Pennsylvania Grade 3 Assessment

Mathematics and Reading

Technical Report

Spring 2004 Operational Test

Submitted to the Pennsylvania Department of Education

September 2004

Table of Contents

Overview 5
Test Design and Sample 6
  Test Structure 6
  Test Sample 7
Descriptive Statistics and Item Analysis 8
  Item-Level Descriptive Statistics 8
  Speededness 9
  Rater Agreement 9
  Differential Item Functioning (DIF) 10
  Item Fit Assessment 11
IRT Calibration and Equating 13
Establishment of the Pennsylvania Grade 3 Score Scale 15
  Test Form Statistics 15
Summary statistics for Content Standards 17
Appendix 65

Tables

Table 1 Test Design 19
Table 2 Number of Items and Score Points by Item Type 20
Table 3 2004 Pennsylvania Grade 3 Sample Characteristics by Ethnicity 21
Table 4 2004 Pennsylvania Grade 3 Sample Characteristics by Gender 22
Table 5 Item Statistics for Reading Form A (N=28,012) * 23
Table 6 Item Statistics for Reading Form B (N=27,689) * 24
Table 7 Item Statistics for Reading Form C (N=27,387) * 25
Table 8 Item Statistics for Reading Form D (N=27,189) * 26
Table 9 Item Statistics for Mathematics Common items (N=110,657) 27
Table 10 Item Statistics for Mathematics Matrix items 28
Table 11 Summary for Differential Item Functioning based on Criteria ± C 30
Table 12 Rater Agreement for Reading Constructed-Response Items 32
Table 13 Rater Agreement for Mathematics Constructed-Response Items 33
Table 14 Scoring Table for Reading Form A 35
Table 15 Scoring Table for Reading Form B 36
Table 16 Scoring Table for Reading Form C 37
Table 17 Scoring Table for Reading Form D 38
Table 18 Scoring Table for Mathematics 39
Table 19 Raw score Descriptive Statistics Based on All Samples 40
Table 20 Raw Score Descriptive Statistics by Ethnicity 41
Table 21 Raw Score Descriptive Statistics by Gender 42
Table 22 Descriptive Statistics for Reported Scale Scores Based on All Samples 43
Table 23 Descriptive Statistics for Reported Scale Scores by Ethnicity 44
Table 24 Descriptive Statistics for Reported Scale Scores by Gender 45
Table 25 Percentiles of Scale Score Ranges 46
Table 26 The Number of Items per Each Reading Standard * 49
Table 27 The Number of Items per each Mathematics Standard * 50
Table 28 Summary Statistics for Reading Standards * 51
Table 29 Summary Statistics for Mathematics Standards * 52
Table 30 Factor Analysis Results for Reading 3 Standards 53
Table 31 Factor Analysis Results for Mathematics 11 Standards 54

Figures

Figure 1 Reading Scale Score and SEM across Four Forms 55
Figure 2 Mathematics Scale Scores and SEM 56
Figure 3 Reading Form A Raw Score and Scale Score Distribution 57
Figure 4 Reading Form B Raw Score and Scale Score Distribution 58
Figure 5 Reading Form C Raw Score and Scale Score Distribution 59
Figure 6 Reading Form D Raw Score and Scale Score Distribution 60
Figure 7 Mathematics Raw Score and Scale Score Distribution 61
Figure 8 Scree Plot for Reading and Mathematics 62
Figure 9 State Mean Scale Score 63

Overview

The purpose of this report is to provide technical information about the 2004 spring operational test administration of the Pennsylvania Grade 3 Reading and Mathematics Assessments. Testing in Reading began in 2002, and testing in Mathematics was included in 2003. The score scales for Reading and Mathematics were set up during the 2003 operational testing. The 2004 score scales were transformed to the 2003 scale.

This report includes an overview of the operational test design, summary of the operational test items, and test form analyses. The report also provides a summary of raw score descriptive statistics at the item and test form levels and a discussion of the procedures used for form calibration and equating. Also included is a summary of the reported test scores, as individual test scores were reported based on the scale scores. Note that several Year-to-Year equating procedures, including the procedure used for 2004 operational testing, were analyzed in the special study that was delivered in August of 2004. Therefore, this report does not include the information for Year-to-Year equating. Note once more that there has been no standard setting for Pennsylvania Grade 3 Reading and Mathematics so that there are no cut score points.

Test Design and Sample

Test Structure

The Pennsylvania Grade 3 Reading and Mathematics Assessments are part of the Pennsylvania School of System Assessment (PSSA) currently being administered in Pennsylvania. This criterion-referenced assessment is intended to measure the three content standards for Reading and the eleven content standards for Mathematics (see Appendix and Tables 24 & 25). Reading and Mathematics were administered in one test booklet. Twelve books were spiraled within classroom. Reading has 4 unique forms repeated three times, whereas each of the 12 Mathematics forms is unique (See Table 1 for the test design). Mathematics consists of common items, which were taken by every student and were common across all forms of the test, and matrix items that are unique to each test form. In addition to common and matrix items, Reading contains embedded field-tested items. In Mathematics, matrix items do not contribute to students' reported scores, but do contribute to aggregated scores for curriculum analysis. In Reading, common and matrix items do contribute to students' scores. Table 2 shows the number of items and score points by item type, multiple choice (MC) items and open ended (OE) items. Note that in Mathematics, 3 OE items are scored using a 5 score point rubric and 9 OE items are scored using a 4 score point rubric. The 3 OE items based on the 4 score point rubric were used for Year-to-Year equating, and the other 9 items will be used for 2005 test construction. From 2005, all OE items will be scored using a 4 score point rubric. For the reported individual scores, the total number-correct score was 50 (sum of the scores of common and matrix items) for Reading and 70 (score of common items) for Mathematics.

Test Sample

In May 2004, approximately 130,000 Pennsylvania Grade 3 students took one of the twelve books that contain both Reading and Mathematics. Responses of all students (i.e., the population) were used for item calibration and classical item analyses.

Table 3 shows the ethnic characteristics for the total number of students by form. The percentages are rounded to whole numbers. As the table indicates, most of the students were White (76%), approximately 15% of the students were African American, and approximately 5% of the students were Hispanic. As expected, these ratios were similar across all test forms because the test forms were spiraled within the classroom. As shown in Table 4, slightly more male students (around 51%) than female students (49%) were in the tested population.

Descriptive Statistics and Item Analysis

Item-Level Descriptive Statistics

Tables 5–10 present item-level descriptive statistics for each of the operational test forms in test book order. For Reading, common items across forms were treated as unique items within each form, whereas for Mathematics one set of statistics was provided for common items instead of providing different 12 statistics for each common item. The statistics for Mathematics common items were computed based on the entire tested population. Statistics for common items were similar across alternative forms. Note that there were four alternative Reading forms and twelve alternative Mathematics forms. These tables contain the following information: item type, item *p*-value, item correlation with the total test score (R-ITT), the percent of examinees that omitted an item, and fit information. The *p*-value for an MC item represents the proportion of students who answered the item correctly. The *p*-value for an OE item represents the proportion of the obtained mean raw score for the item to the number of points possible for the item. A point-biserial correlation between the item score and the total score on the test was also computed for the MC items. For the OE items, a Pearson correlation between the item score and the total score on the test was computed. For the item analysis, the studied item was excluded from the computation of the total score so as to not inflate the correlation artificially. This effect would be most noticeable for OE items worth several points. Note that in 2004 Reading and Mathematics items were evaluated using the following three criteria: a *p*-value below 0.30 for MC items, a point-biserial below 0.15, and an omit–rate above 5%.

Speededness

The degree to which a test is speeded can be evaluated by examining the percentage of students who fail to respond to the last items on the test. The omit rates shown in Tables 5–10 demonstrate that no forms are speeded. There were no differences between omit rates for items at the beginning of the test forms and items at the end of the test forms.

Rater Agreement

In order to monitor the reliability of the scoring of the OE items, approximately 10% of the student papers were submitted to a second rater for scoring. All other responses were read by a single rater. Indices of rater agreement and consistency were obtained using those students who had their OE items read by two raters. Tables 12–13 present the rater agreement statistics for the Reading and Mathematics OE items. These tables provide the percentages of pairs of raters' scores that did not differ (i.e., perfect agreement) and the percentages of pairs of raters' scores that differed by one point (i.e., adjacent agreement) for all OE items over all test forms (with the exception of item #55 on each of the Reading forms). When rater agreement was defined as two rater scores that differed by no more than one point (i.e., agreement), there was high rater agreement in terms of the percentage of agreement, which ranged from 98.6% to 99.7% for Reading and from 94.7 to 99.8% for Mathematics. Again, please note that the maximum possible score points for the OE items was 4 or 5. In addition to the percentage of agreement, the tables present the mean item score and item standard deviation of the item scores assigned by each rater group. Examination of Tables 12–13 shows that the mean score

points awarded by the two rater groups are very close. To further study rater agreement, intraclass correlations (Cronbach's alpha) and Kappa (Fleiss, et. al., 1969) coefficients were calculated and are reported as measures of rater agreement for each OE item. Ordinal rating scales (e.g., 0, 1, 2) used in scoring OE items contain a certain level of chance agreement that is expected. Although the intraclass correlation is reported in this report, it does not take into account chance agreement between the two raters, but Kappa does. Therefore, in general, Kappa will have values equal to or smaller than the intraclass correlation. If agreement is perfect, then Kappa is +1. If agreement is at chance levels, Kappa is 0. Landis and Koch (1977) suggest that values of Kappa greater than .75 indicate "excellent agreement", values between .40 and .74 represent "good agreement" beyond chance, and values below .40 denote "poor agreement." As Tables 12–13 show, Kappa coefficients ranged from 0.61 to 0.74 for Reading and from 0.83 to 0.93 for Mathematics. Note that the intraclass correlation and Kappa for Mathematics were higher than those for Reading. This trend has been often found in other large scale assessment programs. The values of Kappa for Reading and Mathematics meet the criteria of "good agreement" for Reading and the criteria of "excellent agreement" for Mathematics according to Landis and Koch.

Differential Item Functioning (DIF)

An item flagged for DIF is more difficult for a particular group of students than would be expected based on their total test scores, compared to the performance of the other group. The groups compared in the analysis were female and male students, and African–American, Hispanic, and White students. Other ethnic groups were not included in these analyses because their sample sizes were too small.

The statistical procedures used by CTB to identify items thought to exhibit substantial DIF are the same procedures used by ETS and NAEP. For multiple-choice items, the Mantel-Haenszel (χ^2_{MH}) statistic was used to evaluate potential DIF items. In this procedure, the "C"-level DIF items are flagged, where a "C" item indicates a large amount of DIF and has an absolute value of the Mantel-Haenszel (Δ_{MH}) significantly greater than zero (at the .05 level), and $|\Delta_{MH}|$ exceeds 1.5 (Zwick, Donoghue, and Grima, 1993).

For the constructed-response items, both the Mantel χ^2 and the standardized mean difference (SMD) statistics were used to evaluate DIF. Using these procedures, items can be flagged where the Mantel statistic is greater than zero with probability greater than .05, and the absolute value of the SMD is greater than .25. A detailed description of these procedures can be found in Zwick, et al., (1993).

Table 11 presents summary for Differential Item Functioning based on Criteria $\pm$C. Because the DIF statistics were computed based on test form, there were multiple statistics for common items. When a common item was flagged on only a few forms, this item was not flagged. Note that all items flagged based on DIF statistics were reviewed also by content editors to consider content perspective of those items.

Item Fit Assessment

A statistical procedure was used to identify items that did not fit the IRT model. Item model fit information was obtained for each item using a *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:

$$Z_j = \frac{(Q_{1j} - DF_j)}{\sqrt{2DF_j}}$$

where Q_{1j} is the item chi-square statistic,

j is an item, and

DF is the degrees of freedom for a given item *j*.

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 student ability and item parameters. These values, along with the associated chi-squares (Q_1), are computed for ten intervals corresponding to deciles of the ability distribution (Yen, 1984). Because the value of *Z* increases as the sample size increases, with other things being equal, the critical values for *Z* were established using the following equation (Yen, 1991a):

$$Z_{crit,j} = \frac{4N_j}{1500}$$

where $Z_{crit,j}$ is critical value of *Z* for item *j*, and

N_j is the number of students who responded to item *j*.

Tables 5–10 present items that were flagged statistically for poor fit for each test form. Many items displayed poor fit because the one-parameter (1PL)/one-parameter partial credit (1PPC) approach (See below IRT calibration and equating section) was used to produce *Z* statistics. In the tables, the number "3" represents poor fit. Poor fit can easily happen for many items because the 1PL model does not consider the guessing factor.

IRT Calibration and Equating

Student item responses were calibrated using the combination of two IRT models. The 1PL was used to scale the SR items, and the 1PPC model was employed to scale the OE items. The 1PL defines an SR item in terms of the item difficulty (b_i). That is, the item discrimination (a_i) does not vary over items. In this model, the probability that a student with scale score θ responds correctly to item *i* is

$$P_i(\theta) = \frac{1}{1 + \exp[-1.7a_i(\theta - b)]}$$

The 2PPC model defines an OE item in terms of an item discrimination and a location parameter for each score point (Muraki, 1990, 1992):

$$P_{jk}(\theta) = P(x_j = k-1|\theta) = \frac{\exp Z_{jk}}{\sum_{i=1}^{m_j} \exp Z_{ji}}, k = 1,...,m_{j,}$$

where m_j is the number of score levels,

$Z_{jk} = A_{jk}\,\theta + C_{jk}$,

$C_{jk} = -\sum_{i=0}^{k-1} \gamma_{ji}$, where $\gamma_{j0} = 0$,

where γ_{ji} is a parameter freely estimated from the data.

