New York State Regents Examination in Geometry

2011 Field Test Analysis, Equating Procedure, and Scaling of Operational Test Forms

Technical Report

Prepared for the New York State Education Department
by Pearson

December 2011

Copyright

Table of Contents

List of Tables

Section I: Introduction

PURPOSE

The purpose of this report is to document the psychometric work on the New York State Regents Examination in Geometry in 2011. Specifically, contained within this report are procedures for and results of field test analysis, equating, and scaling of operational test forms that were conducted by Pearson. Information on test development can be found in the test design and development report for the New York State Regents Examination in Geometry.

Section II: Field Test Analysis

In May 2011, field testing was conducted for the New York State Regents Examination in Geometry to better understand the psychometric quality of the items. The results of this testing are used to help determine which items will be selected for use on operational tests.

Target student samples for participation in this testing were selected such that each would represent the student population expected to take the operational test. The Need/Resource Capacity Categories were used as variables in the sampling plan. See Table 1 for the seven Need/Resource Capacity Categories and their definitions.

Table 1. Need/Resource Capacity Category Definitions

Need/Resource Capacity (N/RC) Category	Definition
High N/RC Districts: New York City	New York City
Large Cities	Buffalo, Rochester, Syracuse, Yonkers
Urban-Suburban	Districts at or above the 70th percentile on the index with at least 100 students per square mile or enrollment greater than 2500
Rural	All districts at or above the 70th percentile with fewer than 50 students per square mile or enrollment of fewer than 2500
Average N/RC Districts	All districts between the 20th and 70th percentiles on the index
Low N/RC Districts	All districts below the 20th percentile on the index
Charter Schools	Each charter school is a district

The data collected from field testing were scored by two entities. The multiple-choice items were scored by the New York State Education Department and the constructed-response items were scored by Measurement Incorporated. Therefore, it was necessary to combine data files for data analysis. Both classical and item response theory analyses were conducted using the data to evaluate the quality of the test items.

FILE MERGING AND DATA CLEANUP

Field test forms contained multiple-choice and constructed-response item types. Response data were contained in two separate files. The multiple-choice data file contained 19,325 student records and the constructed-response data file contained 18,688 student records. The two files were combined by merging the multiple-choice records and the constructed-response records by unique test booklet numbers. After the exclusion rules were applied, the resulting field test data file contained 18,117 records.

Multiple-choice response data were then compared to the answer key. All item responses not matching the answer key were assigned scores of 0. The responses matching the answer key were assigned scores of 1. With respect to the constructed-response items, scores from 0 to the maximum point value available for each tested item were kept while out of range values were assigned scores of 0. For IRT calibrations, blanks (i.e., missing data) were assigned scores of 0 to be consistent with how operational test items are scored.

The final data file contained both the scored and unscored student responses. Unscored data were used to calculate the percentage of students who selected the various answer choices for the multiple-choice items or the percentage of students who received the range of possible raw score points for the constructed-response items. Thus, the frequency of students leaving items blank can be calculated. The scored data were used for all other analyses.

CLASSICAL ANALYSIS

Classical Test Theory is based on the assumption that an observed test score x is composed of both true score t and error score e. This assumption is expressed as follows:

$$x = t + e$$

In other words, error is associated with measuring a student's true score. For example, the choice of test items or the administration conditions might influence student responses making a student's observed score higher or lower than the student's true score. The error is considered random. After repeated administrations,

the mean of the error scores is virtually zero. Thus, a student's observed score is expected to equal his or her true score. This expectation is expressed as follows:

$$E(x) = t$$

Using a Classical Test Theory framework, field test data can be analyzed to provide information about the quality of test items. Item difficulties, point-biserial correlations, reliability estimates, and various statistics related to rater agreement have been calculated and are summarized in the following section.

Item Difficulty

Item difficulty is an indication of students' performance on a specific item. Because this examination contains polytomous items, item means are not appropriate for comparing difficulty across items. Instead weighted item means were calculated by dividing an item's mean by the maximum points possible for that item.

For multiple-choice items, the item difficulty is the proportion of students who answer an item correctly. If 90% of the student responses to a multiple-choice item are correct, then this item is considered easier than a multiple-choice item with correct responses by 30% of the students.

Point-Biserial Correlation

The point-biserial correlation is another classical statistic that can be used to evaluate items. For multiple-choice items, it is the correlation between students' performance on a given item (correct or incorrect) and overall performance scores. This statistic is used to evaluate how well an item identifies students who understand the concept being measured, and can be generalized for constructed-response items. The possible range for the point-biserial correlation is −1 to 1, with higher values being more desirable.

Table 2 presents a summary of the classical item analysis for each of the field test forms. The first three columns identify the form number, the number of students who took each form, and the number of items on each field test form, respectively. The remaining columns are divided into two sections (i.e., item difficulty and point-biserial correlations). Recall that for constructed-response items, item means were divided by the maximum number of points possible in order to place them in the same metric as the multiple-choice items. All except two items had difficulties that were equal to or below 0.90. With respect to the point-biserial correlations, few of the correlations fell below 0.25.

Table 2. Classical Item Analysis

Form	N-Count	No. of Items	Item Difficulty			Point-Biserial		
			<0.50	0.50 to 0.90	>0.90	<0.25	0.25 to 0.50	>0.50
931	977	10	4	6	0	0	7	3
932	965	10	4	6	0	0	6	4
933	970	10	6	4	0	0	7	3
934	966	10	7	3	0	0	6	4
935	965	10	4	6	0	1	7	2
936	932	10	6	4	0	1	6	3
937	948	10	6	4	0	0	6	4
938	947	10	5	5	0	1	6	3
939	951	10	6	4	0	1	6	3
940	948	10	7	3	0	0	7	3
941	928	10	3	7	0	0	7	3
942	973	10	7	3	0	1	5	4
943	968	10	6	4	0	1	6	3
944	974	10	5	5	0	0	5	5
945	959	10	8	2	0	0	6	4
946	915	10	6	4	0	2	4	4
947	940	10	4	6	0	0	6	4
948	954	10	4	4	2	1	6	3
949	937	10	8	2	0	0	7	3

In addition to the summary information provided in Table 2, all of the classical item statistics are provided in Appendix A. "Max" is the maximum number of possible points. "N-Count" refers to the number of student records in the analysis. "Alpha" contains the internal consistency statistics discussed below. For multiple-choice items, "B" represents the proportion of students who left the item blank and "M1" through "M4" are the proportions of students who selected each of the four answer choices. For constructed-response items, "B" represents the proportion of students who left the item blank and "M0" through "M6" are the proportions of students who received scores 0 through 6. "Mean" is the average of the scores received by the students. The final column contains the point-biserial correlation for each item. There are some instances of items missing statistics; this occurs when an item was not scored.

Test Reliability

Classical analysis can also be used to measure the reliability of the test. Reliability is the consistency of the results obtained from a measurement with respect to time or among items or subjects that constitute a test. As such, test reliability can be estimated in a variety of ways. Internal consistency indices are a measure of how consistently examinees respond to items within a test. Two factors influence estimates of internal consistency: test length and homogeneity of items. In general, the more items on the examination the higher the reliability, and the more similar the items are the higher the reliability.

Cronbach's α (alpha) (Cronbach, 1951) has an important use as a measure of the internal consistency of a test. This formula is the extension of an earlier version, the Kuder-Richardson Formula 20 (KR-20), which is the equivalent for dichotomous items.

Table 3 contains the internal consistency statistics for all of the field test forms. These statistics ranged from 0.56 to 0.71 and are based solely on the items in the individual field test forms. It is expected that these statistics associated with the operational tests would be greater because there are more items on the operational test forms.

Table 3. Test and Scoring Reliability

Form Number	Test Reliability	Scoring Reliability
931	0.57	0.77
932	0.66	0.89
933	0.63	0.94
934	0.67	0.94
935	0.56	0.97
936	0.67	0.93
937	0.67	0.84
938	0.57	0.87
939	0.57	0.96
940	0.69	0.94
941	0.64	0.96
942	0.60	0.94
943	0.60	0.95
944	0.71	0.88
945	0.65	0.88
946	0.61	0.93
947	0.67	0.90
948	0.62	0.91
949	0.64	0.93

Scoring Reliability

One concern with constructed-response items is the reliability of the scoring process (i.e., consistency of the score assignment). Constructed-response items must be read by scorers who assign scores based on a comparison between the rubric and student responses. Consistency in the way scores are assigned is a critical part of the reliability of the assessment. To measure this consistency, 10% of the test booklets are scored a second time (i.e., second read scores) and compared to the original set of scores (i.e., first read scores).

As an overall measure of scoring reliability, the Pearson Correlation Coefficient between the first and second scores for each of the constructed-response items was computed. This statistic is often used as an overall indicator of scoring reliability, and generally ranges from 0 to near 1. Table 3 contains the results from these analyses in the column headed "Scoring Reliability." The correlations ranged from 0.77 to 0.97, indicating high scoring reliability.

Inter-rater Agreement

For each constructed-response item, the difference between the first and second reads was computed. When examining inter-rater agreement statistics, it should be kept in mind that the maximum number of points per item varies, as shown in the "Score Points" column of the following tables.

Table 4 contains the proportion of occurrence of these differences for each item. There were no instances of the first read and second read differing by more than 2.

Table 4. Point Differences Between First and Second Reads

Form	Item	Score Points	Difference (First Read minus Second Read)						
			−3	−2	−1	0	1	2	3
931	08	2	0.00	0.01	0.07	0.86	0.03	0.02	0.00
931	09	2	0.00	0.01	0.10	0.76	0.13	0.01	0.00
931	10	6	0.00	0.05	0.07	0.74	0.10	0.03	0.00
932	08	4	0.00	0.05	0.10	0.81	0.03	0.01	0.00
932	09	2	0.00	0.00	0.04	0.92	0.04	0.00	0.00
932	10	2	0.00	0.00	0.06	0.91	0.03	0.00	0.00
933	08	2	0.00	0.01	0.02	0.87	0.08	0.01	0.00
933	09	2	0.00	0.00	0.04	0.94	0.02	0.01	0.00
933	10	6	0.00	0.04	0.06	0.80	0.05	0.04	0.00
934	08	2	0.00	0.00	0.02	0.95	0.02	0.00	0.00
934	09	2	0.00	0.01	0.03	0.93	0.03	0.01	0.00
934	10	6	0.00	0.03	0.07	0.81	0.06	0.04	0.00
935	08	2	0.00	0.01	0.02	0.88	0.08	0.01	0.00
935	09	2	0.00	0.01	0.08	0.85	0.06	0.00	0.00
935	10	6	0.00	0.02	0.05	0.86	0.05	0.01	0.00
936	08	2	0.00	0.01	0.04	0.92	0.04	0.00	0.00
936	09	4	0.00	0.02	0.10	0.80	0.07	0.01	0.00
936	10	4	0.00	0.02	0.09	0.76	0.11	0.02	0.00
937	08	2	0.00	0.00	0.02	0.93	0.03	0.02	0.00
937	09	2	0.00	0.00	0.05	0.86	0.09	0.00	0.00
937	10	6	0.00	0.04	0.13	0.73	0.08	0.03	0.00
938	08	2	0.00	0.01	0.11	0.76	0.12	0.01	0.00
938	09	2	0.00	0.00	0.10	0.82	0.08	0.01	0.00
938	10	6	0.00	0.04	0.11	0.70	0.15	0.00	0.00

Table 4. Point Differences Between First and Second Reads *(continued)*

Form	Item	Score Points	Difference (First Read minus Second Read)						
			−3	−2	−1	0	1	2	3
939	08	2	0.00	0.01	0.04	0.91	0.04	0.01	0.00
939	09	6	0.00	0.06	0.07	0.74	0.09	0.04	0.00
939	10	4	0.00	0.02	0.03	0.86	0.07	0.01	0.00
940	08	2	0.00	0.00	0.05	0.87	0.08	0.00	0.00
940	09	4	0.00	0.03	0.05	0.87	0.04	0.02	0.00
940	10	4	0.00	0.01	0.08	0.77	0.12	0.01	0.00
941	08	2	0.00	0.00	0.04	0.93	0.03	0.00	0.00
941	09	2	0.00	0.00	0.03	0.92	0.04	0.01	0.00
941	10	6	0.00	0.02	0.05	0.87	0.05	0.01	0.00
942	08	2	0.00	0.00	0.08	0.85	0.08	0.00	0.00
942	09	2	0.00	0.00	0.09	0.80	0.11	0.00	0.00
942	10	6	0.00	0.02	0.06	0.83	0.08	0.01	0.00
943	08	2	0.00	0.00	0.01	0.98	0.01	0.00	0.00
943	09	4	0.00	0.03	0.05	0.78	0.09	0.06	0.00
943	10	4	0.00	0.03	0.02	0.85	0.07	0.02	0.00
944	08	2	0.00	0.00	0.05	0.86	0.08	0.01	0.00
944	09	4	0.00	0.05	0.12	0.68	0.13	0.01	0.00
944	10	4	0.00	0.02	0.05	0.85	0.07	0.02	0.00
945	08	2	0.00	0.00	0.10	0.77	0.11	0.02	0.00
945	09	2	0.00	0.01	0.04	0.87	0.08	0.01	0.00
945	10	6	0.00	0.01	0.15	0.73	0.08	0.04	0.00
946	08	2	0.00	0.02	0.03	0.90	0.03	0.02	0.00
946	09	2	0.00	0.00	0.04	0.92	0.04	0.00	0.00
946	10	6	0.00	0.02	0.08	0.77	0.09	0.04	0.00
947	08	2	0.00	0.00	0.05	0.91	0.04	0.00	0.00
947	09	4	0.00	0.03	0.14	0.75	0.07	0.02	0.00
947	10	4	0.00	0.02	0.12	0.73	0.12	0.01	0.00
948	08	2	0.00	0.00	0.05	0.88	0.06	0.00	0.00
948	09	4	0.00	0.02	0.21	0.59	0.15	0.03	0.00
948	10	4	0.00	0.04	0.12	0.74	0.08	0.02	0.00
949	08	2	0.00	0.01	0.06	0.90	0.04	0.00	0.00
949	09	4	0.00	0.02	0.07	0.80	0.10	0.01	0.00

Table 4. Point Differences Between First and Second Reads *(continued)*

Form	Item	Score Points	Difference (First Read minus Second Read)						
			−3	−2	−1	0	1	2	3
949	10	4	0.00	0.04	0.09	0.74	0.09	0.04	0.00

Table 5 contains additional summary information regarding the first and second reads. In the fifth column the percent of exact matches between the first and second scores is provided. "Adj." is the percentage of differences with a magnitude of 1. "Total" is the sum of the two prior columns and contains values between 89.8% and 100%. These values indicate a high degree of agreement.

