b>Grade 3</b>					
Males	0.76	0.67	0.84	0.08	0.09
Females	0.76	0.66	0.84	0.08	0.08
Asian	0.76	0.62	0.84	0.07	0.10
African American	0.76	0.65	0.84	0.08	0.09
Hispanic	0.74	0.63	0.82	0.08	0.10
White	0.78	0.63	0.85	0.07	0.07
<b>Grade 4</b>					
Males	0.77	0.67	0.83	0.08	0.09
Females	0.76	0.66	0.83	0.08	0.09
Asian	0.81	0.68	0.86	0.07	0.07
African American	0.75	0.65	0.83	0.08	0.09
Hispanic	0.77	0.66	0.83	0.08	0.08
White	0.83	0.67	0.87	0.07	0.06
<b>Grade 5</b>					
Males	0.78	0.70	0.85	0.08	0.07
Females	0.78	0.68	0.85	0.08	0.07
Asian	0.83	0.73	0.89	0.07	0.04
African American	0.77	0.67	0.84	0.08	0.08
Hispanic	0.78	0.68	0.85	0.08	0.07
White	0.83	0.69	0.88	0.07	0.05
<b>Grade 6</b>					
Males	0.74	0.65	0.82	0.09	0.09
Females	0.75	0.65	0.82	0.09	0.09
Asian	0.81	0.68	0.86	0.06	0.08
African American	0.74	0.63	0.81	0.09	0.10
Hispanic	0.74	0.63	0.81	0.09	0.10
White	0.85	0.70	0.89	0.07	0.04
<b>Grade 7</b>					
Males	0.71	0.60	0.80	0.10	0.10
Females	0.72	0.60	0.80	0.10	0.10
Asian	0.83	0.69	0.88	0.06	0.07
African American	0.70	0.57	0.79	0.11	0.11
Hispanic	0.75	0.66	0.79	0.14	0.08
White	0.86	0.71	0.90	0.05	0.05

**Table E2. Classification Consistency and Accuracy Rates for All Cut Scores and Examinee Subgroups: Mathematics (continued)**

Grade/Subgroup	Classification Consistency		Classification Accuracy		
	Consistency	Kappa	Accuracy	False Positive Errors	False Negative Errors
<b>Grade 8</b>					
Males	0.71	0.58	0.79	0.10	0.10
Females	0.73	0.59	0.80	0.10	0.09
Asian	0.82	0.69	0.88	0.06	0.06
African American	0.71	0.56	0.79	0.11	0.10
Hispanic	0.72	0.58	0.80	0.10	0.10
White	0.83	0.69	0.88	0.07	0.05
<b>Grade 10</b>					
Males	0.64	0.50	0.74	0.13	0.13
Females	0.66	0.52	0.75	0.12	0.13
Asian	0.81	0.70	0.87	0.06	0.07
African American	0.64	0.48	0.73	0.13	0.14
Hispanic	0.66	0.52	0.75	0.13	0.12
White	0.79	0.66	0.85	0.08	0.07

**Table E3. Classification Consistency and Accuracy Rates for All Cut Scores and Examinee Subgroups: Science/Biology**

Grade/Subgroup	Classification Consistency		Classification Accuracy		
	Consistency	Kappa	Accuracy	False Positive Errors	False Negative Errors
<b>Grade 5</b>					
Males	0.71	0.58	0.79	0.11	0.10
Females	0.70	0.57	0.79	0.11	0.10
Asian	0.74	0.62	0.81	0.11	0.08
African American	0.70	0.54	0.78	0.11	0.11
Hispanic	0.71	0.56	0.80	0.11	0.09
White	0.78	0.61	0.85	0.08	0.08
<b>Grade 8</b>					
Males	0.68	0.54	0.76	0.12	0.13
Females	0.68	0.53	0.76	0.12	0.13
Asian	0.70	0.53	0.77	0.09	0.14
African American	0.67	0.51	0.75	0.12	0.13
Hispanic	0.67	0.53	0.75	0.12	0.13
White	0.81	0.68	0.86	0.06	0.08
<b>High School</b>					
Males	0.68	0.52	0.75	0.12	0.13
Females	0.68	0.51	0.75	0.12	0.13
Asian	0.75	0.53	0.82	0.08	0.10
African American	0.67	0.50	0.74	0.13	0.14
Hispanic	0.69	0.53	0.76	0.11	0.13
White	0.78	0.63	0.85	0.08	0.07

**Table E4. Classification Consistency and Accuracy Rates for All Cut Scores and Examinee Subgroups: Composition**

Grade/Subgroup	Classification Consistency		Classification Accuracy		
	Consistency	Kappa	Accuracy	False Positive Errors	False Negative Errors
<b>Grade 4</b>					
Males	0.60	0.46	0.70	0.12	0.18
Females	0.58	0.44	0.68	0.14	0.18
Asian	0.60	0.44	0.69	0.14	0.17
African American	0.59	0.44	0.68	0.13	0.19
Hispanic	0.58	0.44	0.68	0.14	0.18
White	0.62	0.43	0.72	0.16	0.13
<b>Grade 7</b>					
Males	0.63	0.49	0.73	0.14	0.14
Females	0.65	0.50	0.75	0.12	0.14
Asian	0.73	0.56	0.81	0.08	0.11
African American	0.63	0.49	0.73	0.13	0.14
Hispanic	0.62	0.46	0.72	0.13	0.15
White	0.73	0.54	0.81	0.07	0.12
<b>Grade 10</b>					
Males	0.60	0.45	0.71	0.16	0.13
Females	0.60	0.45	0.71	0.16	0.13
Asian	0.65	0.47	0.74	0.14	0.12
African American	0.59	0.44	0.70	0.16	0.14
Hispanic	0.61	0.46	0.71	0.15	0.13
White	0.73	0.51	0.80	0.12	0.07