The 1PPC model for the OE items can be considered a special case of the 2PPC model. In the 1PPC model, the discrimination does not vary over items. This is the same discrimination parameter that is applied to all test items. In the above equation for the 2PPC model, the following equation replaces A_{jk} with

$A_k = \alpha\ (k-1)$, $k = 1, 2,\ldots m_j$,

where α represents a common discrimination parameter for all items.

The IRT calibrations were implemented using CTB's PARDUX software (Burket, 1991). PARDUX simultaneously estimates parameters for MC and OE items using marginal maximum likelihood procedures implemented via the EM algorithm (Bock and Aitkin, 1981; Thissen, 1982). Because the twelve test forms were spiraled within classrooms, the groups of students who took the different forms can be considered randomly equivalent. Using the anchor items (i.e., items common to all forms), student item response data from alternate test forms were calibrated together. All items across all test forms converged during item calibration.

Establishment of the Pennsylvania Grade 3 Score Scale

The 2004 Pennsylvania score scale was transformed to the 2003 scale using anchor items by Year-to-Year equating. After transformation to the 2003 score scale, scoring tables for Reading and Mathematics were generated. Tables 14–18 show the scoring tables. Note that raw number-correct scores and scale scores have a one-to-one relationship because the 1PL/1PPC model was used for item calibration and scaling. These scale scores and standard errors of measurement (SEM) on the scoring tables were plotted in Figures 1 and 2. Also, Figures 3-7 show the distributions of raw scores and scale scores for the four Reading forms and for Mathematics. For Reading, all scale scores and SEMs across all four forms appeared to be similar. Because only common items are used for scoring across all alternative forms for Mathematics, only one curve line for each scale score and SEM appears in Figure 2.

Test Form Statistics

Table 19 presents raw score descriptive statistics for each test form. Listed by test form are the number of students, mean raw score, average *p*-value, standard deviation, minimum score, maximum score, reliability coefficient (Cronbach's alpha), and the SEM. The average *p*-value was calculated by dividing the mean raw score by the maximum possible raw score points for the test. The maximum possible raw score was 50 for Reading and 70 for Mathematics. The p-values, which ranged from 0.65 to 0.66 for Reading and 0.73 to 0.75 for Mathematics, showed rather moderate difficulties. The table also shows the information when both common and matrix items combined. Examination of Table 19 reveals that for Reading, the maximum mean raw score form difference between any two forms was 0.64 for Form A (32.34) and Form B (32.98). All Reading and Mathematics forms had high reliability, with the lowest reliability

coefficient being 0.883. The reliability and the SEM for all alternative forms were similar.

Tables 20 and 21 present the mean and standard deviation of raw score for each form by ethnicity and gender, respectively. Note that for Mathematics, two raw scores were used for computation. One is the raw score for common items only and the other is the raw score for both common items and matrix items. The results show the mean (average) performance of each subgroup on each test form. As the results indicate, for both Reading and Mathematics, White students generally performed better than African American and Hispanic students. Note that there were also substantially more White students than other ethnic groups. For Reading, female students generally performed better than male students, and for Mathematics, there was no difference in performance between female and male students.

Table 22 presents descriptive statistics for reported scale scores based on all samples. For 2004, the state mean was 1303 for Reading and 1349 for Mathematics. In 2003, the state mean was 1303 for Reading and 1306 for Mathematics. The state means for 2003 and 2004 are plotted in Figure 9. The mean and standard deviation were similar across alternative forms. The values of skewness and kurtosis showed that the distribution of scale scores across Reading and Mathematics forms is close to the normal distribution. Tables 23 and 24 show the scale score means and standard deviations of the ethnic and gender subgroups for each form. The results show how each subgroup performed compared to other students by test form. The pattern found in the raw score statistics was also found in the scale score statistics. That is, for both Reading and Mathematics, White students generally performed better than African American and

Hispanic students. Like the difference in raw score, there was almost one standard deviation (200) difference between the score of White students and the other two ethnic groups. For Reading, female students generally performed better than male students, and for Mathematics, there was no difference in performance between female and male students. The distributions of raw scores and scale scores can be found in Figures 3–7. The upper plot shows the raw score distribution and the lower plot shows the scale score distribution. Because Reading and Mathematics were relatively easy for Pennsylvania students, the distribution of raw score appeared to be positively skewed. Every distribution of scale scores appeared to be a normal distribution because the distribution of scale scores is generally transformed to be a normal distribution during item calibration.

To help test score interpretation, percentiles of scale scores are provided in Table 24. These percentiles are based on Pennsylvania students who took Grade 3 Reading and Mathematics.

Summary statistics for Content Standards

Reading consists of 3 content standards and Mathematics consists of 11 content standards. Tables 28 and 29 show the raw mean, standard deviation, and mean p-value for each standard. Note that raw score is reported for content standard. Raw scores for Mathematics were computed across all forms because only common items are used for individual student reports. For Mathematics, the mean p-values vary a lot across standards.

Factor analysis was done to examine the structure of the content standards for 2004 Pennsylvania Grade 3 Reading and Mathematics. Principal axis factor analysis was applied with one factor extraction.

Table 30, Table 31, and Figure 8 show factor analysis results for Reading and Mathematics. Communalities, total variance explained, factor matrix with one factor, and scree plot were reported. As can be seen in Figure 8 for the scree plot, a one factor model appeared to be appropriate for Reading and Mathematics.

Table 1
Test Design

Test Book	Reading Form	Mathematics Form
1	A	A
2	B	B
3	C	C
4	D	D
5	A	E
6	B	F
7	C	G
8	D	H
9	A	I
10	B	J
11	C	K
12	D	L

Table 2
Number of Items and Score Points by Item Type

Content		Common Items			Matrix Items			Field-Tested Items		
		Total	MC	OE	Total	MC	OE*	Total	MC	OE
Reading	Number of Items	18	17	1	26	25	1	11	10	1
	Score Points	21	17	4	29	25	4	13	10	3
Mathematics	Number of Items	62	60	2	11	10	1			
	Score Points	70	60	10	14/15	10	4/5			

* Mathematics Forms have one Matrix OE item worth 4 score points, with the exception of forms B, H and K which is worth 5 score points.

Table 3
2004 Pennsylvania Grade 3 Sample Characteristics by Ethnicity

Content	Form	Number of Students *	Percent White	Percent African American	Percent Hispanic	Percent Others
Reading	A	30,624	76	15	5	4
	B	30,334	76	15	5	4
	C	29,940	76	15	5	4
	D	29,715	76	15	5	4
	Total	120,613	76	15	5	4
Mathematics	A	10,527	75	15	5	4
	B	10,599	76	15	5	4
	C	10,524	76	14	6	4
	D	10,443	76	15	6	4
	E	10,303	76	15	5	4
	F	10,173	76	15	5	4
	G	10,046	76	15	6	4
	H	9,942	76	15	5	4
	I	9,771	76	14	5	4
	J	9,535	76	15	5	4
	K	9,351	76	15	5	4
	L	9,311	76	15	5	4
	Total	120,525	76	15	5	4

*Students of unspecified ethnicity are not counted.

Table 4
2004 Pennsylvania Grade 3 Sample Characteristics by Gender

Content	Form	Number of Students *	Percent Male	Percent Female
Reading	A	30,695	51	49
	B	30,433	51	49
	C	30,025	51	49
	D	29,787	52	48
	Total	120,940	51	49
Mathematics	A	10,555	51	49
	B	10,632	51	49
	C	10,558	52	48
	D	10,466	51	49
	E	10,318	51	49
	F	10,203	52	48
	G	10,063	51	49
	H	9,972	52	48
	I	9,799	51	49
	J	9,571	50	50
	K	9,385	51	49
	L	9,330	51	49
	Total	120,852	51	49

*Students of unspecified gender are not counted.

Table 5
Item Statistics for Reading Form A (N=28,012) *

Item	Type	P-Val	R-ITT	Omit	FIT	Item	Type	P-Val	R-ITT	Omit	FIT
1	MC	0.61	0.25	0.00	3	29	MC	0.59	0.17	0.00	3
2	MC	0.91	0.36	0.00	1	30	MC	0.68	0.35	0.00	3
3	MC	0.85	0.42	0.01	1	31	MC	0.78	0.44	0.01	1
4	MC	0.88	0.31	0.00	1	32	MC	0.64	0.44	0.00	1
5	MC	0.75	0.36	0.01	1	33	MC	0.72	0.59	0.01	3
6	MC	0.52	0.27	0.00	3	34	MC	0.59	0.52	0.01	3
7	MC	0.76	0.49	0.01	1	35	MC	0.76	0.56	0.00	3
8	MC	0.49	0.35	0.01	3	36	MC	0.73	0.41	0.01	1
9	MC	0.77	0.46	0.02	1	37	MC	0.37	0.30	0.00	3
10	MC	0.81	0.53	0.00	3	38	MC	0.58	0.47	0.00	3
11	MC	0.81	0.37	0.01	1	39	MC	0.79	0.55	0.00	3
12	MC	0.80	0.52	0.00	3	40	MC	0.88	0.50	0.00	3
13	MC	0.81	0.51	0.00	3	41	MC	0.78	0.58	0.00	3
14	MC	0.61	0.34	0.00	3	42	MC	0.72	0.48	0.00	1
15	MC	0.85	0.45	0.00	1	43	MC	0.64	0.46	0.01	1
16	MC	0.82	0.52	0.00	3	44	MC	0.61	0.45	0.01	1
17	MC	0.81	0.41	0.00	1	45	MC	0.89	0.51	0.00	3
18	OE	0.38	0.50	0.02	1	46	MC	0.94	0.43	0.01	3
19	MC	0.61	0.50	0.00	3	47	MC	0.51	0.44	0.01	1
20	MC	0.46	0.29	0.00	3	48	MC	0.92	0.48	0.00	3
21	MC	0.79	0.56	0.00	3	49	MC	0.81	0.49	0.00	1
22	MC	0.90	0.37	0.00	1	50	MC	0.77	0.52	0.00	1
23	MC	0.82	0.48	0.01	1	51	MC	0.84	0.60	0.01	3
24	MC	0.86	0.47	0.00	1	52	MC	0.63	0.50	0.00	1
25	MC	0.84	0.28	0.00	3	53	MC	0.86	0.52	0.00	3
26	MC	0.52	0.37	0.01	3	54	MC	0.87	0.55	0.00	3
27	OE	0.36	0.46	0.02	1	55	OE	0.04	0.13	**	3
28	MC	0.82	0.46	0.00	1						

* OE omit rates are considered ‘Blanks’.

** FT OE omit rates were not calculated

Table 6
Item Statistics for Reading Form B (N=27,689) *

Item	Type	P-Val	R-ITT	Omit	FIT	Item	Type	P-Val	R-ITT	Omit	FIT
1	MC	0.60	0.25	0.00	3	29	MC	0.65	0.42	0.00	1
2	MC	0.92	0.37	0.00	1	30	MC	0.73	0.40	0.01	1
3	MC	0.84	0.42	0.01	1	31	MC	0.66	0.52	0.01	3
4	MC	0.88	0.32	0.00	1	32	MC	0.68	0.43	0.01	1
5	MC	0.75	0.37	0.01	1	33	MC	0.65	0.39	0.01	1
6	MC	0.52	0.26	0.00	3	34	MC	0.52	0.41	0.01	1
7	MC	0.76	0.49	0.01	1	35	MC	0.91	0.43	0.01	1
8	MC	0.49	0.34	0.01	3	36	MC	0.93	0.42	0.01	3
9	MC	0.78	0.46	0.02	1	37	MC	0.91	0.49	0.01	3
10	MC	0.81	0.53	0.01	3	38	MC	0.87	0.29	0.00	1
11	MC	0.81	0.36	0.01	1	39	MC	0.51	0.39	0.00	1
12	MC	0.80	0.50	0.00	3	40	MC	0.83	0.34	0.00	1
13	MC	0.82	0.51	0.00	3	41	MC	0.89	0.47	0.01	3
14	MC	0.61	0.35	0.00	3	42	MC	0.82	0.53	0.00	3
15	MC	0.85	0.45	0.00	1	43	MC	0.52	0.32	0.00	3
16	MC	0.82	0.52	0.00	3	44	MC	0.69	0.32	0.01	3
17	MC	0.81	0.40	0.00	1	45	MC	0.82	0.50	0.00	3
18	OE	0.38	0.51	0.02	1	46	MC	0.81	0.49	0.03	1
19	MC	0.80	0.53	0.00	3	47	MC	0.21	0.07	0.00	3
20	MC	0.68	0.45	0.00	1	48	MC	0.61	0.38	0.01	1
21	MC	0.72	0.39	0.00	1	49	MC	0.73	0.47	0.01	1
22	MC	0.66	0.45	0.01	1	50	MC	0.72	0.42	0.01	1
23	MC	0.46	0.30	0.00	3	51	MC	0.83	0.52	0.00	3
24	MC	0.66	0.39	0.00	1	52	MC	0.61	0.50	0.00	3
25	MC	0.84	0.45	0.00	1	53	MC	0.62	0.50	0.00	3
26	MC	0.46	0.27	0.00	3	54	MC	0.79	0.56	0.00	3
27	OE	0.43	0.50	0.02	1	55	OE	0.04	0.13	**	3
28	MC	0.73	0.38	0.00	1						

* OE omit rates are considered ‘Blanks’.