Table 5. First and Second Read Descriptive Statistics and Agreement

Form	Item	Score Points	Total N-Count	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intraclass Correla-tion	Wt. Kappa
				Exact	Adj.	Total	First Read	Second Read	First Read	Second Read		
931	08	2	165	86.1	10.3	96.4	0.2	0.2	0.51	0.51	0.52	0.44
931	09	2	177	76.3	22.6	98.9	1.2	1.1	0.85	0.88	0.82	0.73
931	10	6	168	74.4	17.3	91.7	0.4	0.4	0.79	0.82	0.61	0.46
932	08	4	168	81.0	13.1	94.0	0.3	0.4	0.75	0.89	0.75	0.58
932	09	2	171	91.8	8.2	100.0	0.8	0.8	0.76	0.77	0.93	0.90
932	10	2	168	91.1	8.9	100.0	1.1	1.1	0.89	0.88	0.94	0.91
933	08	2	165	87.3	10.9	98.2	0.3	0.2	0.65	0.63	0.78	0.69
933	09	2	168	94.0	5.4	99.4	1.0	1.0	0.81	0.81	0.94	0.93
933	10	6	165	80.0	11.5	91.5	1.7	1.7	1.99	2.01	0.94	0.86
934	08	2	170	95.3	4.7	100.0	0.3	0.3	0.62	0.61	0.94	0.89
934	09	2	174	93.1	5.7	98.9	0.3	0.3	0.62	0.64	0.87	0.84
934	10	6	166	80.7	12.7	93.4	1.0	1.0	1.80	1.82	0.94	0.84
935	08	2	170	88.2	10.0	98.2	1.3	1.2	0.89	0.93	0.90	0.85
935	09	2	182	85.2	14.3	99.5	1.5	1.5	0.77	0.73	0.86	0.78
935	10	6	164	86.0	11.0	97.0	2.5	2.5	2.52	2.54	0.98	0.94
936	08	2	166	91.6	7.8	99.4	0.7	0.7	0.83	0.83	0.93	0.90
936	09	4	163	79.8	17.2	96.9	1.1	1.2	1.42	1.44	0.93	0.84
936	10	4	160	76.3	20.0	96.3	1.2	1.2	1.55	1.56	0.93	0.82
937	08	2	166	92.8	4.8	97.6	1.0	1.0	0.92	0.92	0.92	0.90
937	09	2	160	85.6	14.4	100.0	0.8	0.7	0.77	0.77	0.88	0.82
937	10	6	160	72.5	21.2	93.8	0.4	0.5	0.86	0.92	0.71	0.53
938	08	2	174	75.9	23.0	98.9	0.8	0.8	0.89	0.88	0.82	0.72
938	09	2	168	81.5	17.9	99.4	1.1	1.1	0.77	0.74	0.82	0.77
938	10	6	164	69.5	26.2	95.7	0.8	0.9	1.44	1.56	0.91	0.73
939	08	2	166	91.0	7.2	98.2	0.9	0.9	0.93	0.93	0.92	0.89

Table 5. First and Second Read Descriptive Statistics and Agreement *(continued)*

Form	Item	Score Points	Total N-Count	Agreement (%) Exact	Agreement (%) Adj.	Agreement (%) Total	Raw Score Mean First Read	Raw Score Mean Second Read	Raw Score Standard Deviation First Read	Raw Score Standard Deviation Second Read	Intraclass Correlation	Wt. Kappa
939	09	6	157	73.9	15.9	89.8	2.7	2.7	2.61	2.60	0.96	0.87
939	10	4	161	86.3	10.6	96.9	0.9	0.9	1.43	1.40	0.94	0.87
940	08	2	165	86.7	13.3	100.0	0.6	0.5	0.85	0.82	0.91	0.83
940	09	4	157	86.6	8.9	95.5	0.5	0.6	1.22	1.23	0.91	0.81
940	10	4	165	77.0	20.6	97.6	1.9	1.9	1.56	1.57	0.94	0.85
941	08	2	166	93.4	6.6	100.0	0.2	0.2	0.55	0.55	0.89	0.81
941	09	2	162	92.0	7.4	99.4	1.0	0.9	0.84	0.84	0.93	0.90
941	10	6	156	87.2	10.3	97.4	0.7	0.8	1.74	1.74	0.97	0.88
942	08	2	170	84.7	15.3	100.0	0.8	0.8	0.95	0.92	0.91	0.84
942	09	2	170	80.0	20.0	100.0	0.9	0.9	0.81	0.78	0.84	0.77
942	10	6	169	82.8	13.6	96.4	1.2	1.2	1.92	1.87	0.96	0.89
943	08	2	165	98.2	1.8	100.0	0.1	0.1	0.37	0.40	0.94	0.89
943	09	4	169	77.5	13.6	91.1	1.4	1.3	1.70	1.71	0.92	0.82
943	10	4	163	84.7	9.8	94.5	1.7	1.7	1.75	1.79	0.95	0.89
944	08	2	172	86.0	12.8	98.8	1.0	0.9	0.88	0.89	0.89	0.84
944	09	4	164	68.3	25.6	93.9	0.9	1.0	1.30	1.33	0.86	0.71
944	10	4	172	84.9	11.6	96.5	0.5	0.5	1.12	1.11	0.90	0.79
945	08	2	174	77.0	21.3	98.3	0.8	0.7	0.84	0.84	0.80	0.72
945	09	2	169	87.0	11.8	98.8	0.9	0.9	0.89	0.90	0.90	0.85
945	10	6	171	72.5	22.8	95.3	1.2	1.3	1.39	1.44	0.90	0.78
946	08	2	164	90.2	6.1	96.3	0.8	0.8	0.92	0.92	0.88	0.86
946	09	2	159	92.5	7.5	100.0	0.9	0.9	0.92	0.91	0.95	0.92
946	10	6	155	77.4	16.8	94.2	1.0	1.0	1.90	1.77	0.94	0.82
947	08	2	164	90.9	9.1	100.0	0.6	0.6	0.86	0.88	0.94	0.89
947	09	4	170	74.7	20.6	95.3	1.7	1.8	1.49	1.51	0.91	0.82
947	10	4	164	73.2	23.2	96.3	0.7	0.7	0.83	0.86	0.73	0.63
948	08	2	165	88.5	11.5	100.0	0.5	0.5	0.74	0.73	0.89	0.83
948	09	4	168	58.9	36.3	95.2	1.4	1.4	1.51	1.50	0.88	0.72
948	10	4	171	74.3	19.9	94.2	1.5	1.5	1.53	1.57	0.91	0.81
949	08	2	164	89.6	9.8	99.4	0.8	0.8	0.85	0.86	0.92	0.88
949	09	4	156	80.1	17.3	97.4	1.5	1.5	1.58	1.59	0.94	0.87
949	10	4	163	74.2	18.4	92.6	1.4	1.4	1.71	1.66	0.92	0.81

* Adj. = difference of 1

Constructed-Response Item Means and Standard Deviations

The average score for each constructed-response item was computed based on the first and second reads. In addition, the standard deviation of the scores was computed.

Table 5 contains the means and standard deviations for the first and second read scores. The largest difference between the item means for the first and second read scores was 0.1, while there were minimal differences among standard deviation statistics.

Intraclass Correlation

The intraclass correlation was computed for each item. This correlation is an estimate of the reliability of scoring based on an average of the first and second read scores. Correlations greater than 0.60 are considered very strong because they explain more than one-third of the variance in scores. All items except one had intraclass correlations greater than or equal to 0.61 (See Table 5). Consistent with other information provided in the table, these values indicate a very high level of scoring reliability.

Weighted Kappa

Weighted Kappa (Cohen, 1968) was calculated for each item based on the first and second reads. This statistic produces an estimate of the reliability of the score classifications relative to what would be expected to occur by chance.

Weighted Kappa is an estimate of the reliability of the score classifications. That is, the Kappa statistic is a measure of reproducibility for categorical data. Guidelines for the evaluation of this statistic are:

- $k > 0.75$ denotes excellent reproducibility
- $0.4 < k \leq 0.75$ denotes good reproducibility
- $0 < k \leq 0.4$ denotes marginal reproducibility

The results found in Table 5 show a high degree of consistency between the first and second reads. The Weighted Kappa statistics ranged from 0.44 to 0.94, which in all cases indicates good-to-excellent reproducibility.

Based on the scoring reliability analyses, there is strong evidence that the scoring of the constructed-response items was performed in a highly reliable manner.

ITEM RESPONSE THEORY (IRT) STATISTICS

As discussed above, the item mean is a statistic used to evaluate item difficulty. However, many different test forms are used during field testing and different samples of students are responding to these items. The average ability of the different samples of students varies and a direct comparison of item means across test forms may lead to inaccurate interpretations. Therefore, Item Response Theory (IRT) was also used to evaluate item difficulty.

Specifically, the Rasch Partial Credit Model (PCM) (Masters, 1982) was used. With the use of this model, the difficulty of items and the ability of examinees are placed on the same metric. Thus, the difficulty of an item and the ability of a person can be meaningfully compared across field test forms. Also, the use of this model provides greater flexibility in situations where different samples or test forms are used because the parameters generated are generally not considered to be sample dependent or test dependent. A description of this model, results of item calibration, and item fit evaluation are presented below.

The PCM provides an overall difficulty estimate for each item. Specifically, for constructed-response items when there are several points possible, individual estimates of difficulty for each of the possible score points are also calculated (i.e., step values). Each step value represents the difficulty of a student receiving a particular score point, given that he or she has already received the prior score point. For example, if a 3-point item had step values of −1.0, 1.0, and 0.0, one could say that it is relatively easy to obtain a score of 1. However, it is much more difficult to obtain a 2 given the student has the ability to score a 1 because the difference in difficulty between a 1 and a 2 is much greater than the difference between a 0 and a 1. Also, the difference between a 2 and a 3 is not as great as the difference between a 1 and a 2. Thus, with this example, a small step is needed to go from a 0 to a 1, a large step is needed to move from a 1 to a 2, and a moderate step is needed to proceed from a 2 to a 3.

Item Calibration

As discussed above, the use of Rasch item difficulty statistics provides an advantage over the use of classical item means because they can be compared across test forms. Students from different samples responded to the various test forms. Although the samples were selected to be similar with respect to student ability, there are differences. By equating the test forms (See the Equating Procedure section below), the Rasch item difficulties account for those differences and these statistics can be compared across test forms.

Rasch item difficulty values generally range from −3.00 to +3.00. An item with a Rasch difficulty greater than +2.0 is considered very difficult and should be examined carefully. If the item is measuring an important concept that students are having difficulty with, then the item can be useful. However, if the item is measuring a trivial

concept or is written in a confusing manner, then it might not be appropriate to use on an operational test form. Likewise, any item with a Rasch difficulty less than −2.0 is considered very easy and usually provides little information regarding student achievement. The vast majority of test items should range between −2.0 and +2.0. This range represents approximately two standard deviations around the average difficulty of 0. Thus, one would expect that, based on chance, roughly 5% of the items will fall outside of that range and, therefore, these are items that should be closely examined for content.

Item Fit Evaluation

The INFIT statistic is used to determine whether items are functioning in a way that is congruent with the assumptions of the Rasch model. Under these assumptions, how a student will respond to an item depends on the proficiency of the student and the difficulty of the item, both of which are on the same measurement scale. If an item is as difficult as a student is able, the student will have a 50% chance of getting the item correct. If a student is more able than an item is difficult, under the assumptions of the Rasch model, that student has a greater than 50% chance of correctly answering the item. On the other hand, if the item is more difficult than the student is able, he or she has a less than 50% chance of correctly responding to the item. Rasch fit statistics estimate the extent to which an item is functioning in this predicted manner. Items showing a poor fit with the Rasch model typically have values outside the range of 0.7 to 1.3.

Table 6 contains a summary of the Partial Credit Model item analysis for each of the field test forms. The first column lists the form numbers. The next two columns list the number of students who participated and the number of items on each field test form, respectively. The remaining columns are divided into two sections. The first section pertains to the Rasch item difficulties, while the second pertains to the INFIT statistics. Nearly all of the items fell within the moderate −2.0 to +2.0 difficulty range, and only one item had an INFIT statistic outside the typical range.

