

Pennsylvania Grade 3 Assessment

Mathematics and Reading

Technical Report Spring 2006 Operational Test

Submitted to the Pennsylvania Department of Education

August 2006



Table of Contents

Part 1: Overview	4
Part 2: Test Design and Sample	5
2.1 Test Structure	5
2.2 Test Sample	5
Part 3: Test Development Process	7
3.1 Development of the 2006 PSSA-Grade 3 Assessment.....	7
3.1.1 The Linking Plan and Item Selection	8
Part 4: Item Analysis.....	11
4.1 Classical Item Analysis	11
4.2 Speededness	12
Part 5: IRT Calibration and Equating	13
5.1 Test Fairness for Using Matrix Items.....	14
5.2 Scoring Tables for Raw Score to Scale Score	15
Part 6: Test Results	16
6.1 Summary Statistics for Raw Scores	16
6.2 Summary Statistics for Scale Scores	17
6.3 Percents at Each Performance Level	18
Part 7: Reliability and Validity	20
7.1 Reliability	20
7.1.1 Internal Consistency	20
7.1.2 Inter-Rater Agreement	21
7.2 Validity.....	22
7.2.1 Content Validity.....	23
7.2.2 Construct Validity.....	24
7.2.3 Differential Item Functioning (DIF)	25
7.2.4 Item Fit Assessment.....	26
7.2.5 OE Item Scoring	27
References.....	98
Appendix: Reading and Mathematics Academic Standards	99

Table of Tables

Table 1 2005 Test Design	28
Table 2 Number of Items and Score Points by Item.....	29
Table 3 2005 Pennsylvania Grade 3 Sample Characteristics by Ethnicity	31
Table 4 2005 Pennsylvania Grade 3 Sample Characteristics by Gender.....	32
Table 5 2005 Pennsylvania Grade 3 Sample Characteristics by Disability	33
Table 6 2005 Pennsylvania Grade 3 Sample Characteristics by English Language Learner (ELL)	34
Table 7 2005 Pennsylvania Grade 3 Sample Characteristics by Economically Disadvantaged.....	35
Table 8 2005 Pennsylvania Grade 3 Sample Characteristics by Migrant Status.....	36
Table 9 Item Statistics for Reading Form A (N=31,455)*.....	37
Table 10 Item Statistics for Reading Form B (N=23,957)*.....	38
Table 11 Item Statistics for Reading Form C (N=23,785)*	39
Table 12 Item Statistics for Reading Form D (N=23,577)*	40
Table 13 Item Statistics for Reading Form E (N=23,577)*.....	41
Table 14 Item Statistics for Mathematics Common items (N=126,631).....	42
Table 15 Item Statistics for Mathematics Matrix items	43
Table 16 Scoring Table for Reading Form A.....	45
Table 17 Scoring Table for Reading Form B.....	46
Table 18 Scoring Table for Reading Form C	47
Table 19 Scoring Table for Reading Form D	48
Table 20 Scoring Table for Reading Form E.....	49
Table 21 Scoring Table for Mathematics	50
Table 22 Raw score Descriptive Statistics Based on All Samples	51
Table 23 Raw Score Descriptive Statistics by Ethnicity	52
Table 24 Raw Score Descriptive Statistics by Gender.....	54
Table 25 Raw Score Descriptive Statistics by Disability Status	55
Table 26 Raw Score Descriptive Statistics by ELP Status.....	56
Table 27 Raw Score Descriptive Statistics by Economic Disadvantage Status	57
Table 28 Raw Score Descriptive Statistics by Migrant Status.....	58
Table 29 Scale Score Descriptive Statistics Based on All Samples	59
Table 30 Scale Score Descriptive Statistics by Ethnicity.....	60
Table 31 Scale Score Descriptive Statistics by Gender	62
Table 32 Scale Score Descriptive Statistics by ELP Status	63
Table 33 Scale Score Descriptive Statistics by Disability Status.....	64
Table 34 Scale Score Descriptive Statistics by Economic Disadvantage Status	65
Table 35 Scale Score Descriptive Statistics by Migrant Status	66
Table 36 Percentiles of Scale Score Ranges.....	67
Table 37 Percent at Each of Reading Performance Level	70
Table 38 Percent at Each Mathematics Performance Level.....	71
Table 39 Reliability and SEM for Reading and Mathematics.....	72
Table 40 Rater Agreement for Reading Constructed-Response Items.....	73
Table 41 Rater Agreement for Mathematics Constructed-Response Items.....	73
Table 42 Number of Items per Each Reading Assessment Anchor *	74
Table 43 Number of Items per each Mathematics Assessment Anchor*.....	75
Table 44 Summary Statistics for Reading Reporting Categories/Assessment Anchors	76
Table 45 Summary Statistics for Mathematics Reporting Categories/Assessment Anchors (N = 126,880)*	77
Table 46 Correlation Between Mathematics Reporting Categories	78
Table 47 Correlation Between Reading Reporting Categories	78
Table 48 Correlation Between Math Assessment Anchors	79
Table 49 Correlation Between Reading Assessment Anchors	80
Table 50 Factor Analysis.....	81
Table 51 Summary for Differential Item Functioning based on Criteria $\pm C$	82

Table of Figures

Figure 1 <i>Reading B parameters with all anchor items</i>	85
Figure 2 <i>Reading B parameters of anchor items after dropping item #17</i>	86
Figure 3 <i>Mathematics B parameters with all anchor items</i>	87
Figure 4 <i>Mathematics B parameters of anchor items after dropping item # 59</i>	88
Figure 5 <i>Mathematics Scale Score and SEM</i>	89
Figure 6 <i>Reading Scale Score and SEM</i>	90
Figure 7 <i>Reading Form 1 Raw Score and Scale Score Distribution</i>	91
Figure 8 <i>Reading Form 2 Raw Score and Scale Score Distribution</i>	92
Figure 9 <i>Reading Form 3 Raw Score and Scale Score Distribution</i>	93
Figure 10 <i>Reading Form 4 Raw Score and Scale Score Distribution</i>	94
Figure 11 <i>Reading Form 5 Raw Score and Scale Score Distribution</i>	95
Figure 12 <i>Mathematics Raw Score and Scale Score Distribution</i>	96
Figure 13 <i>State Mean Scale Scores</i>	97

Part 1: Overview

In April 2006, approximately 130,000 Pennsylvania Grade 3 students participated in the 2006 spring administration of the Pennsylvania Grade 3 Reading and Mathematics Assessments. This report provides technical information about the assessments, including an overview of the operational test design, a 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, a discussion of the procedures used for calibrating and equating forms, and a summary of scale scores.

Part 2: Test Design and Sample

2.1 Test Structure

The Pennsylvania Grade 3 Reading and Mathematics Assessments are part of the Pennsylvania System of School Assessment (PSSA) currently being administered in Pennsylvania. Last year, in an effort to improve the way its standard-based assessment fosters increases in student learning, Pennsylvania implemented the use of Assessment Anchors to replace the previously used content standards. Assessment Anchors are subcategories of broader Reporting Categories, and are designed to focus the PSSA test on specific and meaningful knowledge and skills. We refer tests which score students with reference to specific knowledge and skills as criterion-referenced tests.

The criterion-referenced Reading assessment includes two reporting categories and five assessment anchors. The criterion-referenced Mathematics assessment includes five reporting categories and eleven assessment anchors (See Tables 42 and 43).

Reading and Mathematics were administered together in one test book. Sixteen books were spiraled within classrooms. Reading had five unique forms which were repeated across 16 test books. Form 1 occurred four times; the other forms occurred 3 times each. Mathematics had 10 unique forms, forms A to F repeated (see Table 1 for the test design).

Both Reading and Mathematics tests consist of common items, which were taken by every student and were common across all forms, as well as the matrix and embedded field-tested (FT) items, which are unique to each form. For Reading, both common and matrix items contribute to students' scores. Mathematics matrix items do not contribute to students' reported scores, but do contribute to aggregated scores used for curriculum analysis. Table 2 shows the number of items and score points for each. Both Reading and Mathematics tests consists of multiple-choice (MC) items and open-ended (OE) items. Note that in Reading, OE items are scored using a three point scoring rubric and for Mathematics, OE items are scored using a four point scoring rubric. The maximum number-correct score for Reading was either 45 or 46, and 61 for Mathematics.

2.2 Test Sample

Table 3 shows the ethnic characteristics of the examinees, split by form. As the table indicates, most of the students were White (74%), 16% were African American, 7% of the students were Hispanic, 3% were Asian, and less than one percent were American Indian. As expected, these ratios were similar across all test forms. As shown in Table 4, slightly more male students (51%) than female students (49%) were in the tested population. Table 5 shows Disability status. As indicated in the table, approximately 15% of tested students were Disabled. Table 6 shows the English Language Learner (ELL) population. As indicated there, between 3 and 4 percent of the tested population had an ELL status. Table 7 indicates the size of the Economically Disadvantaged population. As shown in the table, approximately 37% of students

were Economically Disadvantaged. Table 8 shows Migrant status. As indicated there, less than one half of one percent of students were Migrants.

Part 3: Test Development Process

3.1 Development of the 2006 PSSA-Grade 3 Assessment

The following is a description of the process followed in the development of the 2006 PSSA-Grade 3 Operational Assessment:

- ❑ CTB and PDE (Pennsylvania Department of Education) held a joint review of recent revisions to the Pennsylvania Assessment Anchors in Mathematics and Reading. Alignment issues of depth and breadth, as well as interpretation of specific statements, were discussed in detail, resulting in a shared understanding of the revisions.
- ❑ CTB and PDE jointly reviewed and revised the test blueprints for the two content areas.
- ❑ CTB and PDE jointly reviewed and revised the scoring rubric formats for open-ended (OE) items in Mathematics and Reading.
- ❑ CTB selected reading passages for consideration by PDE. PDE and selected Pennsylvania educators reviewed the passages for content and bias/sensitivity issues, resulting in a final list of passages for which to write test items.
- ❑ Reading items and Mathematics items (both MC items and OE items with rubrics) were written, content edited, and style edited by CTB. The Reading items, accompanied by their passages, and the Mathematics items were sent to PDE for an initial review. Revisions requested by PDE were incorporated into the items by CTB.
- ❑ The Reading items and Mathematics items were reviewed by Pennsylvania educators at a content review held in Pennsylvania. A bias committee also reviewed the items at that time. Any items rejected at either of the reviews were deleted from further consideration. CTB subsequently revised items passing the reviews for which revisions had been requested. Items that passed the reviews with no requests for revisions remained as they had been presented to the committees.
- ❑ CTB selected Reading items and Mathematics items for field testing; PDE reviewed and verified the selections.
- ❑ CTB produced test books in which the Reading and Mathematics field test items were embedded in the test forms. PDE reviewed and approved the test books at second pages.
- ❑ Pennsylvania students participated in the Reading and Mathematics field tests; MC and OE items were scored; score results and individual item statistics were relayed to the CTB development team.
- ❑ CTB selected operational Reading passage/item sets (subsequently referred to as passage sets) and Mathematics items from the field test and from previous Matrix and Common operational tests, using individual item data, the test blueprint, the PDE-approved linking plan, passage and item content, and test characteristic curves to guide the selection (see below).
- ❑ PDE reviewed and approved the operational test item selections for both content areas.
- ❑ CTB produced first and second pages of the 2006 operational test book forms; PDE reviewed and approved the second pages. CTB also produced the Teacher's Administration Manual and the Assessment Coordinator's Handbook, which was reviewed and approved by PDE.

- ❑ CTB produced final pages of the test book forms and the manuals; the test books and manuals were printed and shipped to Pennsylvania school districts.
- ❑ CTB's Braille publisher reviewed Form A of the 2006 operational test for Braille issues and made recommendations to CTB regarding the accessibility of each item for Braille students. PDE reviewed the Braille publisher's recommendations and approved the final selection of items for the Braille edition of the test. The Braille publisher transcribed the items and produced the Braille edition test books. The books were shipped to Pennsylvania school districts.

3.1.1 The Linking Plan and Item Selection

When 2006 test was constructed, selection of anchor items for 2005 and 2006, and anchor items for 2006 and 2007 were discussed with PDE.

Mathematics

PDE's requirements for selection of 2005-to-2006 Mathematics **linking** items were adhered to:

- ❑ Ten MC items from the Common set were selected.
- ❑ Fifteen MC and three OE items from the Matrix set were selected. The fifteen MC Matrix items were independently selected, in sets of five consecutive items, from three different forms. The three OE Matrix items were independently selected, with each item from a different form.
- ❑ The five consecutive Matrix items from a form were chosen from the beginning of the session, and were subsequently placed in the same place in the 2006 test, in order to maintain consistency and avoid context effects.
- ❑ The 2005 items selected for linking appeared in the 2006 assessment in the same location (Common vs. Matrix) and order as in 2005.

The proportions of score points in each Reporting Category across all unique Common and Matrix items were also considered when selecting items for linking. The goal in selecting linking items was to choose Common items and sets of Matrix items that yielded similar percentages across reporting categories so that a representative sample of the curricular content would be used for linking.

For **non-linking** Mathematics items, 2006 Common items were selected from among 2003, 2004, and 2005 Matrix items. These 2006 non-linking Common items may be considered as potential items for linking 2006 and 2007. The 2006 Matrix items were selected from the 2003, 2004, and 2005 field test items or Matrix items. For 2006 Matrix items, 2005 field test items had priority over other items. Both Common items and Matrix items were selected so as to fulfill the requirements of the test blueprint for each Reporting Category. Items were also selected so that the test would contain a representative sampling of the next lower curricular

level, the Assessment Anchor. DRC (Data Recognition Corporation) selected items for the field test portion.

Reading

PDE's requirements for selection of 2005-to-2006 Reading **linking** items were adhered to:

- ❑ Three linking passage sets were used to make a 2:1 ratio of literary passages to information passages. (CTB used four linking passage sets; after calibration, one passage set was dropped.)
- ❑ An approximately 75:25 ratio of Comprehension items to Skills for Interpretation/Analysis items was used in the linking passage sets.
- ❑ Repetition of the same Common passage sets for several years in a row was avoided as much as possible when considering passage sets for linking.

The 2006 test forms, both **linking** and **non-linking** passage sets, were made similar to 2005 forms with respect to test form difficulty and test blueprint requirements. Four passages—S, W1, W2, and W4—needed to be selected for 2006. (See the table below.)

Reading Passage Set Linking Plan

2005					2006					2007				
Form1	Form2	Form3	Form4	Form5	Form1	Form2	Form3	Form4	Form5	Form1	Form2	Form3	Form4	Form5
O	O	O	O	O	S	S	S	S	S	?	?	?	?	?
U4	U4	U4	U4	U4	U4	U4	U4	U4	U4	?	?	?	?	?
P	P	P	P	P	Y1	Y1	Y1	Y1	Y1	Y1	Y1	Y1	Y1	Y1
U1	*U2	U3	U1	*U2	Y2	Y3	Y4	Y5	*U2	Y2	Y3	Y4	Y5	?
T1	X2	X3	T4	X5	W1	W2	X3	W4	X5	?	?	?	?	?
Y1	Y2	Y3	Y4	Y5	C1	C2	C3	C4	C5	?	?	?	?	?

Red: 2004-2005 Year-to-Year linking passage sets (O, T1, and T4)

Blue: 2005-2006 Year-to-Year linking passage sets (U4, X3, and X5)

Purple: potential 2006-2007 Year-to-Year linking passage sets (Y1-Y5)

Part 4: Item Analysis

4.1 Classical Item Analysis

Tables 9 to 15 present item-level descriptive statistics for each test form. The tables show item type, item p -value, an item correlation with the total test score, the omit rate, and item fit information. The p -value for a MC item represents the proportion of students who answered the item correctly. If all students answered a given MC item correctly, its p -value would be 1.0. If only 30% of students answered the question correctly, the p -value would be .30. So, the lower the p -value, the more difficult the item is. The item p -value is a good indication of difficulty; it takes student performance into account, and it makes comparing items in terms of a common statistic very simple. The p -value for an OE item represents the mean proportion of possible raw score points that students actually obtained for the item. A p -value of .33 for a given OE item would indicate that, on average, students obtained one-third of the possible points for the item. If the p -value were .75, this would indicate a much easier item, where, on average, students obtained 75% of the maximum possible points for the item. As such, for OE items as well, p -value indicates difficulty and the lower the p -value, the more difficult the item is.

For MC items, the item correlation with the total test score is a point-biserial correlation. The point-biserial correlation indicates the correlation between the item score and the total score on the test. If an item were to show a correlation of .80, this would indicate a strong relationship between the item score and a total test score. If the correlation for a given item were only .10, this would indicate that the performance on the item is weakly related to the total test score. The point-biserial correlation is only appropriate for dichotomous level data (yes/no, right/wrong), so for the OE items, a Pearson correlation between the item score and the total score on the test was computed. The Pearson correlation can be interpreted the same way: it is a correlation between the score for a given OE item and the total test score. For item analysis, the studied item was excluded from the computation of the total score so as to not artificially inflate the correlation statistic. This effect would be most noticeable for OE items worth several points.

Tables 9 to 13 display the data for Reading. Split by the five forms, the tables show five different statistics for each Reading common item. For Mathematics, the statistics for common items are in Table 14 and the statistics for the unique matrix items in the ten unique Mathematics forms can be found in Table 15.

Items were flagged for further investigation when certain thresholds were reached. The p -value was flagged when the statistic fell below 0.30 for MC items. This would indicate a difficult item, where fewer than 30% of students obtained the correct answer. The item-to-total score correlation was flagged where the coefficient was below 0.15. This would indicate a weak correlation between the likelihood of a correct answer choice and the total test score. The omit-rate was flagged when it was above 5%. This could indicate an especially difficult item, or if located near the end of the test, it could indicate a speeded test, where students did not have enough time. Note that item-level descriptive statistics were not given for the suppressed item (Mathematics common item #44).

4.2 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 9 to 15 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.

Part 5: IRT Calibration and Equating

Student item responses were calibrated using the combination of two IRT models. The one-parameter logistic (1PL) was used to scale the SR items, and the one-parameter partial credit (1PPC) model was employed to scale the OE items. The 1PL defines an SR item in terms of the item difficulty (b_i). 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):

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

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 $\gamma_{j0} = 0$, and 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 two-parameter partial credit (2PPC model). As stated, in the 1PPC model, the discrimination does not vary over items, the same discrimination parameter is applied to all test items. In the above equation for the 2PPC model, the following equation replaces A_{jk}

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

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 expected maximum (EM) algorithm (Bock and Aitkin, 1981; Thissen, 1982). Because the 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.

After the 2006 items were calibrated using PARDUX, the scale of 2006 items was transformed to the 2005 scale using anchor items. The procedure used was based on the Stocking and Lord (1983) procedure for multiple choice items, and the Stocking and Lord extension for open ended items. To check the stability of anchor items, item b parameters of 2005 anchor items and estimated item b parameters of 2006 anchor items were plotted in Figure 1. The horizontal axis represents 2005 item b parameters and the vertical axis represents 2006 item b parameters. As can be seen in the figure, most 2005 item parameters were aligned well with 2006 item parameters.

As can also be seen in Figure 1 for Reading, item #17 was much deviated from the diagonal line. This large deviation implies that this item functioned much differently in 2005 and 2006 administrations. This item was dropped from the anchor set so that this item was not used for the Stocking and Lord transformation procedure. Figure 2 shows the b parameters of Reading anchor items after dropping the item #17. Figure 3 shows the b parameters of anchor items for Mathematics. Item #59 was much deviated from the line, and dropped from the anchor set. Neither Reading item #17 nor Mathematics item #59 were used as anchor items.

5.1 Test Fairness for Using Matrix Items

Equating is one of the procedures used to help assure test fairness. Whenever alternate test forms are administered, equating is required to place scores for the different forms on the same scale. If the same test form is used repeatedly, test items may easily become known to future examinees. So, in many testing programs, multiple forms of the test are used to prevent test disclosure. Although multiple test forms are built to have similar characteristics, such as content, format, and level of difficulty, test forms cannot be exactly equivalent. For this reason, examinees who take an easier test form will have an advantage over those who take a harder test form. Equating addresses this issue by placing scores from different test forms on the same scale. Equating thus assures test fairness, which is important for both test takers and test score users.

The precondition for equating is that alternate forms should have similar characteristics, such that all forms measure the same content. If this precondition is not satisfied, no equating procedure could place scores for different test forms on the same scale. Pennsylvania Grade 3 Assessments have been built by content experts and psychometric researchers such that matrix items in alternate forms have very similar content, format, and level of difficulty. Therefore, all matrix items in alternate forms measure the same Pennsylvania Standards.

The equating process used two strong equating designs to ensure that Pennsylvania Grade 3 matrix items on alternate forms were as similar as possible. First, a common item design was used. The items which were common to all forms were also similar to the total set of items on each form. So, the common items are considered a short version of all forms. These common items provide a basis for determining the performance of each student. Based on information yielded from the common items, the performance for matrix items can be estimated, using a psychometric model.

Second, the alternate test forms were administered under a random groups design. That means each form was taken by a similar group of students. The characteristics, such as gender, ethnicity, and achievement level for the group of students who took Form A are very similar to those of students who took Form B. The similarity of random groups has been supported in the literature and proven in many testing programs. Based on the fact that random groups have very similar performance for each set of matrix items, the performance for each set of matrix items can be estimated, using a psychometric model.

5.2 Scoring Tables for Raw Score to Scale Score

The 2006 Pennsylvania score scale was transformed to the 2005 scale using anchor items and Year-to-Year equating. After transformation to the 2005 score scale, scoring tables for Reading and Mathematics were generated. Tables 16 to 21 show the scoring tables. These scale scores, and the standard errors of measurement (SEM) on the scoring tables, are plotted in Figures 5 and 6. Also, Figures 7 to 11 show the distributions of raw scores and scale scores for Reading and Mathematics by form. For Reading, all scale scores and SEMs across all five 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 6.

Part 6: Test Results

6.1 Summary Statistics for Raw Scores

Table 22 presents raw score descriptive statistics for each test form. The tables show the number of students, mean raw score, test difficulty, standard deviation, minimum score, and maximum score.

In terms of the measurements applied in the raw score table, note first that the mean raw score, or the mean number of items correct, should be understood by content area, form, and maximum score points. The distinction between common and matrix items should also be kept in mind. Test difficulty is computed as mean raw score / total score points. Test difficulty ranges from 0 to 1.0. Consider an example. If the mean number of items correct on a test were very low, such as 15.00 on a test where the maximum possible score was 65, the test difficulty would be 0.23, thereby indicating a difficult test. If the mean raw score were very high, such as 60 on the same test, test difficulty would be 0.92, thereby indicating an easier test. The smaller the test difficulty statistic is, the more difficult the test.