** FT OE omit rates were not calculated

Table 7
Item Statistics for Reading Form C (N=27,387) *

Item	Type	P-Val	R-ITT	Omit	FIT	Item	Type	P-Val	R-ITT	Omit	FIT
1	MC	0.61	0.25	0.00	3	29	MC	0.84	0.45	0.00	1
2	MC	0.91	0.36	0.00	1	30	MC	0.83	0.50	0.01	1
3	MC	0.85	0.42	0.01	1	31	MC	0.76	0.48	0.01	1
4	MC	0.88	0.30	0.00	1	32	MC	0.40	0.29	0.01	3
5	MC	0.76	0.36	0.01	1	33	MC	0.62	0.43	0.00	1
6	MC	0.52	0.27	0.00	3	34	MC	0.79	0.42	0.01	1
7	MC	0.76	0.48	0.01	1	35	MC	0.46	0.30	0.01	3
8	MC	0.49	0.35	0.01	3	36	MC	0.52	0.34	0.01	3
9	MC	0.78	0.45	0.02	1	37	MC	0.37	0.29	0.00	3
10	MC	0.81	0.52	0.00	3	38	MC	0.59	0.48	0.00	3
11	MC	0.81	0.37	0.01	1	39	MC	0.79	0.55	0.00	3
12	MC	0.79	0.51	0.00	3	40	MC	0.88	0.49	0.00	3
13	MC	0.82	0.52	0.00	3	41	MC	0.78	0.58	0.00	3
14	MC	0.61	0.34	0.00	3	42	MC	0.72	0.48	0.00	1
15	MC	0.85	0.44	0.00	1	43	MC	0.63	0.47	0.01	1
16	MC	0.82	0.52	0.00	3	44	MC	0.62	0.45	0.01	1
17	MC	0.81	0.40	0.00	1	45	MC	0.71	0.58	0.00	3
18	OE	0.38	0.51	0.02	1	46	MC	0.33	0.30	0.01	3
19	MC	0.85	0.46	0.00	1	47	MC	0.75	0.56	0.00	3
20	MC	0.71	0.50	0.00	1	48	MC	0.79	0.54	0.00	3
21	MC	0.79	0.47	0.03	1	49	MC	0.74	0.56	0.00	3
22	MC	0.84	0.51	0.04	3	50	MC	0.82	0.56	0.01	3
23	MC	0.81	0.56	0.01	3	51	MC	0.72	0.51	0.00	3
24	MC	0.73	0.52	0.00	1	52	MC	0.53	0.41	0.00	3
25	MC	0.79	0.50	0.00	1	53	MC	0.88	0.58	0.00	3
26	MC	0.66	0.48	0.01	1	54	MC	0.64	0.52	0.01	1
27	OE	0.43	0.56	0.02	1	55	OE	0.05	0.14	**	1
28	MC	0.71	0.50	0.00	1						

* OE omit rates are considered ‘Blanks’.

** FT OE omit rates were not calculated

Table 8
Item Statistics for Reading Form D (N=27,189) *

Item	Type	P-Val	R-ITT	Omit	FIT	Item	Type	P-Val	R-ITT	Omit	FIT
1	MC	0.60	0.26	0.00	3	29	MC	0.78	0.47	0.01	1
2	MC	0.92	0.37	0.00	1	30	MC	0.43	0.19	0.00	3
3	MC	0.84	0.43	0.01	1	31	MC	0.77	0.45	0.01	1
4	MC	0.87	0.31	0.00	1	32	MC	0.57	0.39	0.01	3
5	MC	0.75	0.37	0.01	1	33	MC	0.49	0.33	0.00	3
6	MC	0.52	0.26	0.00	3	34	MC	0.84	0.52	0.00	3
7	MC	0.76	0.49	0.01	1	35	MC	0.81	0.43	0.01	1
8	MC	0.49	0.35	0.01	3	36	MC	0.89	0.38	0.00	1
9	MC	0.77	0.46	0.02	1	37	MC	0.85	0.42	0.00	1
10	MC	0.81	0.52	0.01	3	38	MC	0.30	0.29	0.00	3
11	MC	0.81	0.36	0.01	1	39	MC	0.87	0.53	0.00	3
12	MC	0.80	0.51	0.00	3	40	MC	0.88	0.47	0.00	1
13	MC	0.81	0.51	0.00	3	41	MC	0.76	0.52	0.00	1
14	MC	0.61	0.35	0.01	3	42	MC	0.75	0.48	0.00	1
15	MC	0.85	0.47	0.00	1	43	MC	0.69	0.49	0.00	1
16	MC	0.82	0.53	0.00	3	44	MC	0.81	0.45	0.01	1
17	MC	0.81	0.41	0.00	1	45	MC	0.78	0.57	0.00	3
18	OE	0.38	0.51	0.02	1	46	MC	0.32	0.24	0.01	3
19	MC	0.61	0.50	0.00	3	47	MC	0.87	0.58	0.00	3
20	MC	0.46	0.30	0.00	3	48	MC	0.89	0.51	0.01	3
21	MC	0.79	0.55	0.00	3	49	MC	0.78	0.47	0.00	1
22	MC	0.90	0.38	0.00	1	50	MC	0.87	0.56	0.00	3
23	MC	0.82	0.50	0.01	1	51	MC	0.90	0.51	0.01	3
24	MC	0.86	0.46	0.00	1	52	MC	0.75	0.58	0.00	3
25	MC	0.84	0.27	0.00	3	53	MC	0.31	0.32	0.00	3
26	MC	0.52	0.37	0.01	3	54	MC	0.78	0.43	0.00	1
27	OE	0.36	0.46	0.02	1	55	OE	0.05	0.15	**	1
28	MC	0.35	0.31	0.01	3						

* OE omit rates are considered ‘Blanks’.

** FT OE omit rates were not calculated

Table 9
Item Statistics for Mathematics Common items (N=110,657)

Item	Type	P-Val	R-ITT	Omit	Fit	Item	Type	P-Val	R-ITT	Omit	Fit
1	MC	0.88	0.42	0.00	1	32	MC	0.76	0.55	0.01	3
2	MC	0.95	0.27	0.00	1	33	MC	0.72	0.42	0.00	1
3	MC	0.91	0.41	0.01	1	34	MC	0.82	0.44	0.00	1
4	MC	0.96	0.32	0.00	1	35	MC	0.57	0.45	0.03	1
5	MC	0.99	0.18	0.00	1	36	MC	0.77	0.36	0.00	1
6	MC	0.82	0.54	0.00	3	37	OE	0.61	0.51	0.01	3
7	MC	0.91	0.31	0.01	1	38	MC	0.91	0.34	0.00	1
8	MC	0.73	0.23	0.00	3	39	MC	0.94	0.42	0.00	3
9	MC	0.83	0.38	0.00	1	40	MC	0.90	0.42	0.00	1
10	MC	0.75	0.53	0.00	3	41	MC	0.62	0.47	0.00	1
11	MC	0.61	0.47	0.00	1	42	MC	0.73	0.56	0.00	3
12	MC	0.90	0.34	0.00	1	43	MC	0.76	0.52	0.00	3
13	MC	0.89	0.41	0.00	1	44	MC	0.92	0.40	0.00	1
14	MC	0.83	0.42	0.01	1	45	MC	0.73	0.44	0.00	1
15	MC	0.83	0.42	0.00	1	46	MC	0.92	0.44	0.00	3
16	OE	0.47	0.60	0.01	1	47	MC	0.63	0.29	0.00	3
17	MC	0.97	0.25	0.00	1	48	MC	0.79	0.46	0.00	1
18	MC	0.93	0.29	0.00	1	49	MC	0.61	0.40	0.00	1
19	MC	0.81	0.48	0.00	1	50	MC	0.67	0.42	0.01	1
20	MC	0.69	0.28	0.00	3	51	MC	0.51	0.35	0.00	3
21	MC	0.86	0.26	0.00	3	52	MC	0.74	0.60	0.00	3
22	MC	0.76	0.53	0.00	3	53	MC	0.77	0.24	0.03	3
23	MC	0.97	0.25	0.00	1	54	MC	0.61	0.36	0.00	3
24	MC	0.98	0.25	0.00	1	55	MC	0.91	0.42	0.00	1
25	MC	0.69	0.49	0.00	1	56	MC	0.79	0.34	0.01	1
26	MC	0.87	0.39	0.00	1	57	MC	0.70	0.42	0.00	1
27	MC	0.93	0.35	0.00	1	58	MC	0.81	0.40	0.00	1
28	MC	0.81	0.53	0.00	3	59	MC	0.72	0.28	0.00	3
29	MC	0.84	0.39	0.03	1	60	MC	0.70	0.46	0.00	1
30	MC	0.77	0.50	0.00	3	61	MC	0.91	0.50	0.00	3
31	MC	0.72	0.47	0.00	1	62	MC	0.74	0.47	0.00	1

Table 10
Item Statistics for Mathematics Matrix items

Form	Item	Type	P-Val	R-ITT	Omit	FIT	Form	Item	Type	P-Val	R-ITT	Omit	FIT
A (N = 9603)	63	MC	0.88	0.39	0.00	1	D (N = 9560)	63	MC	0.85	0.52	0.00	3
	64	MC	0.80	0.40	0.00	1		64	MC	0.95	0.36	0.00	1
	65	MC	0.97	0.23	0.01	1		65	MC	0.94	0.40	0.01	1
	66	MC	0.80	0.50	0.00	3		66	MC	0.85	0.40	0.01	1
	67	MC	0.91	0.48	0.00	3		67	MC	0.80	0.43	0.02	1
	68	MC	0.68	0.42	0.00	1		68	MC	0.84	0.48	0.00	3
	69	MC	0.68	0.37	0.00	1		69	MC	0.58	0.21	0.01	3
	70	MC	0.75	0.43	0.00	3		70	MC	0.55	0.16	0.01	3
	71	MC	0.81	0.51	0.00	3		71	MC	0.85	0.32	0.00	1
	72	MC	0.68	0.44	0.01	1		72	MC	0.38	0.12	0.01	3
	73	OE	0.53	0.53	0.01	3		73	OE	0.87	0.48	0.01	3
B (N = 9626)	63	MC	0.96	0.25	0.00	1	E (N = 9480)	63	MC	0.93	0.18	0.00	3
	64	MC	0.92	0.38	0.01	1		64	MC	0.84	0.31	0.00	3
	65	MC	0.85	0.51	0.02	3		65	MC	0.93	0.41	0.00	3
	66	MC	0.85	0.42	0.00	1		66	MC	0.83	0.34	0.01	1
	67	MC	0.83	0.36	0.00	1		67	MC	0.81	0.53	0.00	3
	68	MC	0.82	0.41	0.00	1		68	MC	0.86	0.43	0.00	1
	69	MC	0.78	0.43	0.01	1		69	MC	0.93	0.44	0.00	3
	70	MC	0.77	0.32	0.01	3		70	MC	0.78	0.21	0.00	3
	71	MC	0.80	0.52	0.00	3		71	MC	0.59	0.47	0.00	1
	72	MC	0.35	0.32	0.01	1		72	MC	0.44	0.49	0.00	3
	73	OE	0.51	0.63	0.01	1		73	OE	0.73	0.54	0.02	3
C (N = 9536)	63	MC	0.64	0.33	0.00	3	F (N = 9412)	63	MC	0.95	0.31	0.00	1
	64	MC	0.93	0.37	0.00	1		64	MC	0.82	0.30	0.00	3
	65	MC	0.85	0.37	0.00	1		65	MC	0.86	0.37	0.00	1
	66	MC	0.86	0.36	0.00	1		66	MC	0.76	0.48	0.00	1
	67	MC	0.92	0.37	0.01	1		67	MC	0.82	0.47	0.01	1
	68	MC	0.69	0.52	0.00	3		68	MC	0.87	0.37	0.01	1
	69	MC	0.67	0.39	0.00	1		69	MC	0.71	0.37	0.00	3
	70	MC	0.92	0.34	0.00	1		70	MC	0.91	0.28	0.00	1
	71	MC	0.68	0.50	0.01	1		71	MC	0.73	0.48	0.00	1
	72	MC	0.90	0.50	0.01	3		72	MC	0.81	0.28	0.01	3
	73	OE	0.56	0.61	0.01	1		73	OE	0.70	0.61	0.01	3