Table 6. Partial Credit Model Item Analysis

Form	N-Count	No. of Items	Rasch <−2.0	Rasch −2.0 to 2.0	Rasch >2.0	INFIT <−0.70	INFIT −0.70 to 1.30	INFIT >1.30
931	977	10	0	8	2	0	10	0
932	965	10	0	9	1	0	10	0
933	970	10	0	8	2	0	10	0
934	966	10	0	8	2	0	10	0
935	965	10	0	10	0	0	10	0
936	932	10	0	10	0	0	9	1
937	948	10	0	9	1	0	10	0
938	947	10	1	9	0	0	10	0
939	951	10	0	10	0	0	10	0
940	948	10	0	10	0	0	10	0
941	928	10	0	9	1	0	10	0
942	973	10	0	10	0	0	10	0
943	968	10	0	9	1	0	10	0
944	974	10	0	10	0	0	10	0
945	959	10	0	10	0	0	10	0
946	915	10	0	9	1	0	10	0
947	940	10	0	9	1	0	10	0
948	954	10	2	8	0	0	10	0
949	937	10	0	9	1	0	10	0

All of the individual IRT item statistics are provided in Appendix B. The column entitled "RID" contains the Rasch item difficulty statistics. S1–S6 contain the step values for the constructed-response items. Finally, "INFIT" contains the INFIT statistic for each item.

DIFFERENTIAL ITEM FUNCTIONING (DIF) STATISTICS

Statistical procedures are employed to observe whether, on the basis of data, there exists the possibility of unfair treatment of different populations. DIF statistics are used to identify items for which members of a focal group have a different probability of getting the items correct than members of a reference group after the groups have been matched on ability level on the test.

For the multiple-choice items, the Mantel-Haenszel Delta (MHD) DIF statistics were computed (Dorans & Holland, 1992) to classify test items in three levels of DIF for each

comparison: negligible DIF (A), moderate DIF (B), and large DIF (C). An item was flagged if it exhibited a B or C category of DIF, using the following rules derived from National Assessment of Educational Progress (NAEP) guidelines (Allen, Carlson, & Zalanak, 1999):

- MHD not significantly different from 0 (based on alpha = 0.05) **or** |MHD| < 1.0 are classified as A.

- MHD significantly different from 0 and {|MHD| ≥ 1.0 and < 1.5} **or**
  MHD not significantly different from 0 and |MHD| ≥ 1.0 are classified as B.

- |MHD| ≥ 1.5 and significantly different from 0 are classified as C.

For the constructed-response items, the effect size of the standardized mean difference (SMD) was used to flag DIF. The SMD reflects the size of the differences in performance on constructed-response items between student groups matched on the total score. It is the difference between the unweighted item mean of the focal group and the weighted item mean of the reference group. The weights applied to the reference group are applied so that the weighted number of reference group students is the same as in the focal group (within the same ability group). The SMD is divided by the total group item standard deviation to get a measure of the effect size (ES) for the SMD. The SMD effect size groups each item into one of three categories: negligible DIF (AA), moderate DIF (BB), and large DIF (CC). Only categories BB and CC were flagged in the results.

- Probability is > 0.05 **or** if |ES| is ≤ 0.17, classified as AA.

- Probability is > 0.05 and if 0.17 < |ES| ≤ 0.25, classified as BB.

- Probability is > 0.05 and if |ES| is > 0.25, classified as CC.

Although DIF statistics are typically conducted by gender and ethnicity, the low n-counts for ethnic subgroups did not allow for these statistics to be meaningful. The n-counts for gender allowed for comparisons to be made, but were still somewhat low, so resulting statistics should be interpreted with caution.

The DIF statistics for gender are shown in Appendix C. Flagging of items appears in the "DIF Category" column and if an item is flagged, the "Favored Group" column indicates which gender is favored.

Section III: Equating Procedure

The 2011 field test administration for the New York State Regents Examination in Geometry consisted of 17 field test forms numbered 933–949 and two anchor forms labeled 931 and 932. The field test forms contained multiple-choice and constructed-response items. Each student participating in the field test was administered one of the 19 test forms. The test forms were spiraled within the classroom so that the groups of students taking each form were equivalent. A complete listing of these field test forms can be seen in Appendix A where item type (e.g., multiple-choice, constructed-response) and the maximum points for each item are displayed.

The anchor forms were equated to the item bank using a common-item equating design. The anchor item difficulty parameters were fixed to their 2010 item bank values. This places the item difficulty estimates and the ability estimates of the students taking the anchor forms onto the item bank scale. After the anchor forms were placed onto the bank scale, the average of the two mean ability estimates for the two forms was computed using ability estimates of nonextreme students. This average ability estimate was used to equate the remaining field test forms, as well as update the item parameters for the anchor forms.

As a part of the anchor item equating, an item-stability check was performed. After fixing all of the items to their 2010 bank values, any item with a displacement value with a magnitude greater than 0.30 was no longer fixed and the test form was reanalyzed. If more than one item had a displacement value with a magnitude greater than 0.30, then the item with the largest displacement was freed and the test form was reanalyzed. In a stepwise fashion, this procedure was repeated until all remaining fixed anchor items had displacements with magnitudes less than, or equal to, 0.30.

Applying the anchor item-stability check to each of the anchor forms resulted in one item having a displacement value with a magnitude greater than 0.30. For form 931, one item was freed and for form 932 no items were freed. This indicates a strong level of stability in the items used on the anchor forms.

The equated mean ability estimate for form 931 was 0.32 and for form 932 the mean was 0.44. This produced a target mean ability estimate of 0.38 which was used for the remainder of the equating process.

After the anchor forms were equated and the target mean was computed, the field test forms were equated using the equivalent groups design. The first step was to calibrate each form separately where all the item parameters were free to estimate (without constraint). From those initial calibrations, the mean ability estimates for each field test form were obtained. The second step was to determine the equating constant for each form by subtracting the mean ability for a given field test form from the target mean ability calculated from the anchor forms (i.e., forms 931 and 932). The respective

equating constant was then added to each of the item parameters on a given form. If the resulting mean of the ability estimates for those students did not equal that of the target mean, then the procedure was repeated until the mean abilities for each of the field test forms equaled the target mean ability. Table 7 shows the mean abilities and constants used for the equating.

Table 7. Initial Mean Abilities and Equating Constants

Form Number	Mean Ability	Constant
931	−0.22	0.57
932	0.04	0.33
933	−0.32	0.67
934	−0.41	0.75
935	0.16	0.21
936	−0.03	0.39
937	−0.18	0.54
938	0.08	0.29
939	−0.11	0.47
940	−0.51	0.85
941	−0.09	0.45
942	−0.09	0.45
943	−0.32	0.67
944	−0.12	0.48
945	−0.57	0.91
946	−0.36	0.70
947	0.04	0.33
948	0.52	−0.13
949	−0.37	0.71

The equated item parameters for the field test items can now be compared across test forms. since the equating process places all items on the same scale. In addition, when items are combined to form unique operational test forms, raw score-to-scale score tables can be generated based on these parameters. The following section contains a description of the development of the operational test forms and scoring tables.

Section IV: Scaling of Operational Test Forms

Operational test items are selected based on content coverage, content accuracy, and statistical quality. The sets of items on each operational test conform to the coverage suggested by content experts. These expert judgments are based on the learning standards established by the New York State Education Department. With respect to statistical quality, classical and Rasch statistics are examined to determine how well items function. Also, items are selected such that they range in difficulty in order to measure students across ability levels. Appendix D contains the 2011 operational test maps with content information regarding each item included on the January 2011, June 2011, and August 2011 operational test forms.

In order to limit wide fluctuations of raw scores that correspond to scale scores of 65 and 85 across administrations, the average Rasch item difficulty for the operational test is considered. For this examination, an average Rasch difficulty of approximately 0.279 is used as a target for each administration. In most cases, meeting this target will provide raw scores of similar magnitude to other forms. However, differences with these scores also occur due to the distribution of the Rasch item difficulty parameters.

Scoring tables display the relationship between raw scores on the operational test and assigned scale scores. Appendix E contains the scoring tables used for the January 2011, June 2011, and August 2011 operational test forms. Four steps are taken in order to produce these tables and resulting conversion charts.

The first step is to develop a raw score (i.e., number of points on the test form) to theta (i.e., student ability) to scale score relationship for the baseline operational test form. This relationship is determined when standards are set and then used for every administration moving forward until the standards are revisited. The baseline form was determined by the New York State Education Department to be June 2009. The raw score-to-theta relationship from that examination was used, and then scale scores are calculated based on the raw score cuts according to the following formula:

$$p(x) = m_3 x^3 + m_2 x^2 + m_1 x + m_0$$

The raw score of zero was assigned a scale score of zero and the maximum raw score was assigned a scale score of 100. The raw scores corresponding to the scale scores of 65 and 85 were also fixed. The polynomial relationship shown above was then used to assign all scale scores to the remaining raw scores. The resulting values for m_1–m_3 are the transformation constants used to produce the final raw score-to-scale score table.

The second step is to develop a raw score-to-theta relationship for the new operational test form, using the field test equated PCM item parameters. This is

accomplished by doing a calibration where all items are anchored to their field test parameters. One modification that is made is that for 6-point items, a constant based on existing bank values is used in place of the field test parameters. The number of points on the test form (i.e., raw score) expected across student ability levels is based on the difficulty of the items on the form. Thus, given a particular student ability level (i.e., theta), if the points are more difficult to earn on the new test than the points on the June 2009 test, the number of points expected of this student on the new test will be less than the number of points expected of this student on the baseline form.

The third step is to use linear interpolation to determine the raw score-to-theta-to-scale score relationship for the new test. The theta values associated with scale scores of 65 and 85 on the baseline form are used along with the raw score-to-theta relationship developed in the previous step. In other words, the baseline 65 and 85 theta values are used as reference points and linear interpolation assigns the other scale scores.

Finally, a conversion chart is created based on the scoring table generated in the third step. Scale scores are rounded to the nearest whole number in all cases except for 0, 65, 85, and 100. A raw score of zero is assigned a scale score of zero. The maximum raw score is assigned a scale score of 100. With respect to the 65 and 85 scale scores, the raw scores with scale scores of 65 or 85, after rounding, are assigned those values.

References

Allen, N. L., Carlson, J. E., and Zalanak, C. A. (1999). *The NAEP 1996 technical report.* Washington, DC: National Center for Education Statistics.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*, 213–220.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Dorans, N. J., & Holland, P. W. (1992). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.) *Differential item functioning: Theory and practice (*35–66). Hillsdale, NJ: Erlbaum.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.

Appendix A: Classical Item Analysis

Table 8. Classical Item Analysis

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	M6	Mean	Point-Biserial
2011_Geom_FT	931	MC	01	1	977	0.57	0.00		0.17	0.55	0.17	0.11			0.55	0.48
2011_Geom_FT	931	MC	02	1	977	0.57	0.00		0.78	0.03	0.03	0.16			0.78	0.36
2011_Geom_FT	931	MC	03	1	977	0.57	0.01		0.58	0.11	0.20	0.11			0.58	0.30
2011_Geom_FT	931	MC	04	1	977	0.57	0.01		0.05	0.21	0.17	0.57			0.57	0.52
2011_Geom_FT	931	MC	05	1	977	0.57	0.01		0.39	0.25	0.22	0.12			0.25	0.29
2011_Geom_FT	931	MC	06	1	977	0.57	0.01		0.22	0.05	0.60	0.12			0.60	0.42
2011_Geom_FT	931	MC	07	1	977	0.57	0.01		0.42	0.04	0.49	0.04			0.49	0.42
2011_Geom_FT	931	CR	08	2	977	0.57	0.13	0.77	0.04	0.06					0.16	0.43
2011_Geom_FT	931	CR	09	2	977	0.57	0.03	0.27	0.21	0.49					1.19	0.61
2011_Geom_FT	931	CR	10	6	977	0.57	0.10	0.66	0.16	0.06	0.01	0.00	0.01	0.00	0.40	0.61
2011_Geom_FT	932	MC	01	1	965	0.66	0.01		0.17	0.03	0.06	0.73			0.73	0.46
2011_Geom_FT	932	MC	02	1	965	0.66	0.00		0.05	0.32	0.61	0.02			0.61	0.50
2011_Geom_FT	932	MC	03	1	965	0.66	0.00		0.15	0.53	0.22	0.10			0.53	0.31
2011_Geom_FT	932	MC	04	1	965	0.66	0.00		0.68	0.10	0.12	0.10			0.68	0.49
2011_Geom_FT	932	MC	05	1	965	0.66	0.00		0.15	0.14	0.17	0.55			0.55	0.51
2011_Geom_FT	932	MC	06	1	965	0.66	0.00		0.14	0.07	0.45	0.34			0.45	0.34
2011_Geom_FT	932	MC	07	1	965	0.66	0.02		0.32	0.56	0.07	0.03			0.56	0.44
2011_Geom_FT	932	CR	08	4	965	0.66	0.03	0.81	0.07	0.02	0.06	0.01			0.33	0.55
2011_Geom_FT	932	CR	09	2	965	0.66	0.04	0.38	0.40	0.18					0.76	0.64
2011_Geom_FT	932	CR	10	2	965	0.66	0.09	0.32	0.22	0.37					0.96	0.67
2011_Geom_FT	933	MC	01	1	970	0.63	0.01		0.12	0.37	0.35	0.15			0.35	0.41
2011_Geom_FT	933	MC	02	1	970	0.63	0.00		0.06	0.12	0.16	0.65			0.65	0.46