The maximum possible raw score for Reading was either 45 or 46, depending on the form. Mean raw scores for Reading ranged from 33.01 to 34.12. The maximum difference in mean scores by form was between forms D and B. The difference was 1.11 raw score points. Test difficulty in Reading ranged from 0.72 to 0.74. Standard deviations were stable across all forms. The minimum and maximum observed scores were also stable across forms.

For Mathematics, the maximum possible raw score was 61 for common items and 75 for common and matrix items. The mean raw score for common items was 50.58. For common and matrix items together, the mean raw score ranged from 60.24 to 62.12. There, the maximum difference in mean scores by form was between forms F and J. The difference was 1.88 score points. Test difficulty in Mathematics ranged from 0.80 to 0.83. Standard deviations were stable across all forms. The minimum and maximum observed scores were also stable across forms.

Tables 23 to 28 present the raw score mean, standard deviation, and test difficulty by NCLB subgroups. 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 performance of each subgroup on each test form.

Looking at common and matrix items in Reading, the scores of White students, as a group, were highest, followed by Asian students, American Indian students, African American students, and Hispanic students. However, the scores of White students and Asian students were very close. As a group, female students scored higher than male students in Reading. The difference in scores across gender was small. There were wider differences by ELP status. As a group, those students who were Proficient in English scored higher than English Language Learners. Migrant students scored lower, as a group, than Non-migrant students. There were differences in scores by Disability status as well. Disabled students scored lower, as a group, than Not Disabled students. Reading scores varied by economic status. Students who were Not

Economically Disadvantaged score higher, as a group, than those students who were Economically Disadvantaged.

For common items only in Mathematics, Asian students had the highest scores, followed by White students and American Indian students. Here, the scores of Asian students and White students were again very close, and the scores of American Indian students were also close to those of Asian and White students. Hispanic students scored in the fourth position, relative to other ethnicities, and, as a group, African American students had the lowest score. Male students scored higher than female students, though the difference in scores was very small. In Mathematics, as in Reading, there were wider differences by ELP status. Those students who were Proficient in English scored higher, as a group, than those students who were English Language Learners. Migrant students scored lower, as a group, than Non-migrant students. Differences in scores also existed by Disability status. Those students who were Disabled scored lower, as a group, than Not Disabled students. Scores in Mathematics varied by economic status. Those students who were Not Economically Disadvantaged scored higher in Mathematics, as a group, than those students who were Economically Disadvantaged.

6.2 Summary Statistics for Scale Scores

Table 29 presents descriptive statistics for scale scores. The 2006 the state mean for Reading was 1329. The mean and standard deviation were similar across alternative forms. In 2005, the mean was 1327, and in 2004 it was 1296. In other words, scores trended upward. The state mean for 2006 in Mathematics was 1396. The mean and standard deviation were similar across alternative forms here as well. Last year, the mean was 1365, and the year prior it was 1341. As was observed in Reading, scores trended upwards. To facilitate score interpretation, percentiles of scale scores are provided in Table 36. Means for 2004, 2005, and 2006 are plotted in Figure13.

Tables 30 and 31 show descriptive statistics for scale scores by NCLB subgroups. Results varied across NCLB subgroups. For Reading, Asian students had the highest scores, followed by White students, American Indian students, African American students, and Hispanic students. Scores also varied by gender. Female students, as a group, scored higher than male students. There was also a difference in scores by ELP status. As a group, those students who were Proficient in English scored higher than English Language Learners. Differences in scores by Migrant status were also observed. Those students who were Migrants scored lower than those students who were not. Students who were Disabled scored lower, as a group, than those students who were Not Disabled. There were differences in scores by economic status as well. Those students who were Not Economically Disadvantaged scored higher, as a group, students who were Economically Disadvantaged.

In Mathematics, for common items only, Asian students had the highest scores, followed by White students, American Indian students, Hispanic students, and African American students. Differences were also observed by gender. Male students, as a group, scored higher than female students. Scores varied by ELP status as well. Those students who were Proficient in English scored higher, as a group, than English Language Learners. Students who were Not Migrants scored higher, as a group, than students who were Migrants. Those students who were Disabled

scored lower, as a group, than students who were Not Disabled. Scores in Mathematics varied by economic status. Those students who were Not Economically Disadvantaged scored higher, as a group, than those students who were Economically Disadvantaged.

The distributions of raw scores and scale scores can be found in Figures 7 to 12. 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.

6.3 Percents at Each Performance Level

As mentioned previously, standard setting was conducted last year in order to establish cut scores for performance levels. Four performance levels were established: Below Basic, Basic, Proficient, and Advanced. Most students were either Proficient or Advanced. However, results varied across NCLB subgroups. The Advanced level of performance was most common among Asian students, White students, and American Indian students. Most Asian, White and American Indian students were either Advanced or Proficient. A relatively small proportion of White, Asian, and American Indian students were Below Basic performance. However, as Table 37 shows, most African American students were either Proficient or at the in the lowest performance category, and about equally likely to be in either. A relatively small proportion of African American students were at the Advanced level. The same was true for Hispanic students. Relatively few Hispanic students were at the Advanced level, most were either Below Basic or Proficient. Relative to other ethnicities, the lowest level of performance was much more common among Hispanic students and African American students, and the Advanced level was much less common. In terms of gender, most males and most females were either Proficient or Advanced. Females were more likely than males to be Advanced and less likely to be Below Basic. Performance varied by Disability status. As a group Disabled students did not score as highly as those students who were Not Disabled. There were also differences in performance by ELP status. While most students who were Proficient in English were either Proficient or Advanced, nearly half of English Language Learners, were Below Basic performance and the Proficient and Advanced levels were much less common than among their Proficient counterparts. Differences in performance were also observed by economic status. Students who were Not Economically Disadvantaged, were much more likely to be Advanced and much less likely to be at either the Basic or Below Basic level than those students who were Economically Disadvantaged. There were differences in performance level by Migrant status. Most Non-migrant students were either Proficient or Advanced and the Below Basic level was relatively uncommon. For Migrant students however, nearly half were Below Basic performance, about one fourth was Proficient, and the Advanced level was relatively uncommon.

Table 38 shows the percentage of students in each performance level. In Mathematics, most students were either Advanced or Proficient. Grouped and compared by ethnicity, most Asian, White, and American Indian students were Advanced. The Basic and Below Basic performance levels were not as common. Among African American and Hispanic students, performance was more evenly dispersed across all categories, including the lowest level and the Basic level. There were smaller differences in Mathematics performance by gender. Most males and females were Advanced, the Proficient level was common, and the lowest level and the

Basic level were less common. Slightly more males than females were Advanced. Those students who were Disabled did not score as well as those students who were Not Disabled. Mathematics performance varied by ELP status. Among students who were Proficient, performance was often Advanced or Proficient, and the lower levels of performance were not as common. Among English Language Learners, scores were more evenly dispersed across categories, and the below Basic and Basic levels were common. Differences in performance were observed by economic status. Those students who were Not Economically Disadvantaged were much more likely to score at the Advanced level and much less likely to score at the lowest level than Economically Disadvantaged students. Migrant students were much more likely to score at the lowest performance level, and much less likely to score at the highest level than Non-migrant students.

Part 7: Reliability and Validity

7.1 Reliability

PSSA-Grade 3 is a reliable assessment; it provides data regarding student performance that may be generalized. In order to provide evidence of reliability, a number of item analyses were conducted, and they were reported as measures of the consistency of test results. The following analyses were reported in Part 4 Item Analysis and Part 5 IRT Calibration and Equating:

- ❑ Item-level descriptive statistics were calculated, including p-values, item correlation with total test score (R-ITT), percent of omitted items, and fit information.
- ❑ Tests for speededness were conducted.
- ❑ Two IRT models—1PL, which is used to scale MC items, and 1PPC, used to scale OE items—were used, allowing both MC and OE items to be placed on the same scale.
- ❑ The standard error of measurement was calculated.

Also, Reliability of the 2006 Spring PSSA Grade 3 assessments was estimated in two ways: internal consistency was assessed for all multiple choice items and inter-rater agreement was assessed for all writing tests.

7.1.1 Internal Consistency

Cronbach's alpha is a frequently used measure of internal consistency for tests consisting of multiple choice (or open ended) items. Cronbach's alpha is computed as

$$\hat{\alpha} = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right),$$

where k = number of items, σ_x^2 = the total score variance, and σ_i^2 = the variance of item i (Crocker & Algina, 1986). Then, standard error of measurement (SEM) is defined as follows:

$$SEM = SD \sqrt{1 - \text{reliability}},$$

where SD represents standard deviation.

Table 39 shows Cronbach's alpha and standard error of measurement (SEM) for Reading and Mathematics. All Reading and Mathematics forms showed high reliability, and the reliability and the SEM for all alternative forms were similar. For Reading, the reliability ranges from 0.91 to 0.92, and SEM ranged from 2.45 to 2.61. For Mathematics, reliability ranged from 0.92 to 0.93 for common and matrix items, and reliability was 0.91 for common items only. SEM for common and matrix items ranged from 2.96 to 3.10 among the common and matrix items, and it was 2.65 for common items only.

7.1.2 Inter-Rater Agreement

Reliability for open ended items is typically examined by calculating indices of inter-rater agreement: the degree of reliability with which different human raters assign scores to student responses. All responses were read by a single rater, and approximately 5% of the common items and Reading matrix items and 10% of the Mathematics matrix items were submitted to a second rater for scoring.

Tables 40 and 41 present the rater agreement statistics for Reading and Mathematics OE items. The tables provide, first, mean scores for rater 1, then mean scores for rater 2, and the standard deviations of both. Next, the tables show modes of agreement between raters. “Perfect” agreement is defined as scores that are exactly the same. “Adjacent” agreement is defined as scores differing by one point. “Discrepant” cases are those cases where scores from two raters differed by more than one raw score point. Intraclass correlation, kappa, and weighted kappa are also reported.

As a central summarizing measure of inter-rater agreement we can say that rater agreement occurs where scores differ by no more than one score point. Defined as such, there was a very high degree of agreement: ranging from 95.29 to 99.59 for Reading, and from 95.21 to 99.76 for Mathematics. In addition, Tables 40 and 41 show that the mean score points awarded by the two raters were also very close.

In addition to using the percentage of rater agreement as a central summarizing measure of rater agreement, other measures are also supplied in this report. To further study rater agreement, Cohen’s kappa (Cohen, 1960), weighted kappa, and intraclass correlation are reported as 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. Kappa does. 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”.

Cohen’s Kappa is computed as:

$$\kappa = \frac{\sum P_{ii} - \sum P_{i\cdot} \cdot P_{\cdot i}}{1 - \sum P_{i\cdot} \cdot P_{\cdot i}},$$

where $\sum P_{ii}$ is the observed proportion of agreement and $\sum P_{i\cdot} \cdot P_{\cdot i}$ is the chance proportion of agreement (Brennan & Prediger, 1981). Kappa is commonly used to summarize the agreement between raters.

As Table 40 shows, Kappa coefficients ranged from 0.40 to 0.65 for Reading and for Mathematics from 0.58 to 0.92. Note that Kappa for Mathematics was higher than for Reading. This trend has also been often found in other large scale assessment programs. According to the criteria established by Landis and Koch, the values of Kappa for Reading constitute “good agreement,” and the values for Mathematics range from “good agreement” to “excellent agreement.”

The difference between weighted kappa and (unweighted) kappa is that weighted kappa is calculated based on the average score, while kappa is computed based on a single score. Weighted kappa, k , is used in many contexts as a measure of association in square contingency tables. Weighted kappa is computed as:

$$k = \frac{\sum_{i=0}^k \sum_{j=0}^k w_{ij} \frac{n_{ij}}{n_{++}} - \sum_{i=0}^k \sum_{j=0}^k w_{ij} \frac{n_{i+} n_{+j}}{n_{++}^2}}{1 - \sum_{i=0}^k \sum_{j=0}^k w_{ij} \frac{n_{i+} n_{+j}}{n_{++}^2}}, \text{ where } w_{ij} = 1 - \frac{(i-j)^2}{K^2}$$

If agreement is perfect, $k=1$. If agreement is what would be expected by chance, $k=0$. Always, $0 \leq k \leq 1$. For a full explanation of the formula applied here, refer to Rich Patz’s 1998 unpublished paper, “Calculating Handscoring Reliability Coefficients.” Also, a full discussion of the intraclass correlation computation can be found there as well. Weighted kappa ranged from 0.63 to 0.80 for Reading. For Mathematics, weighted kappa ranged from 0.81 to 0.96.

Intraclass correlation ρ_{IC} is defined by the percent of overall score variance accounted for by the variance of mean response score:

$$\rho_{IC} = \frac{Var_n(\bar{X}_n)}{Var_n(X_{n1}, X_{n2})} = \frac{\frac{1}{N-1} \sum_{n=1}^N (\bar{X}_n - \bar{X}_{..})^2}{\frac{1}{2(N-1)} \sum_{n=1}^N [(X_{n1} - \bar{X}_{..})^2 + (X_{n2} - \bar{X}_{..})^2]}$$

Here, score1 and score2 are X_{n1} and X_{n2} . If agreement is perfect $\rho_{IC} = 1$. Always, $0 \leq \rho_{IC} \leq 1$.

As Table 40 shows, for Reading the intraclass correlation ranged from 0.81 to 0.90. For Mathematics, the intraclass correlation ranged from 0.90 to 0.98. Note that the intraclass for Mathematics was higher than for Reading. This trend has been often found in other large scale assessment programs.

7.2 Validity

The Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) defines validity as “the degree to which evidence and theory

support the interpretations of test scores entailed by proposed users of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests.” The purpose of test score validation is not to validate the test itself, but to validate interpretations of the test scores for particular purposes or uses. Test score validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization and continuing throughout the entire assessment process. Every aspect of an assessment provides evidence in support of its validity (or evidence to the contrary), including design, content specifications, item development, psychometric quality, and inferences made from the results. The 2006 Spring PSSA Grade 3 tests were designed and developed to provide fair and accurate ability scores that support appropriate, meaningful, and useful educational decisions.

In addition to the evidence provided in Part 2 (Test Design), Part 3 (Test Development Process), Part 4 (Item Analysis), Part 5 (IRT Calibration and Equating), and Reliability in Part 7 (Reliability and Validity) additional evidence to support the validity of the 2006 PSSA Grade 3 Assessments is provided by the following:

- Content Validity
- Two types of evidence for construct validity were produced. First, correlations between subscale scores (such as reporting category and assessment anchors), were estimated. Second, factor analysis was conducted using students’ responses for operational items.
- Identification of any items that displayed differential item functioning for subgroups of ethnicity and gender.
- Identification of any items that displayed item fit considerations after item calibration
- Scoring of OE items

7.2.1 Content Validity

The PSSA Grade 3 is a valid assessment; it measures what it purports to measure, namely Pennsylvania student achievement. Appropriate, meaningful, and useful inferences may be made from the test results. PSSA Grade 3 assessments were constructed through the following process.

Test Content

- ☐ PSSA Grade 3 measures both knowledge and cognitive processes that were determined to be aligned to the Pennsylvania Assessment Anchors. Each test item’s content alignment was reviewed by CTB/McGraw-Hill, as well as reviewed and verified by groups of Pennsylvania educators in a formal review process.
- ☐ PSSA Grade 3 adequately sampled the knowledge domain as defined by the Pennsylvania Assessment Anchors. The sampling of the domain was articulated in the test blueprints for each content area, which were reviewed and approved by Pennsylvania educators.
- ☐ Items were reviewed for grade-level appropriateness by CTB/McGraw-Hill and also by groups of Pennsylvania educators in a formal review process; the items were verified to be grade-level appropriate in terms of both knowledge and cognitive processes.

- ❑ Content alignment between all operational forms of PSSA Grade 3 in each content area was reviewed by CTB/McGraw-Hill, as well as reviewed and verified by Pennsylvania educators in a formal review process.
- ❑ Bias review committees verified that the test items are free of bias, including ethnicity, gender, religion, age, disability, and socioeconomic factors. The bias review process helped to ensure that the test results would be a measure of what the student knows and is able to do rather than a measure of irrelevant factors such as demographics.

Test Construction Process

- ❑ Test items were field tested to ensure that invalid test items would be eliminated before operational tests were constructed. Both individual item analysis and DIF analysis were utilized from field test data for the item selection process.
- ❑ Field test data was also used to construct test characteristic curves in order to ensure that the different forms of the operational test were comparable.
- ❑ Each year different operational forms of the PSSA-Grade 3 are administered. Longitudinal comparability of results was achieved by using test characteristic curves from previous years to guide new operational test construction.

Other Considerations

- ❑ During the development of the PSSA-Grade 3, Pennsylvania adhered to universal design practices to ensure that the assessment was accessible to all students, including students with disabilities and limited English proficiency.
- ❑ A Braille edition of the test was developed and reviewed by Braille experts for technical braille issues as well as accessibility issues.
- ❑ Two IRT models—1PL, which is used to scale MC items, and 1PPC, used to scale OE items—were used, allowing both MC and OE items to be placed on the same scale.

7.2.2 Construct Validity

Construct validity indicates how well tests measure the skills or constructs they intend to measure, and it is the central concept underlying the PSSA Grade 3 assessment validation process. Achievement tests are typically designed to measure student proficiency on a single continuum (or unidimensional construct). Although a well-designed achievement test might encompass several sub-content areas, the test as a whole should coherently assess a single construct, e.g., Mathematics achievement. To establish meaningfulness of a test form for a given content, the test should have appropriate correlation coefficients within Reporting Categories. If the correlation coefficient is very high between two Reporting Categories, it indicates that the two Categories measure the same trait, while low correlation coefficients indicate two Categories measure traits which are a little different.

Reading consists of two Reporting Categories (RCs) which contain five Assessment Anchors (AAs). Mathematics consists of five RCs which contain 11 AAs. Tables 42 and 43 show the number of items in each AA for Reading and Mathematics. Tables 44 and 45 give the raw score mean, standard deviation, and test difficulty for each RC and AA. Note that raw scores

are reported for both RCs and AAs. Raw scores for Mathematics were computed across all forms because only common items are used for individual student reports. Note that for both Reading and Mathematics, the mean p-values vary across RCs and AAs.

Tables 46 to 47 show correlations among Reporting Categories. In general, the size of the correlation coefficient is influenced by the length of the test, the number of items, or score points. Correlations between assessment anchors are presented in Tables 48 to 49. Where there were less than three items, statistics were not reported.

Factor analysis is a statistical technique commonly used to identify the latent constructs underlying test items. For a test to be scalable and adequately analyzed using a unidimensional Item Response Theory (IRT) model, like what is used for PSSA, the test should be essentially unidimensional. Factor analysis was conducted to examine the structure of both the Reporting Categories and Assessment Anchors for 2006 Pennsylvania Grade 3 Reading and Mathematics.

Table 50 displays the factor analysis results. Previous research shows that the examination of first two Eigenvalues can be useful in determining the existence of a dominant factor. The results indicate the presence of a single construct underlying the test.

In Reading the ratios of the first two Eigenvalues of Reporting Categories range from 12.73 to 16.26. That is, the variance of the first factor is approximately 13 to 16 times larger than the variance of the second largest factor. In Mathematics, the ratio was smaller, at 11.77. Within the context of the strength of the IRT as a unidimensional model, in general, these ratios can be understood as indicating that the content assessments in the PSSA Grade 3 assessments are sufficiently unidimensional. In general, the first factor accounting for over 90% of the estimated total common variance strongly suggests the presence of a single dominant factor underlying test items. For both Reading and Mathematics, as indicated in the percentage column, the first factor accounting for estimated total common variance is over 93%.

7.2.3 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. DIF was conducted for both ethnicity and gender. DIF was not conducted for the American Indian population as the population size was 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 Zwirk, et al., (1993).

Table 51 presents a 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 the content perspective on those items.

7.2.4 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 9 to 15 present items that were flagged for poor fit for each test form. In the tables, the number “3” represents poor fit. Many items displayed poor fit because the one-parameter (1PL)/one-parameter partial credit (1PPC) approach (See IRT calibration and equating section) was used to produce Z statistics and the 1PL model does not consider the guessing factor. The flagging of an item does not require that the item not be used. This item fit is just one of the criteria for selecting sound operational items.

7.2.5 OE Item Scoring

PSSA-Grade 3 OE items were scored by extensively-trained individuals at a CTB scoring center. Standardized scoring procedures were utilized throughout the scoring process. Before scoring commenced, groups of Pennsylvania educators reviewed, revised, and verified the rubrics in a formal review process.