Table 10 (Cont.)
Item Statistics for Mathematics Matrix items

Form	Item	Type	P-Val	R-ITT	Omit	FIT	Form	Item	Type	P-Val	R-ITT	Omit	FIT
G (N = 9289)	63	MC	0.85	0.38	0.00	1	J (N = 8755)	63	MC	0.98	0.17	0.00	1
	64	MC	0.88	0.40	0.00	1		64	MC	0.90	0.42	0.00	1
	65	MC	0.79	0.50	0.01	3		65	MC	0.74	0.50	0.00	1
	66	MC	0.69	0.35	0.01	1		66	MC	0.55	0.43	0.00	1
	67	MC	0.62	0.44	0.02	1		67	MC	0.82	0.41	0.01	1
	68	MC	0.97	0.31	0.00	1		68	MC	0.24	0.23	0.01	3
	69	MC	0.63	0.49	0.00	1		69	MC	0.93	0.26	0.00	1
	70	MC	0.90	0.49	0.01	3		70	MC	0.73	0.39	0.00	1
	71	MC	0.72	0.12	0.02	3		71	MC	0.78	0.41	0.00	1
	72	MC	0.61	0.34	0.00	3		72	MC	0.85	0.55	0.00	3
	73	OE	0.48	0.56	0.02	3		73	OE	0.69	0.43	0.01	3
H (N = 9145)	63	MC	0.72	0.34	0.00	3	K (N = 8668)	63	MC	0.96	0.31	0.00	1
	64	MC	0.85	0.39	0.00	1		64	MC	0.97	0.33	0.00	1
	65	MC	0.96	0.23	0.01	1		65	MC	0.84	0.47	0.01	1
	66	MC	0.69	0.49	0.00	3		66	MC	0.88	0.36	0.03	1
	67	MC	0.79	0.53	0.01	3		67	MC	0.98	0.24	0.00	1
	68	MC	0.63	0.36	0.01	3		68	MC	0.90	0.36	0.01	1
	69	MC	0.75	0.41	0.01	1		69	MC	0.72	0.38	0.00	1
	70	MC	0.96	0.29	0.00	1		70	MC	0.82	0.41	0.00	1
	71	MC	0.94	0.41	0.00	3		71	MC	0.41	0.34	0.01	3
	72	MC	0.97	0.27	0.00	1		72	MC	0.65	0.45	0.00	1
	73	OE	0.57	0.59	0.01	3		73	OE	0.60	0.55	0.01	3
I (N = 9022)	63	MC	0.94	0.31	0.00	1	L (N = 8561)	63	MC	0.90	0.35	0.00	1
	64	MC	0.89	0.42	0.00	1		64	MC	0.80	0.37	0.00	1
	65	MC	0.53	0.41	0.02	1		65	MC	0.79	0.27	0.03	3
	66	MC	0.84	0.42	0.00	1		66	MC	0.91	0.57	0.00	3
	67	MC	0.87	0.37	0.01	1		67	MC	0.75	0.34	0.00	3
	68	MC	0.87	0.46	0.00	3		68	MC	0.86	0.48	0.00	1
	69	MC	0.75	0.46	0.00	1		69	MC	0.80	0.54	0.00	3
	70	MC	0.93	0.39	0.00	1		70	MC	0.55	0.52	0.00	3
	71	MC	0.68	0.33	0.00	3		71	MC	0.62	0.40	0.00	1
	72	MC	0.78	0.47	0.00	1		72	MC	0.74	0.43	0.01	1
	73	OE	0.69	0.52	0.01	3		73	OE	0.77	0.64	0.01	1

Table 11
Summary for Differential Item Functioning based on Criteria ± C

Content	Form	Focal	Item	Type	SS Ref	SS Foc	Delta	Criteria
MA	5	Hispanic	23	MC	7160	450	1.937	+C
MA	7	Hispanic	23	MC	6952	488	1.586	+C
MA	8	Afr. Amer.	68	MC	6842	1375	1.552	+C
MA	9	Hispanic	5	MC	6769	445	1.705	+C
MA	9	Hispanic	63	MC	6769	445	2.143	+C
MA	10	Afr. Amer.	41	MC	6531	1313	1.583	+C
MA	1	Afr. Amer.	32	MC	7117	1519	-1.750	-C
MA	1	Hispanic	32	MC	7117	484	-1.977	-C
MA	1	Hispanic	46	MC	7117	484	-1.576	-C
MA	1	Other	32	MC	7117	2486	-1.777	-C
MA	2	Afr. Amer.	32	MC	7173	1515	-1.883	-C
MA	2	Hispanic	32	MC	7173	465	-2.160	-C
MA	2	Other	32	MC	7173	2453	-1.976	-C
MA	3	Afr. Amer.	32	MC	7158	1393	-2.100	-C
MA	3	Hispanic	32	MC	7158	492	-1.944	-C
MA	3	Other	32	MC	7158	2378	-1.964	-C
MA	4	Afr. Amer.	32	MC	7149	1435	-1.893	-C
MA	4	Hispanic	32	MC	7149	510	-1.869	-C
MA	4	Other	32	MC	7149	2411	-1.905	-C
MA	5	Afr. Amer.	25	MC	7160	1452	-1.685	-C
MA	5	Afr. Amer.	32	MC	7160	1452	-2.305	-C
MA	5	Afr. Amer.	46	MC	7160	1452	-1.585	-C
MA	5	Hispanic	6	MC	7160	450	-1.612	-C
MA	5	Hispanic	17	MC	7160	450	-1.590	-C
MA	5	Hispanic	32	MC	7160	450	-1.702	-C
MA	5	Other	25	MC	7160	2320	-1.530	-C
MA	5	Other	32	MC	7160	2320	-2.080	-C
MA	5	Other	46	MC	7160	2320	-1.522	-C
MA	6	Afr. Amer.	32	MC	7088	1421	-2.161	-C
MA	6	Hispanic	32	MC	7088	447	-1.982	-C
MA	6	Other	32	MC	7088	2324	-2.010	-C
MA	7	Afr. Amer.	25	MC	6952	1413	-1.503	-C
MA	7	Afr. Amer.	32	MC	6952	1413	-1.767	-C
MA	7	Afr. Amer.	39	MC	6952	1413	-1.904	-C
MA	7	Afr. Amer.	70	MC	6952	1413	-2.125	-C
MA	7	Hispanic	70	MC	6952	488	-1.678	-C
MA	7	Other	32	MC	6952	2337	-1.502	-C
MA	7	Other	39	MC	6952	2337	-1.642	-C
MA	7	Other	70	MC	6952	2337	-2.097	-C
MA	8	Afr. Amer.	32	MC	6842	1375	-2.034	-C
MA	8	Hispanic	4	MC	6842	477	-1.643	-C
MA	8	Hispanic	6	MC	6842	477	-1.572	-C
MA	8	Hispanic	32	MC	6842	477	-1.794	-C

Table 11 (Cont.)
Summary for Differential Item Functioning

Content	Form	Focal	Item	Type	SS Ref	SS Foc	Delta	Criteria
MA	8	Other	32	MC	6842	2303	-1.898	-C
MA	9	Afr. Amer.	25	MC	6769	1348	-1.520	-C
MA	9	Afr. Amer.	32	MC	6769	1348	-2.078	-C
MA	9	Afr. Amer.	46	MC	6769	1348	-1.739	-C
MA	9	Hispanic	32	MC	6769	445	-1.786	-C
MA	9	Other	32	MC	6769	2253	-1.871	-C
MA	9	Other	46	MC	6769	2253	-1.602	-C
MA	10	Afr. Amer.	32	MC	6531	1313	-1.917	-C
MA	10	Afr. Amer.	72	MC	6531	1313	-2.256	-C
MA	10	Hispanic	4	MC	6531	446	-1.752	-C
MA	10	Hispanic	32	MC	6531	446	-1.621	-C
MA	10	Hispanic	72	MC	6531	446	-2.025	-C
MA	10	Other	32	MC	6531	2224	-1.773	-C
MA	10	Other	72	MC	6531	2224	-2.173	-C
MA	11	Afr. Amer.	32	MC	6463	1320	-1.854	-C
MA	11	Hispanic	32	MC	6463	451	-2.175	-C
MA	11	Hispanic	69	MC	6463	451	-1.609	-C
MA	11	Other	32	MC	6463	2205	-1.998	-C
MA	12	Afr. Amer.	32	MC	6371	1332	-1.697	-C
MA	12	Hispanic	32	MC	6371	438	-1.642	-C
MA	12	Other	32	MC	6371	2190	-1.665	-C
RD	1	Hispanic	3	MC	21014	1359	-2.310	-C
RD	1	Hispanic	53	MC	21014	1359	-1.591	-C
RD	2	Hispanic	3	MC	20747	1343	-2.339	-C
RD	3	Hispanic	3	MC	20524	1412	-2.452	-C
RD	3	Other	3	MC	20524	6863	-1.598	-C
RD	4	Hispanic	3	MC	20333	1412	-2.484	-C
RD	4	Other	3	MC	20333	6856	-1.652	-C
MA	4	Female	23	MC	4676	4813	1.596	+C
MA	8	Female	23	MC	4628	4441	1.844	+C
MA	1	Female	70	MC	4755	4764	-2.081	-C
MA	3	Female	40	MC	4732	4706	-1.535	-C
MA	3	Female	52	MC	4732	4706	-1.525	-C
MA	5	Female	52	MC	4725	4687	-1.668	-C
MA	7	Female	65	MC	4616	4596	-1.52	-C
MA	9	Female	40	MC	4486	4455	-1.702	-C
MA	11	Female	40	MC	4302	4292	-1.569	-C
RD	1	Female	53	MC	13914	13871	-1.531	-C

Table 12
Rater Agreement for Reading Constructed-Response Items

Form	Item	Mean of G1	Mean of G2	SD of G1	SD of G2	Percent of Perfect Agreement	Percent of Adjacent Agreement	Percent of Agreement	Intraclass Correlation	Kappa
A	18	1.45	1.45	0.61	0.61	75.12	24.60	99.72	0.83	0.65
A	27	1.34	1.38	0.60	0.61	77.57	22.00	99.57	0.84	0.68
B	18	1.45	1.45	0.61	0.61	75.12	24.60	99.72	0.83	0.65
B	27	1.65	1.66	0.70	0.72	64.32	34.37	98.69	0.80	0.61
C	18	1.45	1.45	0.61	0.61	75.12	24.60	99.72	0.83	0.65
C	27	1.63	1.62	0.70	0.69	76.19	23.43	99.62	0.87	0.74
D	18	1.45	1.45	0.61	0.61	75.12	24.60	99.72	0.83	0.65
D	27	1.33	1.35	0.59	0.61	76.08	23.23	99.31	0.82	0.64

G1: Rater group 1 G2: Rater group 2
Percent of Agreement is the sum of percents of perfect and adjacent agreements.

Table 13
Rater Agreement for Mathematics Constructed-Response Items

Form	Item	N	Mean of G1	Mean of G2	SD of G1	SD of G2	Percent of Perfect Agreement	Percent of Adjacent Agreement	Percent of Agreement	Intraclass Correlation	Kappa
A	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
A	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
A	73	1158	2.00	2.01	1.00	1.01	76.9	21.4	98.3	0.93	0.86
B	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
B	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
B	73	1116	2.38	2.37	1.30	1.30	73.8	23.3	97.1	0.95	0.90
C	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
C	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
C	73	1167	2.13	2.15	0.95	0.95	77.1	20.7	97.9	0.92	0.84
D	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
D	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
D	73	1112	3.40	3.40	0.96	0.97	95.4	3.7	99.1	0.98	0.96
E	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
E	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
E	73	1221	2.86	2.88	1.25	1.25	86.0	11.8	97.8	0.96	0.93
F	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
F	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
F	73	1065	2.78	2.77	1.26	1.26	88.3	10.2	98.5	0.97	0.95
G	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
G	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
G	73	1090	1.88	1.89	1.25	1.24	74.5	24.0	98.5	0.95	0.90

G1: Rater group 1 G2: Rater group 2
Percent of Agreement is the sum of percents of perfect and adjacent agreements.

Table 13 (Cont.)
Rater Agreement for Mathematics Constructed-Response Items

Form	Item	N	Mean of G1	Mean of G2	SD of G1	SD of G2	Percent of Perfect Agreement	Percent of Adjacent Agreement	Percent of Agreement	Intraclass Correlation	Kappa
H	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
H	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
H	73	1079	2.70	2.70	1.21	1.22	68.2	26.5	94.7	0.91	0.83
I	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
I	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
I	73	1040	2.66	2.67	1.19	1.19	73.2	23.9	97.0	0.93	0.87
J	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
J	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
J	73	1015	2.69	2.68	0.80	0.80	91.5	8.2	99.7	0.96	0.93
K	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
K	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
K	73	1061	2.89	2.89	1.39	1.42	81.2	14.9	96.1	0.96	0.92
L	16	6556	2.24	2.23	1.16	1.15	76.3	22.5	98.8	0.95	0.90
L	37	6556	3.01	3.00	1.24	1.23	77.2	21.2	98.4	0.95	0.91
L	73	991	2.99	3.00	0.92	0.92	86.7	13.1	99.8	0.96	0.92

G1: Rater group 1 G2: Rater group 2
Percent of Agreement is the sum of percents of perfect and adjacent agreements

Table 14
Scoring Table for Reading Form A

NC	SS	SEM	NC	SS	SEM
0	300	187			
1	319	179	26	1154	54
2	461	133	27	1172	55
3	550	112	28	1190	55
4	616	98	29	1209	56
5	669	89	30	1227	56
6	712	82	31	1247	57
7	750	76	32	1267	58
8	783	72	33	1287	59
9	813	69	34	1308	60
10	840	66	35	1331	61
11	866	64	36	1354	63
12	889	62	37	1379	65
13	912	60	38	1405	68
14	933	59	39	1434	71
15	954	58	40	1466	74
16	974	57	41	1501	79
17	993	56	42	1541	84
18	1011	55	43	1587	91
19	1030	55	44	1642	101
20	1048	55	45	1712	114
21	1066	54	46	1801	131
22	1083	54	47	1921	152
23	1101	54	48	1999	164
24	1119	54	49	1999	164
25	1136	54	50	1999	164