Table 8. Classical Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	M6	Mean	Point-Biserial
2011_Geom_FT	933	MC	03	1	970	0.63	0.00		0.15	0.63	0.18	0.04			0.63	0.35
2011_Geom_FT	933	MC	04	1	970	0.63	0.00		0.19	0.06	0.49	0.25			0.49	0.50
2011_Geom_FT	933	MC	05	1	970	0.63	0.01		0.62	0.18	0.10	0.09			0.62	0.55
2011_Geom_FT	933	MC	06	1	970	0.63	0.01		0.16	0.15	0.54	0.14			0.54	0.33
2011_Geom_FT	933	MC	07	1	970	0.63	0.01		0.53	0.18	0.16	0.11			0.18	0.39
2011_Geom_FT	933	CR	08	2	970	0.63	0.03	0.82	0.07	0.09					0.24	0.48
2011_Geom_FT	933	CR	09	2	970	0.63	0.10	0.29	0.27	0.34					0.94	0.55
2011_Geom_FT	933	CR	10	6	970	0.63	0.08	0.33	0.23	0.09	0.02	0.12	0.04	0.09	1.66	0.87
2011_Geom_FT	934	MC	01	1	966	0.67	0.01		0.10	0.41	0.34	0.14			0.41	0.44
2011_Geom_FT	934	MC	02	1	966	0.67	0.01		0.21	0.28	0.17	0.32			0.32	0.46
2011_Geom_FT	934	MC	03	1	966	0.67	0.01		0.09	0.05	0.73	0.12			0.73	0.48
2011_Geom_FT	934	MC	04	1	966	0.67	0.01		0.46	0.16	0.16	0.21			0.46	0.51
2011_Geom_FT	934	MC	05	1	966	0.67	0.01		0.12	0.17	0.36	0.34			0.34	0.43
2011_Geom_FT	934	MC	06	1	966	0.67	0.01		0.13	0.17	0.64	0.05			0.64	0.48
2011_Geom_FT	934	MC	07	1	966	0.67	0.01		0.14	0.12	0.08	0.65			0.65	0.34
2011_Geom_FT	934	CR	08	2	966	0.67	0.07	0.74	0.08	0.11					0.30	0.58
2011_Geom_FT	934	CR	09	2	966	0.67	0.06	0.72	0.12	0.10					0.32	0.59
2011_Geom_FT	934	CR	10	6	966	0.67	0.06	0.57	0.11	0.11	0.01	0.06	0.01	0.07	1.07	0.84
2011_Geom_FT	935	MC	01	1	965	0.56	0.00		0.19	0.12	0.14	0.56			0.56	0.46
2011_Geom_FT	935	MC	02	1	965	0.56	0.01		0.32	0.23	0.37	0.08			0.37	0.45
2011_Geom_FT	935	MC	03	1	965	0.56	0.00		0.06	0.08	0.21	0.65			0.65	0.43
2011_Geom_FT	935	MC	04	1	965	0.56	0.01		0.15	0.07	0.67	0.10			0.67	0.43
2011_Geom_FT	935	MC	05	1	965	0.56	0.01		0.19	0.70	0.06	0.03			0.70	0.41

Table 8. Classical Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	M6	Mean	Point-Biserial
2011_Geom_FT	935	MC	06	1	965	0.56	0.01		0.31	0.37	0.18	0.13			0.37	0.20
2011_Geom_FT	935	MC	07	1	965	0.56	0.02		0.13	0.31	0.26	0.28			0.26	0.31
2011_Geom_FT	935	CR	08	2	965	0.56	0.04	0.27	0.13	0.56					1.25	0.59
2011_Geom_FT	935	CR	09	2	965	0.56	0.03	0.12	0.15	0.71					1.56	0.50
2011_Geom_FT	935	CR	10	6	965	0.56	0.08	0.31	0.14	0.08	0.03	0.08	0.05	0.24	2.36	0.84
2011_Geom_FT	936	MC	01	1	932	0.67	0.00		0.65	0.04	0.15	0.16			0.65	0.50
2011_Geom_FT	936	MC	02	1	932	0.67	0.00		0.08	0.11	0.04	0.76			0.76	0.41
2011_Geom_FT	936	MC	03	1	932	0.67	0.00		0.09	0.78	0.07	0.05			0.78	0.45
2011_Geom_FT	936	MC	04	1	932	0.67	0.02		0.22	0.18	0.15	0.43			0.43	0.45
2011_Geom_FT	936	MC	05	1	932	0.67	0.01		0.13	0.14	0.36	0.36			0.36	0.50
2011_Geom_FT	936	MC	06	1	932	0.67	0.01		0.21	0.14	0.55	0.09			0.55	0.47
2011_Geom_FT	936	MC	07	1	932	0.67	0.03		0.17	0.33	0.37	0.10			0.33	0.16
2011_Geom_FT	936	CR	08	2	932	0.67	0.08	0.46	0.30	0.16					0.62	0.57
2011_Geom_FT	936	CR	09	4	932	0.67	0.11	0.40	0.17	0.07	0.13	0.12			1.20	0.74
2011_Geom_FT	936	CR	10	4	932	0.67	0.08	0.48	0.11	0.08	0.06	0.19			1.21	0.78
2011_Geom_FT	937	MC	01	1	948	0.67	0.00		0.67	0.08	0.07	0.17			0.67	0.46
2011_Geom_FT	937	MC	02	1	948	0.67	0.01		0.18	0.61	0.14	0.07			0.61	0.46
2011_Geom_FT	937	MC	03	1	948	0.67	0.00		0.24	0.34	0.22	0.19			0.34	0.39
2011_Geom_FT	937	MC	04	1	948	0.67	0.00		0.05	0.02	0.85	0.08			0.85	0.36
2011_Geom_FT	937	MC	05	1	948	0.67	0.00		0.07	0.26	0.06	0.61			0.61	0.56
2011_Geom_FT	937	MC	06	1	948	0.67	0.00		0.27	0.30	0.30	0.12			0.27	0.29
2011_Geom_FT	937	MC	07	1	948	0.67	0.01		0.20	0.37	0.10	0.32			0.32	0.35
2011_Geom_FT	937	CR	08	2	948	0.67	0.06	0.39	0.19	0.36					0.91	0.70

Table 8. Classical Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	M6	Mean	Point-Biserial
2011_Geom_FT	937	CR	09	2	948	0.67	0.07	0.37	0.37	0.19					0.75	0.64
2011_Geom_FT	937	CR	10	6	948	0.67	0.07	0.66	0.13	0.09	0.02	0.01	0.01	0.01	0.49	0.67
2011_Geom_FT	938	MC	01	1	947	0.57	0.01		0.41	0.37	0.18	0.04			0.41	0.23
2011_Geom_FT	938	MC	02	1	947	0.57	0.01		0.14	0.37	0.15	0.33			0.33	0.30
2011_Geom_FT	938	MC	03	1	947	0.57	0.00		0.10	0.26	0.51	0.13			0.51	0.41
2011_Geom_FT	938	MC	04	1	947	0.57	0.00		0.08	0.75	0.05	0.12			0.75	0.40
2011_Geom_FT	938	MC	05	1	947	0.57	0.00		0.04	0.05	0.90	0.01			0.90	0.26
2011_Geom_FT	938	MC	06	1	947	0.57	0.01		0.65	0.13	0.05	0.16			0.65	0.35
2011_Geom_FT	938	MC	07	1	947	0.57	0.02		0.36	0.20	0.24	0.19			0.36	0.32
2011_Geom_FT	938	CR	08	2	947	0.57	0.07	0.49	0.14	0.30					0.74	0.68
2011_Geom_FT	938	CR	09	2	947	0.57	0.02	0.25	0.39	0.34					1.07	0.55
2011_Geom_FT	938	CR	10	6	947	0.57	0.09	0.54	0.18	0.09	0.01	0.03	0.03	0.02	0.80	0.77
2011_Geom_FT	939	MC	01	1	951	0.57	0.00		0.08	0.62	0.09	0.21			0.62	0.46
2011_Geom_FT	939	MC	02	1	951	0.57	0.00		0.09	0.12	0.52	0.27			0.52	0.49
2011_Geom_FT	939	MC	03	1	951	0.57	0.00		0.07	0.27	0.46	0.19			0.46	0.36
2011_Geom_FT	939	MC	04	1	951	0.57	0.00		0.22	0.16	0.17	0.45			0.45	0.32
2011_Geom_FT	939	MC	05	1	951	0.57	0.00		0.13	0.21	0.55	0.11			0.55	0.39
2011_Geom_FT	939	MC	06	1	951	0.57	0.00		0.34	0.50	0.08	0.08			0.50	0.09
2011_Geom_FT	939	MC	07	1	951	0.57	0.01		0.42	0.33	0.13	0.11			0.42	0.35
2011_Geom_FT	939	CR	08	2	951	0.57	0.09	0.48	0.08	0.34					0.77	0.57
2011_Geom_FT	939	CR	09	6	951	0.57	0.08	0.32	0.11	0.11	0.04	0.04	0.05	0.25	2.37	0.84
2011_Geom_FT	939	CR	10	4	951	0.57	0.04	0.58	0.09	0.09	0.04	0.16			1.03	0.64
2011_Geom_FT	940	MC	01	1	948	0.69	0.01		0.14	0.51	0.22	0.12			0.51	0.50

Table 8. Classical Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	M6	Mean	Point-Biserial
2011_Geom_FT	940	MC	02	1	948	0.69	0.01		0.07	0.24	0.51	0.17			0.51	0.49
2011_Geom_FT	940	MC	03	1	948	0.69	0.00		0.43	0.14	0.24	0.19			0.43	0.39
2011_Geom_FT	940	MC	04	1	948	0.69	0.01		0.28	0.25	0.33	0.14			0.25	0.42
2011_Geom_FT	940	MC	05	1	948	0.69	0.02		0.22	0.11	0.20	0.45			0.45	0.45
2011_Geom_FT	940	MC	06	1	948	0.69	0.01		0.05	0.08	0.56	0.29			0.56	0.44
2011_Geom_FT	940	MC	07	1	948	0.69	0.02		0.10	0.27	0.31	0.31			0.27	0.36
2011_Geom_FT	940	CR	08	2	948	0.69	0.07	0.62	0.07	0.24					0.55	0.66
2011_Geom_FT	940	CR	09	4	948	0.69	0.04	0.79	0.05	0.03	0.02	0.06			0.43	0.68
2011_Geom_FT	940	CR	10	4	948	0.69	0.07	0.21	0.20	0.23	0.08	0.22			1.76	0.75
2011_Geom_FT	941	MC	01	1	928	0.64	0.00		0.20	0.22	0.51	0.06			0.51	0.45
2011_Geom_FT	941	MC	02	1	928	0.64	0.00		0.07	0.54	0.08	0.30			0.54	0.37
2011_Geom_FT	941	MC	03	1	928	0.64	0.01		0.56	0.22	0.10	0.11			0.56	0.53
2011_Geom_FT	941	MC	04	1	928	0.64	0.01		0.14	0.13	0.19	0.53			0.53	0.50
2011_Geom_FT	941	MC	05	1	928	0.64	0.01		0.10	0.56	0.23	0.09			0.56	0.46
2011_Geom_FT	941	MC	06	1	928	0.64	0.01		0.73	0.09	0.09	0.08			0.73	0.42
2011_Geom_FT	941	MC	07	1	928	0.64	0.02		0.13	0.58	0.17	0.11			0.58	0.45
2011_Geom_FT	941	CR	08	2	928	0.64	0.11	0.76	0.05	0.08					0.21	0.47
2011_Geom_FT	941	CR	09	2	928	0.64	0.08	0.32	0.29	0.31					0.92	0.60
2011_Geom_FT	941	CR	10	6	928	0.64	0.08	0.72	0.07	0.02	0.02	0.02	0.02	0.06	0.69	0.76
2011_Geom_FT	942	MC	01	1	973	0.60	0.00		0.09	0.77	0.03	0.10			0.77	0.37
2011_Geom_FT	942	MC	02	1	973	0.60	0.01		0.27	0.46	0.17	0.09			0.46	0.55
2011_Geom_FT	942	MC	03	1	973	0.60	0.00		0.10	0.28	0.56	0.05			0.56	0.33
2011_Geom_FT	942	MC	04	1	973	0.60	0.00		0.09	0.08	0.69	0.13			0.69	0.30