Table 1
2005 Test Design

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

Table 2
Number of Items and Score Points by Item

Content	Form		Common Items			Matrix Items		
			Total	MC	OE	Total	MC	OE
Reading	A	Number of Items	25	24	1	17	16	1
		Score Points	27	24	3	19	16	3
	B	Number of Items	25	24	1	17	16	1
		Score Points	27	24	3	19	16	3
	C	Number of Items	25	24	1	17	16	1
		Score Points	27	24	3	19	16	3
	D	Number of Items	25	24	1	17	16	1
		Score Points	27	24	3	19	16	3
	E	Number of Items	25	24	1	16	15	1
		Score Points	27	24	3	18	15	3
Mathematics	A	Number of Items	55	53	2	11	10	1
		Score Points	61	53	8	14	10	4
	B	Number of Items	55	53	2	11	10	1
		Score Points	61	53	8	14	10	4
	C	Number of Items	55	53	2	11	10	1
		Score Points	61	53	8	14	10	4
	D	Number of Items	55	53	2	11	10	1
		Score Points	61	53	8	14	10	4
	E	Number of Items	55	53	2	11	10	1
		Score Points	61	53	8	14	10	4
	F	Number of Items	55	53	2	11	10	1
		Score Points	61	53	8	14	10	4
	G	Number of Items	55	53	2	11	10	1
		Score Points	61	53	8	14	10	4

Table 2 Cont'd
Number of Items and Score Points by Item

Content	Form		Common Items			Matrix Items		
			Total	MC	OE	Total	MC	OE
Mathematics	H	Number of Items	55	53	2	11	10	1
		Score Points	61	53	8	14	10	4
	I	Number of Items	55	53	2	11	10	1
		Score Points	61	53	8	14	10	4
	J	Number of Items	55	53	2	11	10	1
		Score Points	61	53	8	14	10	4

Table 3
2005 Pennsylvania Grade 3 Sample Characteristics by Ethnicity

Content	Form	Number of Students *	Caucasian		African American		Hispanic		Asian		Native American	
			Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Reading	A	31,136	22,987	73.83	4,921	15.80	2,215	7.11	917	2.95	96	0.31
	B	23,772	17,527	73.73	3,784	15.92	1,680	7.07	713	3.00	68	0.29
	C	23,558	17,509	74.32	3,657	15.52	1,659	7.04	667	2.83	66	0.28
	D	23,358	17,259	73.89	3,694	15.81	1,654	7.08	666	2.85	85	0.36
	E	23,057	17,078	74.07	3,656	15.86	1,628	7.06	624	2.71	71	0.31
	Total	124,881	92,360	73.96	19,712	15.78	8,836	7.08	3,587	2.87	386	0.31
Mathematics	A	15,985	11,848	74.12	2,486	15.55	1,124	7.03	476	2.98	51	0.32
	B	15,741	11,569	73.50	2,524	16.03	1,107	7.03	494	3.14	47	0.30
	C	15,628	11,605	74.26	2,461	15.75	1,088	6.96	434	2.78	40	0.26
	D	15,442	11,367	73.61	2,474	16.02	1,111	7.19	435	2.82	55	0.36
	E	15,283	11,287	73.85	2,446	16.00	1,084	7.09	421	2.75	45	0.29
	F	15,151	11,139	73.52	2,435	16.07	1,091	7.20	441	2.91	45	0.30
	G	8,031	5,958	74.19	1,260	15.69	573	7.13	219	2.73	21	0.26
	H	7,930	5,904	74.45	1,196	15.08	571	7.20	233	2.94	26	0.33
	I	7,916	5,892	74.43	1,220	15.41	543	6.86	231	2.92	30	0.38
	J	7,774	5,791	74.49	1,210	15.56	544	7.00	203	2.61	26	0.33
	Total	124,881	92,360	73.96	19,712	15.78	8,836	7.08	3,587	2.87	386	0.31

*Students of unspecified ethnicity are not counted.

Table 4
2005 Pennsylvania Grade 3 Sample Characteristics by Gender

Content	Form	Number of Students *	Male		Female	
			Count	Percent	Count	Percent
Reading	A	31,479	15,363	48.80	16,116	51.20
	B	24,003	11,902	49.59	12,101	50.41
	C	23,832	11,694	49.07	12,138	50.93
	D	23,647	11,595	49.03	12,052	50.97
	E	23,337	11,399	48.85	11,938	51.15
	Total	126,298	61,953	49.05	64,345	50.95
Mathematics	A	16,161	7,854	48.60	8,307	51.40
	B	15,889	7,869	49.52	8,020	50.48
	C	15,814	7,745	48.98	8,069	51.02
	D	15,633	7,673	49.08	7,960	50.92
	E	15,469	7,544	48.77	7,925	51.23
	F	15,318	7,509	49.02	7,809	50.98
	G	8,114	4,033	49.70	4,081	50.30
	H	8,018	3,949	49.25	4,069	50.75
	I	8,014	3,922	48.94	4,092	51.06
	J	7,868	3,855	49.00	4,013	51.00
	Total	126,298	61,953	49.05	64,345	50.95

* Students of unspecified gender are not counted.

Table 5
2005 Pennsylvania Grade 3 Sample Characteristics by Disability

Content	Form	Number of Students*	No		Yes	
			Count	Percent	Count	Percent
Reading	A	31,895	27,274	85.51	4,621	14.49
	B	24,286	20,937	86.21	3,349	13.79
	C	24,103	20,804	86.31	3,299	13.69
	D	23,900	20,658	86.44	3,242	13.56
	E	23,623	20,307	85.96	3,316	14.04
	Total	127,807	109,980	86.05	17,827	13.95
Mathematics	A	16,372	13,854	84.62	2,518	15.38
	B	16,073	13,799	85.85	2,274	14.15
	C	15,996	13,762	86.03	2,234	13.97
	D	15,802	13,633	86.27	2,169	13.73
	E	15,655	13,448	85.90	2,207	14.10
	F	15,523	13,420	86.45	2,103	13.55
	G	8,213	7,138	86.91	1,075	13.09
	H	8,107	7,042	86.86	1,065	13.14
	I	8,098	7,025	86.75	1,073	13.25
	J	7,968	6,859	86.08	1,109	13.92
	Total	127,807	109,980	86.05	17,827	13.95

*"Disabled" refers to students with any of the following disabilities: autism, deaf-blindness, deafness, emotional disturbance, hearing impairment, mental retardation, multiple disabilities, orthopedic impairment, other health impairment, specific learning disability, speech or language impairment, traumatic brain injury, visual impairment including blindness.

Table 6
2005 Pennsylvania Grade 3 Sample Characteristics by English Language Learner (ELL)

Content	Form	Number of Students*	No		Yes	
			Count	Percent	Count	Percent
Reading	A	31,895	30,706	96.27	1,189	3.73
	B	24,286	23,357	96.17	929	3.83
	C	24,103	23,273	96.56	830	3.44
	D	23,900	23,021	96.32	879	3.68
	E	23,623	22,813	96.57	810	3.43
	Total	127,807	123,170	96.37	4,637	3.63
Mathematics	A	16,372	15,766	96.30	606	3.70
	B	16,073	15,462	96.20	611	3.80
	C	15,996	15,459	96.64	537	3.36
	D	15,802	15,214	96.28	588	3.72
	E	15,655	15,110	96.52	545	3.48
	F	15,523	14,940	96.24	583	3.76
	G	8,213	7,895	96.13	318	3.87
	H	8,107	7,814	96.39	293	3.61
	I	8,098	7,807	96.41	291	3.59
	J	7,968	7,703	96.67	265	3.33
	Total	127,807	123,170	96.37	4,637	3.63

*"English Language Learners" includes students who are in their first year of enrollment as those NOT in their first year of enrollment.

Table 7
2005 Pennsylvania Grade 3 Sample Characteristics by Economically Disadvantaged

Content	Form	Number of Students	No		Yes	
			Count	Percent	Count	Percent
Reading	A	31,895	20,084	62.97	11,811	37.03
	B	24,286	15,240	62.75	9,046	37.25
	C	24,103	15,114	62.71	8,989	37.29
	D	23,900	15,023	62.86	8,877	37.14
	E	23,623	14,828	62.77	8,795	37.23
	Total	127,807	80,289	62.82	47,518	37.18
Mathematics	A	16,372	10,244	62.57	6,128	37.43
	B	16,073	10,074	62.68	5,999	37.32
	C	15,996	9,994	62.48	6,002	37.52
	D	15,802	9,917	62.76	5,885	37.24
	E	15,655	9,774	62.43	5,881	37.57
	F	15,523	9,840	63.39	5,683	36.61
	G	8,213	5,166	62.90	3,047	37.10
	H	8,107	5,120	63.16	2,987	36.84
	I	8,098	5,106	63.05	2,992	36.95
	J	7,968	5,054	63.43	2,914	36.57
	Total	127,807	80,289	62.82	47,518	37.18

Table 8
2005 Pennsylvania Grade 3 Sample Characteristics by Migrant Status

Content	Form	Number of Students *	No		Yes	
			Count	Percent	Count	Percent
Reading	A	31,895	31,774	99.62	121	0.38
	B	24,286	24,208	99.68	78	0.32
	C	24,103	24,028	99.69	75	0.31
	D	23,900	23,816	99.65	84	0.35
	E	23,623	23,552	99.70	71	0.30
	Total	127,807	127,378	99.66	429	0.34
Mathematics	A	16,372	16,313	99.64	59	0.36
	B	16,073	16,015	99.64	58	0.36
	C	15,996	15,940	99.65	56	0.35
	D	15,802	15,746	99.65	56	0.35
	E	15,655	15,607	99.69	48	0.31
	F	15,523	15,461	99.60	62	0.40
	G	8,213	8,193	99.76	20	0.24
	H	8,107	8,088	99.77	19	0.23
	I	8,098	8,070	99.65	28	0.35
	J	7,968	7,945	99.71	23	0.29
	Total	127,807	127,378	99.66	429	0.34

*"Migrant" includes all students who are migrants at the school, district, and/or state level (i.e. they initially enrolled in the school, district, or state of residence after October 1, 2004).

Table 9
Item Statistics for Reading Form A (N=31,455)*

Item	Type	P-Val	Corr	Omit	FIT	Item	Type	P-Val	Corr	Omit	FIT
1	MC	0.82	0.49	0.08%	1	22	MC	0.92	0.56	0.11%	3
2	MC	0.81	0.41	0.17%	1	23	MC	0.87	0.45	0.53%	1
3	MC	0.93	0.49	0.08%	3	24	MC	0.89	0.52	0.52%	3
4	MC	0.69	0.34	0.12%	3	25	MC	0.61	0.38	0.88%	3
5	MC	0.76	0.50	0.18%	1	26	MC	0.66	0.41	0.11%	1
6	MC	0.82	0.48	0.15%	1	27	MC	0.64	0.45	0.32%	1
7	MC	0.56	0.36	0.26%	3	28	MC	0.63	0.47	0.22%	1
8	MC	0.84	0.53	0.63%	3	29	MC	0.79	0.47	0.65%	1
9	MC	0.76	0.53	0.15%	1	30	MC	0.51	0.40	0.12%	1
10	MC	0.84	0.58	0.21%	3	31	MC	0.54	0.39	0.36%	3
11	MC	0.89	0.51	1.51%	3	32	MC	0.47	0.29	0.23%	3
12	MC	0.78	0.45	0.13%	1	33	MC	0.79	0.44	1.49%	1
13	MC	0.88	0.54	0.24%	3	34	CR	0.52	0.49	0.62%	3
14	MC	0.74	0.58	0.14%	3	35	MC	0.88	0.51	0.10%	3
15	MC	0.34	0.27	0.28%	3	36	MC	0.82	0.48	0.48%	1
16	MC	0.79	0.45	0.34%	1	37	MC	0.56	0.40	0.12%	3
17	CR	0.58	0.56	0.58%	3	38	MC	0.91	0.49	0.11%	3
18	MC	0.85	0.46	0.05%	1	39	MC	0.75	0.48	0.16%	1
19	MC	0.80	0.43	0.34%	1	40	MC	0.84	0.59	0.33%	3
20	MC	0.73	0.42	0.17%	1	41	MC	0.65	0.49	0.11%	1
21	MC	0.75	0.41	0.33%	1	42	MC	0.85	0.53	0.62%	3

* OE omit rates are considered blank, FT OE omit rates were not calculated.

Table 10

*Item Statistics for Reading Form B (N=23,957)**

Item	Type	P-Val	Corr	Omit	FIT	Item	Type	P-Val	Corr	Omit	FIT
1	MC	0.83	0.49	0.05%	1	22	MC	0.92	0.57	0.08%	3
2	MC	0.81	0.42	0.19%	1	23	MC	0.88	0.45	0.46%	1
3	MC	0.94	0.49	0.08%	3	24	MC	0.90	0.53	0.48%	3
4	MC	0.69	0.34	0.10%	3	25	MC	0.62	0.37	0.93%	3
5	MC	0.76	0.50	0.23%	1	26	MC	0.86	0.56	0.11%	3
6	MC	0.83	0.48	0.19%	1	27	MC	0.65	0.34	0.73%	3
7	MC	0.55	0.36	0.28%	3	28	MC	0.71	0.36	0.09%	3
8	MC	0.84	0.54	0.68%	3	29	MC	0.84	0.46	0.18%	1
9	MC	0.77	0.52	0.14%	1	30	MC	0.58	0.36	0.44%	3
10	MC	0.84	0.59	0.23%	3	31	MC	0.89	0.46	0.11%	1
11	MC	0.89	0.51	1.54%	3	32	MC	0.80	0.60	0.12%	3
12	MC	0.79	0.45	0.10%	1	33	MC	0.72	0.48	0.18%	1
13	MC	0.89	0.55	0.21%	3	34	CR	0.55	0.54	0.52%	3
14	MC	0.74	0.58	0.10%	3	35	MC	0.65	0.50	0.07%	1
15	MC	0.34	0.26	0.27%	3	36	MC	0.75	0.52	0.38%	3
16	MC	0.79	0.44	0.31%	1	37	MC	0.79	0.53	0.11%	1
17	CR	0.58	0.55	0.68%	3	38	MC	0.73	0.54	0.11%	3
18	MC	0.85	0.47	0.05%	1	39	MC	0.84	0.54	0.11%	1
19	MC	0.80	0.43	0.28%	1	40	MC	0.72	0.48	0.20%	1
20	MC	0.73	0.43	0.14%	1	41	MC	0.57	0.39	0.40%	3
21	MC	0.76	0.43	0.28%	1	42	MC	0.88	0.60	0.45%	3

* OE omit rates are considered blank, FT OE omit rates were not calculated.

Table 11
Item Statistics for Reading Form C (N=23,785)*

Item	Type	P-Val	Corr	Omit	FIT	Item	Type	P-Val	Corr	Omit	FIT
1	MC	0.84	0.48	0.06%	1	22	MC	0.92	0.54	0.08%	3
2	MC	0.81	0.42	0.19%	1	23	MC	0.87	0.45	0.48%	1
3	MC	0.94	0.49	0.11%	3	24	MC	0.90	0.52	0.38%	3
4	MC	0.69	0.33	0.15%	3	25	MC	0.62	0.38	0.83%	3
5	MC	0.77	0.48	0.19%	1	26	MC	0.82	0.46	0.16%	1
6	MC	0.83	0.47	0.19%	1	27	MC	0.76	0.59	0.76%	3
7	MC	0.56	0.36	0.29%	3	28	MC	0.73	0.47	0.13%	1
8	MC	0.84	0.54	0.71%	3	29	MC	0.52	0.41	0.49%	1
9	MC	0.77	0.53	0.13%	1	30	MC	0.77	0.48	0.25%	3
10	MC	0.84	0.59	0.20%	3	31	MC	0.65	0.48	0.60%	3
11	MC	0.89	0.51	1.51%	3	32	MC	0.84	0.53	0.13%	1
12	MC	0.79	0.44	0.12%	1	33	MC	0.41	0.35	0.24%	3
13	MC	0.89	0.54	0.23%	3	34	CR	0.55	0.53	0.56%	3
14	MC	0.74	0.56	0.08%	3	35	MC	0.84	0.50	0.03%	1
15	MC	0.35	0.26	0.23%	3	36	MC	0.69	0.46	0.12%	1
16	MC	0.80	0.44	0.34%	1	37	MC	0.80	0.50	1.49%	1
17	CR	0.59	0.56	0.68%	3	38	MC	0.84	0.55	2.25%	3
18	MC	0.85	0.47	0.04%	1	39	MC	0.79	0.56	0.58%	3
19	MC	0.80	0.45	0.29%	1	40	MC	0.70	0.48	0.08%	1
20	MC	0.73	0.42	0.15%	1	41	MC	0.77	0.50	0.26%	1
21	MC	0.76	0.43	0.30%	1	42	MC	0.61	0.42	0.39%	3

* OE omit rates are considered blank, FT OE omit rates were not calculated.

Table 12

*Item Statistics for Reading Form D (N=23,577)**

Item	Type	P-Val	Corr	Omit	FIT	Item	Type	P-Val	Corr	Omit	FIT
1	MC	0.83	0.49	0.06%	1	22	MC	0.92	0.56	0.11%	3
2	MC	0.81	0.39	0.15%	1	23	MC	0.87	0.46	0.49%	1
3	MC	0.94	0.49	0.06%	3	24	MC	0.90	0.51	0.46%	3
4	MC	0.70	0.35	0.10%	3	25	MC	0.62	0.39	0.89%	3
5	MC	0.76	0.50	0.22%	1	26	MC	0.52	0.49	1.33%	3
6	MC	0.83	0.46	0.14%	1	27	MC	0.71	0.49	0.09%	1
7	MC	0.56	0.36	0.19%	3	28	MC	0.81	0.50	0.26%	1
8	MC	0.84	0.52	0.56%	3	29	MC	0.68	0.49	0.88%	3
9	MC	0.76	0.54	0.11%	1	30	MC	0.80	0.44	0.11%	1
10	MC	0.84	0.58	0.17%	3	31	MC	0.61	0.30	0.20%	3
11	MC	0.89	0.51	1.59%	3	32	MC	0.77	0.31	0.18%	3
12	MC	0.79	0.45	0.08%	1	33	MC	0.76	0.44	0.60%	1
13	MC	0.89	0.54	0.21%	3	34	CR	0.51	0.52	0.68%	1
14	MC	0.74	0.57	0.08%	3	35	MC	0.81	0.51	0.09%	3
15	MC	0.34	0.26	0.24%	3	36	MC	0.69	0.43	0.21%	1
16	MC	0.79	0.45	0.28%	1	37	MC	0.72	0.35	0.24%	3
17	CR	0.59	0.55	0.69%	3	38	MC	0.65	0.42	0.42%	3
18	MC	0.85	0.47	0.06%	1	39	MC	0.45	0.27	0.12%	3
19	MC	0.81	0.43	0.28%	1	40	MC	0.63	0.33	0.20%	3
20	MC	0.73	0.43	0.14%	1	41	MC	0.82	0.46	0.21%	1
21	MC	0.75	0.43	0.31%	1	42	MC	0.52	0.24	0.28%	3

* OE omit rates are considered blank, FT OE omit rates were not calculated.

Table 13
Item Statistics for Reading Form E (N=23,320)*

Item	Type	P-Val	Corr	Omit	FIT	Item	Type	P-Val	Corr	Omit	FIT
1	MC	0.83	0.48	0.09%	1	22	MC	0.92	0.56	0.12%	3
2	MC	0.81	0.40	0.19%	1	23	MC	0.88	0.46	0.49%	1
3	MC	0.93	0.49	0.10%	3	24	MC	0.90	0.52	0.52%	3
4	MC	0.69	0.35	0.14%	3	25	MC	0.62	0.38	1.04%	3
5	MC	0.77	0.49	0.19%	1	26	MC	0.80	0.51	0.15%	1
6	MC	0.83	0.48	0.18%	1	27	MC	0.78	0.51	1.54%	1
7	MC	0.56	0.35	0.30%	3	28	MC	0.59	0.31	0.18%	3
8	MC	0.84	0.53	0.65%	3	29	MC	0.72	0.44	0.24%	1
9	MC	0.76	0.53	0.18%	1	30	MC	0.81	0.53	0.39%	3
10	MC	0.84	0.58	0.23%	3	31	MC	0.61	0.45	0.22%	3
11	MC	0.89	0.50	1.61%	3	32	MC	0.78	0.54	0.21%	3
12	MC	0.79	0.46	0.11%	1	33	CR	0.58	0.45	0.45%	3
13	MC	0.89	0.55	0.24%	3	34	MC	0.75	0.36	0.06%	3
14	MC	0.74	0.58	0.09%	3	35	MC	0.69	0.38	0.11%	1
15	MC	0.34	0.27	0.26%	3	36	MC	0.64	0.49	0.29%	1
16	MC	0.79	0.46	0.36%	1	37	MC	0.67	0.39	0.58%	3
17	CR	0.58	0.55	0.60%	3	38	MC	0.62	0.38	0.29%	1
18	MC	0.84	0.46	0.04%	1	39	MC	0.55	0.42	0.11%	1
19	MC	0.80	0.44	0.27%	1	40	MC	0.94	0.41	0.10%	1
20	MC	0.73	0.43	0.12%	1	41	MC	0.88	0.50	0.39%	1
21	MC	0.76	0.42	0.28%	1						

* OE omit rates are considered blank, FT OE omit rates were not calculated.

Table 14

*Item Statistics for Mathematics Common items (N=126,631)**

Item	Type	P-Val	Corr	Omit	Fit	Item	Type	P-Val	Corr	Omit	Fit
1	MC	0.93	0.44	0.05%	3	29	MC	0.82	0.41	0.14%	1
2	MC	0.96	0.32	0.20%	1	30	MC	0.92	0.36	0.11%	1
3	MC	0.98	0.31	0.06%	1	31	MC	0.85	0.36	2.01%	1
4	MC	0.94	0.33	0.07%	1	32	MC	0.97	0.36	0.06%	1
5	MC	0.90	0.55	0.08%	3	33	MC	0.93	0.42	0.23%	1
6	MC	0.86	0.34	0.09%	1	34	MC	0.91	0.37	0.19%	1
7	MC	0.93	0.46	0.55%	3	35	MC	0.77	0.51	0.19%	3
8	MC	0.95	0.42	0.13%	3	36	MC	0.84	0.38	0.24%	1
9	MC	0.92	0.38	0.09%	1	37	MC	0.77	0.47	0.15%	1
10	MC	0.87	0.40	0.14%	1	38	MC	0.68	0.40	0.30%	1
11	MC	0.81	0.37	0.27%	1	39	MC	0.90	0.36	0.41%	1
12	MC	0.89	0.45	0.12%	1	40	MC	0.89	0.40	0.58%	1
13	MC	0.90	0.44	0.48%	1	41	MC	0.85	0.49	0.26%	1
14	MC	0.89	0.37	0.15%	1	42	MC	0.82	0.37	0.41%	1
15	MC	0.93	0.43	0.10%	1	43	MC	0.71	0.46	0.32%	3
16	MC	0.90	0.26	0.14%	1	44					
17	MC	0.87	0.55	0.11%	3	45	MC	0.74	0.38	0.44%	1
18	MC	0.86	0.32	0.32%	1	46	MC	0.65	0.46	0.26%	1
19	MC	0.86	0.50	0.41%	3	47	MC	0.58	0.39	0.32%	1
20	MC	0.80	0.47	0.23%	1	48	MC	0.54	0.34	0.40%	1
21	MC	0.84	0.40	0.33%	1	49	MC	0.87	0.45	0.24%	1
22	MC	0.86	0.39	0.66%	1	50	MC	0.76	0.43	4.63%	1
23	MC	0.78	0.29	0.19%	3	51	MC	0.83	0.29	0.26%	1
24	MC	0.82	0.52	0.16%	3	52	MC	0.82	0.37	0.20%	1
25	MC	0.90	0.44	0.13%	1	53	MC	0.83	0.51	0.28%	3
26	MC	0.75	0.40	0.17%	3	54	MC	0.74	0.45	0.15%	1
27	MC	0.86	0.44	0.60%	1	55	MC	0.84	0.42	0.21%	1
28	CR	0.85	0.40	0.06%	3	56	CR	0.66	0.38	0.21%	3

*Item 44 is suppressed, so no item statistics are presented.