Table 15
Scoring Table for Reading Form B

NC	SS	SEM	NC	SS	SEM
0	300	192			
1	339	174	26	1143	54
2	470	128	27	1161	55
3	552	107	28	1179	55
4	613	95	29	1197	55
5	662	86	30	1216	56
6	703	80	31	1235	57
7	739	75	32	1254	57
8	772	71	33	1274	58
9	801	68	34	1295	59
10	828	66	35	1317	61
11	853	64	36	1340	62
12	877	62	37	1364	64
13	899	60	38	1390	67
14	921	59	39	1418	69
15	941	58	40	1449	73
16	961	57	41	1483	77
17	981	56	42	1521	82
18	999	56	43	1565	89
19	1018	55	44	1618	98
20	1036	55	45	1684	110
21	1054	54	46	1767	125
22	1072	54	47	1875	144
23	1090	54	48	1999	164
24	1107	54	49	1999	164
25	1125	54	50	1999	164

Table 16
Scoring Table for Reading Form C

NC	SS	SEM	NC	SS	SEM
0	300	196			
1	342	176	26	1150	54
2	478	130	27	1168	54
3	563	109	28	1185	55
4	625	96	29	1204	55
5	676	87	30	1222	56
6	718	80	31	1241	57
7	754	75	32	1261	57
8	786	71	33	1281	58
9	816	68	34	1302	60
10	842	65	35	1324	61
11	867	63	36	1347	63
12	890	61	37	1372	65
13	912	60	38	1398	67
14	933	58	39	1426	70
15	953	57	40	1457	73
16	973	56	41	1492	78
17	991	56	42	1531	83
18	1010	55	43	1576	90
19	1028	54	44	1629	99
20	1046	54	45	1695	111
21	1063	54	46	1780	127
22	1080	54	47	1893	147
23	1098	54	48	1999	164
24	1115	54	49	1999	164
25	1132	54	50	1999	164

Table 17
Scoring Table for Reading Form D

NC	SS	SEM	NC	SS	SEM
0	300	182			
1	311	178	26	1142	56
2	449	131	27	1161	56
3	535	110	28	1180	57
4	600	97	29	1199	57
5	651	88	30	1219	58
6	694	81	31	1240	59
7	731	76	32	1261	60
8	764	72	33	1283	61
9	794	69	34	1306	62
10	821	66	35	1330	64
11	846	64	36	1355	65
12	870	62	37	1381	67
13	893	60	38	1410	70
14	914	59	39	1440	72
15	935	58	40	1473	76
16	955	57	41	1510	80
17	975	57	42	1551	86
18	994	56	43	1599	92
19	1013	56	44	1655	102
20	1031	55	45	1725	114
21	1050	55	46	1814	130
22	1068	55	47	1934	152
23	1086	55	48	1999	163
24	1105	55	49	1999	163
25	1123	55	50	1999	163

Table 18
Scoring Table for Mathematics

NC	SS	SEM	NC	SS	SEM
0	200	142			
1	200	142	36	1066	50
2	200	142	37	1079	50
3	265	127	38	1093	50
4	343	112	39	1107	50
5	405	101	40	1120	50
6	456	93	41	1134	50
7	500	86	42	1148	50
8	539	81	43	1162	50
9	573	77	44	1176	51
10	604	74	45	1190	51
11	633	71	46	1204	51
12	659	68	47	1219	52
13	684	66	48	1234	52
14	707	64	49	1249	53
15	729	62	50	1265	54
16	750	61	51	1281	54
17	770	60	52	1297	55
18	789	59	53	1314	56
19	808	57	54	1332	57
20	825	57	55	1351	59
21	843	56	56	1370	60
22	860	55	57	1391	62
23	876	54	58	1412	64
24	892	54	59	1436	66
25	908	53	60	1461	69
26	923	53	61	1488	72
27	938	52	62	1518	76
28	953	52	63	1552	81
29	967	51	64	1591	88
30	982	51	65	1638	96
31	996	51	66	1695	109
32	1010	51	67	1770	127
33	1024	50	68	1881	159
34	1038	50	69	1999	196
35	1052	50	70	1999	196

Table 19
Raw score Descriptive Statistics Based on All Samples

Content	Form	N Count	Mean	Mean P-Value	SD	Min	Max	Alpha	SEM
Reading	A	30,947	32.34	0.65	8.80	4	48	0.896	2.84
	B	30,667	32.98	0.66	8.42	3	49	0.883	2.88
	C	30,299	32.61	0.65	9.06	3	49	0.902	2.84
	D	30,041	32.59	0.65	8.28	4	48	0.887	2.78
Mathematics (Both Common items and Matrix items)	A	10,649	61.75	0.74	14.18	0	84	0.924	3.91
	B	10,711	62.43	0.73	14.11	0	84	0.922	3.94
	C	10,661	62.23	0.74	14.00	0	84	0.921	3.93
	D	10,543	63.03	0.75	13.63	0	83	0.916	3.95
	E	10,392	63.07	0.75	13.73	0	84	0.918	3.93
	F	10,282	63.23	0.75	13.81	0	84	0.920	3.91
	G	10,149	61.86	0.74	13.76	0	84	0.917	3.96
	H	10,051	63.23	0.74	13.71	0	85	0.917	3.95
	I	9,884	62.88	0.75	14.09	0	84	0.921	3.96
	J	9,648	62.40	0.74	13.65	0	84	0.920	3.86
	K	9,469	63.13	0.74	13.84	0	84	0.917	3.99
	L	9,428	62.44	0.74	14.62	0	83	0.928	3.92
Mathematics (Common Items only)	Total	121,867	52.58	0.75	10.98	9	70	0.920	3.11

Table 20
Raw Score Descriptive Statistics by Ethnicity

		White				African American				Hispanic			
Content	Form	N Count	Raw Score Mean	Raw Score SD	Mean P-Value	N Count	Raw Score Mean	Raw Score SD	Mean P-Value	N Count	Raw Score Mean	Raw Score SD	Mean P-Value
Reading	A	23,270	34.01	8.01	0.68	4,530	26.21	8.75	0.52	1,614	25.61	9.05	0.51
	B	23,076	34.60	7.57	0.69	4,481	27.09	8.60	0.54	1,559	26.04	8.94	0.52
	C	22,747	34.29	8.25	0.69	4,368	26.46	9.09	0.53	1,642	25.82	9.58	0.52
	D	22,534	34.22	7.43	0.68	4,382	26.68	8.32	0.53	1,632	26.11	8.89	0.52
Mathematics (Both Common items and Matrix items)	A	7,932	64.63	12.42	0.77	1,597	51.05	14.92	0.61	575	53.10	15.22	0.63
	B	8,047	65.14	12.44	0.77	1,600	52.05	14.94	0.61	530	52.21	15.49	0.61
	C	8,008	64.80	12.35	0.77	1,508	52.01	15.15	0.62	583	52.44	15.57	0.62
	D	7,907	65.73	11.87	0.78	1,535	52.56	14.55	0.63	583	54.11	14.71	0.64
	E	7,875	65.62	12.06	0.78	1,525	53.07	14.42	0.63	523	53.56	15.63	0.64
	F	7,778	65.88	12.09	0.78	1,485	52.77	14.84	0.63	513	53.17	15.19	0.63
	G	7,640	64.54	11.96	0.77	1,467	51.21	14.74	0.61	560	52.74	15.43	0.63
	H	7,555	65.85	12.04	0.77	1,459	52.55	15.13	0.62	546	54.73	14.36	0.64
	I	7,447	65.57	12.40	0.78	1,404	52.53	15.12	0.63	515	52.42	15.44	0.62
	J	7,237	64.94	11.95	0.77	1,389	52.03	15.32	0.62	509	53.74	14.40	0.64
	K	7,088	65.72	12.03	0.77	1,389	53.15	14.88	0.63	496	53.86	15.64	0.63
	L	7,065	65.25	12.79	0.78	1,381	51.90	15.67	0.62	499	52.32	15.66	0.62
Mathematics (Common items only)	Total	91,579	54.74	9.57	0.78	17,739	44.02	11.89	0.63	6,432	44.81	12.08	0.64

Table 21
Raw Score Descriptive Statistics by Gender

Content	Form	Male				Female			
		N Count	Raw Score Mean	Raw Score SD	Mean P-Value	N Count	Raw Score Mean	Raw Score SD	Mean P-Value
Reading	A	15,775	31.69	8.98	0.63	14,920	33.06	8.55	0.66
	B	15,461	32.18	8.72	0.64	14,972	33.83	8.02	0.68
	C	15,443	31.62	9.30	0.63	14,582	33.70	8.65	0.67
	D	15,344	31.82	8.56	0.64	14,443	33.46	7.86	0.67
Mathematics (Both Common items and Matrix items)	A	5,435	62.00	14.22	0.74	5,120	61.65	13.97	0.73
	B	5,375	62.56	14.43	0.74	5,257	62.35	13.73	0.73
	C	5,455	62.29	14.32	0.74	5,103	62.27	13.59	0.74
	D	5,307	63.01	13.94	0.75	5,159	63.17	13.14	0.75
	E	5,306	63.17	14.00	0.75	5,012	63.06	13.34	0.75
	F	5,267	63.45	13.92	0.76	4,936	63.10	13.57	0.75
	G	5,159	61.96	14.11	0.74	4,904	61.86	13.26	0.74
	H	5,229	63.20	14.17	0.74	4,743	63.35	13.14	0.75
	I	5,026	63.00	14.31	0.75	4,773	62.85	13.74	0.75
	J	4,800	62.73	13.67	0.75	4,771	62.17	13.48	0.74
	K	4,815	63.43	14.01	0.75	4,570	62.97	13.42	0.74
	L	4,798	62.63	14.69	0.75	4,532	62.49	14.24	0.74
Mathematics (Common Items only)	Total	61,972	52.76	11.09	0.75	58,880	52.45	10.83	0.75

Table 22
Descriptive Statistics for Reported Scale Scores Based on All Samples

Content	Form	N Count	Scale Score Mean	Scale Score SD	Skewness	Kurtosis
Reading	A	30,947	1,302	198	-0.20	-0.35
	B	30,667	1,301	192	-0.19	0.14
	C	30,299	1,305	206	-0.13	-0.25
	D	30,041	1,302	196	-0.20	-0.18
	Total	121,954	1,303	198	-0.18	-0.16
Mathematics	A	10,649	1,343	215	-0.03	0.00
	B	10,711	1,347	215	-0.02	0.00
	C	10,661	1,349	213	-0.04	0.04
	D	10,543	1,349	214	-0.01	-0.03
	E	10,392	1,351	213	0.02	0.05
	F	10,282	1,351	213	0.02	0.05
	G	10,149	1,352	212	-0.01	0.06
	H	10,051	1,351	210	0.03	0.05
	I	9,884	1,350	213	-0.04	0.07
	J	9,648	1,351	213	-0.03	0.09
	K	9,469	1,348	212	-0.03	0.04
	L	9,428	1,345	214	-0.07	0.04
	Total	121,867	1,349	213	-0.02	0.04

Table 23
Descriptive Statistics for Reported Scale Scores by Ethnicity

Content	Test Form	White				African American				Hispanic			
		Mean	SD	MIN	MAX	Mean	SD	MIN	MAX	Mean	SD	MIN	MAX
Reading	A	1,338	186	712	1,999	1,167	178	616	1,921	1,156	184	669	1,801
	B	1,337	180	552	1,999	1,171	174	613	1,875	1,150	182	613	1,999
	C	1,342	195	676	1,999	1,170	185	625	1,893	1,157	194	563	1,780
	D	1,339	183	600	1,999	1,166	175	694	1,725	1,155	189	651	1,814
	Total	1,339	186	552	1,999	1,168	178	613	1,921	1,155	187	563	1,999
Mathematics	A	1,385	200	573	1,999	1,184	193	684	1,881	1,214	197	684	1,770
	B	1,387	199	684	1,999	1,193	202	573	1,999	1,198	190	770	1,770
	C	1,385	199	659	1,999	1,197	193	707	1,881	1,205	201	659	1,770
	D	1,390	199	633	1,999	1,185	190	659	1,881	1,208	195	729	1,770
	E	1,388	199	684	1,999	1,199	190	659	1,999	1,210	213	659	1,881
	F	1,390	199	729	1,999	1,193	195	684	1,999	1,203	195	729	1,770
	G	1,391	196	729	1,999	1,194	195	604	1,881	1,214	203	659	1,881
	H	1,389	196	750	1,999	1,193	192	684	1,881	1,218	195	750	1,999
	I	1,390	198	684	1,999	1,197	195	684	1,999	1,193	205	604	1,999
	J	1,389	197	604	1,999	1,194	202	659	1,999	1,210	199	633	1,881
	L	1,385	196	707	1,999	1,197	197	604	1,999	1,212	211	750	1,881
	K	1,385	196	604	1,999	1,189	198	659	1,881	1,192	202	684	1,770
	Total	1,388	198	573	1,999	1,193	195	573	1,999	1,206	200	604	1,999

.

