

**New York State Examination in Grade 8  
Intermediate-Level Science**

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

**Technical Report**



Prepared for the New York State Education Department  
by Pearson

**March 2012**

# Copyright

---

Developed and published under contract with the New York State Education Department by Pearson. Copyright © 2011 by the New York State Education Department.

## **Secure Materials.**

All rights reserved. No part of this document may be reproduced or transmitted by any means. Use of these materials is expressly limited to the New York State Education Department.

## Table of Contents

---

Table of Contents.....	i
List of Tables.....	ii
Section I: Introduction.....	1
Purpose .....	1
Section II: Field Test Analysis .....	1
Data Cleanup .....	2
Classical Analysis.....	2
Item Difficulty .....	3
Point-Biserial Correlation.....	3
Test Reliability.....	4
Scoring Reliability.....	5
Inter-rater Agreement.....	5
Constructed-Response Item Means and Standard Deviations.....	11
Intraclass Correlation .....	11
Weighted Kappa.....	12
Item Response Theory (IRT) Statistics.....	12
Item Calibration.....	13
Item Fit Evaluation .....	13
Differential Item Functioning (DIF) Statistics .....	14
Section III: Equating Procedure.....	16
Section IV: Scaling of Operational Test Forms.....	17
References.....	19
Appendix A: Classical Item Analysis .....	20
Appendix B: Partial Credit Model Item Analysis .....	31
Appendix C: DIF Statistics.....	42
Appendix D: Operational Test Map .....	48
Appendix E: Scoring Table.....	53

## List of Tables

---

Table 1. Need/Resource Capacity Category Definitions .....	1
Table 2. Classical Item Analysis.....	4
Table 3. Test and Scoring Reliability .....	5
Table 4. Point Differences Between First and Second Reads.....	6
Table 5. First and Second Read Descriptive Statistics and Agreement .....	9
Table 6. Partial Credit Model Item Analysis.....	14
Table 7. Initial Mean Abilities and Equating Constants.....	17
Table 8. Classical Item Analysis.....	21
Table 9. Partial Credit Model Item Analysis.....	32
Table 10. DIF Statistics .....	43
Table 11. Operational Test Map for June 2011 .....	49
Table 12. Scoring Table for June 2011 .....	54

## Section I: Introduction

---

### PURPOSE

The purpose of this report is to document the psychometric work on the New York State Examination in Grade 8 Intermediate-Level Science for 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 Examination in Grade 8 Intermediate-Level Science.

## Section II: Field Test Analysis

---

In May 2011, field testing was conducted for the New York State Examination in Grade 8 Intermediate-Level Science 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**

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

The data collected from field testing were scored by the New York State Education Department. Both classical and item response theory analyses were conducted using the data to evaluate the quality of the test items.

## **DATA CLEANUP**

Field test forms contained multiple-choice and constructed-response item types. Response data were contained in one file that contained 9,492 student records. After the exclusion rules were applied, the resulting field test data file contained 8,684 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. For all items except six, item difficulties were equal to or below 0.90. With respect to the point-biserial correlations, only two of these 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
401	937	23	7	13	3	0	21	2
402	970	23	3	20	0	0	20	3
403	949	23	5	16	1	0	21	1
404	994	30	6	23	0	1	20	8
405	981	23	7	16	0	1	16	6
406	968	25	8	15	0	0	18	5
407	953	25	4	19	2	0	22	3
408	956	25	5	19	0	0	19	5
409	976	25	7	18	0	0	18	7

\* For some forms, the item counts in the “Item Difficulty” and “Point-Biserial” columns may not sum to the value in the “No. of Items” column due to DNS (do not score) items.

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  $\alpha$  (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 0.75 to 0.85 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**

<b>Form Number</b>	<b>Test Reliability</b>	<b>Scoring Reliability</b>
401	0.78	0.91
402	0.81	0.85
403	0.75	0.90
404	0.85	n/a
405	0.80	0.92
406	0.80	0.89
407	0.79	0.83
408	0.81	0.82
409	0.84	0.92

### 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 0.82 to 0.92, 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. The majority of the differences between the first read and the second read were 0.