Table 15
Item Statistics for Mathematics Matrix items

Form	Item	Type	P-Val	Corr	Omit	FIT	Form	Item	Type	P-Val	Corr	Omit	FIT
A (N = 16,199)	57	MC	0.73	0.47	0.09%	1	D (N = 15,666)	57	MC	0.99	0.25	0.03%	1
	58	MC	0.89	0.47	0.12%	1		58	MC	0.63	0.23	0.08%	3
	59	MC	0.84	0.38	1.07%	1		59	MC	0.57	0.36	0.15%	3
	60	MC	0.86	0.31	0.16%	1		60	MC	0.70	0.42	0.11%	1
	61	MC	0.90	0.59	0.31%	3		61	MC	0.80	0.49	0.19%	1
	62	MC	0.56	0.39	0.27%	1		62	MC	0.90	0.23	0.13%	3
	63	MC	0.95	0.41	0.23%	1		63	MC	0.88	0.59	0.16%	3
	64	MC	0.84	0.32	0.18%	3		64	MC	0.82	0.49	0.37%	1
	65	MC	0.82	0.48	0.15%	1		65	MC	0.90	0.33	0.08%	1
	66	MC	0.86	0.39	0.17%	1		66	MC	0.95	0.41	0.12%	1
	67	CR	0.77	0.60	0.27%	1		67	CR	0.74	0.48	0.28%	3
B (N = 15,914)	57	MC	0.59	0.51	0.06%	3	E (N = 15,502)	57	MC	0.93	0.38	0.17%	1
	58	MC	0.84	0.39	0.18%	1		58	MC	0.91	0.38	0.82%	1
	59	MC	0.86	0.35	0.23%	1		59	MC	0.90	0.53	0.12%	3
	60	MC	0.85	0.54	0.33%	3		60	MC	0.83	0.35	0.16%	1
	61	MC	0.63	0.35	0.11%	3		61	MC	0.91	0.41	0.18%	1
	62	MC	0.84	0.45	0.14%	1		62	MC	0.81	0.28	0.11%	3
	63	MC	0.70	0.42	0.26%	1		63	MC	0.82	0.53	0.17%	3
	64	MC	0.70	0.51	0.30%	3		64	MC	0.82	0.44	0.21%	1
	65	MC	0.92	0.39	0.11%	1		65	MC	0.67	0.30	0.54%	3
	66	MC	0.93	0.42	0.27%	1		66	MC	0.74	0.27	0.30%	3
	67	CR	0.72	0.42	0.64%	3		67	CR	0.50	0.37	0.66%	3
C (N = 15,855)	57	MC	0.90	0.49	0.09%	3	F (N = 15,382)	57	MC	0.88	0.45	0.12%	1
	58	MC	0.75	0.46	0.13%	1		58	MC	0.84	0.51	0.16%	3
	59	MC	0.88	0.40	0.48%	1		59	MC	0.84	0.38	0.15%	1
	60	MC	0.62	0.38	0.41%	1		60	MC	0.82	0.34	0.03%	1
	61	MC	0.70	0.47	0.14%	3		61	MC	0.80	0.39	0.05%	1
	62	MC	0.80	0.33	0.09%	3		62	MC	0.76	0.44	0.21%	1
	63	MC	0.97	0.34	0.23%	1		63	MC	0.75	0.50	0.36%	1
	64	MC	0.93	0.43	0.27%	1		64	MC	0.42	0.17	0.27%	3
	65	MC	0.85	0.29	0.22%	3		65	MC	0.53	0.39	0.22%	1
	66	MC	0.52	0.38	0.40%	1		66	MC	0.69	0.35	0.20%	1
	67	CR	0.81	0.44	0.40%	3		67	CR	0.59	0.47	0.18%	3

Table 15 Cont'd
Item Statistics for Mathematics Matrix items

Form	Item	Type	P-Val	Corr	Omit	FIT	Form	Item	Type	P-Val	Corr	Omit	FIT
G (N = 8151)	57	CR	0.74	0.58	0.12%	3	I (N = 8028)	57	MC	0.97	0.39	0.04%	1
	58	MC	0.94	0.44	0.10%	3		58	MC	0.95	0.44	0.09%	1
	59	MC	0.88	0.43	0.12%	1		59	MC	0.97	0.26	0.05%	1
	60	MC	0.71	0.22	0.11%	3		60	MC	0.83	0.33	0.19%	1
	61	MC	0.88	0.46	0.16%	1		61	MC	0.94	0.38	0.11%	1
	62	MC	0.67	0.45	0.27%	1		62	MC	0.50	0.38	0.46%	1
	63	MC	0.72	0.50	0.18%	3		63	MC	0.83	0.41	0.11%	1
	64	MC	0.75	0.45	0.38%	1		64	MC	0.97	0.37	0.20%	1
	65	MC	0.97	0.28	0.11%	1		65	MC	0.75	0.48	0.16%	3
	66	MC	0.97	0.28	0.10%	1		66	MC	0.74	0.50	0.45%	3
	67	MC	0.72	0.24	0.44%	3		67	CR	0.51	0.51	0.32%	1
H (N = 8028)	57	MC	0.91	0.35	0.05%	1	J (N = 7906)	57	MC	0.97	0.27	0.05%	1
	58	MC	0.92	0.38	0.02%	1		58	MC	0.83	0.32	0.01%	1
	59	MC	0.86	0.22	0.11%	3		59	MC	0.96	0.38	0.11%	1
	60	MC	0.56	0.40	0.16%	1		60	MC	0.94	0.36	0.04%	1
	61	MC	0.84	0.54	0.16%	3		61	MC	0.79	0.56	0.03%	3
	62	MC	0.88	0.53	0.16%	3		62	MC	0.85	0.35	0.22%	1
	63	MC	0.83	0.48	0.09%	1		63	MC	0.90	0.15	0.10%	3
	64	MC	0.93	0.42	0.49%	1		64	MC	0.78	0.43	0.05%	1
	65	MC	0.65	0.47	0.17%	1		65	MC	0.78	0.51	0.10%	1
	66	MC	0.82	0.42	0.25%	1		66	MC	0.79	0.37	0.24%	1
	67	CR	0.59	0.49	0.16%	3		67	CR	0.74	0.59	0.19%	1

Table 16
Scoring Table for Reading Form A

NC	SS	SEM	NC	SS	SEM
0	300	202			
1	354	174	26	1133	57
2	480	126	27	1152	57
3	556	105	28	1171	58
4	613	92	29	1191	58
5	658	84	30	1211	59
6	697	78	31	1232	60
7	730	74	32	1253	61
8	761	70	33	1275	62
9	789	68	34	1299	64
10	815	65	35	1323	65
11	839	63	36	1349	67
12	862	62	37	1376	70
13	884	61	38	1405	73
14	905	60	39	1438	76
15	925	59	40	1474	81
16	945	58	41	1514	87
17	965	57	42	1562	95
18	984	57	43	1622	107
19	1003	57	44	1702	128
20	1021	56	45	1831	177
21	1040	56	46	1999	277
22	1058	56			
23	1077	56			
24	1095	56			
25	1114	57			

Table 17
Scoring Table for Reading Form B

NC	SS	SEM	NC	SS	SEM
0	300	197			
1	346	174	26	1108	56
2	471	125	27	1127	57
3	546	104	28	1146	57
4	602	92	29	1165	58
5	647	84	30	1185	59
6	685	78	31	1206	60
7	718	73	32	1227	61
8	748	70	33	1249	63
9	775	67	34	1273	64
10	800	65	35	1298	66
11	824	63	36	1325	69
12	846	61	37	1354	72
13	868	60	38	1386	76
14	888	59	39	1421	80
15	908	58	40	1461	86
16	927	57	41	1508	94
17	946	56	42	1565	104
18	965	56	43	1637	119
19	983	56	44	1737	143
20	1001	55	45	1897	194
21	1019	55	46	1999	238
22	1036	55			
23	1054	55			
24	1072	55			
25	1090	56			

Table 18
Scoring Table for Reading Form C

NC	SS	SEM	NC	SS	SEM
0	300	202			
1	354	174	26	1121	56
2	480	126	27	1140	57
3	556	104	28	1159	58
4	611	92	29	1179	58
5	656	84	30	1199	59
6	695	78	31	1220	60
7	728	73	32	1242	62
8	758	70	33	1265	63
9	785	67	34	1288	65
10	811	65	35	1314	67
11	834	63	36	1341	69
12	857	61	37	1370	72
13	878	60	38	1401	75
14	899	59	39	1436	79
15	919	58	40	1474	84
16	938	57	41	1519	90
17	957	57	42	1571	99
18	976	56	43	1635	111
19	994	56	44	1720	132
20	1012	56	45	1855	180
21	1030	56	46	1999	259
22	1048	55			
23	1067	56			
24	1085	56			
25	1103	56			

Table 19
Scoring Table for Reading Form D

NC	SS	SEM	NC	SS	SEM
0	300	208			
1	364	174	26	1141	57
2	490	126	27	1160	57
3	566	105	28	1179	58
4	623	92	29	1199	59
5	668	84	30	1219	60
6	707	78	31	1241	61
7	741	74	32	1262	62
8	771	70	33	1285	63
9	799	68	34	1309	65
10	824	65	35	1334	67
11	849	63	36	1361	69
12	872	62	37	1390	71
13	893	60	38	1421	75
14	914	59	39	1455	79
15	935	59	40	1494	83
16	954	58	41	1537	90
17	974	57	42	1589	98
18	993	57	43	1652	111
19	1011	56	44	1737	132
20	1030	56	45	1872	179
21	1048	56	46	1999	248
22	1067	56			
23	1085	56			
24	1103	56			
25	1122	56			

Table 20
Scoring Table for Reading Form E

NC	SS	SEM	NC	SS	SEM
0	300	199			
1	349	175	26	1131	57
2	475	126	27	1151	58
3	552	105	28	1170	58
4	608	93	29	1191	59
5	654	84	30	1211	60
6	693	79	31	1233	61
7	727	74	32	1255	63
8	757	71	33	1279	64
9	785	68	34	1304	66
10	811	66	35	1330	68
11	836	64	36	1358	71
12	859	62	37	1389	74
13	881	61	38	1423	78
14	902	60	39	1461	83
15	923	59	40	1504	89
16	943	58	41	1555	98
17	962	57	42	1617	110
18	981	57	43	1701	131
19	1000	57	44	1835	179
20	1019	56	45	1999	273
21	1038	56			
22	1056	56			
23	1075	56			
24	1093	57			
25	1112	57			

Table 21
Scoring Table for Mathematics

NC	SS	SEM	NC	SS	SEM
0	200	161			
1	200	161	36	1075	49
2	257	144	37	1088	49
3	353	118	38	1102	50
4	422	103	39	1116	50
5	475	92	40	1131	50
6	519	85	41	1145	51
7	557	79	42	1160	52
8	590	74	43	1175	52
9	620	71	44	1191	53
10	647	68	45	1208	54
11	672	65	46	1225	56
12	696	63	47	1243	57
13	717	61	48	1262	59
14	738	59	49	1283	61
15	758	58	50	1304	63
16	776	57	51	1328	66
17	794	56	52	1354	69
18	812	55	53	1382	72
19	829	54	54	1413	77
20	845	53	55	1449	82
21	861	53	56	1491	89
22	877	52	57	1541	99
23	892	52	58	1605	113
24	907	51	59	1692	137
25	921	51	60	1835	187
26	936	50	61	1999	278
27	950	50			
28	964	50			
29	978	50			
30	992	49			
31	1006	49			
32	1020	49			
33	1034	49			
34	1047	49			
35	1061	49			

Table 22
Raw score Descriptive Statistics Based on All Samples

Content	Form	N Count	Mean	Test Difficulty	SD	Min	Max
Reading	A	31,505	33.28	0.72	9.11	2	46
	B	24,011	34.12	0.74	8.95	2	46
	C	23,833	33.78	0.73	9.09	2	46
	D	23,609	33.01	0.72	8.74	3	46
	E	23,357	33.04	0.73	8.66	1	45
Mathematics (Both Common items and Matrix items)	A	16,225	61.83	0.82	11.20	7	75
	B	15,943	61.37	0.82	11.16	12	75
	C	15,881	61.72	0.82	10.96	10	75
	D	15,685	61.64	0.82	11.01	9	75
	E	15,530	60.91	0.81	10.82	8	75
	F	15,403	60.24	0.80	11.21	12	75
	G	8,165	61.78	0.82	10.89	13	75
	H	8,040	61.39	0.82	10.87	11	75
	I	8,037	61.20	0.82	10.91	11	75
	J	7,916	62.12	0.83	11.09	10	75
Mathematics (Common Items only)	Total	126,880	50.58	0.83	9.00	5	61

Table 23

Raw Score Descriptive Statistics by Ethnicity

Content	Form	White				African American				Hispanic			
		N Count	Raw Score Mean	Raw Score SD	Test Difficulty	N Count	Raw Score Mean	Raw Score SD	Test Difficulty	N Count	Raw Score Mean	Raw Score SD	Test Difficulty
Reading	A	22,825	34.90	0.76	8.19	4,839	28.32	0.62	10	2,152	27.60	0.60	10.13
	B	17,405	35.80	0.78	7.86	3,723	28.89	0.63	10	1,643	28.24	0.61	10.31
	C	17,378	35.46	0.77	8.09	3,607	28.45	0.62	10	1,613	27.70	0.60	10.27
	D	17,137	34.68	0.75	7.77	3,628	28.13	0.61	9	1,606	26.91	0.59	9.60
	E	16,953	34.64	0.77	7.64	3,607	28.21	0.63	9	1,589	27.47	0.61	9.92
Mathematics (Both Common items and Matrix items)	A	11,777	63.97	0.85	9.50	2,456	54.62	0.73	13	1,111	55.16	0.74	13.66
	B	11,505	63.58	0.85	9.37	2,494	53.92	0.72	13	1,095	54.74	0.73	13.20
	C	11,538	63.81	0.85	9.26	2,442	54.61	0.73	13	1,075	54.89	0.73	13.23
	D	11,315	63.71	0.85	9.31	2,445	55.29	0.74	13	1,099	54.84	0.73	13.24
	E	11,225	63.12	0.84	8.97	2,417	53.57	0.71	13	1,073	54.50	0.73	12.87
	F	11,091	62.47	0.83	9.38	2,404	52.85	0.70	13	1,080	53.52	0.71	13.12
	G	5,937	63.95	0.85	9.14	1,251	54.64	0.73	13	568	54.63	0.73	13.13
	H	5,868	63.39	0.85	9.09	1,181	54.60	0.73	13	567	54.28	0.72	13.38
	I	5,858	63.32	0.84	9.22	1,206	53.88	0.72	13	538	54.65	0.73	12.72
	J	5,762	64.22	0.86	9.44	1,200	55.16	0.74	13	540	55.56	0.74	13.02
Mathematics (Common items only)	Total	91,895	52.31	0.86	7.53	19,515	44.83	0.73	11	8,755	45.20	0.74	10.91

Table 23 Cont'd
Raw Score Descriptive Statistics by Ethnicity

Content	Form	Asian				Native American			
		N Count	Raw Score Mean	Raw Score SD	Test Difficulty	N Count	Raw Score Mean	Raw Score SD	Test Difficulty
Reading	A	897	34.76	0.76	8.79	94	34.10	0.74	9.20
	B	699	35.78	0.78	7.77	66	34.59	0.75	8.42
	C	652	35.36	0.77	7.96	65	33.15	0.72	9.29
	D	647	34.40	0.75	7.57	83	31.92	0.69	8.80
	E	600	34.58	0.77	8.09	71	31.69	0.70	9.87
Mathematics (Both Common items and Matrix items)	A	475	64.69	0.86	10.27	50	63.88	0.85	9.37
	B	492	64.44	0.86	10.08	45	63.49	0.85	12.07
	C	433	65.65	0.88	8.96	39	63.15	0.84	10.39
	D	433	65.30	0.87	9.20	54	59.72	0.80	11.82
	E	419	63.94	0.85	9.67	45	59.38	0.79	12.23
	F	437	64.35	0.86	10.28	44	61.27	0.82	13.59
	G	219	64.11	0.85	9.73	21	62.05	0.83	9.44
	H	233	63.96	0.85	10.04	26	59.69	0.80	13.64
	I	231	63.55	0.85	9.92	30	61.27	0.82	11.00
	J	201	65.38	0.87	9.28	26	63.88	0.85	10.61
Mathematics (Common items only)	Total	3,573	53.18	0.87	7.84	380	50.94	0.84	9.57

Table 24
Raw Score Descriptive Statistics by Gender

Content	Form	Male				Female			
		N Count	Raw Score Mean	Raw Score SD	Test Difficulty	N Count	Raw Score Mean	Raw Score SD	Test Difficulty
Reading	A	15,907	32.53	0.71	9.47	15,231	34.16	0.74	8.57
	B	11,961	33.39	0.73	9.32	11,799	34.96	0.76	8.42
	C	11,987	32.83	0.71	9.44	11,597	34.83	0.76	8.57
	C	11,897	32.22	0.70	9.15	11,485	33.92	0.74	8.13
	D	11,800	32.20	0.72	8.99	11,299	34.03	0.76	8.13
Mathematics (Both Common items and Matrix items)	A	8,232	62.25	0.83	11.22	7,808	61.52	0.82	11.08
	B	7,952	61.63	0.82	11.26	7,824	61.22	0.82	11.02
	C	8,002	62.00	0.83	10.98	7,710	61.55	0.82	10.88
	D	7,895	61.88	0.83	11.31	7,639	61.53	0.82	10.61
	E	7,865	61.16	0.82	10.83	7,499	60.77	0.81	10.75
	F	7,745	60.66	0.81	11.14	7,474	59.96	0.80	11.14
	G	4,057	62.24	0.83	10.91	4,021	61.44	0.82	10.79
	H	4,028	61.70	0.82	10.98	3,933	61.13	0.82	10.74
	I	4,056	61.68	0.82	10.85	3,905	60.79	0.81	10.88
	J	3,977	62.56	0.83	11.02	3,845	61.85	0.82	10.98
Mathematics (Common Items only)	Total	63,837	50.88	0.83	9.05	61,679	50.37	0.83	8.89

Table 25
Raw Score Descriptive Statistics by Disability Status

Content	Form	Not Disabled				Disabled			
		N	Mean	SD	Test Difficulty	N	Mean	SD	Test Difficulty
Reading	A	27,051	34.64	7.99	0.75	4,454	25.04	10.94	0.54
	B	20,784	35.36	7.84	0.77	3,227	26.15	11.26	0.57
	C	20,656	35.04	7.99	0.76	3,177	25.55	11.28	0.56
	D	20,491	34.21	7.73	0.74	3,118	25.13	10.68	0.55
	E	20,156	34.26	7.67	0.76	3,201	25.36	10.42	0.56
Mathematics (Both Common items and Matrix items)	A	13,796	63.30	9.82	0.84	2,429	53.50	14.43	0.71
	B	13,742	62.68	9.97	0.84	2,201	53.22	14.28	0.71
	C	13,711	63.01	9.61	0.84	2,170	53.60	14.81	0.71
	D	13,584	63.00	9.62	0.84	2,101	52.87	14.72	0.70
	E	13,391	62.16	9.59	0.83	2,139	53.12	14.24	0.71
	F	13,362	61.50	10.06	0.82	2,041	51.95	14.38	0.69
	G	7,118	62.93	9.80	0.84	1,047	54.01	14.24	0.72
	H	7,018	62.60	9.69	0.83	1,022	53.04	14.29	0.71
	I	6,999	62.49	9.67	0.83	1,038	52.53	14.29	0.70
	J	6,839	63.47	9.77	0.85	1,077	53.59	14.65	0.71
Mathematics (Common Items only)	Total	109,600	51.64	7.94	0.85	17,280	43.87	11.97	0.72

*"Disabled" refers to students with any of the following disabilities: autism, deaf-blindness, deafness, emotional disturbance, hearing impairment, mental retardation, multiple disabilities, orthopedic impairment, other health impairment, specific learning disability, speech or language impairment, traumatic brain injury, visual impairment including blindness.

Table 26
Raw Score Descriptive Statistics by ELP Status

Content	Form	English Language Proficient				English Language Learners			
		N	Mean	SD	Test Difficulty	N	Mean	SD	Test Difficulty
Reading	A	30,392	33.60	8.91	0.73	1,113	24.56	9.98	0.53
	B	23,132	34.45	8.72	0.75	879	25.57	10.51	0.56
	C	23,055	34.05	8.91	0.74	778	25.62	10.43	0.56
	D	22,781	33.32	8.53	0.72	828	24.36	9.84	0.53
	E	22,606	33.32	8.46	0.74	751	24.47	10.03	0.54
Mathematics (Both Common items and Matrix items)	A	15,625	62.18	10.91	0.83	600	52.60	14.13	0.70
	B	15,337	61.69	10.91	0.82	606	53.25	14.05	0.71
	C	15,349	62.00	10.74	0.83	532	53.73	13.78	0.72
	D	15,101	61.98	10.72	0.83	584	52.94	14.27	0.71
	E	14,989	61.24	10.56	0.82	541	51.90	13.58	0.69
	F	14,822	60.61	10.92	0.81	581	50.67	13.90	0.68
	G	7,851	62.16	10.58	0.83	314	52.46	14.03	0.70
	H	7,749	61.75	10.55	0.82	291	51.77	14.39	0.69
	I	7,747	61.55	10.62	0.82	290	51.88	13.98	0.69
	J	7,654	62.41	10.87	0.83	262	53.61	13.82	0.71
Mathematics (Common Items only)	Total	122,274	50.85	8.78	0.83	4,606	43.46	11.59	0.71

***"English Language Learners" includes students who are in their first year of enrollment as those NOT in their first year of enrollment.