Table 8. Classical Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	M6	Mean	Point-Biserial
2011_Geom_FT	942	MC	05	1	973	0.60	0.00		0.10	0.14	0.34	0.42			0.42	0.22
2011_Geom_FT	942	MC	06	1	973	0.60	0.01		0.26	0.27	0.33	0.14			0.26	0.26
2011_Geom_FT	942	MC	07	1	973	0.60	0.01		0.24	0.24	0.45	0.06			0.45	0.41
2011_Geom_FT	942	CR	08	2	973	0.60	0.02	0.47	0.12	0.40					0.91	0.66
2011_Geom_FT	942	CR	09	2	973	0.60	0.02	0.31	0.33	0.33					0.99	0.56
2011_Geom_FT	942	CR	10	6	973	0.60	0.06	0.53	0.15	0.06	0.06	0.06	0.07	0.02	1.17	0.82
2011_Geom_FT	943	MC	01	1	968	0.60	0.01		0.32	0.49	0.14	0.04			0.49	0.32
2011_Geom_FT	943	MC	02	1	968	0.60	0.00		0.70	0.15	0.13	0.02			0.70	0.52
2011_Geom_FT	943	MC	03	1	968	0.60	0.00		0.54	0.15	0.07	0.23			0.54	0.44
2011_Geom_FT	943	MC	04	1	968	0.60	0.01		0.13	0.08	0.27	0.51			0.51	0.49
2011_Geom_FT	943	MC	05	1	968	0.60	0.01		0.07	0.09	0.70	0.13			0.70	0.44
2011_Geom_FT	943	MC	06	1	968	0.60	0.01		0.32	0.35	0.06	0.27			0.35	0.43
2011_Geom_FT	943	MC	07	1	968	0.60	0.02		0.22	0.38	0.15	0.23			0.23	0.17
2011_Geom_FT	943	CR	08	2	968	0.60	0.13	0.80	0.04	0.03					0.10	0.43
2011_Geom_FT	943	CR	09	4	968	0.60	0.03	0.53	0.05	0.12	0.05	0.21			1.30	0.73
2011_Geom_FT	943	CR	10	4	968	0.60	0.10	0.45	0.05	0.15	0.03	0.22			1.33	0.76
2011_Geom_FT	944	MC	01	1	974	0.71	0.01		0.13	0.41	0.39	0.06			0.39	0.35
2011_Geom_FT	944	MC	02	1	974	0.71	0.01		0.07	0.08	0.69	0.15			0.69	0.45
2011_Geom_FT	944	MC	03	1	974	0.71	0.01		0.64	0.09	0.13	0.14			0.64	0.36
2011_Geom_FT	944	MC	04	1	974	0.71	0.00		0.20	0.09	0.60	0.11			0.60	0.37
2011_Geom_FT	944	MC	05	1	974	0.71	0.01		0.42	0.16	0.30	0.10			0.42	0.55
2011_Geom_FT	944	MC	06	1	974	0.71	0.01		0.20	0.10	0.15	0.54			0.54	0.32
2011_Geom_FT	944	MC	07	1	974	0.71	0.02		0.15	0.23	0.09	0.51			0.51	0.52

Table 8. Classical Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	M6	Mean	Point-Biserial
2011_Geom_FT	944	CR	08	2	974	0.71	0.06	0.37	0.22	0.34					0.91	0.64
2011_Geom_FT	944	CR	09	4	974	0.71	0.07	0.54	0.18	0.09	0.06	0.06			0.79	0.77
2011_Geom_FT	944	CR	10	4	974	0.71	0.05	0.69	0.09	0.04	0.04	0.09			0.65	0.79
2011_Geom_FT	945	MC	01	1	959	0.65	0.00		0.25	0.11	0.15	0.48			0.25	0.26
2011_Geom_FT	945	MC	02	1	959	0.65	0.00		0.15	0.52	0.23	0.09			0.23	0.26
2011_Geom_FT	945	MC	03	1	959	0.65	0.01		0.53	0.17	0.10	0.19			0.53	0.55
2011_Geom_FT	945	MC	04	1	959	0.65	0.00		0.21	0.21	0.15	0.43			0.43	0.46
2011_Geom_FT	945	MC	05	1	959	0.65	0.01		0.31	0.36	0.22	0.10			0.36	0.40
2011_Geom_FT	945	MC	06	1	959	0.65	0.00		0.27	0.02	0.58	0.13			0.58	0.41
2011_Geom_FT	945	MC	07	1	959	0.65	0.01		0.12	0.12	0.35	0.41			0.35	0.31
2011_Geom_FT	945	CR	08	2	959	0.65	0.00	0.47	0.25	0.29					0.82	0.58
2011_Geom_FT	945	CR	09	2	959	0.65	0.13	0.37	0.19	0.32					0.83	0.68
2011_Geom_FT	945	CR	10	6	959	0.65	0.10	0.33	0.23	0.18	0.07	0.06	0.02	0.01	1.22	0.82
2011_Geom_FT	946	MC	01	1	915	0.61	0.01		0.20	0.27	0.35	0.17			0.35	0.20
2011_Geom_FT	946	MC	02	1	915	0.61	0.00		0.08	0.14	0.10	0.68			0.68	0.40
2011_Geom_FT	946	MC	03	1	915	0.61	0.01		0.12	0.11	0.15	0.61			0.61	0.51
2011_Geom_FT	946	MC	04	1	915	0.61	0.00		0.25	0.65	0.05	0.04			0.25	0.14
2011_Geom_FT	946	MC	05	1	915	0.61	0.01		0.15	0.12	0.10	0.62			0.62	0.46
2011_Geom_FT	946	MC	06	1	915	0.61	0.01		0.52	0.44	0.02	0.01			0.52	0.47
2011_Geom_FT	946	MC	07	1	915	0.61	0.01		0.44	0.18	0.21	0.16			0.18	0.29
2011_Geom_FT	946	CR	08	2	915	0.61	0.08	0.52	0.08	0.32					0.72	0.60
2011_Geom_FT	946	CR	09	2	915	0.61	0.12	0.39	0.22	0.28					0.77	0.65
2011_Geom_FT	946	CR	10	6	915	0.61	0.12	0.54	0.14	0.03	0.04	0.02	0.04	0.07	1.03	0.82

Table 8. Classical Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	M6	Mean	Point-Biserial
2011_Geom_FT	947	MC	01	1	940	0.67	0.00		0.04	0.76	0.14	0.06			0.76	0.44
2011_Geom_FT	947	MC	02	1	940	0.67	0.01		0.26	0.11	0.53	0.09			0.53	0.41
2011_Geom_FT	947	MC	03	1	940	0.67	0.00		0.07	0.06	0.17	0.69			0.69	0.41
2011_Geom_FT	947	MC	04	1	940	0.67	0.01		0.17	0.66	0.13	0.04			0.66	0.48
2011_Geom_FT	947	MC	05	1	940	0.67	0.00		0.10	0.48	0.36	0.06			0.36	0.39
2011_Geom_FT	947	MC	06	1	940	0.67	0.01		0.10	0.65	0.17	0.07			0.65	0.58
2011_Geom_FT	947	MC	07	1	940	0.67	0.01		0.06	0.18	0.51	0.24			0.51	0.40
2011_Geom_FT	947	CR	08	2	940	0.67	0.08	0.63	0.08	0.21					0.51	0.54
2011_Geom_FT	947	CR	09	4	940	0.67	0.03	0.26	0.24	0.14	0.08	0.25			1.76	0.77
2011_Geom_FT	947	CR	10	4	940	0.67	0.10	0.48	0.29	0.09	0.02	0.01			0.59	0.66
2011_Geom_FT	948	MC	01	1	954	0.62	0.01		0.13	0.64	0.10	0.13			0.64	0.30
2011_Geom_FT	948	MC	02	1	954	0.62	0.00		0.03	0.05	0.02	0.91			0.91	0.31
2011_Geom_FT	948	MC	03	1	954	0.62	0.01		0.08	0.32	0.45	0.14			0.45	0.46
2011_Geom_FT	948	MC	04	1	954	0.62	0.00		0.12	0.32	0.02	0.54			0.54	0.40
2011_Geom_FT	948	MC	05	1	954	0.62	0.00		0.01	0.04	0.93	0.02			0.93	0.29
2011_Geom_FT	948	MC	06	1	954	0.62	0.02		0.52	0.17	0.22	0.06			0.52	0.35
2011_Geom_FT	948	MC	07	1	954	0.62	0.01		0.19	0.08	0.07	0.65			0.65	0.20
2011_Geom_FT	948	CR	08	2	954	0.62	0.06	0.56	0.19	0.19					0.56	0.66
2011_Geom_FT	948	CR	09	4	954	0.62	0.06	0.37	0.15	0.12	0.11	0.19			1.46	0.79
2011_Geom_FT	948	CR	10	4	954	0.62	0.04	0.46	0.05	0.20	0.11	0.15			1.35	0.76
2011_Geom_FT	949	MC	01	1	937	0.64	0.01		0.09	0.64	0.19	0.08			0.64	0.45
2011_Geom_FT	949	MC	02	1	937	0.64	0.00		0.19	0.24	0.53	0.05			0.53	0.33
2011_Geom_FT	949	MC	03	1	937	0.64	0.01		0.13	0.35	0.39	0.13			0.39	0.43

Table 8. Classical Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	M6	Mean	Point-Biserial
2011_Geom_FT	949	MC	04	1	937	0.64	0.00		0.11	0.35	0.32	0.22			0.35	0.27
2011_Geom_FT	949	MC	05	1	937	0.64	0.00		0.17	0.14	0.42	0.27			0.42	0.37
2011_Geom_FT	949	MC	06	1	937	0.64	0.01		0.48	0.16	0.17	0.18			0.48	0.41
2011_Geom_FT	949	MC	07	1	937	0.64	0.01		0.30	0.42	0.09	0.18			0.18	0.30
2011_Geom_FT	949	CR	08	2	937	0.64	0.06	0.40	0.22	0.32					0.85	0.66
2011_Geom_FT	949	CR	09	4	937	0.64	0.04	0.35	0.20	0.13	0.02	0.24			1.51	0.75
2011_Geom_FT	949	CR	10	4	937	0.64	0.05	0.51	0.07	0.10	0.06	0.20			1.26	0.73

Appendix B: Partial Credit Model Item Analysis

Table 9. Partial Credit Model Item Analysis

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_Geom_FT	931	MC	01	1	977	0.1216							0.96
2011_Geom_FT	931	MC	02	1	977	−1.1508							1.03
2011_Geom_FT	931	MC	03	1	977	−0.0125							1.16
2011_Geom_FT	931	MC	04	1	977	0.0324							0.89
2011_Geom_FT	931	MC	05	1	977	1.6338							1.12
2011_Geom_FT	931	MC	06	1	977	−0.1438							1.00
2011_Geom_FT	931	MC	07	1	977	0.4202							1.03
2011_Geom_FT	931	CR	08	2	977	2.3296	1.4889	−1.4888					0.98
2011_Geom_FT	931	CR	09	2	977	−0.0242	0.2350	−0.2350					0.93
2011_Geom_FT	931	CR	10	6	977	2.5245	−0.5671	−0.4231	0.6324	0.5541	−1.0360	0.8397	0.77
2011_Geom_FT	932	MC	01	1	965	−0.8884							0.95
2011_Geom_FT	932	MC	02	1	965	−0.2079							0.93
2011_Geom_FT	932	MC	03	1	965	0.1769							1.21
2011_Geom_FT	932	MC	04	1	965	−0.5560							0.95
2011_Geom_FT	932	MC	05	1	965	0.1155							0.96
2011_Geom_FT	932	MC	06	1	965	0.6044							1.16
2011_Geom_FT	932	MC	07	1	965	0.0487							1.03
2011_Geom_FT	932	CR	08	4	965	2.6321	0.3624	−0.1028	−1.9981	1.7385			0.93
2011_Geom_FT	932	CR	09	2	965	0.9571	−0.8346	0.8346					0.90
2011_Geom_FT	932	CR	10	2	965	0.4337	0.1693	−0.1693					0.90
2011_Geom_FT	933	MC	01	1	970	1.1043							1.05
2011_Geom_FT	933	MC	02	1	970	−0.4029							0.99
2011_Geom_FT	933	MC	03	1	970	−0.2992							1.10

Table 9. Partial Credit Model Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_Geom_FT	933	MC	04	1	970	0.4030							0.95
2011_Geom_FT	933	MC	05	1	970	−0.2508							0.84
2011_Geom_FT	933	MC	06	1	970	0.1700							1.18
2011_Geom_FT	933	MC	07	1	970	2.1271							1.01
2011_Geom_FT	933	CR	08	2	970	2.1503	0.9459	−0.9459					1.07
2011_Geom_FT	933	CR	09	2	970	0.4671	−0.1565	0.1565					1.11
2011_Geom_FT	933	CR	10	6	970	1.2031	−0.7671	0.1516	0.8915	−1.5949	1.5012	−0.1823	0.64
2011_Geom_FT	934	MC	01	1	966	0.7817							1.09
2011_Geom_FT	934	MC	02	1	966	1.2562							1.05
2011_Geom_FT	934	MC	03	1	966	−0.9877							0.91
2011_Geom_FT	934	MC	04	1	966	0.5127							0.99
2011_Geom_FT	934	MC	05	1	966	1.1559							1.08
2011_Geom_FT	934	MC	06	1	966	−0.4328							0.96
2011_Geom_FT	934	MC	07	1	966	−0.5229							1.18
2011_Geom_FT	934	CR	08	2	966	2.0325	0.8314	−0.8314					1.00
2011_Geom_FT	934	CR	09	2	966	2.0432	0.3161	−0.3161					1.01
2011_Geom_FT	934	CR	10	6	966	1.7002	0.1139	−0.7989	2.4702	−2.4471	1.7846	−1.1228	0.70
2011_Geom_FT	935	MC	01	1	965	0.1102							0.95
2011_Geom_FT	935	MC	02	1	965	0.9935							0.98
2011_Geom_FT	935	MC	03	1	965	−0.3362							0.98
2011_Geom_FT	935	MC	04	1	965	−0.4273							0.99
2011_Geom_FT	935	MC	05	1	965	−0.6401							0.99
2011_Geom_FT	935	MC	06	1	965	0.9883							1.23
2011_Geom_FT	935	MC	07	1	965	1.5788							1.07