Table 24
Descriptive Statistics for Reported Scale Scores by Gender

Content	Test Form	Male				Female			
		Mean	SD	MIN	MAX	Mean	SD	MIN	MAX
Reading	A	1,286	198	669	1,999	1,319	196	616	1,999
	B	1,282	193	552	1,999	1,321	188	613	1,999
	C	1,282	205	563	1,999	1,331	204	676	1,999
	D	1,284	199	600	1,999	1,322	190	694	1,999
	Total	1,284	199	552	1,999	1,323	195	613	1,999
Mathematics	A	1,348	217	573	1,999	1,340	212	684	1,999
	B	1,352	219	573	1,999	1,343	210	684	1,999
	C	1,352	214	659	1,999	1,346	211	659	1,999
	D	1,350	216	659	1,999	1,348	210	633	1,999
	E	1,356	216	659	1,999	1,347	211	659	1,999
	F	1,355	215	684	1,999	1,347	211	729	1,999
	G	1,355	213	604	1,999	1,350	210	659	1,999
	H	1,354	212	684	1,999	1,348	208	750	1,999
	I	1,355	213	684	1,999	1,346	213	604	1,999
	J	1,355	215	604	1,999	1,346	211	633	1,999
	L	1,353	213	604	1,999	1,343	211	684	1,999
	K	1,348	216	684	1,999	1,344	211	604	1,999
	Total	1,353	215	573	1,999	1,346	211	604	1,999

Table 25
Percentiles of Scale Score Ranges

Percentile	Reading Score Range	Mathematics Score Range
1	300-853	200-868
2	854-879	869-907
3	880-902	908-938
4	903-922	939-963
5	923-946	964-983
6	947-955	984-1003
7	956-968	1004-1020
8	969-987	1021-1036
9	988-1004	1037-1050
10	1005-1012	1051-1063
11	1013-1027	1064-1076
12	1028-1043	1077-1088
13	1044-1050	1089-1098
14	1051-1065	1099-1109
15	1066-1077	1110-1119
16	1078-1086	1120-1129
17	1087-1098	1130-1137
18	1099-1106	1138-1146
19	1107-1117	1147-1154
20	1118-1124	1155-1162
21	1125-1135	1163-1170
22	1136-1144	1171-1178
23	1145-1151	1179-1185
24	1152-1159	1186-1193
25	1160-1165	1194-1200
26	1166-1174	1201-1207
27	1175-1182	1208-1214
28	1183-1190	1215-1220
29	1191-1197	1221-1227
30	1198-1203	1228-1233
31	1204-1210	1234-1239
32	1211-1218	1240-1245
33	1219-1224	1246-1251
34	1225-1228	1252-1257
35	1229-1236	1258-1263
36	1237-1244	1264-1269
37	1245-1247	1270-1275
38	1248-1252	1276-1280
39	1253-1260	1281-1286
40	1261-1268	1287-1291

Table 25 (Cont.)
Percentiles of Scale Score Ranges

Percentile	Reading Score Range	Mathematics Score Range
41	1269	1292-1297
42	1270-1276	1298-1302
43	1277-1283	1303-1307
44	1284-1289	1308-1313
45	1290-1292	1314-1318
46	1293-1298	1319-1323
47	1299-1304	1324-1329
48	1305-1311	1330-1334
49	1312-1314	1335-1339
50	1315-1317	1340-1344
51	1318-1324	1345-1349
52	1325-1331	1350-1354
53	1332-1337	1355-1359
54	1338-1340	1360-1365
55	1341-1342	1366-1370
56	1343-1350	1371-1375
57	1351-1356	1376-1380
58	1357-1362	1381-1385
59	1363-1365	1386-1390
60	1366-1367	1391-1395
61	1368-1374	1396-1401
62	1375-1381	1402-1406
63	1382-1387	1407-1412
64	1388-1391	1413-1417
65	1392-1393	1418-1423
66	1394-1399	1424-1428
67	1400-1405	1429-1433
68	1406-1412	1434-1439
69	1413-1417	1440-1444
70	1418-1421	1445-1450
71	1422-1424	1451-1455
72	1425-1431	1456-1461
73	1432-1438	1462-1467
74	1439-1444	1468-1473
75	1445-1449	1474-1479
76	1450-1453	1480-1485
77	1454-1460	1486-1491
78	1461-1466	1492-1498
79	1467-1473	1499-1505
80	1474-1479	1506-1511

Table 25 (Cont.)
Percentiles of Scale Score Ranges

Percentile	Reading Score Range	Mathematics Score Range
81	1480-1485	1512-1518
82	1486-1492	1519-1526
83	1493-1500	1527-1534
84	1501-1507	1535-1542
85	1508-1514	1543-1550
86	1515-1521	1551-1558
87	1522-1530	1559-1568
88	1531-1540	1569-1577
89	1541-1549	1578-1587
90	1550-1558	1588-1597
91	1559-1568	1598-1610
92	1569-1583	1611-1624
93	1584-1595	1625-1637
94	1596-1607	1638-1655
95	1608-1626	1656-1676
96	1627-1649	1677-1696
97	1650-1672	1697-1732
98	1673-1719	1733-1771
99	1720-1999	1772-1999

Table 26
The Number of Items per Each Reading Standard *

Objective	Form A		Form B		Form C		Form D	
	MC	OE	MC	OE	MC	OE	MC	OE
1.1.3	10	2	11	2	10	2	11	2
1.2.3	13	0	12	0	13	0	13	0
1.3.3	19	0	19	0	19	0	18	0
Total	42	2	42	2	42	2	42	2

* Table does not include FT items.

Table 27
The Number of Items per each Mathematics Standard *

Standards	A		B		C		D		E		F		G		H		I		J		K		L	
	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE
2.1.3	12	1	12	0	11	0	13	0	11	0	12	0	12	0	12	0	13	0	11	0	11	0	11	0
2.2.3	12	0	12	0	11	1	11	0	11	0	11	0	10	0	11	0	11	0	12	0	11	0	12	0
2.3.3	4	1	5	1	5	1	3	2	4	1	5	1	5	1	4	1	5	1	4	1	4	1	4	1
2.4.3	5	0	5	0	5	0	6	0	6	1	6	1	6	0	5	0	6	0	6	0	6	0	6	0
2.5.3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3
2.6.3	8	0	7	1	7	0	8	0	8	0	8	0	7	1	7	0	7	0	7	0	9	0	8	0
2.7.3	5	0	5	0	6	0	5	0	6	0	6	0	6	0	6	1	6	0	6	0	6	0	5	0
2.8.3	8	0	9	0	9	0	8	0	8	0	7	0	8	0	9	0	8	1	8	0	8	0	8	0
2.9.3	5	1	4	1	4	1	5	1	5	1	4	1	4	1	5	1	3	1	4	2	5	1	4	1
2.10.3	6	0	6	0	6	0	5	0	5	0	5	0	6	0	6	0	6	0	6	0	5	1	6	0
2.11.3	5	0	5	0	6	0	6	0	6	0	6	0	6	0	5	0	5	0	6	0	5	0	6	1
Sum	70	6	70	6	70	6	70	6	70	6	70	6	70	6	70	6	70	6	70	6	70	6	70	6

* All forms contain 3 OE items, with each item mapping to 2 Mathematics Standards. Therefore the OE items for each form sum up to 6 instead of 3.

Table 28
Summary Statistics for Reading Standards *

Standard 1. 1. 3

Content	Form	Total Number of Items	Total Score Points	Mean	Mean P-Value	SD
Reading	A	12	18	9.68	0.54	2.76
	B	13	19	10.92	0.57	2.98
	C	12	18	9.51	0.53	2.95
	D	13	19	10.99	0.58	2.88

Standard 1. 2. 3

Content	Form	Total Number of Items	Total Score Points	Mean	Mean P-Value	SD
Reading	A	13	13	9.39	0.72	3.03
	B	12	12	8.48	0.71	2.66
	C	13	13	9.41	0.72	3.02
	D	13	13	9.76	0.75	2.81

Standard 1. 3. 3

Content	Form	Total Number of Items	Total Score Points	Mean	Mean P-Value	SD
Reading	A	19	19	13.27	0.70	3.92
	B	19	19	13.58	0.71	3.66
	C	19	19	13.70	0.72	3.96
	D	18	18	11.84	0.66	3.44

* Table does not include FT items.

Table 29
Summary Statistics for Mathematics Standards *

Content	Standard	Total Number of Items	Total Score Points	Mean	Mean P-Value	SD
Math	2.1.3	10	10	7.65	0.77	1.89
	2.2.3	10	10	8.28	0.83	2.00
	2.3.3	4	8	4.71	0.59	1.47
	2.4.3	5	5	3.69	0.74	1.26
	2.5.3	2	10	5.34	0.53	1.91
	2.6.3	7	7	5.14	0.73	1.86
	2.7.3	5	5	4.01	0.80	1.13
	2.8.3	7	7	5.18	0.74	1.54
	2.9.3	4	8	5.54	0.69	1.51
	2.10.3	5	5	4.23	0.85	1.16
	2.11.3	5	5	4.14	0.83	1.04

* Table includes common items only

Table 30
Factor Analysis Results for Reading 3 Standards

Communalities

	Initial	Extraction
1	.530	.638
2	.580	.720
3	.571	.702

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.372	79.072	79.072	2.060	68.662	68.662
2	.340	11.338	90.410			
3	.288	9.590	100.000			

Extraction Method: Principal Axis Factoring.

Factor Matrix

Standard	Factor
1	.799
2	.849
3	.838

Table 31
Factor Analysis Results for Mathematics 11 Standards

Communalities

	Initial	Extraction
1	.585	.614
2	.579	.596
3	.726	.481
4	.417	.432
5	.843	.487
6	.607	.640
7	.466	.490
8	.520	.551
9	.725	.357
10	.270	.292
11	.376	.389

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.823	52.938	52.938	5.329	48.444	48.444
2	1.132	10.295	63.233			
3	.704	6.396	69.629			
4	.597	5.428	75.058			
5	.568	5.164	80.222			
6	.532	4.838	85.060			
7	.493	4.480	89.539			
8	.396	3.600	93.139			
9	.340	3.095	96.234			
10	.326	2.961	99.196			
11	.088	.804	100.000			

Factor Matrix

Standards	Factor	Standards	Factor
	1		1
1	.783	7	.700
2	.772	8	.743
3	.693	9	.597
4	.657	10	.540
5	.698	11	.624
6	.800		