Table 27

Raw Score Descriptive Statistics by Economic Disadvantage Status

Content	Form	Not Economically Disadvantaged				Economically Disadvantaged			
		N	Mean	Test Difficulty	SD	N	Mean	Test Difficulty	SD
Reading	A	19,906	35.60	7.74	0.77	11,599	29.30	9.86	0.64
	B	15,103	36.43	7.46	0.79	8,908	30.21	9.87	0.66
	C	14,976	36.21	7.57	0.79	8,857	29.66	9.93	0.64
	D	14,877	35.27	7.39	0.77	8,732	29.16	9.48	0.63
	E	14,699	35.27	7.23	0.78	8,658	29.25	9.53	0.65
Mathematics (Both Common items and Matrix items)	A	10,175	64.68	9.07	0.86	6,050	57.04	12.69	0.76
	B	10,008	64.27	8.96	0.86	5,935	56.48	12.69	0.75
	C	9,933	64.45	8.93	0.86	5,948	57.16	12.42	0.76
	D	9,856	64.31	8.93	0.86	5,829	57.13	12.60	0.76
	E	9,717	63.62	8.67	0.85	5,813	56.39	12.41	0.75
	F	9,779	63.06	9.16	0.84	5,624	55.34	12.66	0.74
	G	5,138	64.59	8.80	0.86	3,027	57.02	12.35	0.76
	H	5,081	64.12	8.70	0.85	2,959	56.70	12.50	0.76
	I	5,074	64.00	8.77	0.85	2,963	56.42	12.44	0.75
	J	5,026	64.87	9.10	0.86	2,890	57.34	12.53	0.76
Mathematics (Common Items only)	Total	79,812	52.81	7.23	0.87	47,068	46.81	10.35	0.77

Table 28
Raw Score Descriptive Statistics by Migrant Status

Content	Form	Non-migrant				Migrant			
		N	Mean	SD	Test Difficulty	N	Mean	SD	Test Difficulty
Reading	A	30,248	33.48	0.73	9.00	1,257	28.40	0.62	10.11
	B	23,101	34.32	0.75	8.84	910	29.22	0.64	10.34
	C	22,938	34.00	0.74	8.94	895	27.96	0.61	10.78
	D	22,709	33.19	0.72	8.63	900	28.35	0.62	10.05
	E	22,383	33.26	0.74	8.52	974	27.99	0.62	10.13
Mathematics (Both Common items and Matrix items)	A	15,566	62.09	0.83	11.05	659	55.81	0.74	12.72
	B	15,316	61.63	0.82	11.00	627	55.12	0.73	13.00
	C	15,315	61.98	0.83	10.77	566	54.77	0.73	13.57
	D	15,067	61.95	0.83	10.75	618	54.23	0.72	14.19
	E	14,857	61.23	0.82	10.55	673	53.88	0.72	13.78
	F	14,768	60.56	0.81	10.97	635	52.86	0.70	13.83
	G	7,854	62.09	0.83	10.66	311	54.04	0.72	13.59
	H	7,696	61.75	0.82	10.53	344	53.31	0.71	14.66
	I	7,731	61.48	0.82	10.74	306	54.31	0.72	12.71
	J	7,585	62.46	0.83	10.85	331	54.37	0.72	13.61
Mathematics (Common Items only)	Total	121,803	50.82	0.83	8.82	5,077	44.91	0.74	11.20

***"Migrant" includes all students who are migrants at the school, district, and/or state level (i.e. they initially enrolled in the school, district, or state of residence after October 1, 2004).

Table 29
Scale Score Descriptive Statistics Based on All Samples

Content	Form	N Count	Scale Score Mean	Scale Score SD	Skewness	Kurtosis
Reading	A	31,505	1325.58	233.98	-0.18	0.14
	B	24,011	1332.36	238.45	-0.28	0.07
	C	23,833	1336.47	241.39	-0.22	0.01
	D	23,609	1326.16	222.09	-0.25	0.17
	E	23,357	1327.95	231.05	-0.21	0.11
	Mean	126,315	1329.47	233.58	-0.22	0.11
Mathematics	A	16,228	1395.51	240.45	0.12	0.00
	B	15,955	1397.16	236.48	0.12	0.04
	C	15,891	1394.63	235.87	0.11	0.07
	D	15,690	1393.89	234.90	0.09	0.14
	E	15,534	1396.31	237.87	0.12	0.04
	F	15,410	1395.41	237.62	0.11	0.07
	G	8,167	1396.52	234.99	0.14	0.04
	H	8,043	1402.00	235.58	0.09	0.07
	I	8,043	1399.32	236.75	0.09	0.07
	J	7,919	1396.92	238.57	0.10	0.07
	Common	126,880	1396.30	237.02	0.11	0.06

Table 30
Scale Score Descriptive Statistics by Ethnicity

Content	Test Form	White					African American					Hispanic				
		N	Mean	SD	MIN	MAX	N	Mean	SD	MIN	MAX	N	Mean	SD	MIN	MAX
Reading	A	22,825	1366.16	221.45	613	1999	4,839	1199.57	221.01	480	1999	2,152	1184.49	226.62	658	1831
	B	17,405	1376.38	222.93	546	1999	3,723	1193.10	227.57	471	1999	1,643	1179.90	233.15	602	1897
	C	17,378	1380.49	227.91	480	1999	3,607	1196.69	228.55	556	1999	1,613	1180.55	233.41	611	1999
	D	17,137	1367.90	208.27	566	1999	3,628	1203.43	211.43	566	1999	1,606	1176.34	212.98	623	1999
	E	16,953	1369.38	216.51	608	1999	3,607	1200.77	221.16	349	1999	1,589	1185.31	230.87	608	1999
	Mean	91,698	1371.73	219.74	480	1999	19,404	1198.74	221.99	349	1999	8,603	1181.50	227.47	602	1999
Mathematics	A	11,779	1436.38	226.52	620	1999	2,456	1251.61	223.25	475	1999	1,112	1270.67	241.59	672	1999
	B	11,510	1438.27	221.54	672	1999	2,497	1253.14	221.41	647	1999	1,098	1274.38	235.72	672	1999
	C	11,540	1435.47	221.28	590	1999	2,446	1252.22	220.96	557	1999	1,077	1263.41	237.66	672	1999
	D	11,316	1434.87	220.43	590	1999	2,447	1264.53	225.03	620	1999	1,099	1258.38	227.96	647	1999
	E	11,226	1439.98	222.75	647	1999	2,418	1247.92	223.60	557	1999	1,073	1268.34	229.51	717	1999
	F	11,094	1438.07	222.08	672	1999	2,407	1249.61	219.84	696	1999	1,081	1262.25	229.96	590	1999
	G	5,937	1439.63	221.56	672	1999	1,252	1251.17	215.69	672	1999	569	1258.19	224.34	672	1835
	H	5,869	1440.64	220.57	717	1999	1,182	1267.69	232.57	620	1999	568	1261.93	225.62	647	1999
	I	5,861	1440.23	221.70	672	1999	1,209	1254.92	227.56	590	1999	538	1272.47	229.49	696	1999
	J	5,763	1438.52	226.52	672	1999	1,201	1255.22	219.00	647	1999	540	1270.16	226.76	672	1999
	Common	91,895	1437.82	222.49	590	1999	19,515	1254.17	222.71	475	1999	8,755	1266.08	231.94	590	1999

Table 30 Cont'd
Scale Score Descriptive Statistics by Ethnicity

Content	Test Form	Asian					Native American				
		N	Mean	SD	MIN	MAX	N	Mean	SD	MIN	MAX
Reading	A	897	1369.32	238.46	556	1999	94	1347.83	239.29	761	1999
	B	699	1380.18	229.75	685	1999	66	1331.42	211.93	718	1737
	C	652	1376.49	222.70	758	1999	65	1316.66	241.32	728	1855
	D	647	1362.38	209.08	707	1999	83	1290.31	204.67	799	1652
	E	600	1374.90	232.07	727	1999	71	1295.01	247.18	693	1835
	Mean	3,495	1372.50	227.42	556	1999	379	1317.14	229.33	693	1999
Mathematics	A	475	1479.29	256.61	794	1999	50	1447.04	234.43	950	1835
	B	492	1481.32	247.39	794	1999	45	1472.71	265.12	696	1999
	C	433	1494.99	238.07	738	1999	39	1438.28	251.00	877	1999
	D	433	1491.44	248.37	861	1999	54	1359.20	247.81	861	1999
	E	419	1484.81	250.36	738	1999	45	1364.18	247.12	861	1999
	F	437	1508.71	248.55	738	1999	44	1465.30	293.24	717	1999
	G	219	1455.20	241.85	758	1999	21	1396.43	212.83	1088	1835
	H	233	1474.91	250.43	794	1999	26	1410.92	314.42	794	1999
	I	231	1466.35	253.94	829	1999	30	1402.03	235.95	845	1835
	J	201	1478.09	232.87	794	1999	26	1424.04	238.25	907	1835
	Common	3,573	1484.52	247.69	738	1999	380	1418.61	256.10	696	1999

Table 31
Scale Score Descriptive Statistics by Gender

Content	Test Form	Male					Female				
		N	Mean	SD	MIN	MAX	N	Mean	SD	MIN	MAX
Reading	A	15,907	1306.66	237.42	480	1999	15,231	1347.94	227.51	556	1999
	B	11,961	1312.26	241.48	471	1999	11,799	1355.19	232.53	546	1999
	C	11,987	1309.49	240.89	480	1999	11,597	1366.17	238.41	480	1999
	D	11,897	1306.48	226.82	566	1999	11,485	1349.00	213.89	623	1999
	E	11,800	1304.20	231.49	349	1999	11,299	1355.49	226.77	608	1999
	Mean	63,552	1307.76	235.83	349	1999	61,411	1354.36	228.07	480	1999
Mathematics	A	8,234	1406.91	243.87	557	1999	7,809	1386.03	236.00	475	1999
	B	7,957	1405.39	238.91	647	1999	7,830	1391.09	233.81	696	1999
	C	8,006	1403.53	237.38	557	1999	7,714	1387.83	233.90	590	1999
	D	7,897	1402.52	239.69	590	1999	7,641	1387.79	229.39	620	1999
	E	7,866	1404.90	239.23	590	1999	7,500	1390.14	236.20	557	1999
	F	7,750	1406.29	239.40	672	1999	7,476	1386.82	234.39	590	1999
	G	4,057	1408.31	238.46	672	1999	4,023	1386.75	230.55	672	1999
	H	4,030	1412.63	239.78	647	1999	3,934	1392.68	231.14	620	1999
	I	4,060	1410.08	237.78	672	1999	3,907	1389.82	235.06	590	1999
	J	3,980	1408.64	239.68	590	1999	3,845	1387.94	235.69	672	1999
	Common	63,837	1406.19	239.56	557	1999	61,679	1388.54	233.73	475	1999

Table 32
Scale Score Descriptive Statistics by ELP Status

Content	Test Form	English Language Proficient					English Language Learners				
		N	Mean	SD	MIN	MAX	N	Mean	SD	MIN	MAX
Reading	A	30,392	1333.20	230.99	480	1999	1,113	1117.71	219.09	556	1999
	B	23,132	1340.41	234.95	471	1999	879	1120.49	232.19	602	1897
	C	23,055	1343.23	238.67	480	1999	778	1135.99	235.48	611	1999
	D	22,781	1333.60	218.84	566	1999	828	1121.49	213.01	623	1737
	E	22,606	1334.97	227.94	552	1999	751	1116.46	224.05	349	1835
	Mean	121,966	1336.87	230.48	471	1999	4,349	1122.05	224.51	349	1999
Mathematics	A	15,628	1401.95	238.23	475	1999	600	1227.87	237.49	620	1999
	B	15,348	1402.79	234.44	647	1999	607	1254.68	243.17	672	1999
	C	15,358	1399.85	234.08	557	1999	533	1244.40	237.92	672	1999
	D	15,106	1400.07	232.47	590	1999	584	1234.08	240.97	647	1999
	E	14,993	1402.43	235.87	557	1999	541	1226.47	230.50	590	1999
	F	14,829	1402.19	234.95	590	1999	581	1222.24	240.03	672	1999
	G	7,852	1403.33	232.43	672	1999	315	1226.68	234.91	672	1999
	H	7,751	1408.76	232.98	620	1999	292	1222.52	233.60	647	1999
	I	7,752	1405.66	233.99	590	1999	291	1230.51	247.54	696	1999
	J	7,657	1402.24	237.15	590	1999	262	1241.64	227.73	758	1835
	Common	122,274	1402.42	234.80	475	1999	4,606	1233.76	237.95	590	1999

***"English Language Learners" includes students who are in their first year of enrollment as those NOT in their first year of enrollment.

Table 33
Scale Score Descriptive Statistics by Disability Status

Content	Test Form	Not Disabled					Disabled				
		N	Mean	SD	MIN	MAX	N	Mean	SD	MIN	MAX
Reading	A	27,051	1357.43	215.74	480	1999	4,454	1132.15	247.01	480	1999
	B	20,784	1362.17	220.38	471	1999	3,227	1140.36	260.15	546	1999
	C	20,656	1366.75	223.30	556	1999	3,177	1139.58	261.12	480	1999
	D	20,491	1354.15	205.53	566	1999	3,118	1142.17	238.48	566	1999
	E	20,156	1357.89	215.27	349	1999	3,201	1139.41	237.81	608	1999
	Mean	109,138	1359.57	216.17	349	1999	17,177	1138.24	249.01	480	1999
Mathematics	A	13,798	1422.46	229.43	620	1999	2,430	1242.46	244.34	475	1999
	B	13,752	1421.23	227.00	647	1999	2,203	1246.91	239.11	672	1999
	C	13,718	1417.74	224.20	717	1999	2,173	1248.78	254.73	557	1999
	D	13,589	1418.64	223.62	590	1999	2,101	1233.80	243.01	590	1999
	E	13,394	1419.96	227.75	590	1999	2,140	1248.27	246.34	557	1999
	F	13,365	1418.36	227.38	590	1999	2,045	1245.40	248.33	672	1999
	G	7,120	1417.41	226.38	672	1999	1,047	1254.44	243.20	672	1999
	H	7,020	1424.49	225.97	647	1999	1,023	1247.69	242.29	620	1999
	I	7,004	1423.19	225.93	590	1999	1,039	1238.43	245.20	696	1999
	J	6,840	1421.21	227.11	672	1999	1,079	1242.98	251.73	590	1999
	Common	109,600	1420.21	226.51	590	1999	17,280	1244.66	245.90	475	1999

***"Disabled" refers to students with any of the following disabilities: autism, deaf-blindness, deafness, emotional disturbance, hearing impairment, mental retardation, multiple disabilities, orthopedic impairment, other health impairment, specific learning disability, speech or language impairment, traumatic brain injury, visual impairment including blindness.

Table 34
Scale Score Descriptive Statistics by Economic Disadvantage Status

Content	Test Form	Not Economically Disadvantaged					Economically Disadvantaged				
		N	Mean	SD	MIN	MAX	N	Mean	SD	MIN	MAX
Reading	A	19,906	1384.99	215.97	613	1999	11,599	1223.63	228.49	480	1999
	B	15,103	1394.57	218.13	602	1999	8,908	1226.88	234.31	471	1999
	C	14,976	1401.48	220.92	480	1999	8,857	1226.54	234.59	480	1999
	D	14,877	1383.47	203.75	566	1999	8,732	1228.52	217.96	566	1999
	E	14,699	1387.13	211.22	608	1999	8,658	1227.46	228.54	349	1999
	Mean	79,561	1390.02	214.33	480	1999	46,754	1226.42	228.87	349	1999
Mathematics	A	10,177	1453.49	225.24	696	1999	6,051	1298.00	233.61	475	1999
	B	10,013	1454.20	220.35	696	1999	5,942	1301.03	231.56	647	1999
	C	9,937	1451.66	221.33	672	1999	5,954	1299.47	228.57	557	1999
	D	9,859	1449.53	219.91	590	1999	5,831	1299.81	229.41	590	1999
	E	9,719	1453.47	222.56	647	1999	5,815	1300.76	231.91	557	1999
	F	9,783	1451.78	221.97	590	1999	5,627	1297.40	232.06	590	1999
	G	5,138	1455.26	221.44	717	1999	3,029	1296.87	223.49	672	1999
	H	5,082	1458.76	220.19	696	1999	2,961	1304.58	229.20	620	1999
	I	5,076	1455.93	219.83	620	1999	2,967	1302.47	233.27	590	1999
	J	5,028	1454.22	226.60	647	1999	2,891	1297.27	225.77	590	1999
	Common	79,812	1453.30	221.94	590	1999	47,068	1299.64	230.37	475	1999

Table 35
Scale Score Descriptive Statistics by Migrant Status

Content	Test Form	Non-Migrant					Migrant				
		N	Mean	SD	MIN	MAX	N	Mean	SD	MIN	MAX
Reading	A	30,248	1330.65	232.67	480	1999	1,257	1203.56	232.51	613	1999
	B	23,101	1337.38	236.98	471	1999	910	1204.94	240.19	602	1897
	C	22,938	1342.15	239.22	480	1999	895	1190.77	250.96	556	1855
	D	22,709	1330.69	220.53	566	1999	900	1211.89	230.80	623	1999
	E	22,383	1333.52	229.05	552	1999	974	1199.98	239.82	349	1999
	Mean	121,379	1334.64	231.91	471	1999	4,936	1202.31	238.50	349	1999
Mathematics	A	15,569	1400.81	239.93	475	1999	659	1270.22	217.75	620	1999
	B	15,328	1402.39	235.79	672	1999	627	1269.37	216.52	647	1999
	C	15,323	1399.60	234.47	557	1999	568	1260.62	234.08	717	1999
	D	15,070	1399.51	232.82	590	1999	620	1257.29	243.99	590	1999
	E	14,861	1402.42	235.83	557	1999	673	1261.37	242.84	590	1999
	F	14,775	1401.38	235.76	590	1999	635	1256.49	238.63	672	1999
	G	7,856	1402.37	233.43	672	1999	311	1248.55	226.05	696	1999
	H	7,699	1408.66	233.01	620	1999	344	1252.85	243.65	696	1999
	I	7,734	1404.78	235.08	590	1999	309	1262.73	237.61	758	1999
	J	7,588	1403.65	236.93	590	1999	331	1242.64	223.75	672	1999
	Common	121,803	1401.99	235.50	475	1999	5,077	1259.81	232.59	590	1999

***"Migrant" includes all students who are migrants at the school, district, and/or state level (i.e. they initially enrolled in the school, district, or state of residence after October 1, 2004).

Table 36
Percentiles of Scale Score Ranges

Percentile	Reading Score Range	Mathematics Score Range
1	300-788	200-872
2	789-824	873-921
3	825-857	922-958
4	858-881	959-988
5	882-904	989-1014
6	905-926	1015-1036
7	927-948	1037-1054
8	949-974	1055-1072
9	975-993	1073-1086
10	994-1011	1087-1100
11	1012-1026	1101-1113
12	1027-1042	1114-1125
13	1043-1057	1126-1136
14	1058-1074	1137-1146
15	1075-1087	1147-1156
16	1088-1098	1157-1165
17	1099-1112	1166-1174
18	1113-1124	1175-1182
19	1125-1135	1183-1190
20	1136-1146	1191-1198
21	1147-1155	1199-1206
22	1156-1165	1207-1213
23	1166-1172	1214-1221
24	1173-1183	1222-1228
25	1184-1190	1229-1235
26	1191-1197	1236-1242
27	1198-1207	1243-1248
28	1208-1214	1249-1254
29	1215-1222	1255-1261
30	1223-1231	1262-1267
31	1232-1236	1268-1273
32	1237-1243	1274-1279
33	1244-1251	1280-1285
34	1252-1255	1286-1291
35	1256-1263	1292-1296
36	1264-1272	1297-1302
37	1273-1275	1303-1307
38	1276-1281	1308-1313
39	1282-1287	1314-1319
40	1288-1297	1320-1324

Table 36 Cont'd
Percentiles of Scale Score Ranges

Percentile	Reading Score Range	Mathematics Score Range
41	1298-1298	1325-1330
42	1299-1303	1331-1336
43	1304-1309	1337-1341
44	1310-1316	1342-1347
45	1317-1323	1348-1352
46	1324-1326	1353-1357
47	1327-1331	1358-1363
48	1332-1336	1364-1368
49	1337-1343	1369-1373
50	1344-1349	1374-1378
51	1350-1354	1379-1384
52	1355-1357	1385-1389
53	1358-1360	1390-1394
54	1361-1368	1395-1399
55	1369-1373	1400-1405
56	1374-1379	1406-1410
57	1380-1386	1411-1415
58	1387-1389	1416-1421
59	1390	1422-1426
60	1390-1399	1427-1432
61	1400-1403	1433-1437
62	1404-1406	1438-1443
63	1407-1413	1444-1448
64	1414-1419	1449-1454
65	1420-1422	1455-1460
66	1423-1428	1461-1466
67	1429-1436	1467-1472
68	1437-1437	1473-1478
69	1438-1448	1479-1484
70	1449-1456	1485-1490
71	1457-1458	1491-1496
72	1459-1460	1497-1503
73	1461-1464	1504-1509
74	1465-1469	1510-1516
75	1470-1474	1517-1522
76	1475-1490	1523-1529
77	1491-1500	1530-1535
78	1501-1505	1536-1542
79	1506-1508	1543-1551
80	1509-1512	1552-1559

Table 36 Cont'd
Percentiles of Scale Score Ranges

Percentile	Reading Score Range	Mathematics Score Range
81	1513-1516	1560-1568
82	1517-1524	1569-1576
83	1525-1540	1577-1584
84	1541-1555	1585-1593
85	1556-1560	1594-1601
86	1561-1563	1602-1612
87	1564-1566	1613-1625
88	1567-1572	1626-1637
89	1573-1593	1638-1649
90	1594-1617	1650-1662
91	1618-1622	1663-1674
92	1623-1635	1675-1687
93	1636-1636	1688-1709
94	1637-1662	1710-1737
95	1663-1701	1738-1765
96	1702-1715	1766-1793
97	1716-1733	1794-1821
98	1734-1833	1822-1877
99	1834-1999	1878-1999

Table 37
Percent at Each of Reading Performance Level

		Below Basic	Basic	Proficient	Advanced
All students		16.24	14.98	37.50	31.27
Ethnicity	Caucasian	10.85	12.79	39.28	37.09
	African American	32.41	22.35	32.44	12.81
	Hispanic	36.16	20.69	31.51	11.64
	Asian	11.62	14.22	36.14	38.03
	Native American	15.83	17.94	37.47	28.76
Gender	Male	18.81	15.64	37.33	28.22
	Female	13.27	14.17	37.83	34.72
Disabled	Yes	11.57	14.47	39.62	34.35
	No	45.95	18.26	24.06	11.74
ELL	Yes	15.12	14.78	37.98	32.12
	No	47.64	20.60	24.14	7.61
Economically Disadvantaged	Yes	9.01	11.73	39.24	40.03
	No	28.56	20.53	34.54	16.38
Migrant	Yes	15.55	14.78	37.75	31.92
	No	33.35	19.98	31.30	15.38

* The total sum may not be 100 due to rounding percent.