**Table 4. Point Differences Between First and Second Reads**

			Difference (First Read minus Second Read)						
Form	Item	Score Points	-3	-2	-1	0	1	2	3
401	41	2	0.00	0.00	0.03	0.95	0.03	0.00	0.00
401	42	1	0.00	0.00	0.01	0.97	0.01	0.00	0.00
401	43	1	0.00	0.00	0.03	0.87	0.10	0.00	0.00
401	44	1	0.00	0.00	0.01	0.92	0.07	0.00	0.00
401	45	1	0.00	0.00	0.01	0.97	0.01	0.00	0.00
401	46	1	0.00	0.00	0.04	0.94	0.03	0.00	0.00
401	47	1	0.00	0.00	0.12	0.86	0.03	0.00	0.00
401	48	1	0.00	0.00	0.03	0.96	0.01	0.00	0.00
401	49	1	0.00	0.00	0.03	0.95	0.03	0.00	0.00
401	50	1	0.00	0.00	0.03	0.95	0.03	0.00	0.00
401	51	1	0.00	0.00	0.04	0.96	0.00	0.00	0.00
402	41	1	0.00	0.00	0.04	0.86	0.10	0.00	0.00
402	42	1	0.00	0.00	0.02	0.96	0.02	0.00	0.00
402	43	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
402	44	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
402	45	1	0.00	0.00	0.01	0.98	0.01	0.00	0.00
402	46	1	0.00	0.00	0.05	0.88	0.07	0.00	0.00
402	47	1	0.00	0.00	0.08	0.84	0.09	0.00	0.00
402	48	1	0.00	0.00	0.01	0.99	0.00	0.00	0.00
402	49	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
403	41	2	0.00	0.02	0.05	0.92	0.01	0.00	0.00
403	43	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
403	44	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
403	45	1	0.00	0.00	0.01	0.99	0.00	0.00	0.00
403	46	1	0.00	0.00	0.05	0.89	0.06	0.00	0.00
403	47	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
403	48	1	0.00	0.00	0.00	0.96	0.04	0.00	0.00
403	49	1	0.00	0.00	0.01	0.98	0.01	0.00	0.00
403	50	1	0.00	0.00	0.12	0.84	0.04	0.00	0.00
403	51	1	0.00	0.00	0.06	0.86	0.07	0.00	0.00
405	41	1	0.00	0.00	0.01	0.97	0.02	0.00	0.00
405	42	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00

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

			Difference (First Read minus Second Read)						
Form	Item	Score Points	-3	-2	-1	0	1	2	3
405	43	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00
405	44	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
405	45	1	0.00	0.00	0.01	0.99	0.00	0.00	0.00
405	46	1	0.00	0.00	0.04	0.88	0.07	0.00	0.00
405	47	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
405	48	1	0.00	0.00	0.00	0.97	0.03	0.00	0.00
405	49	1	0.00	0.00	0.03	0.88	0.09	0.00	0.00
405	50	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
405	51	1	0.00	0.00	0.02	0.96	0.02	0.00	0.00
406	41	1	0.00	0.00	0.02	0.97	0.01	0.00	0.00
406	43	1	0.00	0.00	0.01	0.95	0.04	0.00	0.00
406	44	1	0.00	0.00	0.07	0.81	0.12	0.00	0.00
406	45	1	0.00	0.00	0.02	0.98	0.00	0.00	0.00
406	46	1	0.00	0.00	0.02	0.96	0.02	0.00	0.00
406	47	1	0.00	0.00	0.04	0.92	0.03	0.00	0.00
406	48	1	0.00	0.00	0.01	0.98	0.01	0.00	0.00
406	49	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
406	50	1	0.00	0.00	0.04	0.92	0.04	0.00	0.00
406	51	1	0.00	0.00	0.02	0.96	0.01	0.00	0.00
406	52	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
406	53	1	0.00	0.00	0.04	0.95	0.01	0.00	0.00
407	41	1	0.00	0.00	0.01	0.99	0.00	0.00	0.00
407	42	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00
407	43	1	0.00	0.00	0.07	0.91	0.02	0.00	0.00
407	44	1	0.00	0.00	0.04	0.92	0.04	0.00	0.00
407	45	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00
407	46	1	0.00	0.00	0.03	0.93	0.04	0.00	0.00
407	47	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
407	48	1	0.00	0.00	0.12	0.70	0.18	0.00	0.00
407	49	1	0.00	0.00	0.01	0.97	0.02	0.00	0.00
407	50	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
407	51	1	0.00	0.00	0.10	0.79	0.11	0.00	0.00