Figure 1
Reading Scale Score and SEM across Four Forms

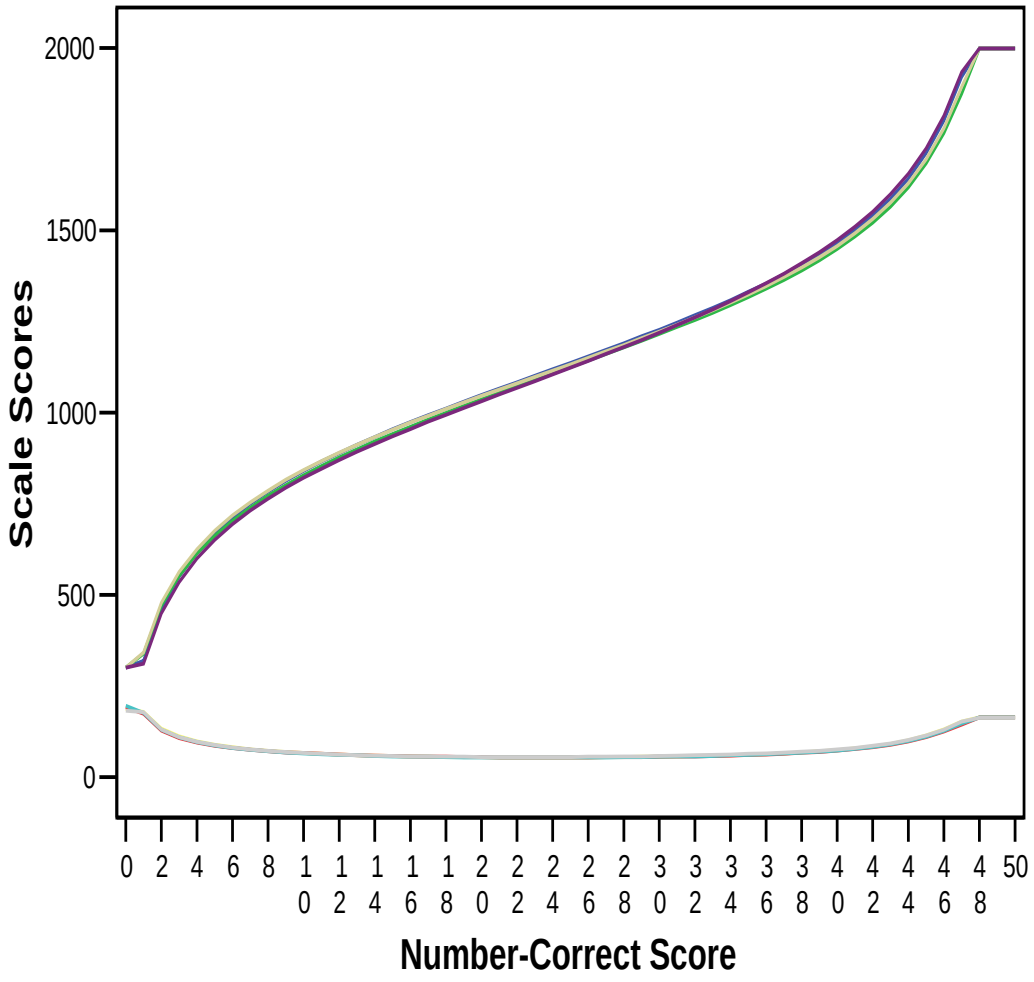


Figure 2
Mathematics Scale Scores and SEM

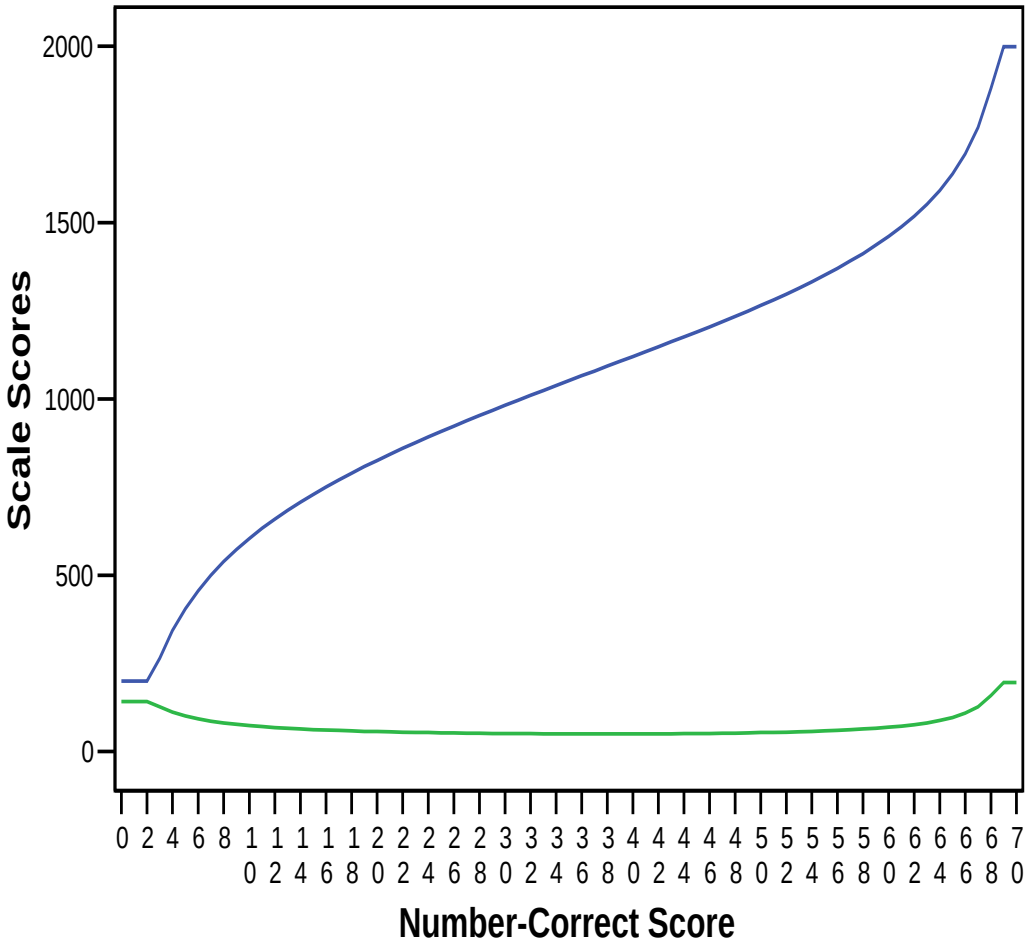
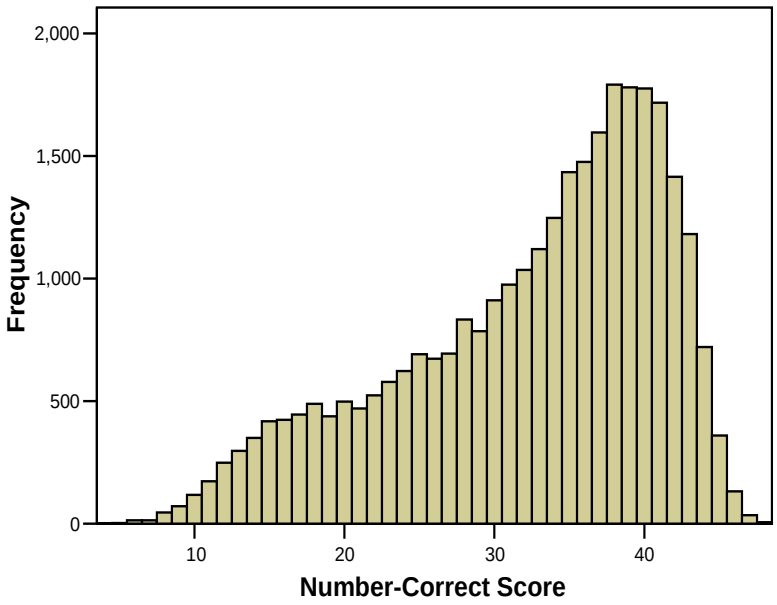