Table 38
Percent at Each Mathematics Performance Level

		Below Basic	Basic	Proficient	Advanced
All students		7.07	10.53	28.16	54.24
Ethnicity	Caucasian	3.68	7.72	26.79	61.81
	African American	17.94	20.03	33.47	28.56
	Hispanic	17.94	19.10	31.95	31.01
	Asian	4.06	6.47	21.94	67.53
	Native American	8.16	10.26	22.89	58.68
Gender	Male	6.98	9.70	26.90	56.41
	Female	6.98	11.27	29.35	52.40
Disabled	Yes	4.59	9.33	27.97	58.11
	No	22.85	18.15	29.32	29.69
ELL	Yes	6.48	10.14	28.12	55.26
	No	22.82	20.97	29.09	27.12
Economically Disadvantaged	Yes	3.14	6.78	25.46	64.62
	No	13.74	16.90	32.73	36.63
Migrant	Yes	6.60	10.18	28.02	55.21
	No	18.57	19.03	31.44	30.96

* The total sum may not be 100 due to rounding percent.

Table 39
Reliability and SEM for Reading and Mathematics

Content	Form	N Count	Alpha	SEM
Reading	A	31,505	0.92	2.61
	B	24,011	0.92	2.45
	C	23,833	0.92	2.50
	D	23,609	0.91	2.57
	E	23,357	0.92	2.51
Mathematics (Both Common items and Matrix items)	A	16,228	0.93	2.96
	B	15,955	0.93	3.10
	C	15,891	0.92	3.06
	D	15,690	0.92	3.08
	E	15,534	0.92	3.01
	F	15,410	0.93	3.07
	G	8,167	0.92	3.04
	H	8,043	0.93	2.97
	I	8,043	0.92	3.07
	J	7,919	0.93	3.04
Mathematics (Common Items only)	Total	126,880	0.91	2.65

Table 40

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	Discrepant Agreement	Intraclass Correlation	Kappa	Weighted Kappa
	17	1.75	1.74	0.86	0.87	70.29	28.83	0.88	0.89	0.56	0.78
A	34	1.56	1.56	1.05	1.06	63.31	31.98	4.71	0.88	0.51	0.76
B	34	1.61	1.59	0.70	0.69	71.34	28.26	0.41	0.85	0.49	0.69
C	34	1.66	1.67	0.82	0.82	66.17	32.25	1.59	0.86	0.49	0.71
D	34	1.48	1.46	0.80	0.78	60.89	36.84	2.27	0.81	0.40	0.63
E	33	1.74	1.73	0.84	0.84	76.53	22.07	1.40	0.90	0.65	0.80

G1: Rater group 1, G2: Rater group 2

Percent of Agreement is the sum of percents of perfect and adjacent agreements.

Table 41

Rater Agreement for Mathematics 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	Discrepant Agreement	Intraclass Correlation	Kappa	Weighted Kappa
	28	3.39	3.39	1.06	1.05	95.80	3.15	1.05	0.98	0.92	0.96
	56	2.59	2.58	0.83	0.83	89.02	10.23	0.75	0.95	0.83	0.90
A	67	3.10	3.09	0.87	0.88	85.74	14.02	0.24	0.95	0.79	0.90
B	67	2.89	2.86	1.13	1.12	69.90	25.31	4.79	0.90	0.58	0.81
C	67	3.19	3.20	1.07	1.06	86.83	11.24	1.94	0.95	0.79	0.90
D	67	2.98	2.97	1.16	1.19	84.39	15.09	0.52	0.97	0.77	0.94
E	67	2.01	2.01	0.99	1.00	87.29	12.03	0.68	0.96	0.82	0.93
F	67	2.33	2.34	0.97	0.97	91.75	7.72	0.53	0.97	0.88	0.95
G	57	2.94	2.93	1.12	1.12	84.60	14.20	1.20	0.96	0.78	0.92
H	67	2.35	2.35	0.94	0.93	76.28	20.47	3.25	0.90	0.65	0.81
I	67	2.06	2.05	1.26	1.26	82.66	14.88	2.46	0.96	0.74	0.92
J	67	2.87	2.87	1.20	1.21	89.85	9.52	0.63	0.98	0.85	0.96

G1: Rater group 1, G2: Rater group 2

Percent of Agreement is the sum of percents of perfect and adjacent agreements.

Table 42

*Number of Items per Each Reading Assessment Anchor **

Objective	Form A		Form B		Form C		Form D		Form E	
	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE
3.A.1	17	1	17	1	14	1	16	2	15	1
3.A.2	12	1	14	0	14	1	14	0	14	1
3.B.1	6	0	6	1	9	0	7	0	7	0
3.B.2	1	0	1	0	1	0	1	0	2	0
3.B.3	4	0	2	0	2	0	2	0	1	0
Total	40	2	40	2	40	2	40	2	39	2

* Table does not include FT items.

Table 43

*Number of Items per each Mathematics Assessment Anchor**

Standards	A		B		C		D		E		F		G		H		I		J	
	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE	MC	OE
1.A.1	18	1	17	0	18	0	19	0	20	0	19	0	17	1	18	0	19	0	18	1
1.A.2	5	0	6	0	6	0	5	0	4	0	5	0	5	0	5	0	5	0	4	0
1.A.3	8	0	8	0	7	0	7	0	7	0	7	0	7	0	8	0	7	0	7	0
2.B.1	6	0	7	0	5	0	6	0	6	0	6	0	6	0	5	1	5	0	5	0
2.B.2	2	1	1	1	2	1	2	1	1	1	2	1	2	1	1	1	1	2	2	1
3.C.1	5	1	4	1	6	1	5	1	5	1	5	1	5	1	5	1	6	1	6	1
3.C.2	2	0	1	0	1	0	2	0	2	0	2	0	2	0	2	0	1	0	2	0
4.D.1	6	0	7	0	7	0	6	1	6	1	6	0	6	0	6	0	6	0	7	0
4.D.2	2	0	3	1	3	0	2	0	2	0	3	0	4	0	3	0	3	0	2	0
5.E.1	9	0	9	0	8	1	9	0	9	0	8	1	9	0	10	0	10	0	10	0
5.E.2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Total	63	3	63	3	63	3	63	3	63	3	63	3	63	3	63	3	63	3	63	3

Table 44
Summary Statistics for Reading Reporting Categories/Assessment Anchors

Form	N	Reporting Category	Assessment Anchor	Total Number of Items	Total Score Points	Mean	Test Difficulty	SD
A	31,505	R3A		31	35	25.85	0.74	6.88
			R3A1	18	20	15.46	0.77	4.00
		R3B	R3A2	13	15	10.39	0.69	3.28
				11	11	7.43	0.68	2.58
			R3B1	6	6	4.64	0.77	1.60
			R3B2	1	1	0.34	0.34	0.47
			R3B3	4	4	2.45	0.61	1.16
B	24,011	R3A		32	34	26.47	0.78	6.83
			R3A1	18	20	15.56	0.78	3.98
		R3B	R3A2	14	14	10.92	0.78	3.18
				10	12	7.65	0.64	2.49
			R3B1	7	9	6.05	0.67	1.94
			R3B2	1	1	0.34	0.34	0.47
			R3B3	2	2	1.26	0.63	0.73
C	23,833	R3A		30	34	25.24	0.74	6.69
			R3A1	15	17	12.88	0.76	3.46
		R3B	R3A2	15	17	12.36	0.73	3.59
				12	12	8.54	0.71	2.74
			R3B1	9	9	7.10	0.79	2.22
			R3B2	1	1	0.35	0.35	0.48
			R3B3	2	2	1.10	0.55	0.72
D	23,609	R3A		32	36	26.27	0.73	6.82
			R3A1	18	22	15.93	0.72	4.28
		R3B	R3A2	14	14	10.34	0.74	2.95
				10	10	6.74	0.67	2.30
			R3B1	7	7	5.18	0.74	1.80
			R3B2	1	1	0.34	0.34	0.47
			R3B3	2	2	1.22	0.61	0.71
E	23,357	R3A		31	35	26.05	0.74	6.78
			R3A1	16	18	13.66	0.76	3.61
		R3B	R3A2	15	17	12.39	0.73	3.53
				10	10	6.99	0.70	2.24
			R3B1	7	7	5.02	0.72	1.79
			R3B2	2	2	1.28	0.64	0.55
			R3B3	1	1	0.69	0.69	0.46

Table 45

Summary Statistics for Mathematics Reporting Categories/Assessment Anchors
*(N = 126,880)**

Reporting Category	Assessment Anchor	Total Number of Items	Total Score Points	Mean	Test Difficulty	SD
M3A		26	26	21.28	0.82	4.52
	M3A1	16	16	13.58	0.85	2.64
	M3A2	4	4	3.17	0.79	0.99
	M3A3	6	6	4.54	0.76	1.55
M3B		7	10	8.47	0.85	1.83
	M3B1	5	5	4.20	0.84	1.11
	M3B2	2	5	4.28	0.86	1.11
M3C		6	9	7.14	0.79	1.31
	M3C1	5	8	6.16	0.77	1.28
	M3C2	1	1	0.98	0.98	0.15
M3D		8	8	6.66	0.83	1.50
	M3D1	6	6	5.02	0.84	1.23
	M3D2	2	2	1.64	0.82	0.52
M3E		8	8	7.03	0.88	1.40
	M3E1	8	8	7.03	0.88	1.40

* Table includes common items only

Table 46

Correlation Between Mathematics Reporting Categories

NAME	M3A	M3B	M3C	M3D
M3A				
M3B	0.67			
M3C	0.56	0.47		
M3D	0.71	0.57	0.47	
M3E	0.72	0.59	0.49	0.61

Table 47

Correlation Between Reading Reporting Categories

Form	NAME	R3A
A	R3B	0.82
B	R3B	0.80
C	R3B	0.83
D	R3B	0.78
E	R3B	0.79

Table 48
Correlation Between Math Assessment Anchors

NAME	M3A1	M3A2	M3A3	M3B1	M3B2	M3C1	M3C2	M3D1	M3D2
M3A1									
M3A2	0.61								
M3A3	0.65	0.53							
M3B1	0.63	0.50	0.55						
M3B2	0.42	0.32	0.36	0.35					
M3C1	0.51	0.41	0.46	0.43	0.32				
M3C2	0.29	0.22	0.20	0.23	0.17	0.21			
M3D1	0.64	0.51	0.56	0.54	0.36	0.43	0.24		
M3D2	0.41	0.33	0.42	0.34	0.25	0.32	0.15	0.37	
M3E1	0.70	0.55	0.58	0.58	0.39	0.47	0.29	0.59	0.38

Table 49
Correlation Between Reading Assessment Anchors

Form	R3A1	R3A2	R3B1	R3B2	R3B3
A	0.78319				
A	0.77563	0.71079			
A	0.23482	0.26365	0.22239		
A	0.57087	0.5995	0.52229	0.22113	
A					
B	0.81583				
B	0.76728	0.73542			
B	0.232	0.24771	0.22552		
B	0.44774	0.46532	0.40921	0.18506	
B					
C	0.80195				
C	0.79487	0.76705			
C	0.23725	0.25408	0.22392		
C	0.40729	0.43709	0.39357	0.18701	
C					
D	0.77038				
D	0.7556	0.70294			
D	0.23328	0.24568	0.2274		
D	0.37097	0.37361	0.32464	0.13859	
D					
E	0.80478				
E	0.74558	0.72405			
E	0.38583	0.3894	0.36821		
E	0.31853	0.33458	0.29483	0.1679	
E					

Table 50
Factor Analysis

Content Area	Form	First Eigenvalue	Second Eigenvalue	Percent	Ratio of First Two Eigenvalues
Reading	A	14.11	1.09	0.99	12.93
	B	14.92	0.92	0.99	16.26
	C	14.82	1.16	0.94	12.73
	D	12.90	0.96	1.00	13.40
	E	13.27	0.94	1.00	14.16
Mathematics		13.37	1.14	0.93	11.77

Table 51
Summary for Differential Item Functioning based on Criteria $\pm C$

Content	Form	Focal Group*	Item	Type	N Ref Group	N Focal Group	Delta	Criteria
RD	A	A	7	MC	22,800	894	-1.75	-C
RD	A	A	24	MC	22,800	894	1.86	+C
RD	A	A	39	MC	22,800	894	-2.06	-C
RD	B	A	7	MC	17,373	697	-1.83	-C
RD	B	A	36	MC	17,373	697	-1.59	-C
RD	C	A	7	MC	17,346	652	-2.18	-C
RD	D	A	2	MC	17,116	647	1.68	+C
RD	D	A	7	MC	17,116	647	-1.81	-C
RD	E	A	7	MC	16,927	600	-2.13	-C
RD	E	A	32	MC	16,927	600	1.55	+C
RD	E	A	41	MC	16,927	600	-2.03	-C
RD	E	H	41	MC	16,927	1,588	-1.95	-C
MA	A	F	26	MC	8,220	7,797	-1.85	-C
MA	A	A	2	MC	11,762	474	1.85	+C
MA	A	A	17	MC	11,762	474	1.83	+C
MA	A	A	18	MC	11,762	474	-1.60	-C
MA	A	A	65	MC	11,762	474	-1.85	-C
MA	A	H	32	MC	11,762	1,110	1.64	+C
MA	B	F	3	MC	7,939	7,809	1.59	+C
MA	B	F	26	MC	7,939	7,809	-1.76	-C
MA	B	A	1	MC	11,487	492	1.66	+C
MA	B	A	62	MC	11,487	492	-1.91	-C
MA	B	AA	62	MC	11,487	2,487	-1.57	-C
MA	B	H	62	MC	11,487	1,095	-1.82	-C
MA	C	F	26	MC	7,989	7,697	-1.89	-C
MA	C	A	2	MC	11,522	433	1.85	+C
MA	C	A	13	MC	11,522	433	-1.59	-C
MA	D	F	26	MC	7,882	7,633	-1.88	-C
MA	D	A	15	MC	11,303	433	2.00	+C
MA	D	A	17	MC	11,303	433	2.10	+C
MA	D	A	32	MC	11,303	433	-1.92	-C
MA	D	A	43	MC	11,303	433	1.68	+C
MA	D	A	57	MC	11,303	433	-4.26	-C
MA	E	F	26	MC	7,854	7,482	-1.94	-C
MA	E	A	4	MC	11,206	419	-1.96	-C
MA	E	A	8	MC	11,206	419	-1.70	-C
MA	E	A	13	MC	11,206	419	-1.55	-C
MA	E	A	15	MC	11,206	419	1.60	+C
MA	E	A	43	MC	11,206	419	1.71	+C
MA	E	A	60	MC	11,206	419	-1.89	-C
MA	E	A	61	MC	11,206	419	1.84	+C

*F = Female, AA = African American, H = Hispanic, A =Asian

Table 51 Cont'd
Summary for Differential Item Functioning based on Criteria $\pm C$

Content	Form	Focal Group*	Item	Type	N Ref Group	N Focal Group	Delta	Criteria
MA	F	F	26	MC	7,735	7,466	-1.75	-C
MA	F	A	3	MC	11,083	435	1.97	+C
MA	F	A	9	MC	11,083	435	-1.86	-C
MA	F	A	31	MC	11,083	435	-1.53	-C
MA	F	A	43	MC	11,083	435	1.60	+C
MA	G	F	3	MC	4,049	4,017	1.63	+C
MA	G	F	26	MC	4,049	4,017	-1.73	-C
MA	G	A	2	MC	5,930	218	2.59	+C
MA	G	A	3	MC	5,930	218	2.32	+C
MA	G	A	5	MC	5,930	218	2.34	+C
MA	G	A	16	MC	5,930	218	1.89	+C
MA	G	A	17	MC	5,930	218	1.63	+C
MA	G	A	33	MC	5,930	218	1.88	+C
MA	G	A	43	MC	5,930	218	1.54	+C
MA	G	A	63	MC	5,930	218	1.62	+C
MA	G	A	66	MC	5,930	218	1.97	+C
MA	G	AA	2	MC	5,930	1,249	1.62	+C
MA	G	H	4	MC	5,930	567	-1.51	-C
MA	G	H	39	MC	5,930	567	-1.52	-C
MA	H	F	26	MC	4,019	3,930	-1.78	-C
MA	H	A	2	MC	5,859	233	2.24	+C
MA	H	A	3	MC	5,859	233	-2.44	-C
MA	H	A	4	MC	5,859	233	-1.82	-C
MA	H	A	8	MC	5,859	233	-1.90	-C
MA	H	A	13	MC	5,859	233	-2.15	-C
MA	H	A	15	MC	5,859	233	1.89	+C
MA	H	A	16	MC	5,859	233	2.54	+C
MA	H	A	25	MC	5,859	233	2.54	+C
MA	H	A	32	MC	5,859	233	2.00	+C
MA	H	A	39	MC	5,859	233	2.15	+C
MA	H	A	43	MC	5,859	233	1.51	+C
MA	H	A	58	MC	5,859	233	-2.24	-C
MA	H	AA	2	MC	5,859	1,178	1.75	+C
MA	H	H	33	MC	5,859	567	1.82	+C
MA	I	F	26	MC	4,052	3,900	-1.93	-C
MA	I	A	2	MC	5,851	231	1.53	+C
MA	I	A	6	MC	5,851	231	-1.77	-C
MA	I	A	9	MC	5,851	231	-2.05	-C
MA	I	A	12	MC	5,851	231	1.58	+C
MA	I	A	33	MC	5,851	231	1.79	+C

*F = Female, AA = African American, H = Hispanic, A =Asian

Table 51 Cont'd
Summary for Differential Item Functioning based on Criteria $\pm C$

Content	Form	Focal Group*	Item	Type	N Ref Group	N Focal Group	Delta	Criteria
MA	I	A	52	MC	5,851	231	2.06	+C
MA	I	A	57	MC	5,851	231	2.39	+C
MA	I	A	61	MC	5,851	231	-1.82	-C
MA	I	A	64	MC	5,851	231	-1.52	-C
MA	J	F	26	MC	3,972	3,840	-2.12	-C
MA	J	A	2	MC	5,754	201	5.04	+C
MA	J	A	5	MC	5,754	201	2.19	+C
MA	J	A	10	MC	5,754	201	-1.65	-C
MA	J	A	15	MC	5,754	201	2.62	+C
MA	J	A	57	MC	5,754	201	-1.80	-C
MA	J	A	59	MC	5,754	201	1.88	+C
MA	J	AA	2	MC	5,754	1,198	2.02	+C
MA		F	26	MC	63,711	61,571	-1.82	-C

*F = Female, AA = African American, H = Hispanic, A =Asian.