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

			Difference (First Read minus Second Read)						
Form	Item	Score Points	-3	-2	-1	0	1	2	3
407	52	1	0.00	0.00	0.01	0.99	0.00	0.00	0.00
408	41	1	0.00	0.00	0.02	0.96	0.02	0.00	0.00
408	42	1	0.00	0.00	0.00	0.97	0.03	0.00	0.00
408	43	1	0.00	0.00	0.03	0.90	0.06	0.00	0.00
408	44	1	0.00	0.00	0.01	0.99	0.00	0.00	0.00
408	45	1	0.00	0.00	0.10	0.85	0.05	0.00	0.00
408	46	1	0.00	0.00	0.03	0.95	0.02	0.00	0.00
408	47	1	0.00	0.00	0.03	0.95	0.02	0.00	0.00
408	48	1	0.00	0.00	0.02	0.98	0.00	0.00	0.00
408	49	1	0.00	0.00	0.00	0.98	0.02	0.00	0.00
408	50	1	0.00	0.00	0.02	0.97	0.01	0.00	0.00
408	51	1	0.00	0.00	0.09	0.73	0.18	0.00	0.00
408	52	1	0.00	0.00	0.05	0.89	0.05	0.00	0.00
408	53	1	0.00	0.00	0.03	0.91	0.05	0.00	0.00
409	41	1	0.00	0.00	0.02	0.97	0.01	0.00	0.00
409	42	1	0.00	0.00	0.04	0.93	0.03	0.00	0.00
409	43	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
409	44	1	0.00	0.00	0.03	0.97	0.00	0.00	0.00
409	45	1	0.00	0.00	0.01	0.98	0.01	0.00	0.00
409	46	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00
409	47	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
409	48	1	0.00	0.00	0.00	1.00	0.00	0.00	0.00
409	49	1	0.00	0.00	0.05	0.93	0.01	0.00	0.00
409	50	1	0.00	0.00	0.01	0.98	0.01	0.00	0.00
409	51	1	0.00	0.00	0.01	0.97	0.02	0.00	0.00
409	52	1	0.00	0.00	0.03	0.97	0.00	0.00	0.00
409	53	1	0.00	0.00	0.03	0.91	0.05	0.00	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 97.9% and 100%. These values indicate a high degree of agreement.

**Table 5. First and Second Read Descriptive Statistics and Agreement**

				Agreement (%)			Raw Score Mean		Raw Score Standard Deviation			
Form	Item	Score Points	Total N-Count	Exact	Adj.	Total	First Read	Second Read	First Read	Second Read	Intraclass Correlation	Wt. Kappa
401	41	2	78	94.9	5.1	100.0	1.6	1.6	0.64	0.66	0.94	0.92
401	42	1	77	97.4	2.6	100.0	0.5	0.5	0.50	0.50	0.95	0.95
401	43	1	78	87.2	12.8	100.0	0.5	0.4	0.50	0.49	0.75	0.74
401	44	1	76	92.1	7.9	100.0	0.5	0.5	0.50	0.50	0.85	0.84
401	45	1	77	97.4	2.6	100.0	0.7	0.7	0.46	0.46	0.94	0.94
401	46	1	78	93.6	6.4	100.0	0.8	0.8	0.41	0.40	0.80	0.80
401	47	1	78	85.9	14.1	100.0	0.9	0.9	0.35	0.22	0.24	0.21
401	48	1	77	96.1	3.9	100.0	0.7	0.7	0.45	0.45	0.90	0.90
401	49	1	78	94.9	5.1	100.0	0.6	0.6	0.48	0.48	0.89	0.89
401	50	1	78	94.9	5.1	100.0	0.3	0.3	0.48	0.48	0.89	0.89
401	51	1	77	96.1	3.9	100.0	0.3	0.4	0.47	0.48	0.92	0.91
402	41	1	94	86.2	13.8	100.0	0.9	0.9	0.28	0.35	0.32	0.31
402	42	1	94	95.7	4.3	100.0	0.8	0.8	0.37	0.37	0.84	0.84
402	43	1	89	98.9	1.1	100.0	0.5	0.5	0.50	0.50	0.98	0.98
402	44	1	90	98.9	1.1	100.0	0.4	0.4	0.50	0.49	0.98	0.98
402	45	1	94	97.9	2.1	100.0	0.8	0.8	0.42	0.42	0.94	0.94
402	46	1	92	88.0	12.0	100.0	0.8	0.8	0.41	0.42	0.66	0.65
402	47	1	92	83.7	16.3	100.0	0.8	0.8	0.42	0.43	0.55	0.54
402	48	1	92	98.9	1.1	100.0	0.8	0.8	0.38	0.37	0.96	0.96
402	49	1	91	98.9	1.1	100.0	0.8	0.8	0.40	0.41	0.97	0.97
403	41	2	96	91.7	6.2	97.9	1.7	1.8	0.62	0.53	0.80	0.76
403	43	1	96	99.0	1.0	100.0	0.5	0.4	0.50	0.50	0.98	0.98
403	44	1	96	93.8	6.2	100.0	0.7	0.7	0.44	0.45	0.84	0.84
403	45	1	96	99.0	1.0	100.0	0.7	0.7	0.47	0.47	0.98	0.98
403	46	1	96	88.5	11.5	100.0	0.7	0.7	0.47	0.47	0.74	0.74
403	47	1	96	99.0	1.0	100.0	0.8	0.8	0.40	0.41	0.97	0.97
403	48	1	95	95.8	4.2	100.0	1.0	0.9	0.20	0.28	0.69	0.65
403	49	1	95	97.9	2.1	100.0	0.4	0.4	0.49	0.49	0.96	0.96
403	50	1	93	83.9	16.1	100.0	0.2	0.3	0.43	0.47	0.61	0.60
403	51	1	95	86.3	13.7	100.0	0.5	0.5	0.50	0.50	0.73	0.73
405	41	1	94	96.8	3.2	100.0	0.5	0.5	0.50	0.50	0.94	0.94
405	42	1	94	100.0	0.0	100.0	0.3	0.3	0.46	0.46	1.00	1.00
405	43	1	95	100.0	0.0	100.0	0.7	0.7	0.47	0.47	1.00	1.00
405	44	1	91	98.9	1.1	100.0	0.7	0.7	0.45	0.45	0.97	0.97
405	45	1	93	98.9	1.1	100.0	0.4	0.4	0.49	0.49	0.98	0.98
405	46	1	94	88.3	11.7	100.0	0.8	0.8	0.39	0.41	0.63	0.63