Figure 3
Reading Form A Raw Score and Scale Score Distribution



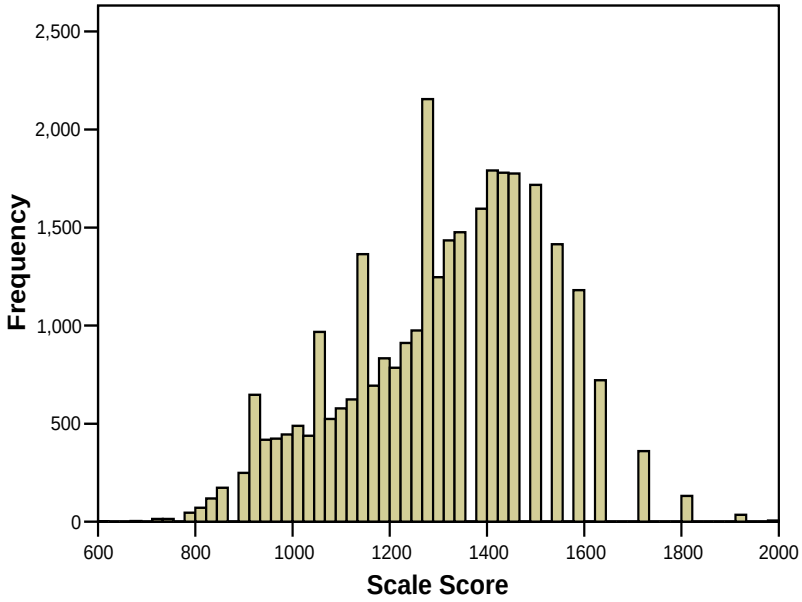


Figure 4
Reading Form B Raw Score and Scale Score Distribution

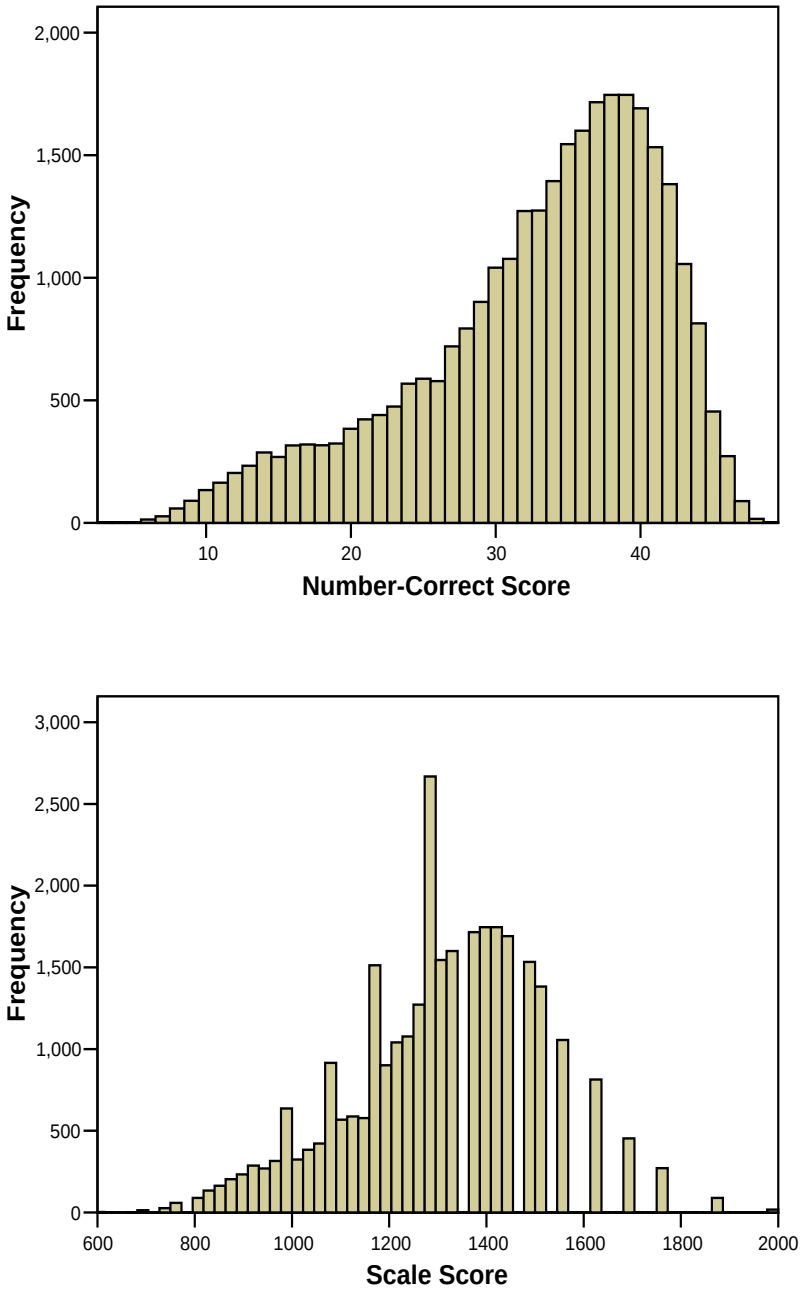


Figure 5
Reading Form C Raw Score and Scale Score Distribution

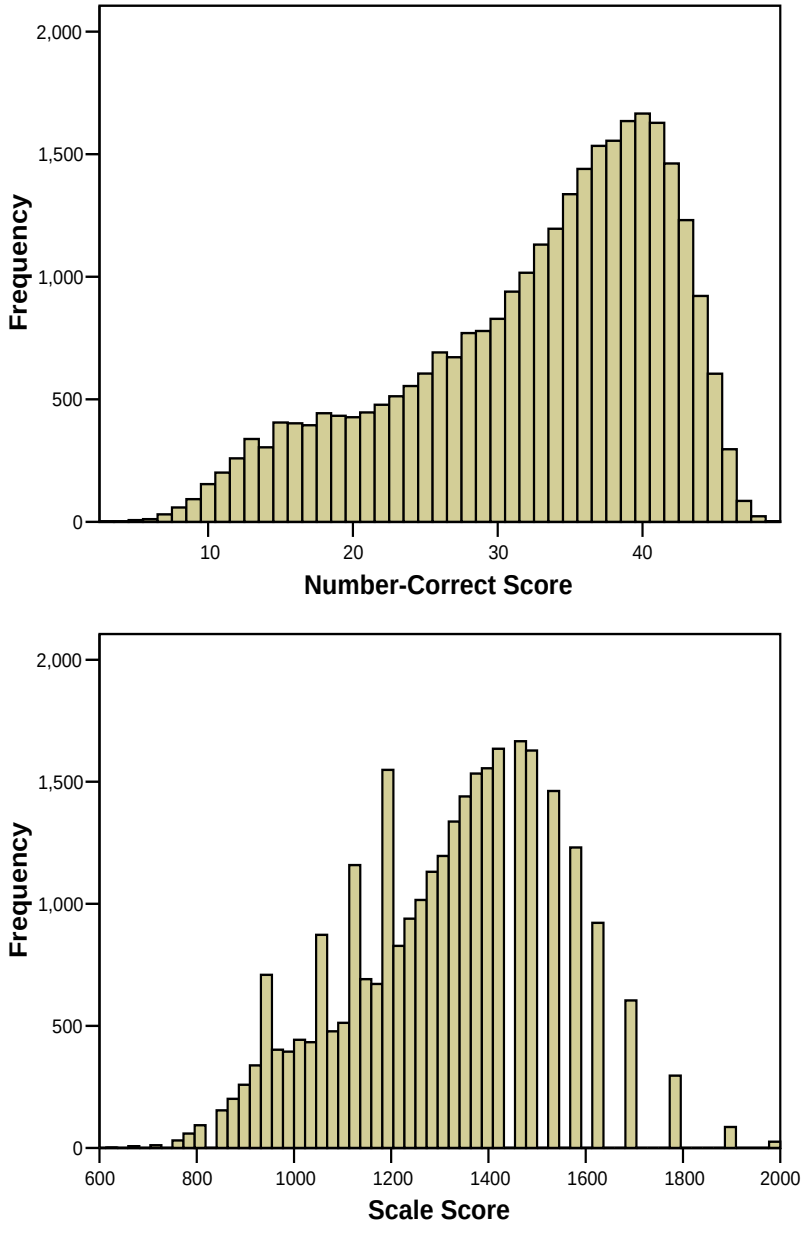


Figure 6
Reading Form D Raw Score and Scale Score Distribution

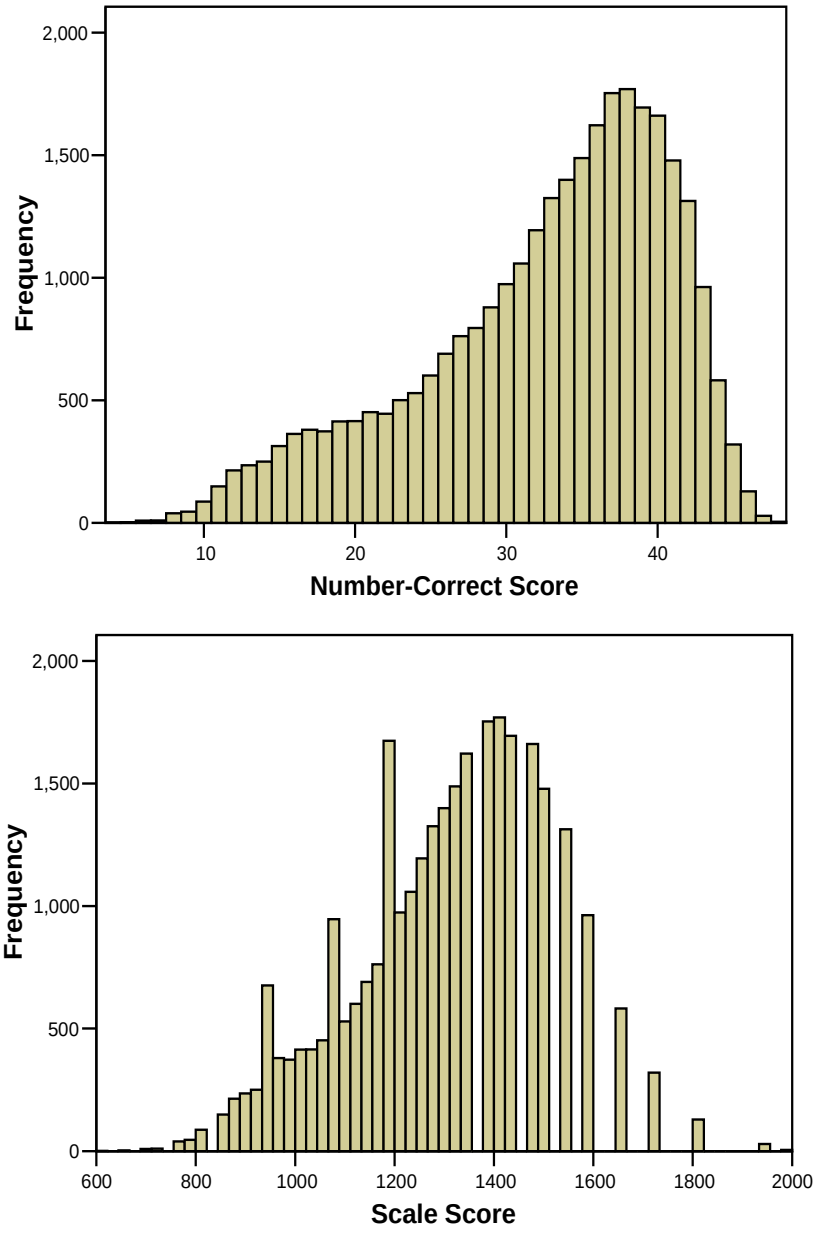


Figure 7
Mathematics Raw Score and Scale Score Distribution

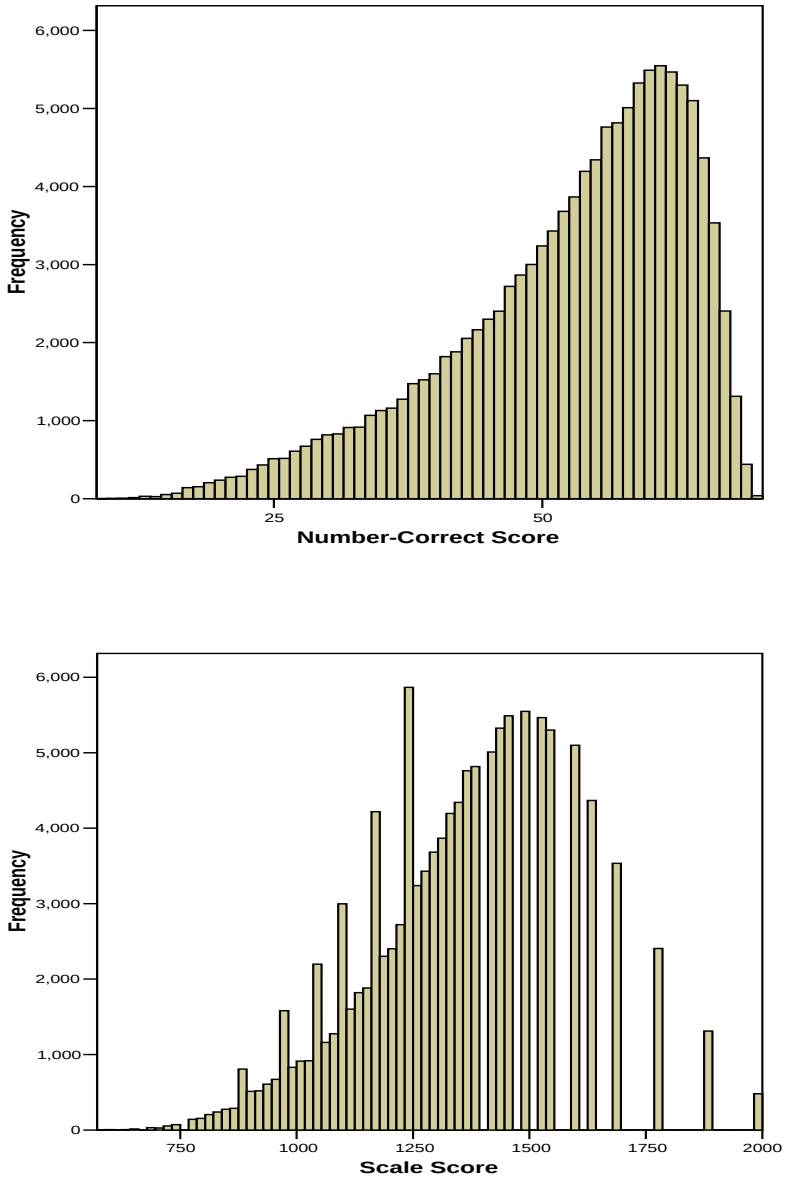
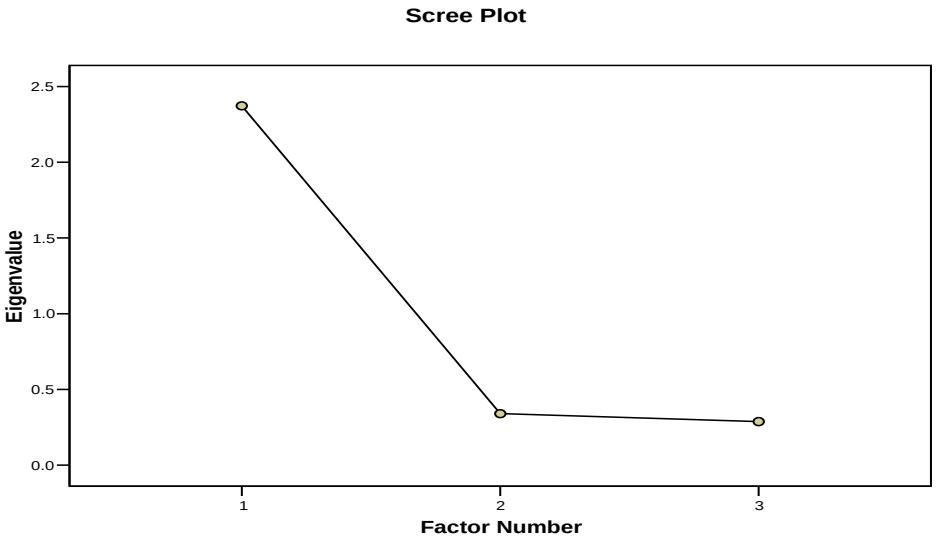


Figure 8
Scree Plot for Reading and Mathematics

Reading



Mathematics

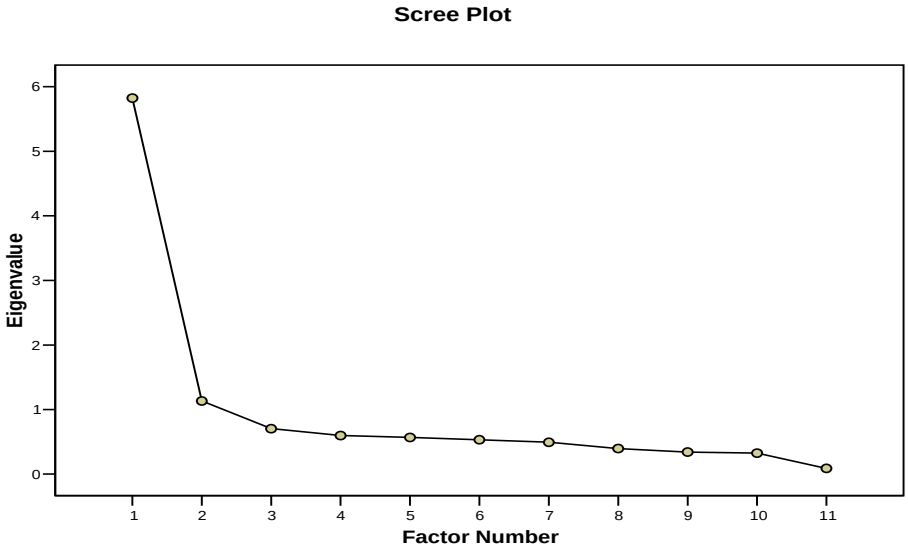
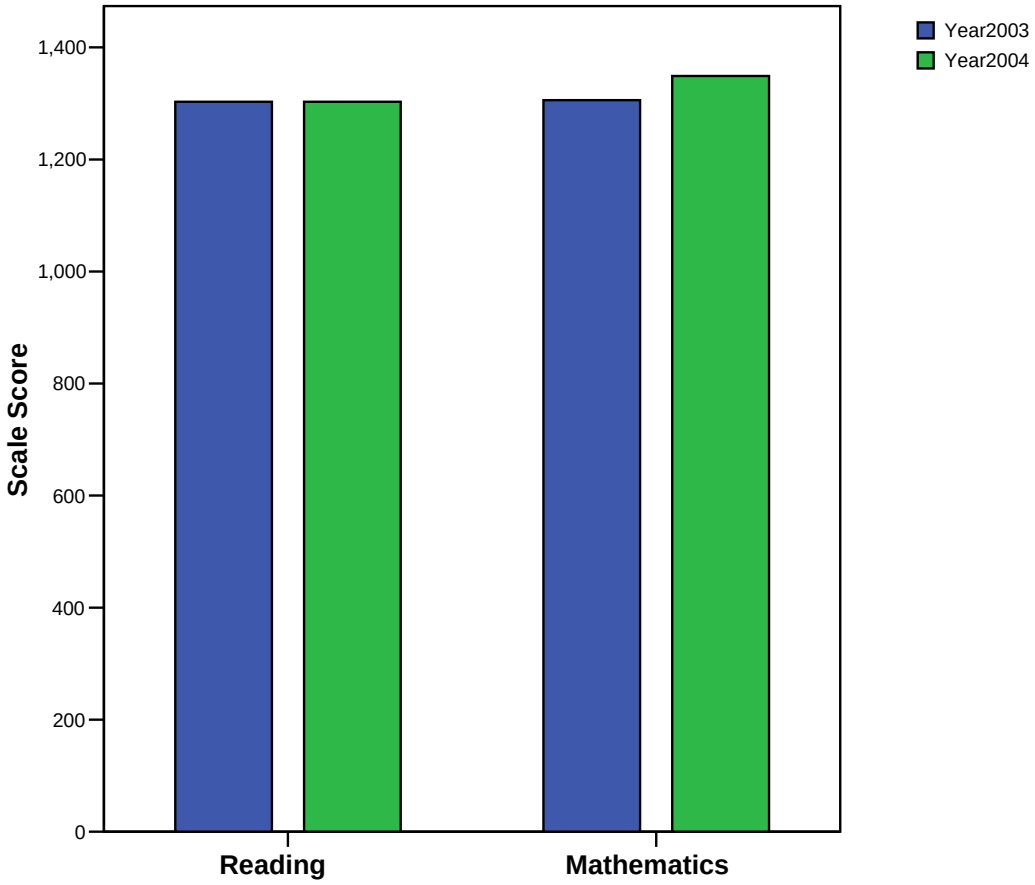


Figure 9
State Mean Scale Score



References

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. (1991). PARDUX [Computer program]. Unpublished.

Fleiss, J.L., Cohen, J., and Everitt, B.S. (1969). Large-Sample Standard Errors of Kappa and Weighted Kappa. *Psychological Bulletin, 72,* 323 –327.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics, *33,* 159–174.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14,* 59–71.

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

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

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 2,* 125–145.

Yen, W. M. (1991a). *How to evaluate FIT for Maryland*. Unpublished manuscript.

Zwick, R., Donoghue, J.R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30,* 233–251.

Appendix

Reading and Mathematics Academic Standards

The Pennsylvania Grade 3 Reading Test measures the following Pennsylvania Academic Standards for Reading:

- 1.1.3, Learning to Read Independently
- 1.2.3, Reading Critically in All Content Areas
- 1.3.3, Reading, Analyzing and Interpreting Literature.

The Pennsylvania Grade 3 Mathematics Test measures the following Pennsylvania Academic Standards for Mathematics:

- 2.1.3, Numbers, Number Systems and Number Relationships
- 2.2.3, Computation and Estimation
- 2.3.3, Measurement and Estimation
- 2.4.3, Mathematical Reasoning and Connections
- 2.5.3, Mathematical Problem Solving and Communication
- 2.6.3, Statistics and Data analysis
- 2.7.3, Probability and Predictions
- 2.8.3, Algebra and Predictions
- 2.9.3, Geometry
- 2.10.3, Trigonometry
- 2.11.3, Concepts of Calculus