Figure 1
Reading B parameters with all anchor items

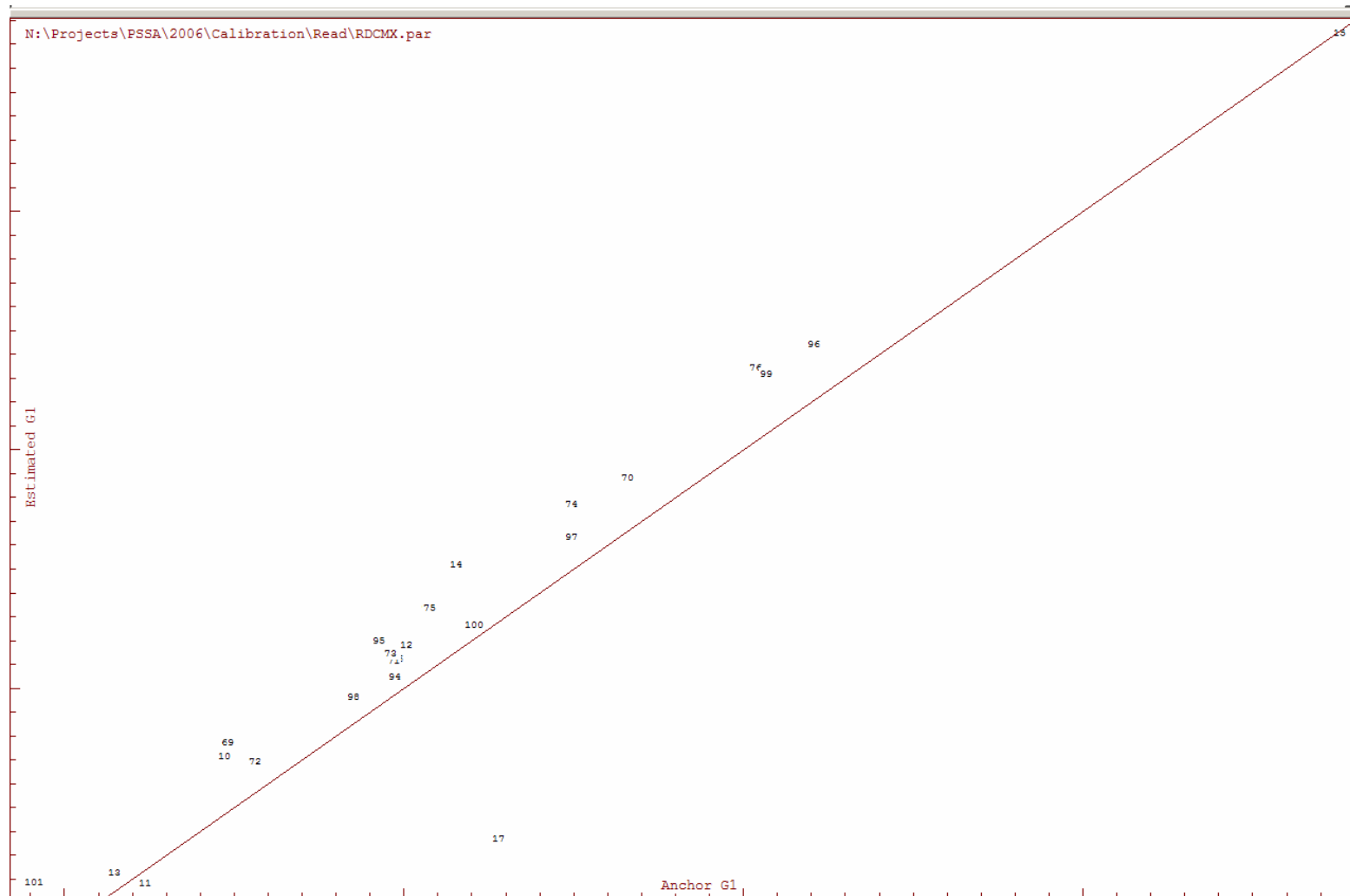


Figure 2
Reading B parameters of anchor items after dropping item #17

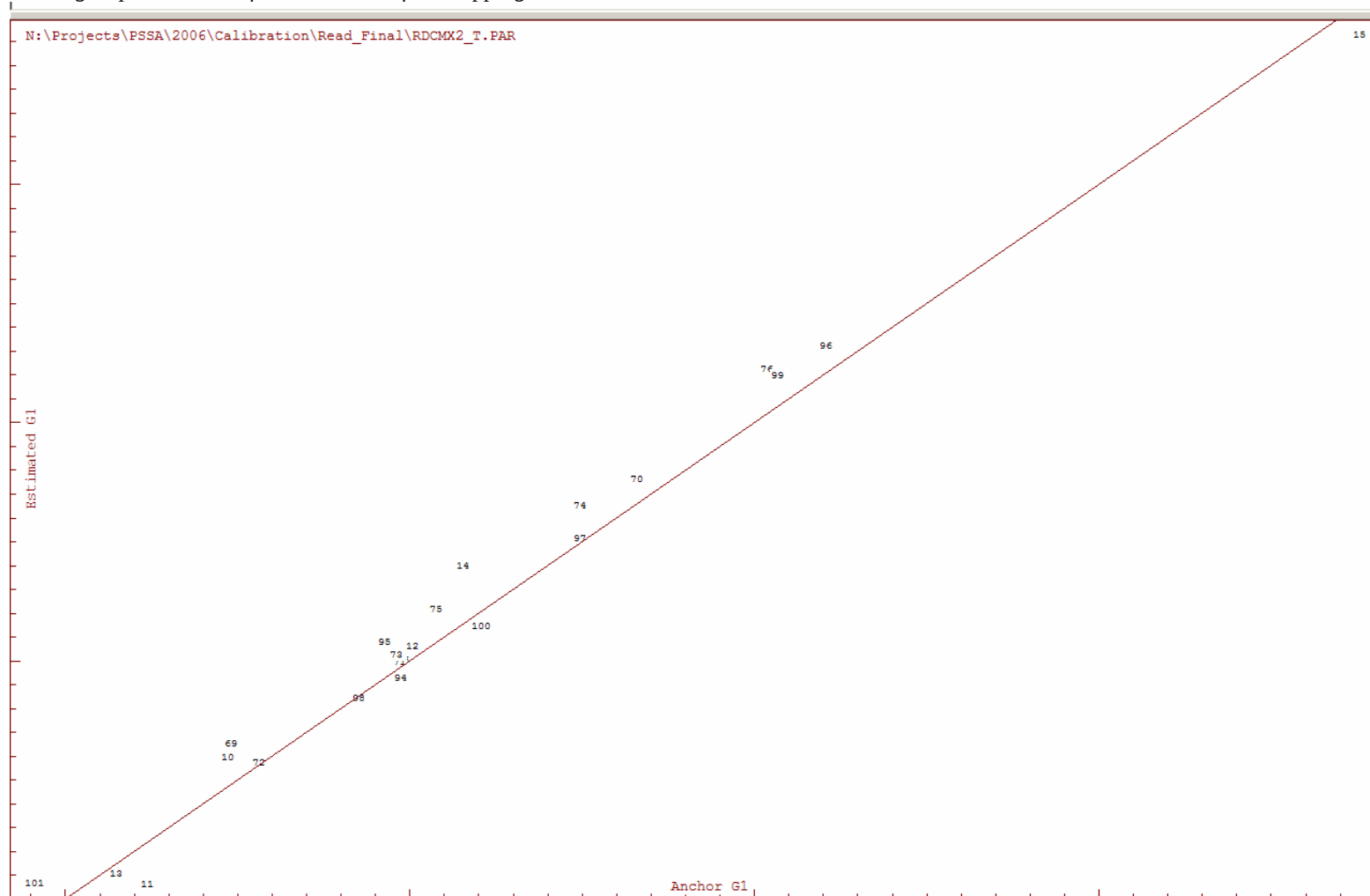


Figure 3
Mathematics B parameters with all anchor items

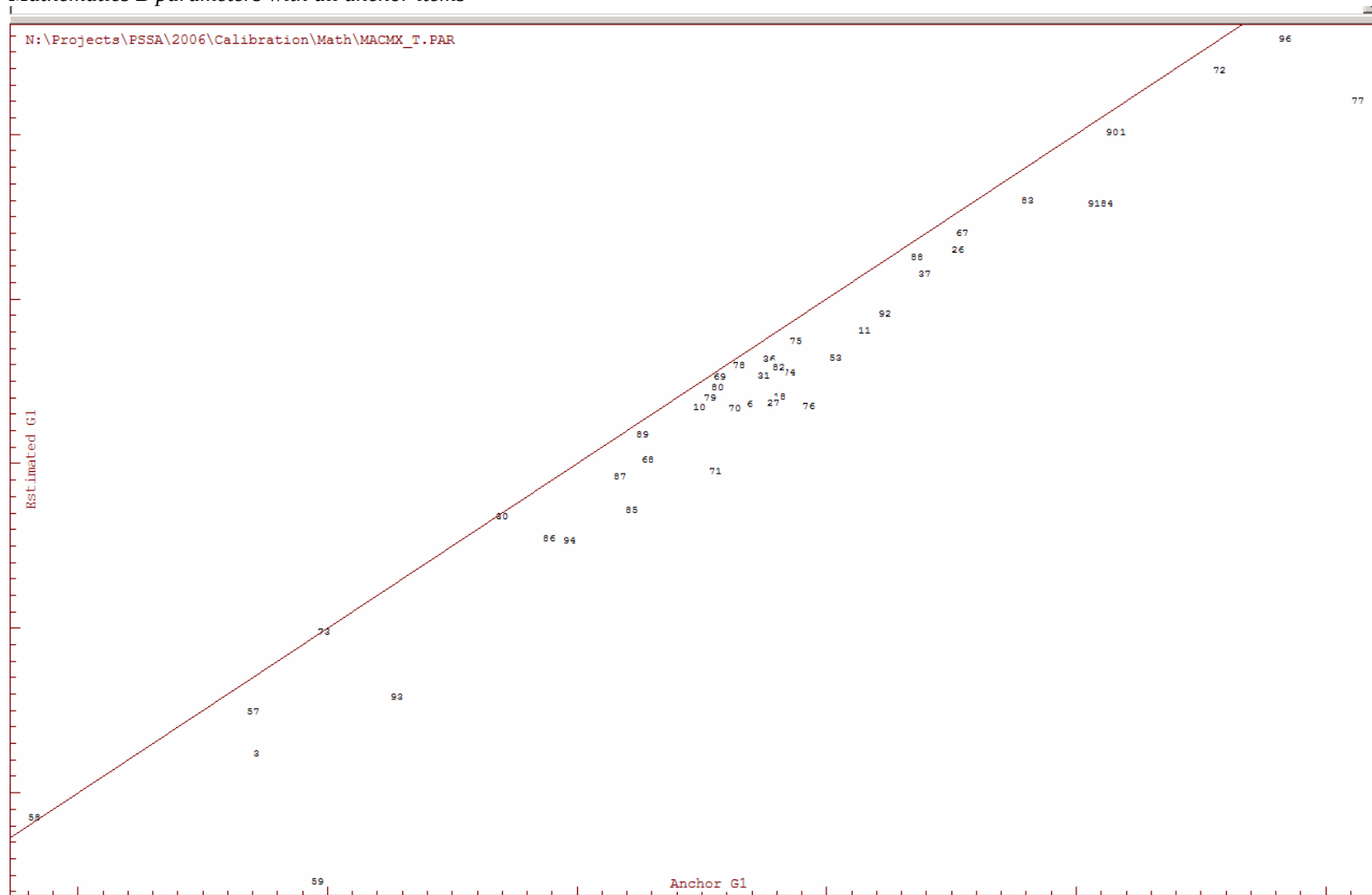


Figure 4
Mathematics B parameters of anchor items after dropping item # 59

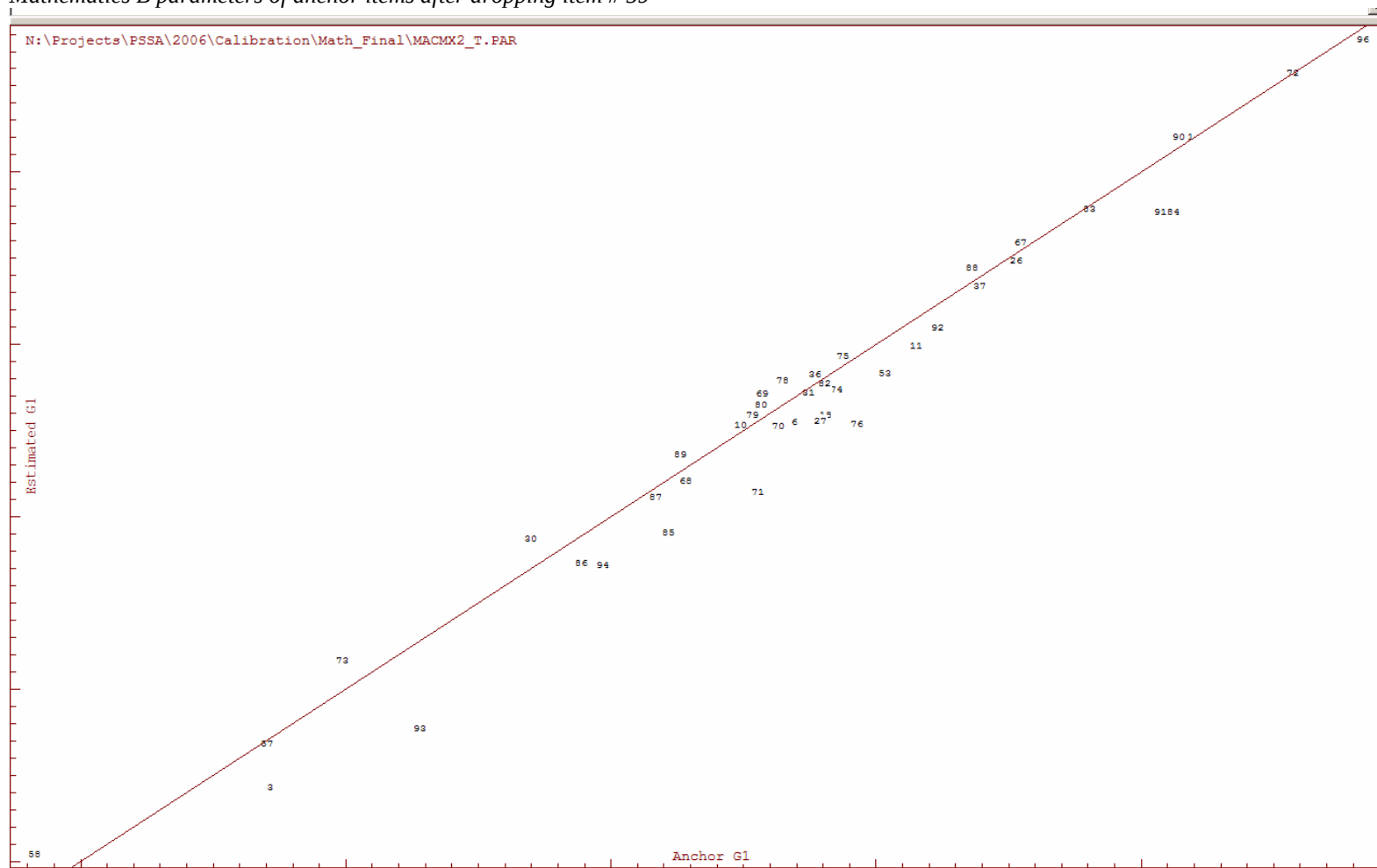


Figure 5
Mathematics Scale Score and SEM

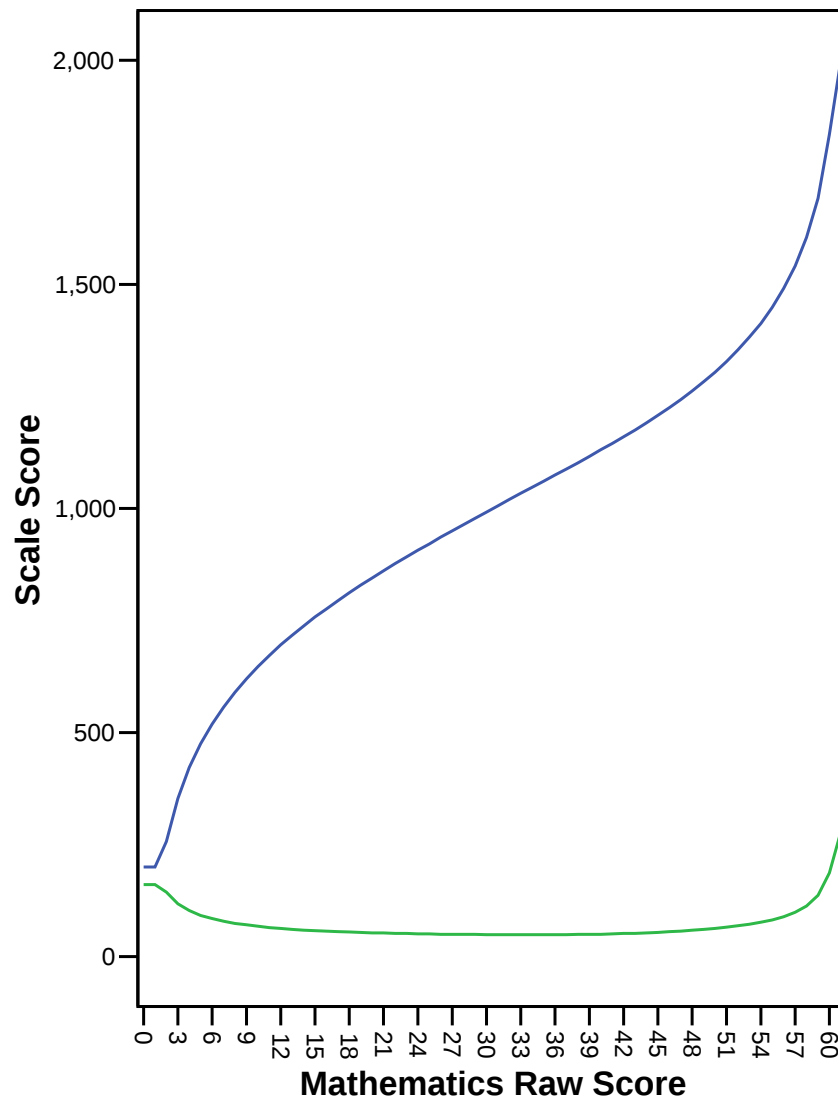


Figure 6
Reading Scale Score and SEM

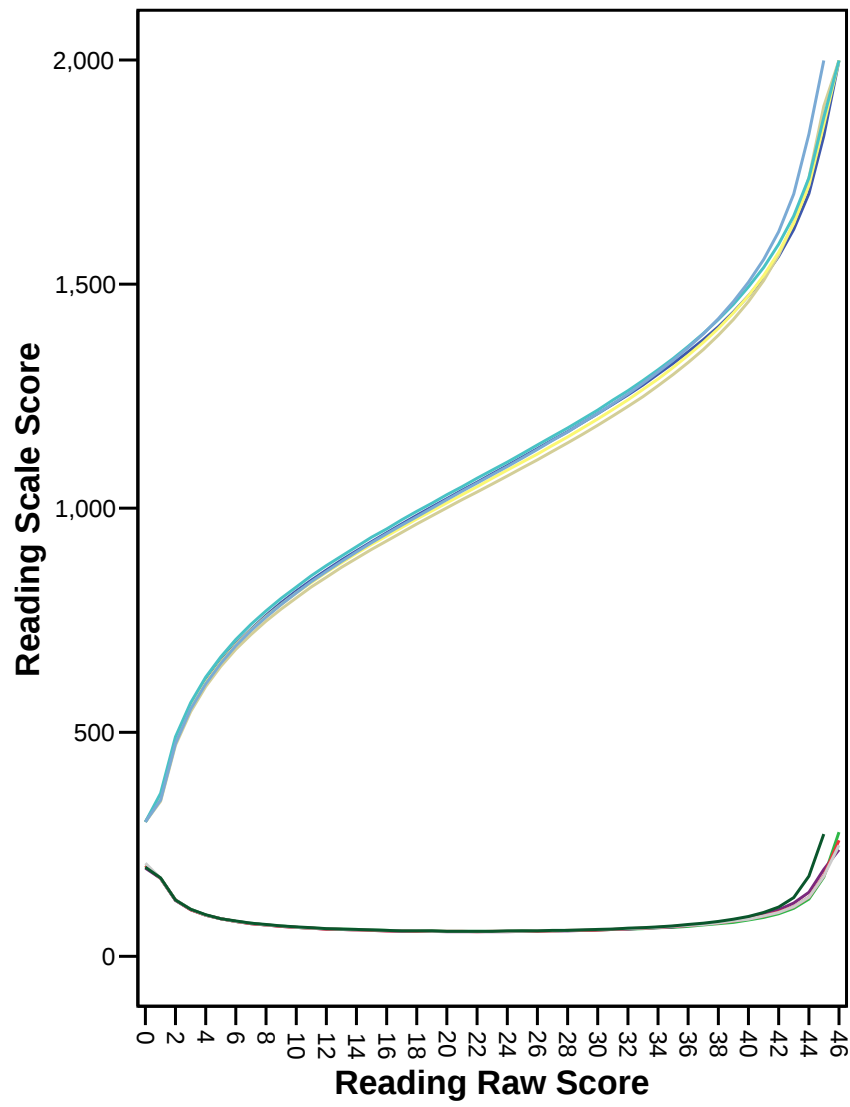


Figure 7
Reading Form 1 Raw Score and Scale Score Distribution

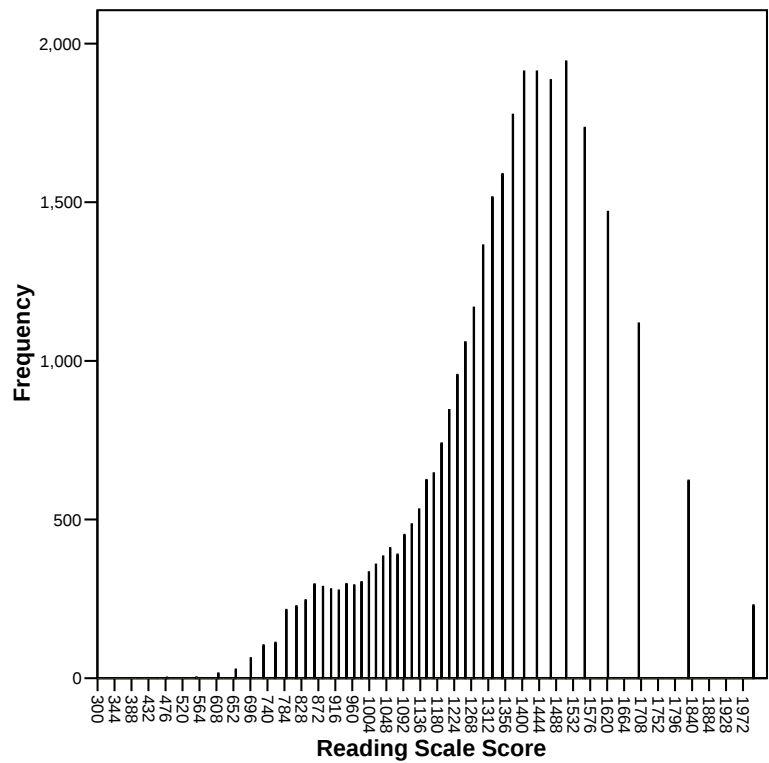
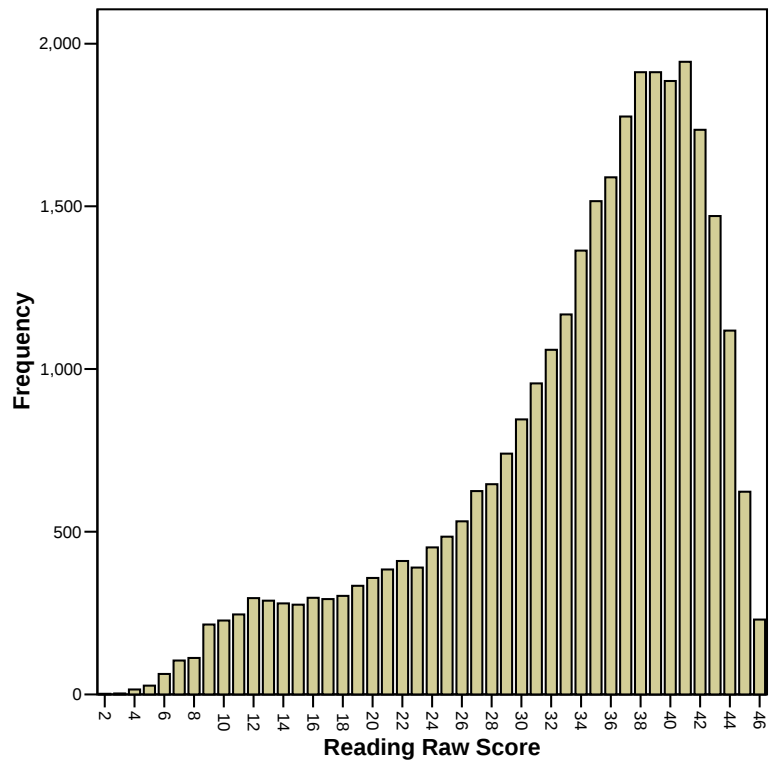