Table 9. Partial Credit Model Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_Geom_FT	935	CR	08	2	965	−0.0726	0.8556	−0.8556					0.98
2011_Geom_FT	935	CR	09	2	965	−0.7900	0.3792	−0.3792					0.97
2011_Geom_FT	935	CR	10	6	965	0.6544	0.2006	0.0876	0.9579	−1.0451	0.6912	−0.8921	0.74
2011_Geom_FT	936	MC	01	1	932	−0.3943							0.94
2011_Geom_FT	936	MC	02	1	932	−1.0480							1.00
2011_Geom_FT	936	MC	03	1	932	−1.2047							0.93
2011_Geom_FT	936	MC	04	1	932	0.7097							1.01
2011_Geom_FT	936	MC	05	1	932	1.0397							0.95
2011_Geom_FT	936	MC	06	1	932	0.0933							0.97
2011_Geom_FT	936	MC	07	1	932	1.1976							1.31
2011_Geom_FT	936	CR	08	2	932	1.2764	−0.4482	0.4482					1.01
2011_Geom_FT	936	CR	09	4	932	1.1561	−0.0261	0.2687	−0.6624	0.4198			1.01
2011_Geom_FT	936	CR	10	4	932	1.0584	0.6793	−0.1707	0.3223	−0.8308			0.86
2011_Geom_FT	937	MC	01	1	948	−0.5590							0.99
2011_Geom_FT	937	MC	02	1	948	−0.2031							1.01
2011_Geom_FT	937	MC	03	1	948	1.1434							1.11
2011_Geom_FT	937	MC	04	1	948	−1.8039							1.03
2011_Geom_FT	937	MC	05	1	948	−0.2031							0.87
2011_Geom_FT	937	MC	06	1	948	1.5532							1.19
2011_Geom_FT	937	MC	07	1	948	1.2652							1.14
2011_Geom_FT	937	CR	08	2	948	0.5327	0.3336	−0.3336					0.85
2011_Geom_FT	937	CR	09	2	948	0.9703	−0.7543	0.7543					0.93
2011_Geom_FT	937	CR	10	6	948	2.6825	−0.6510	−1.1919	0.6499	0.9506	−0.2533	0.4956	0.81
2011_Geom_FT	938	MC	01	1	947	0.7949							1.19

Table 9. Partial Credit Model Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_Geom_FT	938	MC	02	1	947	1.1653							1.11
2011_Geom_FT	938	MC	03	1	947	0.3331							1.01
2011_Geom_FT	938	MC	04	1	947	−0.9029							0.96
2011_Geom_FT	938	MC	05	1	947	−2.1076							0.98
2011_Geom_FT	938	MC	06	1	947	−0.3506							1.06
2011_Geom_FT	938	MC	07	1	947	1.0581							1.10
2011_Geom_FT	938	CR	08	2	947	0.8482	0.7438	−0.7438					0.80
2011_Geom_FT	938	CR	09	2	947	0.2206	−0.6229	0.6229					0.97
2011_Geom_FT	938	CR	10	6	947	1.8538	−0.3610	−0.3198	1.2666	−1.1801	−0.0635	0.6579	0.61
2011_Geom_FT	939	MC	01	1	951	−0.2273							0.95
2011_Geom_FT	939	MC	02	1	951	0.2590							0.91
2011_Geom_FT	939	MC	03	1	951	0.4990							1.04
2011_Geom_FT	939	MC	04	1	951	0.5806							1.06
2011_Geom_FT	939	MC	05	1	951	0.1185							1.01
2011_Geom_FT	939	MC	06	1	951	0.3456							1.26
2011_Geom_FT	939	MC	07	1	951	0.6967							1.02
2011_Geom_FT	939	CR	08	2	951	0.7286	1.3925	−1.3925					0.98
2011_Geom_FT	939	CR	09	6	951	0.6409	0.5647	−0.3429	0.9209	0.0207	−0.0414	−1.1221	0.72
2011_Geom_FT	939	CR	10	4	951	1.0376	1.0134	−0.3773	0.7139	−1.3501			1.03
2011_Geom_FT	940	MC	01	1	948	0.2628							0.95
2011_Geom_FT	940	MC	02	1	948	0.2524							0.96
2011_Geom_FT	940	MC	03	1	948	0.6374							1.10
2011_Geom_FT	940	MC	04	1	948	1.6442							1.02
2011_Geom_FT	940	MC	05	1	948	0.5275							1.03

Table 9. Partial Credit Model Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_Geom_FT	940	MC	06	1	948	−0.0245							1.00
2011_Geom_FT	940	MC	07	1	948	1.5258							1.11
2011_Geom_FT	940	CR	08	2	948	1.1827	1.3870	−1.3870					0.95
2011_Geom_FT	940	CR	09	4	948	1.9741	1.1728	−0.3502	0.0137	−0.8362			0.79
2011_Geom_FT	940	CR	10	4	948	0.5008	−0.6274	−0.4522	1.3017	−0.2222			0.95
2011_Geom_FT	941	MC	01	1	928	0.2832							1.03
2011_Geom_FT	941	MC	02	1	928	0.1231							1.15
2011_Geom_FT	941	MC	03	1	928	0.0452							0.89
2011_Geom_FT	941	MC	04	1	928	0.1950							0.93
2011_Geom_FT	941	MC	05	1	928	0.0173							1.02
2011_Geom_FT	941	MC	06	1	928	−0.9436							0.99
2011_Geom_FT	941	MC	07	1	928	−0.0501							1.03
2011_Geom_FT	941	CR	08	2	928	2.3501	1.0583	−1.0583					1.09
2011_Geom_FT	941	CR	09	2	928	0.5437	−0.3048	0.3048					1.02
2011_Geom_FT	941	CR	10	6	928	1.9367	0.9751	0.2892	−0.1038	−0.5585	0.3229	−0.9249	0.59
2011_Geom_FT	942	MC	01	1	973	−1.0496							0.98
2011_Geom_FT	942	MC	02	1	973	0.5559							0.88
2011_Geom_FT	942	MC	03	1	973	0.0663							1.09
2011_Geom_FT	942	MC	04	1	973	−0.5522							1.09
2011_Geom_FT	942	MC	05	1	973	0.7450							1.21
2011_Geom_FT	942	MC	06	1	973	1.5498							1.14
2011_Geom_FT	942	MC	07	1	973	0.6041							1.02
2011_Geom_FT	942	CR	08	2	973	0.5235	0.9926	−0.9926					0.86
2011_Geom_FT	942	CR	09	2	973	0.3779	−0.3597	0.3597					1.00

Table 9. Partial Credit Model Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_Geom_FT	942	CR	10	6	973	1.6317	−0.1771	−0.1437	−0.7128	−0.3631	−0.2092	1.6059	0.67
2011_Geom_FT	943	MC	01	1	968	0.3553							1.12
2011_Geom_FT	943	MC	02	1	968	−0.7201							0.86
2011_Geom_FT	943	MC	03	1	968	0.1225							1.00
2011_Geom_FT	943	MC	04	1	968	0.2814							0.94
2011_Geom_FT	943	MC	05	1	968	−0.7140							0.98
2011_Geom_FT	943	MC	06	1	968	1.0655							0.99
2011_Geom_FT	943	MC	07	1	968	1.7136							1.21
2011_Geom_FT	943	CR	08	2	968	2.8361	0.9483	−0.9483					0.84
2011_Geom_FT	943	CR	09	4	968	0.8904	1.4589	−1.2016	0.8017	−1.0591			1.01
2011_Geom_FT	943	CR	10	4	968	0.8582	1.4493	−1.3843	1.7037	−1.7687			0.91
2011_Geom_FT	944	MC	01	1	974	0.8689							1.17
2011_Geom_FT	944	MC	02	1	974	−0.6469							0.96
2011_Geom_FT	944	MC	03	1	974	−0.3859							1.08
2011_Geom_FT	944	MC	04	1	974	−0.1591							1.11
2011_Geom_FT	944	MC	05	1	974	0.7209							0.91
2011_Geom_FT	944	MC	06	1	974	0.1125							1.17
2011_Geom_FT	944	MC	07	1	974	0.2616							0.93
2011_Geom_FT	944	CR	08	2	974	0.5269	0.0880	−0.0880					0.96
2011_Geom_FT	944	CR	09	4	974	1.7167	−0.3686	−0.0613	0.0556	0.3744			0.89
2011_Geom_FT	944	CR	10	4	974	1.7592	0.5830	0.0556	−0.1963	−0.4423			0.66
2011_Geom_FT	945	MC	01	1	959	1.5761							1.15
2011_Geom_FT	945	MC	02	1	959	1.6995							1.14
2011_Geom_FT	945	MC	03	1	959	0.1494							0.87

Table 9. Partial Credit Model Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_Geom_FT	945	MC	04	1	959	0.6165							1.00
2011_Geom_FT	945	MC	05	1	959	1.0050							1.04
2011_Geom_FT	945	MC	06	1	959	−0.1059							1.03
2011_Geom_FT	945	MC	07	1	959	1.0588							1.14
2011_Geom_FT	945	CR	08	2	959	0.6845	0.0143	−0.0143					1.04
2011_Geom_FT	945	CR	09	2	959	0.6467	0.3795	−0.3795					0.85
2011_Geom_FT	945	CR	10	6	959	1.7266	−1.2275	−0.9632	0.1920	−0.1556	1.0406	1.1137	0.69
2011_Geom_FT	946	MC	01	1	915	1.0631							1.25
2011_Geom_FT	946	MC	02	1	915	−0.6043							1.00
2011_Geom_FT	946	MC	03	1	915	−0.2315							0.89
2011_Geom_FT	946	MC	04	1	915	1.6048							1.25
2011_Geom_FT	946	MC	05	1	915	−0.2707							0.93
2011_Geom_FT	946	MC	06	1	915	0.2424							0.96
2011_Geom_FT	946	MC	07	1	915	2.0450							1.07
2011_Geom_FT	946	CR	08	2	915	0.8576	1.3816	−1.3816					1.00
2011_Geom_FT	946	CR	09	2	915	0.8107	0.1725	−0.1725					0.88
2011_Geom_FT	946	CR	10	6	915	1.4815	0.2887	0.6318	−0.5437	0.4515	−0.4659	−0.3624	0.57
2011_Geom_FT	947	MC	01	1	940	−1.0412							0.96
2011_Geom_FT	947	MC	02	1	940	0.2217							1.09
2011_Geom_FT	947	MC	03	1	940	−0.6550							1.03
2011_Geom_FT	947	MC	04	1	940	−0.4634							0.98
2011_Geom_FT	947	MC	05	1	940	1.0678							1.08
2011_Geom_FT	947	MC	06	1	940	−0.3997							0.84
2011_Geom_FT	947	MC	07	1	940	0.3162							1.09

Table 9. Partial Credit Model Item Analysis *(continued)*

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_Geom_FT	947	CR	08	2	940	1.3605	1.1294	−1.1294					1.13
2011_Geom_FT	947	CR	09	4	940	0.5485	−0.7057	0.2872	0.7522	−0.3337			0.96
2011_Geom_FT	947	CR	10	4	940	2.3422	−1.4027	−0.1001	0.6969	0.8059			0.88
2011_Geom_FT	948	MC	01	1	954	−0.2671							1.12
2011_Geom_FT	948	MC	02	1	954	−2.2793							0.94
2011_Geom_FT	948	MC	03	1	954	0.6084							0.98
2011_Geom_FT	948	MC	04	1	954	0.1880							1.04
2011_Geom_FT	948	MC	05	1	954	−2.5813							0.92
2011_Geom_FT	948	MC	06	1	954	0.2822							1.09
2011_Geom_FT	948	MC	07	1	954	−0.3423							1.21
2011_Geom_FT	948	CR	08	2	954	1.2842	0.2024	−0.2024					0.85
2011_Geom_FT	948	CR	09	4	954	0.8492	0.1162	−0.1311	0.0546	−0.0398			0.82
2011_Geom_FT	948	CR	10	4	954	1.0004	1.2951	−1.9052	0.5053	0.1048			0.92
2011_Geom_FT	949	MC	01	1	937	−0.3305							0.95
2011_Geom_FT	949	MC	02	1	937	0.2046							1.09
2011_Geom_FT	949	MC	03	1	937	0.8615							0.98
2011_Geom_FT	949	MC	04	1	937	1.0352							1.13
2011_Geom_FT	949	MC	05	1	937	0.6979							1.05
2011_Geom_FT	949	MC	06	1	937	0.4181							1.02
2011_Geom_FT	949	MC	07	1	937	2.0274							1.04
2011_Geom_FT	949	CR	08	2	937	0.6113	0.2427	−0.2427					0.85
2011_Geom_FT	949	CR	09	4	937	0.6620	−0.1199	0.1548	1.7754	−1.8102			0.92
2011_Geom_FT	949	CR	10	4	937	0.8967	1.1833	−0.7762	0.5033	−0.9105			0.99

Appendix C: DIF Statistics

Table 10. DIF Statistics

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
931	01	MC	−0.40	1.26	−0.07		
931	02	MC	−0.37	0.86	−0.05		
931	03	MC	0.98	8.94	0.18		
931	04	MC	0.09	0.06	0.00		
931	05	MC	−0.05	0.02	−0.01		
931	06	MC	−0.24	0.44	−0.05		
931	07	MC	−0.15	0.20	−0.02		
931	08	CR		1.49	−0.08		
931	09	CR		0.88	−0.04		
931	10	CR		3.22	0.08		
932	01	MC	0.01	0.00	0.00		
932	02	MC	−0.19	0.27	−0.02		
932	03	MC	−0.56	2.92	−0.11		
932	04	MC	0.14	0.14	0.03		
932	05	MC	0.32	0.80	0.05		
932	06	MC	−0.21	0.39	−0.04		
932	07	MC	−0.07	0.04	0.00		
932	08	CR		0.06	−0.01		
932	09	CR		0.26	0.02		
932	10	CR		1.24	0.05		
933	01	MC	0.51	2.03	0.09		
933	02	MC	−0.18	0.24	−0.02		
933	03	MC	−0.43	1.50	−0.08		
933	04	MC	−0.23	0.41	−0.03		