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

				Agreement (%)			Raw Score Mean		Raw Score Standard Deviation			
Form	Item	Score Points	Total N-Count	Exact	Adj.	Total	First Read	Second Read	First Read	Second Read	Intraclass Correlation	Wt. Kappa
405	47	1	94	93.6	6.4	100.0	0.7	0.7	0.46	0.47	0.85	0.85
405	48	1	91	96.7	3.3	100.0	0.6	0.6	0.48	0.49	0.93	0.93
405	49	1	89	87.6	12.4	100.0	0.4	0.3	0.49	0.47	0.73	0.73
405	50	1	94	98.9	1.1	100.0	0.8	0.8	0.42	0.43	0.97	0.97
405	51	1	95	95.8	4.2	100.0	0.7	0.7	0.45	0.45	0.89	0.89
406	41	1	93	96.8	3.2	100.0	0.7	0.7	0.45	0.44	0.92	0.92
406	43	1	93	94.6	5.4	100.0	0.4	0.4	0.50	0.49	0.89	0.89
406	44	1	95	81.1	18.9	100.0	0.7	0.7	0.46	0.48	0.57	0.56
406	45	1	95	97.9	2.1	100.0	0.7	0.7	0.47	0.46	0.95	0.95
406	46	1	95	95.8	4.2	100.0	0.8	0.8	0.38	0.38	0.85	0.85
406	47	1	93	92.5	7.5	100.0	0.6	0.6	0.49	0.49	0.84	0.84
406	48	1	95	97.9	2.1	100.0	0.7	0.7	0.48	0.48	0.95	0.95
406	49	1	92	98.9	1.1	100.0	0.3	0.3	0.48	0.48	0.98	0.98
406	50	1	95	91.6	8.4	100.0	0.5	0.5	0.50	0.50	0.83	0.83
406	51	1	81	96.3	3.7	100.0	0.3	0.3	0.46	0.46	0.91	0.91
406	52	1	91	98.9	1.1	100.0	0.5	0.5	0.50	0.50	0.98	0.98
406	53	1	91	94.5	5.5	100.0	0.2	0.2	0.38	0.41	0.83	0.82
407	41	1	95	98.9	1.1	100.0	0.9	1.0	0.22	0.20	0.89	0.88
407	42	1	95	100.0	0.0	100.0	0.6	0.6	0.48	0.48	1.00	1.00
407	43	1	95	90.5	9.5	100.0	0.8	0.9	0.39	0.33	0.65	0.64
407	44	1	95	91.6	8.4	100.0	0.7	0.7	0.48	0.48	0.81	0.81
407	45	1	95	100.0	0.0	100.0	0.8	0.8	0.38	0.38	1.00	1.00
407	46	1	95	92.6	7.4	100.0	0.6	0.6	0.49	0.50	0.85	0.85
407	47	1	95	98.9	1.1	100.0	0.8	0.8	0.41	0.42	0.97	0.97
407	48	1	94	70.2	29.8	100.0	0.4	0.4	0.50	0.49	0.39	0.38
407	49	1	94	96.8	3.2	100.0	0.5	0.5	0.50	0.50	0.94	0.94
407	50	1	94	93.6	6.4	100.0	0.4	0.4	0.49	0.48	0.86	0.86
407	51	1	92	79.3	20.7	100.0	0.7	0.7	0.47	0.48	0.53	0.53
407	52	1	92	98.9	1.1	100.0	0.8	0.8	0.38	0.37	0.96	0.96
408	41	1	93	95.7	4.3	100.0	0.9	0.9	0.30	0.30	0.75	0.75
408	42	1	93	96.8	3.2	100.0	0.8	0.7	0.43	0.45	0.92	0.92
408	43	1	93	90.3	9.7	100.0	0.8	0.8	0.38	0.41	0.69	0.68
408	44	1	93	98.9	1.1	100.0	0.8	0.8	0.41	0.40	0.97	0.97
408	45	1	93	84.9	15.1	100.0	0.8	0.8	0.42	0.39	0.54	0.54
408	46	1	91	94.5	5.5	100.0	0.3	0.3	0.47	0.48	0.88	0.88
408	47	1	92	94.6	5.4	100.0	0.7	0.7	0.47	0.47	0.88	0.88
408	48	1	93	97.8	2.2	100.0	0.9	0.9	0.31	0.28	0.88	0.88