Figure 8
Reading Form 2 Raw Score and Scale Score Distribution

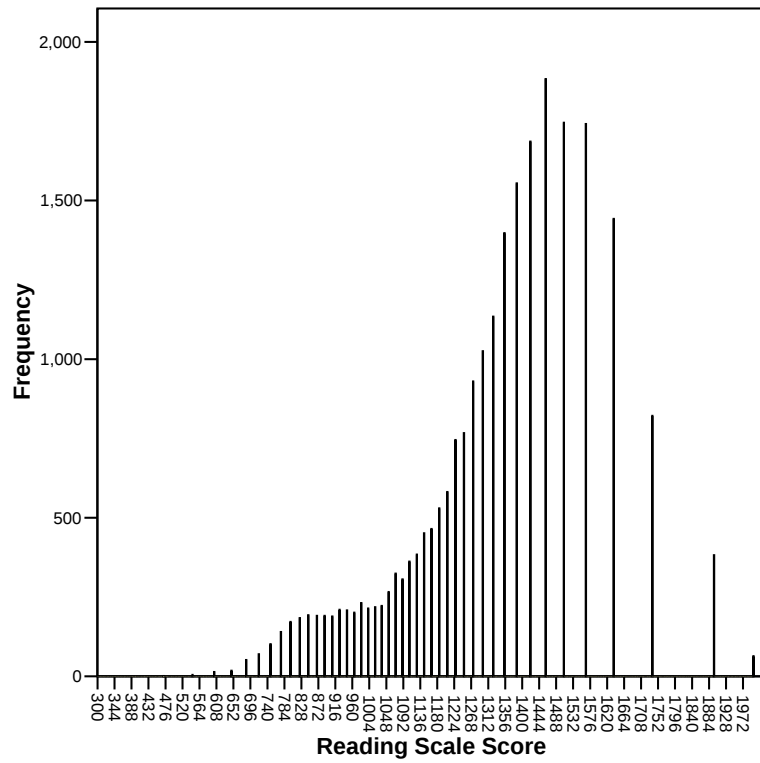
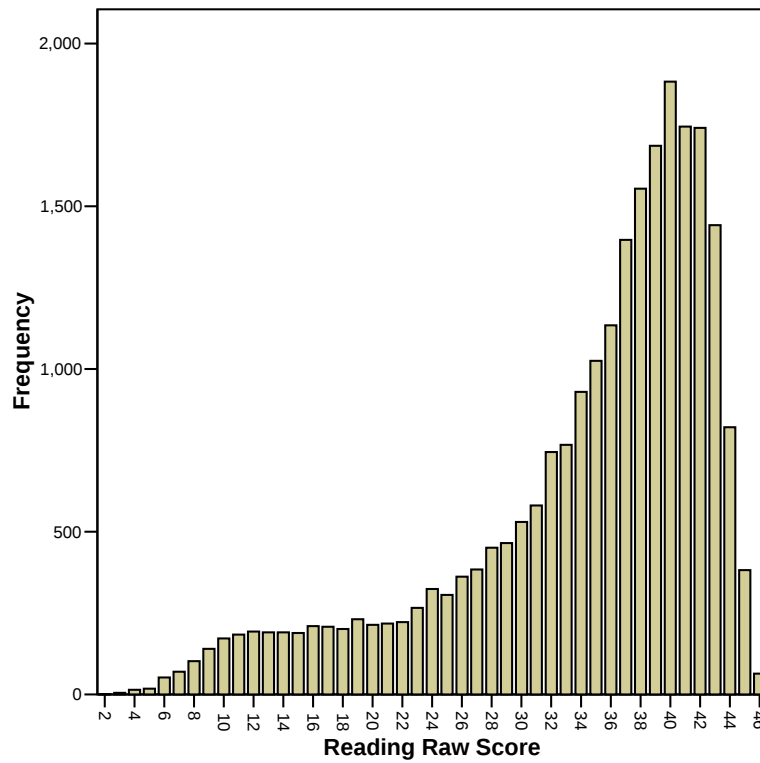


Figure 9
Reading Form 3 Raw Score and Scale Score Distribution

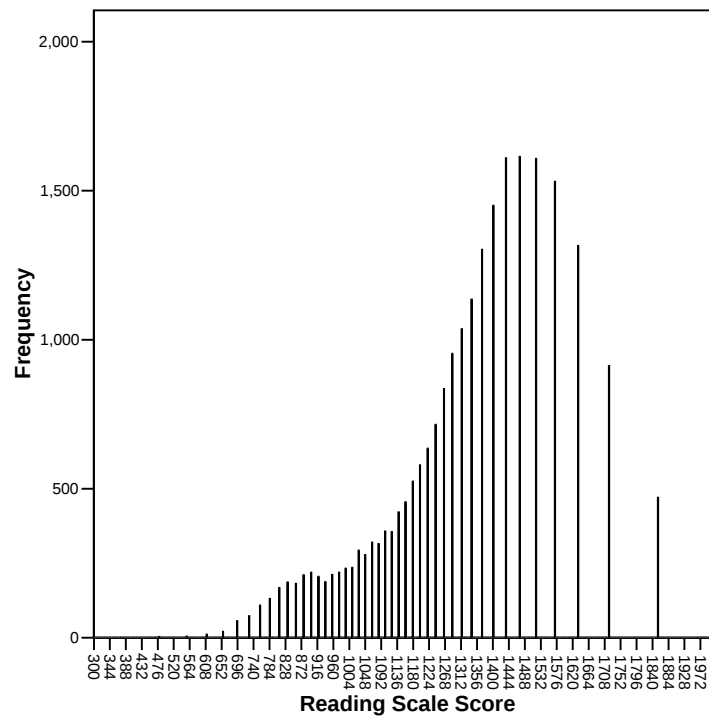
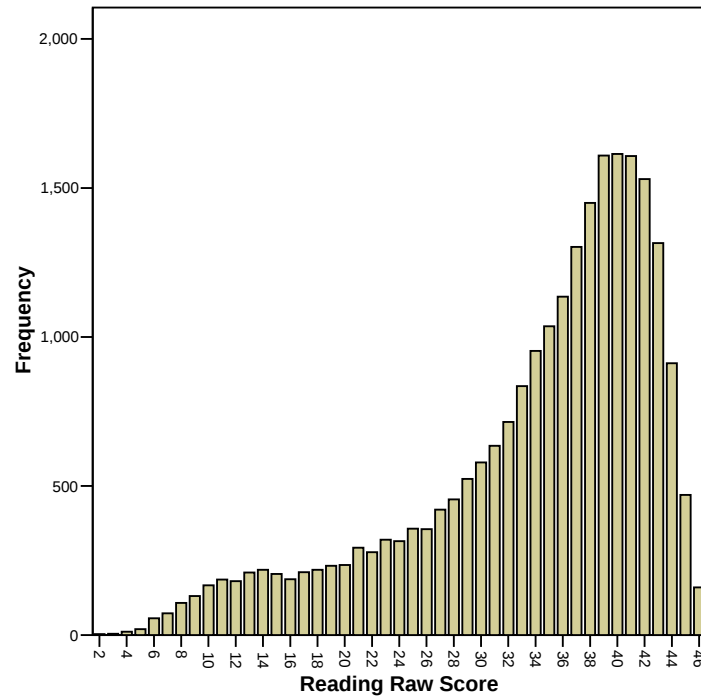


Figure 10
Reading Form 4 Raw Score and Scale Score Distribution

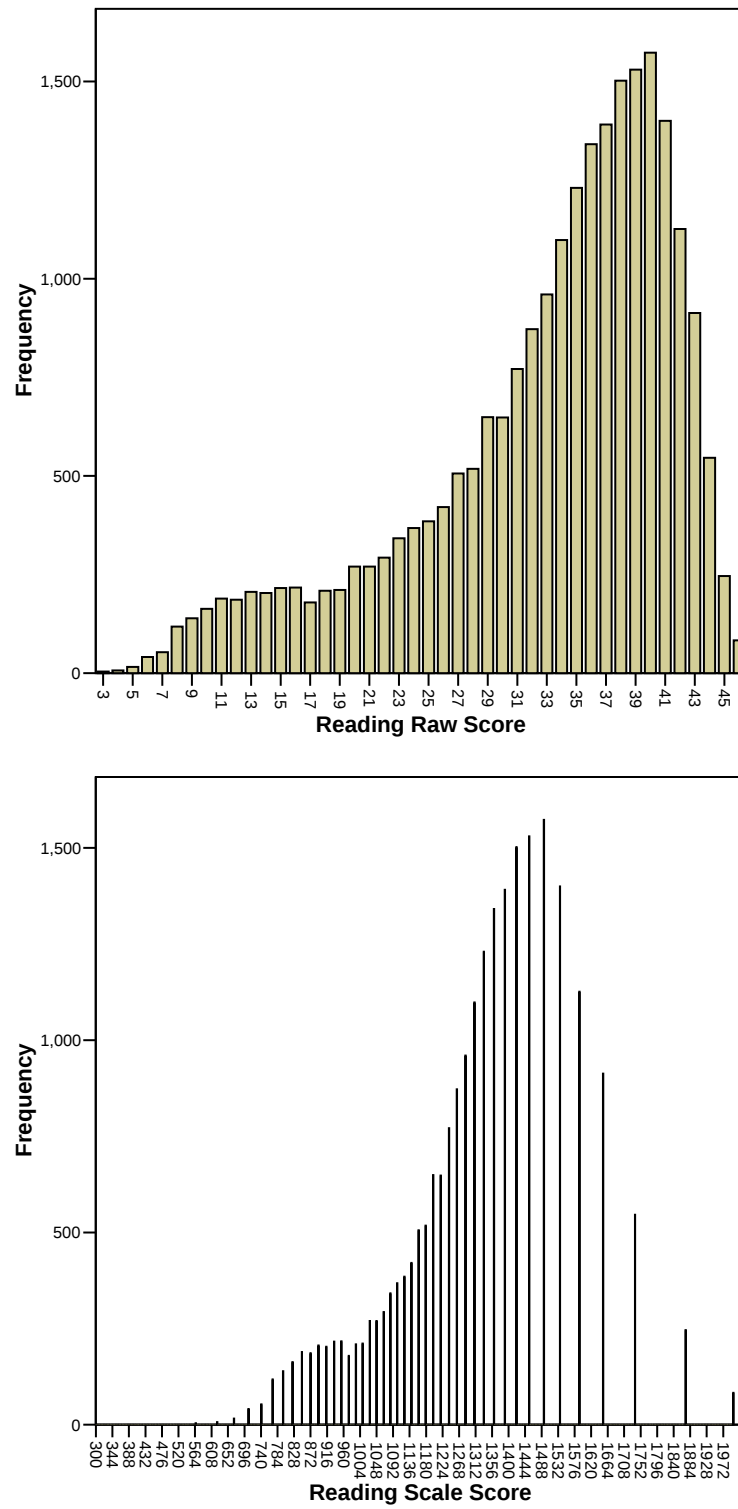


Figure 11
Reading Form 5 Raw Score and Scale Score Distribution

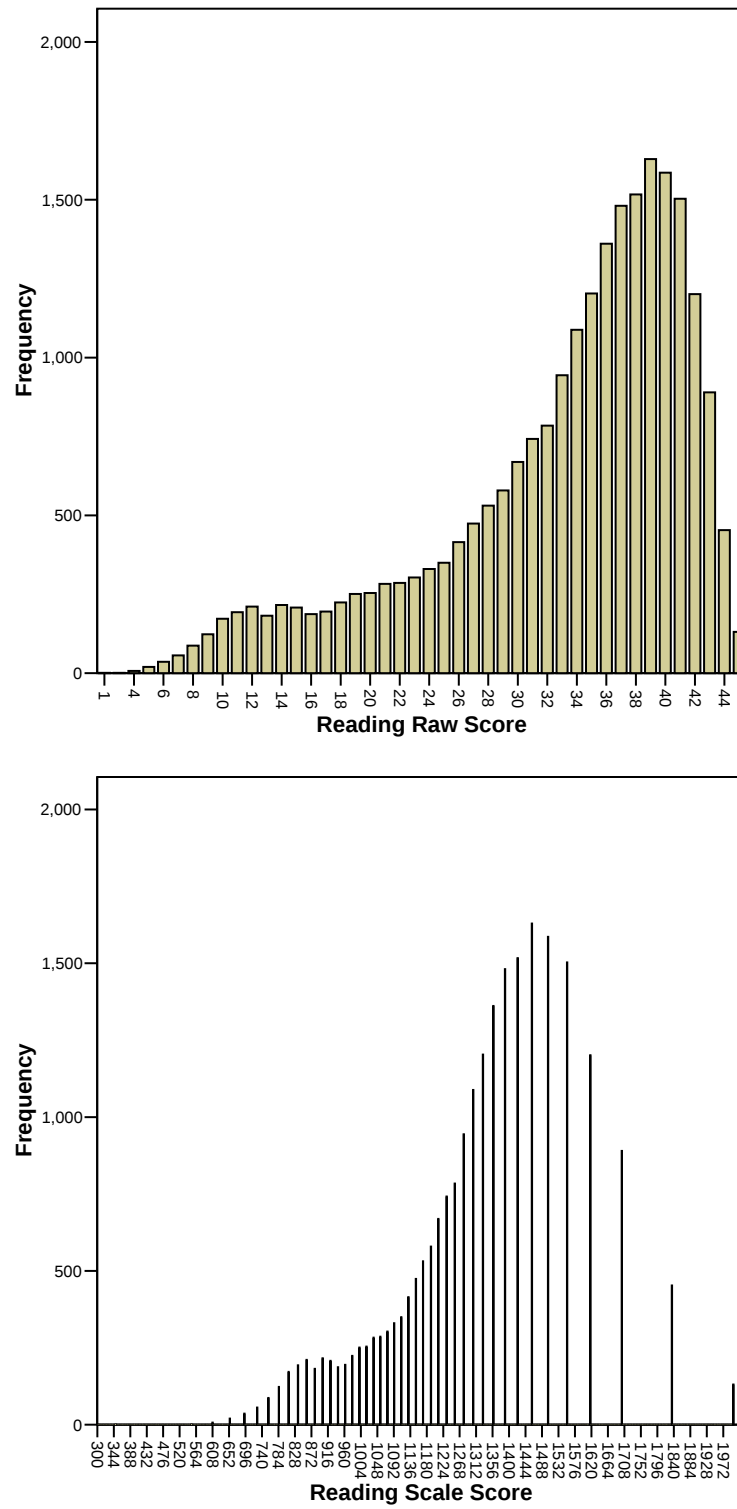


Figure 12
Mathematics Raw Score and Scale Score Distribution

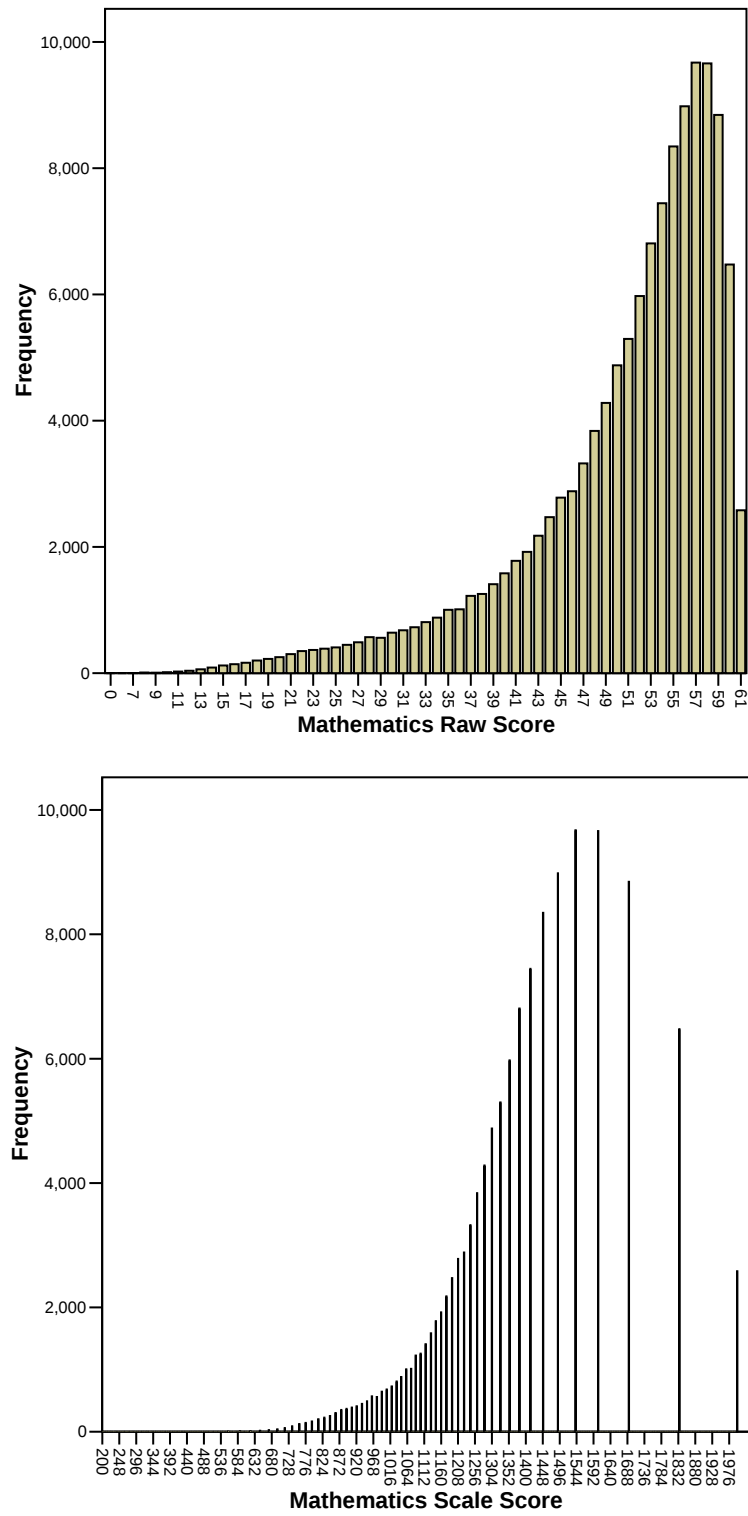
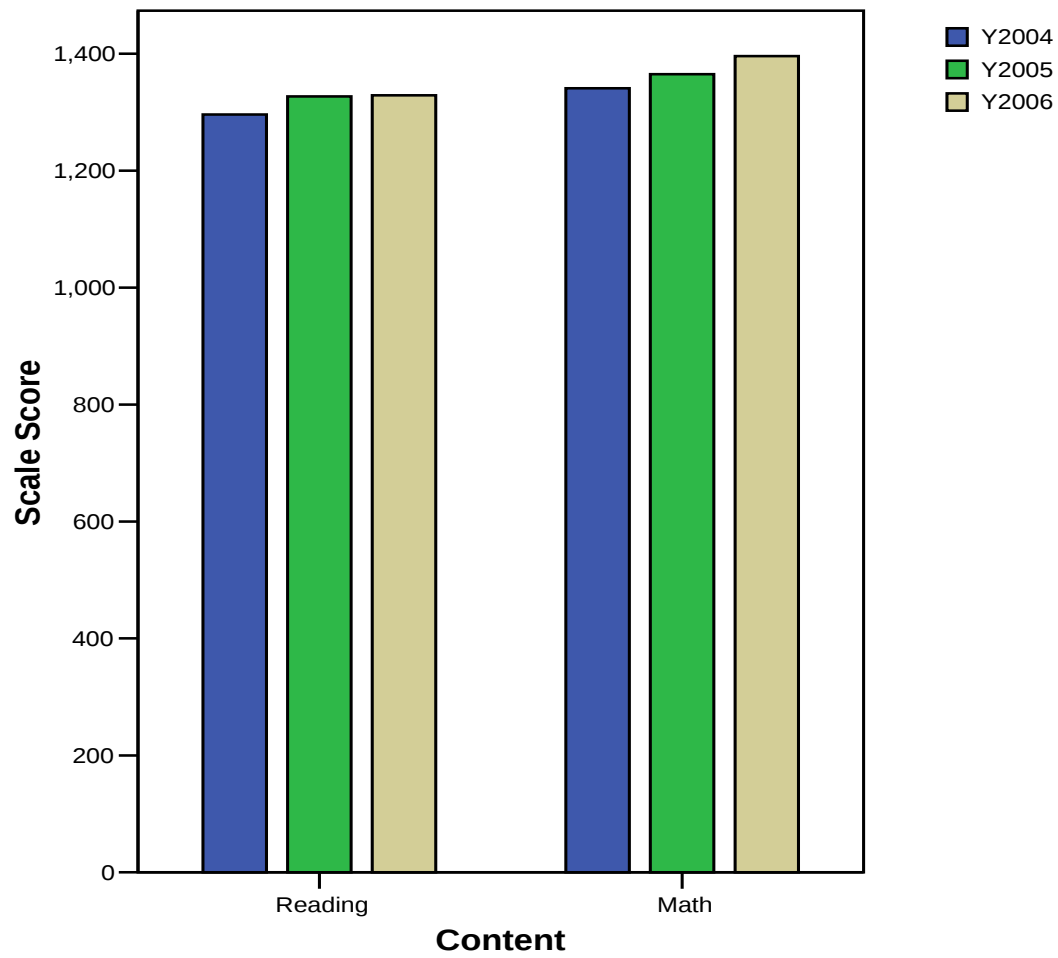


Figure 13
State Mean Scale Scores



References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *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.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41, 687-699.
- Burket, G. R. (1991). PARDUX [Computer program]. Unpublished.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth Group/Thompson Learning.
- Fleiss, J.L., Cohen, J., and Everitt, B.S. (1969). Large-Sample Standard Errors of Kappa and Weighted Kappa. *Psychological Bulletin*, 72, 323 –327.
- Holland, P. W. & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kolen, J. Michael., and Brennan, L. Robert. (1995). *Test equating*. Springer : New York.
- 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.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.
- Stocking, M.L., and Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7.

Appendix:

Reading and Mathematics Academic Standards

The Pennsylvania Grade 3 Reading Test measures the following Pennsylvania Assessment

Anchors and Reporting Categories:

- R3.A Comprehension and Reading Skills
 - R3.A.1 Understanding fiction text appropriate to grade level.
 - R3.A.2 Understanding nonfiction text appropriate to grade level.
- R3.B Interpretation and Analysis of Fiction and Nonfiction Text
 - R3.B.1 Identify components within text.
 - R3.B.2 Identify literary devices.
 - R3.B.3 Identify concepts and organization of nonfiction text.

The Pennsylvania Grade 3 Mathematics Test measures the following Pennsylvania Assessment

Anchors and Reporting Categories:

- MA. Numbers and Operations
 - MA.1 Demonstrate understanding of numbers, ways of representing numbers, relationships among numbers and number systems.
 - MA.2 Understand the meanings of operations, use operations and understand how they relate to each other.
 - MA.3 Compute accurately and fluently and make reasonable estimates.
- MB. Measurement
 - MB.1 Demonstrate an understanding of measurable attributes of objects and figures, and the units, systems and processes of measurement.
 - MB.2 Apply appropriate techniques, tools and formulas to determine measurements.

- MC. Geometry
 - MC.1 Analyze characteristics and properties of 2- and 3-dimensional geometric shapes and demonstrate understanding of geometric relationships.
 - MC.2 Identify and/or apply concepts of transformations or symmetry.
- MD. Algebraic Concepts
 - MD.1 Demonstrate an understanding of patterns, relations and functions.
 - MD.2 Represent and/or analyze mathematical situations using numbers, symbols, words, tables, and/or graphs.
- ME. Data Analysis and Probability
 - ME.1 Formulate or answer questions that can be addressed with data and/or organize, display, interpret, or analyze data.
 - ME.3 Understand and/or apply basic concepts of probability or outcomes.