Table 10. DIF Statistics *(continued)*

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
933	05	MC	−0.46	1.28	−0.06		
933	06	MC	0.21	0.45	0.06		
933	07	MC	−0.44	1.04	−0.08		
933	08	CR		3.02	0.10		
933	09	CR		0.18	−0.03		
933	10	CR		0.11	0.01		
934	01	MC	−0.09	0.06	−0.02		
934	02	MC	−0.18	0.23	−0.04		
934	03	MC	−1.03	5.51	−0.12	B	M
934	04	MC	0.11	0.09	0.02		
934	05	MC	0.42	1.30	0.07		
934	06	MC	−0.38	0.98	−0.05		
934	07	MC	0.12	0.12	0.03		
934	08	CR		2.83	−0.10		
934	09	CR		0.21	0.03		
934	10	CR		2.90	0.06		
935	01	MC	0.16	0.20	0.02		
935	02	MC	−0.13	0.14	−0.02		
935	03	MC	−0.70	3.67	−0.10		
935	04	MC	1.28	11.63	0.20	B	F
935	05	MC	0.12	0.10	0.01		
935	06	MC	−0.60	3.45	−0.13		
935	07	MC	−0.31	0.68	−0.04		
935	08	CR		2.87	−0.07		

Table 10. DIF Statistics *(continued)*

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
935	09	CR		0.68	−0.05		
935	10	CR		3.00	0.05		
936	01	MC	−0.10	0.07	−0.02		
936	02	MC	0.37	0.75	0.05		
936	03	MC	−0.54	1.44	−0.07		
936	04	MC	−0.09	0.06	0.00		
936	05	MC	0.64	2.88	0.10		
936	06	MC	0.41	1.29	0.07		
936	07	MC	−0.46	1.79	−0.08		
936	08	CR		0.67	−0.05		
936	09	CR		0.27	−0.02		
936	10	CR		0.61	0.03		
937	01	MC	0.33	0.74	0.04		
937	02	MC	−0.21	0.34	−0.05		
937	03	MC	−0.03	0.01	−0.01		
937	04	MC	−0.07	0.02	−0.01		
937	05	MC	1.13	8.10	0.17	B	F
937	06	MC	−0.94	6.46	−0.16		
937	07	MC	−0.07	0.03	−0.02		
937	08	CR		0.10	0.01		
937	09	CR		0.80	−0.06		
937	10	CR		0.33	0.04		
938	01	MC	0.21	0.42	0.04		
938	02	MC	0.35	1.03	0.08		

Table 10. DIF Statistics *(continued)*

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
938	03	MC	0.10	0.08	0.03		
938	04	MC	−0.29	0.53	−0.04		
938	05	MC	0.86	2.45	0.11		
938	06	MC	−0.09	0.07	−0.01		
938	07	MC	−0.18	0.27	−0.03		
938	08	CR		7.77	0.12		
938	09	CR		14.02	−0.19	BB	M
938	10	CR		0.00	0.01		
939	01	MC	0.01	0.00	0.01		
939	02	MC	0.29	0.64	0.05		
939	03	MC	−0.96	7.89	−0.18		
939	04	MC	−0.41	1.57	−0.07		
939	05	MC	−0.38	1.22	−0.06		
939	06	MC	0.38	1.46	0.06		
939	07	MC	−0.14	0.18	−0.03		
939	08	CR		0.36	0.03		
939	09	CR		0.09	0.01		
939	10	CR		0.50	0.04		
940	01	MC	−0.20	0.29	−0.03		
940	02	MC	0.19	0.27	0.03		
940	03	MC	0.53	2.36	0.09		
940	04	MC	−0.87	4.76	−0.13		
940	05	MC	0.00	0.00	0.00		
940	06	MC	0.71	4.04	0.12		

Table 10. DIF Statistics *(continued)*

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
940	07	MC	−0.55	2.14	−0.10		
940	08	CR		2.31	−0.08		
940	09	CR		0.31	0.02		
940	10	CR		0.30	0.03		
941	01	MC	−0.62	3.01	−0.10		
941	02	MC	−0.84	5.85	−0.15		
941	03	MC	−0.88	5.12	−0.11		
941	04	MC	1.25	10.44	0.17	B	F
941	05	MC	0.92	6.36	0.14		
941	06	MC	−0.08	0.03	−0.01		
941	07	MC	−0.10	0.07	−0.01		
941	08	CR		0.54	−0.03		
941	09	CR		1.41	−0.05		
941	10	CR		5.72	0.06		
942	01	MC	0.43	1.16	0.06		
942	02	MC	0.13	0.13	0.01		
942	03	MC	−0.52	2.52	−0.10		
942	04	MC	0.03	0.01	0.00		
942	05	MC	−0.35	1.19	−0.06		
942	06	MC	0.30	0.71	0.05		
942	07	MC	0.17	0.27	0.03		
942	08	CR		0.20	−0.03		
942	09	CR		3.61	−0.08		
942	10	CR		3.40	0.06		

Table 10. DIF Statistics *(continued)*

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
943	01	MC	−0.46	1.97	−0.11		
943	02	MC	−0.18	0.18	−0.03		
943	03	MC	−0.49	2.02	−0.08		
943	04	MC	0.38	1.14	0.06		
943	05	MC	0.17	0.19	0.04		
943	06	MC	0.29	0.65	0.06		
943	07	MC	−0.42	1.28	−0.08		
943	08	CR		0.34	−0.04		
943	09	CR		1.19	−0.05		
943	10	CR		4.70	0.10		
944	01	MC	0.67	3.97	0.12		
944	02	MC	−0.26	0.48	−0.04		
944	03	MC	0.25	0.50	0.04		
944	04	MC	−0.17	0.26	−0.03		
944	05	MC	−0.08	0.05	−0.01		
944	06	MC	−0.26	0.60	−0.05		
944	07	MC	0.09	0.07	0.01		
944	08	CR		3.84	−0.10		
944	09	CR		0.74	0.03		
944	10	CR		0.15	0.02		
945	01	MC	−0.79	4.59	−0.15		
945	02	MC	−0.38	0.98	−0.06		
945	03	MC	0.98	6.69	0.14		
945	04	MC	−0.31	0.76	−0.05		

Table 10. DIF Statistics *(continued)*

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
945	05	MC	−0.41	1.36	−0.07		
945	06	MC	−0.55	2.53	−0.09		
945	07	MC	−0.24	0.50	−0.03		
945	08	CR		0.00	0.00		
945	09	CR		3.61	0.09		
945	10	CR		1.35	0.04		
946	01	MC	−0.37	1.22	−0.08		
946	02	MC	1.44	14.05	0.23	B	F
946	03	MC	−0.11	0.07	−0.02		
946	04	MC	−0.61	2.63	−0.11		
946	05	MC	0.11	0.08	0.02		
946	06	MC	−0.21	0.32	−0.03		
946	07	MC	0.49	1.26	0.08		
946	08	CR		6.50	−0.13		
946	09	CR		0.27	−0.02		
946	10	CR		4.37	0.06		
947	01	MC	−0.50	1.42	−0.06		
947	02	MC	0.38	1.21	0.06		
947	03	MC	−0.14	0.13	−0.02		
947	04	MC	−0.86	5.29	−0.15		
947	05	MC	−0.67	3.68	−0.12		
947	06	MC	0.03	0.00	0.01		
947	07	MC	0.48	1.95	0.09		

*DIF Category meanings: A/AA=negligible, B/BB=moderate, C/CC=large

Table 10. DIF Statistics *(continued)*

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
947	08	CR		0.58	0.05		
947	09	CR		0.05	0.01		
947	10	CR		0.64	0.04		
948	01	MC	0.23	0.45	0.05		
948	02	MC	0.07	0.01	0.00		
948	03	MC	0.39	1.26	0.06		
948	04	MC	−1.24	12.98	−0.22	B	M
948	05	MC	0.28	0.18	0.02		
948	06	MC	0.09	0.07	0.02		
948	07	MC	−0.25	0.56	−0.03		
948	08	CR		0.11	0.00		
948	09	CR		2.90	0.07		
948	10	CR		0.22	−0.03		
949	01	MC	−0.39	1.10	−0.06		
949	02	MC	−0.03	0.01	0.00		
949	03	MC	−0.84	5.67	−0.14		
949	04	MC	−0.20	0.33	−0.03		
949	05	MC	−0.08	0.05	−0.01		
949	06	MC	0.07	0.04	0.01		
949	07	MC	−0.60	1.98	−0.10		
949	08	CR		1.19	0.05		
949	09	CR		1.00	−0.05		
949	10	CR		5.84	0.11		

Appendix D: Operational Test Maps

Table 11. Operational Test Map for January 2011

Position	Item Type	Max Points	Weight	Strand	Mean	Point-Biserial	Rasch	S1	S2	S3	S4	S5	S6
1	MC	1	2	Informal and Formal Proofs	0.91	0.41	−2.51						
2	MC	1	2	Transformational Geometry	0.83	0.39	−1.57						
3	MC	1	2	Informal and Formal Proofs	0.88	0.36	−2.09						
4	MC	1	2	Informal and Formal Proofs	0.79	0.44	−1.41						
5	MC	1	2	Geometric Relationships	0.78	0.34	−1.20						
6	MC	1	2	Coordinate Geometry	0.68	0.45	−0.57						
7	MC	1	2	Transformational Geometry	0.80	0.45	−1.35						
8	MC	1	2	Informal and Formal Proofs	0.76	0.54	−1.11						
9	MC	1	2	Geometric Relationships	0.77	0.48	−1.22						
10	MC	1	2	Locus	0.70	0.48	−0.83						
11	MC	1	2	Informal and Formal Proofs	0.64	0.38	−0.48						
12	MC	1	2	Informal and Formal Proofs	0.68	0.46	−0.64						
13	MC	1	2	Informal and Formal Proofs	0.63	0.32	−0.36						
14	MC	1	2	Coordinate Geometry	0.53	0.43	0.19						
15	MC	1	2	Informal and Formal Proofs	0.54	0.55	0.07						
16	MC	1	2	Coordinate Geometry	0.64	0.53	−0.37						
17	MC	1	2	Geometric Relationships	0.45	0.56	0.42						
18	MC	1	2	Informal and Formal Proofs	0.49	0.50	0.29						
19	MC	1	2	Coordinate Geometry	0.50	0.56	0.28						
20	MC	1	2	Constructions	0.44	0.32	0.41						
21	MC	1	2	Coordinate Geometry	0.44	0.56	0.54						
22	MC	1	2	Informal and Formal Proofs	0.40	0.53	0.73						
23	MC	1	2	Informal and Formal Proofs	0.39	0.33	0.87						

Table 11. Operational Test Map for January 2011 *(continued)*

Position	Item Type	Max Points	Weight	Strand	Mean	Point-Biserial	Rasch	S1	S2	S3	S4	S5	S6
24	MC	1	2	Informal and Formal Proofs	0.36	0.40	0.96						
25	MC	1	2	Coordinate Geometry	0.38	0.49	0.83						
26	MC	1	2	Transformational Geometry	0.38	0.34	0.95						
27	MC	1	2	Informal and Formal Proofs	0.35	0.36	1.03						
28	MC	1	2	Geometric Relationships	0.31	0.33	1.24						
29	CR	2	1	Informal and Formal Proofs	1.06	0.66	0.19	1.79	−1.79				
30	CR	2	1	Transformational Geometry	0.95	0.66	0.37	0.26	−0.26				
31	CR	2	1	Informal and Formal Proofs	0.63	0.60	1.05	0.17	−0.17				
32	CR	2	1	Informal and Formal Proofs	0.45	0.52	1.32	0.79	−0.79				
33	CR	2	1	Constructions	0.78	0.56	0.70	4.61	−4.61				
34	CR	2	1	Coordinate Geometry	0.64	0.65	1.10	−0.32	0.32				
35	CR	4	1	Locus	1.15	0.68	1.06	−0.70	0.73	−0.10	0.06		
36	CR	4	1	Informal and Formal Proofs	0.85	0.67	1.57	0.00	−0.76	0.05	0.70		
37	CR	4	1	Informal and Formal Proofs	0.91	0.69	1.04	1.97	−0.87	0.28	−1.38		
38	CR	6	1	Coordinate Geometry	0.97	0.62	1.39	1.55	−1.24	0.34	−0.16	−0.90	0.41