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

				Agreement (%)			Raw Score Mean		Raw Score Standard Deviation			
Form	Item	Score Points	Total N-Count	Exact	Adj.	Total	First Read	Second Read	First Read	Second Read	Intraclass Correlation	Wt. Kappa
408	49	1	91	97.8	2.2	100.0	0.4	0.4	0.49	0.49	0.96	0.95
408	50	1	93	96.8	3.2	100.0	0.9	0.9	0.28	0.27	0.78	0.78
408	51	1	92	72.8	27.2	100.0	0.5	0.4	0.50	0.48	0.45	0.44
408	52	1	92	89.1	10.9	100.0	0.9	0.9	0.35	0.35	0.55	0.55
408	53	1	92	91.3	8.7	100.0	0.5	0.5	0.50	0.50	0.83	0.83
409	41	1	94	96.8	3.2	100.0	0.4	0.4	0.50	0.50	0.94	0.94
409	42	1	94	92.6	7.4	100.0	0.6	0.6	0.50	0.50	0.85	0.85
409	43	1	94	98.9	1.1	100.0	0.6	0.6	0.48	0.48	0.98	0.98
409	44	1	93	96.8	3.2	100.0	0.8	0.8	0.43	0.41	0.91	0.91
409	45	1	94	97.9	2.1	100.0	0.7	0.7	0.46	0.46	0.95	0.95
409	46	1	94	100.0	0.0	100.0	0.8	0.8	0.41	0.41	1.00	1.00
409	47	1	94	93.6	6.4	100.0	0.6	0.6	0.49	0.50	0.87	0.87
409	48	1	92	100.0	0.0	100.0	0.7	0.7	0.45	0.45	1.00	1.00
409	49	1	92	93.5	6.5	100.0	0.5	0.5	0.50	0.50	0.87	0.87
409	50	1	90	97.8	2.2	100.0	0.6	0.6	0.50	0.50	0.95	0.95
409	51	1	94	96.8	3.2	100.0	0.6	0.6	0.50	0.50	0.93	0.93
409	52	1	90	96.7	3.3	100.0	0.2	0.2	0.40	0.43	0.91	0.90
409	53	1	92	91.3	8.7	100.0	0.8	0.8	0.38	0.40	0.71	0.71

\* 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 but nine items 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 0.21 to 1.00 and in all but three cases indicate 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 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 no items had an INFIT statistic outside the typical range.