Table 12. Operational Test Map for June 2011

Position	Item Type	Max Points	Weight	Strand	Mean	Point-Biserial	Rasch	S1	S2	S3	S4	S5	S6
1	MC	1	2	Constructions	0.84	0.33	−1.51						
2	MC	1	2	Informal and Formal Proofs	0.83	0.41	−1.40						
3	MC	1	2	Transformational Geometry	0.78	0.44	−0.97						
4	MC	1	2	Informal and Formal Proofs	0.76	0.27	−1.05						
5	MC	1	2	Informal and Formal Proofs	0.75	0.50	−1.04						
6	MC	1	2	Informal and Formal Proofs	0.76	0.33	−0.85						
7	MC	1	2	Informal and Formal Proofs	0.71	0.38	−0.89						
8	MC	1	2	Geometric Relationships	0.70	0.42	−0.48						
9	MC	1	2	Coordinate Geometry	0.65	0.43	−0.55						
10	MC	1	2	Coordinate Geometry	0.65	0.48	−0.33						
11	MC	1	2	Informal and Formal Proofs	0.63	0.44	−0.42						
12	MC	1	2	Geometric Relationships	0.62	0.28	−0.34						
13	MC	1	2	Coordinate Geometry	0.56	0.52	−0.04						
14	MC	1	2	Coordinate Geometry	0.64	0.53	−0.47						
15	MC	1	2	Coordinate Geometry	0.53	0.33	0.36						
16	MC	1	2	Informal and Formal Proofs	0.58	0.52	−0.06						
17	MC	1	2	Informal and Formal Proofs	0.56	0.29	0.00						
18	MC	1	2	Geometric Relationships	0.56	0.37	0.14						
19	MC	1	2	Informal and Formal Proofs	0.52	0.48	0.42						
20	MC	1	2	Informal and Formal Proofs	0.55	0.37	0.10						
21	MC	1	2	Locus	0.53	0.42	0.12						
22	MC	1	2	Transformational Geometry	0.53	0.21	0.29						
23	MC	1	2	Coordinate Geometry	0.54	0.37	0.06						

Table 12. Operational Test Map for June 2011 *(continued)*

Position	Item Type	Max Points	Weight	Strand	Mean	Point-Biserial	Rasch	S1	S2	S3	S4	S5	S6
24	MC	1	2	Informal and Formal Proofs	0.46	0.37	0.61						
25	MC	1	2	Informal and Formal Proofs	0.43	0.36	0.58						
26	MC	1	2	Transformational Geometry	0.41	0.47	0.74						
27	MC	1	2	Coordinate Geometry	0.34	0.33	0.97						
28	MC	1	2	Informal and Formal Proofs	0.32	0.31	1.20						
29	CR	2	1	Informal and Formal Proofs	0.59	0.46	1.27	−1.02	1.02				
30	CR	2	1	Constructions	0.77	0.61	0.72	3.11	−3.11				
31	CR	2	1	Geometric Relationships	1.16	0.67	−0.04	0.75	−0.75				
32	CR	2	1	Transformational Geometry	0.93	0.39	0.51	−2.47	2.47				
33	CR	2	1	Informal and Formal Proofs	0.80	0.64	0.64	0.27	−0.27				
34	CR	2	1	Coordinate Geometry	0.41	0.62	1.51	1.18	−1.18				
35	CR	4	1	Locus	0.97	0.67	1.13	−0.36	0.56	0.39	−0.60		
36	CR	4	1	Informal and Formal Proofs	1.32	0.67	0.79	0.96	−1.32	2.05	−1.69		
37	CR	4	1	Coordinate Geometry	0.91	0.68	1.43	1.15	−2.10	−0.08	1.03		
38	CR	6	1	Informal and Formal Proofs	1.95	0.74	0.88	0.50	−0.95	2.31	−1.98	0.35	−0.24

Table 13. Operational Test Map for August 2011

Position	Item Type	Max Points	Weight	Strand	Mean	Point-Biserial	Rasch	S1	S2	S3	S4	S5	S6
1	MC	1	2	Informal and Formal Proofs	0.93	0.40	−2.72						
2	MC	1	2	Informal and Formal Proofs	0.78	0.35	−1.10						
3	MC	1	2	Informal and Formal Proofs	0.79	0.34	−1.03						
4	MC	1	2	Transformational Geometry	0.73	0.53	−0.90						
5	MC	1	2	Informal and Formal Proofs	0.65	0.31	−0.52						
6	MC	1	2	Constructions	0.78	0.46	−0.91						
7	MC	1	2	Coordinate Geometry	0.63	0.43	−0.30						
8	MC	1	2	Transformational Geometry	0.70	0.47	−0.52						
9	MC	1	2	Informal and Formal Proofs	0.64	0.38	−0.37						
10	MC	1	2	Coordinate Geometry	0.62	0.49	−0.25						
11	MC	1	2	Informal and Formal Proofs	0.62	0.33	−0.04						
12	MC	1	2	Coordinate Geometry	0.62	0.39	−0.15						
13	MC	1	2	Transformational Geometry	0.57	0.48	0.06						
14	MC	1	2	Informal and Formal Proofs	0.52	0.59	0.34						
15	MC	1	2	Coordinate Geometry	0.59	0.56	0.01						
16	MC	1	2	Geometric Relationships	0.52	0.46	0.18						
17	MC	1	2	Locus	0.54	0.34	0.10						
18	MC	1	2	Coordinate Geometry	0.51	0.36	0.18						
19	MC	1	2	Informal and Formal Proofs	0.50	0.5	0.30						
20	MC	1	2	Geometric Relationships	0.51	0.45	0.25						
21	MC	1	2	Informal and Formal Proofs	0.46	0.39	0.34						
22	MC	1	2	Coordinate Geometry	0.44	0.42	0.72						
23	MC	1	2	Geometric Relationships	0.38	0.39	1.02						

Table 13. Operational Test Map for August 2011 *(continued)*

Position	Item Type	Max Points	Weight	Strand	Mean	Point-Biserial	Rasch	S1	S2	S3	S4	S5	S6
24	MC	1	2	Informal and Formal Proofs	0.38	0.34	0.82						
25	MC	1	2	Informal and Formal Proofs	0.44	0.41	0.53						
26	MC	1	2	Coordinate Geometry	0.36	0.41	1.06						
27	MC	1	2	Informal and Formal Proofs	0.37	0.36	0.84						
28	MC	1	2	Informal and Formal Proofs	0.41	0.44	0.82						
29	CR	2	1	Informal and Formal Proofs	1.35	0.64	−0.45	1.38	−1.38				
30	CR	2	1	Constructions	0.87	0.58	0.73	0.56	−0.56				
31	CR	2	1	Geometric Relationships	1.12	0.65	0.01	0.08	−0.08				
32	CR	2	1	Coordinate Geometry	0.66	0.62	1.00	0.37	−0.37				
33	CR	2	1	Informal and Formal Proofs	0.70	0.68	0.84	0.62	−0.62				
34	CR	2	1	Locus	0.79	0.67	0.83	0.47	−0.47				
35	CR	4	1	Informal and Formal Proofs	1.43	0.74	0.62	1.02	0.72	−0.31	−1.42		
36	CR	4	1	Transformational Geometry	1.54	0.73	0.77	0.97	−1.29	0.63	−0.32		
37	CR	4	1	Informal and Formal Proofs	1.68	0.77	0.64	0.26	0.65	0.07	−0.98		
38	CR	6	1	Coordinate Geometry	1.42	0.84	1.40	−0.77	0.96	−0.65	0.53	0.26	−0.33

Appendix E: Scoring Tables

Table 14. Scoring Table for January 2011

Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score
0	−6.023	0.000		23	−0.825	43.915		46	0.572	69.782		69	1.632	83.527
1	−4.790	0.906		24	−0.748	45.491		47	0.618	70.538		70	1.694	84.142
2	−4.056	3.753		25	−0.673	47.059		48	0.664	71.198		71	1.761	84.846
3	−3.611	5.441		26	−0.599	48.586		49	0.709	71.828		72	1.831	85.564
4	−3.283	7.404		27	−0.528	49.989		50	0.753	72.553		73	1.908	86.294
5	−3.020	9.303		28	−0.458	51.351		51	0.796	73.158		74	1.990	87.087
6	−2.799	11.230		29	−0.389	52.714		52	0.839	73.744		75	2.079	87.885
7	−2.606	13.234		30	−0.323	53.986		53	0.882	74.324		76	2.177	88.693
8	−2.434	15.268		31	−0.257	55.254		54	0.924	74.900		77	2.285	89.562
9	−2.277	17.306		32	−0.193	56.499		55	0.966	75.449		78	2.407	90.485
10	−2.134	19.305		33	−0.131	57.667		56	1.008	76.080		79	2.544	91.483
11	−2.001	21.344		34	−0.070	58.836		57	1.050	76.637		80	2.702	92.496
12	−1.876	23.383		35	−0.010	59.915		58	1.093	77.171		81	2.889	93.571
13	−1.759	25.430		36	0.049	60.989		59	1.136	77.694		82	3.117	94.679
14	−1.647	27.471		37	0.106	62.026		60	1.179	78.245		83	3.410	95.861
15	−1.541	29.441		38	0.163	62.983		61	1.223	78.825		84	3.821	97.108
16	−1.440	31.393		39	0.218	63.932		62	1.268	79.341		85	4.521	98.497
17	−1.343	33.331		40	0.272	64.884		63	1.315	79.931		86	5.729	100.000
18	−1.249	35.174		41	0.324	65.820		64	1.362	80.453				
19	−1.159	36.998		42	0.376	66.657		65	1.412	81.023				
20	−1.071	38.836		43	0.426	67.467		66	1.463	81.616				
21	−0.987	40.617		44	0.476	68.282		67	1.517	82.213				
22	−0.905	42.263		45	0.524	69.077		68	1.573	82.837				

Table 15. Scoring Table for June 2011

Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score
0	−5.739	0.000		23	−0.788	44.669		46	0.520	69.002		69	1.616	83.350
1	−4.516	1.983		24	−0.718	46.102		47	0.566	69.704		70	1.681	84.012
2	−3.796	4.738		25	−0.650	47.537		48	0.612	70.438		71	1.750	84.730
3	−3.364	6.909		26	−0.584	48.905		49	0.658	71.112		72	1.823	85.483
4	−3.050	9.064		27	−0.518	50.164		50	0.703	71.741		73	1.902	86.245
5	−2.800	11.217		28	−0.455	51.411		51	0.747	72.473		74	1.988	87.069
6	−2.591	13.407		29	−0.392	52.656		52	0.792	73.096		75	2.081	87.903
7	−2.410	15.550		30	−0.331	53.826		53	0.836	73.696		76	2.184	88.751
8	−2.250	17.669		31	−0.271	54.985		54	0.880	74.295		77	2.298	89.658
9	−2.105	19.710		32	−0.212	56.159		55	0.923	74.892		78	2.426	90.634
10	−1.973	21.781		33	−0.154	57.235		56	0.967	75.464		79	2.571	91.670
11	−1.851	23.797		34	−0.097	58.314		57	1.011	76.128		80	2.740	92.719
12	−1.737	25.815		35	−0.041	59.367		58	1.056	76.705		81	2.939	93.847
13	−1.630	27.784		36	0.014	60.353		59	1.100	77.267		82	3.181	94.947
14	−1.529	29.662		37	0.068	61.348		60	1.146	77.816		83	3.492	96.185
15	−1.433	31.523		38	0.122	62.288		61	1.192	78.423		84	3.925	97.377
16	−1.342	33.347		39	0.174	63.182		62	1.239	79.005		85	4.651	98.735
17	−1.254	35.076		40	0.226	64.075		63	1.287	79.582		86	5.882	100.000
18	−1.170	36.772		41	0.277	64.976		64	1.337	80.180				
19	−1.088	38.476		42	0.327	65.860		65	1.388	80.745				
20	−1.010	40.134		43	0.376	66.659		66	1.441	81.372				
21	−0.934	41.683		44	0.425	67.437		67	1.497	81.999				
22	−0.860	43.200		45	0.472	68.224		68	1.555	82.635				

Table 16. Scoring Table for August 2011

Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score
0	−5.747	0.000		23	−0.674	47.021		46	0.494	68.586		69	1.483	81.842
1	−4.504	2.032		24	−0.608	48.402		47	0.534	69.228		70	1.545	82.522
2	−3.758	4.884		25	−0.544	49.675		48	0.573	69.794		71	1.610	83.282
3	−3.304	7.274		26	−0.482	50.879		49	0.611	70.422		72	1.681	84.012
4	−2.972	9.690		27	−0.421	52.082		50	0.650	70.998		73	1.757	84.806
5	−2.709	12.159		28	−0.362	53.240		51	0.688	71.538		74	1.839	85.640
6	−2.488	14.623		29	−0.304	54.340		52	0.727	72.125		75	1.929	86.494
7	−2.298	17.029		30	−0.248	55.440		53	0.765	72.721		76	2.027	87.453
8	−2.131	19.347		31	−0.193	56.504		54	0.803	73.254		77	2.136	88.352
9	−1.981	21.656		32	−0.140	57.501		55	0.842	73.782		78	2.258	89.354
10	−1.845	23.901		33	−0.088	58.492		56	0.881	74.313		79	2.397	90.409
11	−1.720	26.128		34	−0.037	59.441		57	0.920	74.851		80	2.556	91.566
12	−1.604	28.263		35	0.013	60.331		58	0.960	75.375		81	2.744	92.745
13	−1.497	30.271		36	0.062	61.222		59	1.001	75.972		82	2.973	94.041
14	−1.395	32.273		37	0.109	62.072		60	1.042	76.537		83	3.268	95.302
15	−1.300	34.178		38	0.155	62.860		61	1.085	77.073		84	3.681	96.745
16	−1.209	35.958		39	0.201	63.636		62	1.128	77.608		85	4.381	98.243
17	−1.123	37.736		40	0.245	64.412		63	1.173	78.166		86	5.591	100.000
18	−1.041	39.481		41	0.288	65.184		64	1.219	78.784				
19	−0.962	41.119		42	0.331	65.929		65	1.267	79.331				
20	−0.886	42.652		43	0.373	66.607		66	1.318	79.968				
21	−0.813	44.163		44	0.414	67.262		67	1.370	80.534				
22	−0.743	45.600		45	0.454	67.921		68	1.425	81.180				