**Table 4. Partial Credit Model Item Analysis**

Form	N-Count	No. of Items	Rasch			INFIT		
			<-2.0	-2.0 to 2.0	>2.0	<-0.70	-0.70 to 1.30	>1.30
401	937	23	2	20	1	0	23	0
402	970	23	0	23	0	0	23	0
403	949	23	0	22	0	0	22	0
404	994	30	0	29	0	0	29	0
405	981	23	0	23	0	0	23	0
406	968	25	0	22	1	0	23	0
407	953	25	1	24	0	0	25	0
408	956	25	0	24	0	0	24	0
409	976	25	0	24	1	0	25	0

\* For some forms, the item counts in the “Item Difficulty” and “Point-Biserial” columns may not sum to the value in the “No. of Items” column due to DNS (do not score) items.

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  $\{|MHD| \geq 1.0 \text{ and } < 1.5\}$  **or** MHD not significantly different from 0 and  $|MHD| \geq 1.0$  are classified as B.
- $|MHD| \geq 1.5$  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  $|ES| \leq 0.17$ , classified as AA.
- Probability is  $> 0.05$  and if  $0.17 < |ES| \leq 0.25$ , classified as BB.
- Probability is  $> 0.05$  and if  $|ES| > 0.25$ , 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 Examination in Grade 8 Intermediate-Level Science consisted of eight field test forms numbered 401–403 and 405–409 and one anchor form labeled 404. The field test forms contained multiple-choice and constructed-response items. Each student participating in the field test was administered one of the nine 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 form was 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 form onto the item bank scale. After the anchor form was 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 to update the item parameters for the anchor form.

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 the anchor form resulted in no items having a displacement value with a magnitude greater than 0.30. This indicates stability in the items used on the anchor form.

The equated mean ability estimate for form 404 was 0.86. This value served as the target mean ability for the remainder of the equating process.

After the anchor form was 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 form (i.e., form 404). 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 5. Initial Mean Abilities and Equating Constants**

Form Number	Mean Ability	Constant
401	0.93	-0.07
402	0.90	-0.04
403	0.84	0.02
405	0.59	0.25
406	0.28	0.55
407	0.91	-0.05
408	0.76	0.10
409	0.55	0.30

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 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 May 2011 operational test map with content information regarding each item included on the form.

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 0.068 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 table used for the May 2011 operational test form. Four steps are taken in order to produce this table and the resulting conversion chart.

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 May 2006. 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:

$$p(x) = m_3x^3 + m_2x^2 + m_1x + m_0$$

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. 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 May 2006 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_G8Sc_FT	401	MC	01	1	937	0.78	0.00		0.95	0.04	0.01	0.00			0.95	0.31
2011_G8Sc_FT	401	MC	02	1	937	0.78	0.00		0.03	0.01	0.80	0.15			0.80	0.29
2011_G8Sc_FT	401	MC	03	1	937	0.78	0.00		0.06	0.10	0.03	0.80			0.80	0.47
2011_G8Sc_FT	401	MC	04	1	937	0.78	0.00		0.01	0.01	0.95	0.02			0.95	0.38
2011_G8Sc_FT	401	MC	05	1	937	0.78	0.00		0.05	0.87	0.07	0.02			0.87	0.34
2011_G8Sc_FT	401	MC	06	1	937	0.78	0.00		0.02	0.05	0.02	0.91			0.91	0.39
2011_G8Sc_FT	401	MC	07	1	937	0.78	0.00		0.67	0.26	0.03	0.03			0.67	0.40
2011_G8Sc_FT	401	MC	08	1	937	0.78	0.00		0.09	0.22	0.01	0.68			0.68	0.40
2011_G8Sc_FT	401	MC	09	1	937	0.78	0.00		0.80	0.05	0.07	0.07			0.80	0.42
2011_G8Sc_FT	401	MC	10	1	937	0.78	0.01		0.07	0.08	0.19	0.66			0.66	0.45
2011_G8Sc_FT	401	MC	11	1	937	0.78	0.00		0.37	0.09	0.12	0.42			0.37	0.43
2011_G8Sc_FT	401	MC	12	1	937	0.78	0.02		0.35	0.39	0.19	0.05			0.35	0.38
2011_G8Sc_FT	401	CR	41	2	937	0.78	0.02	0.14	0.28	0.56					1.40	0.53
2011_G8Sc_FT	401	CR	42	1	937	0.78	0.01	0.57	0.42						0.42	0.45
2011_G8Sc_FT	401	CR	43	1	937	0.78	0.04	0.55	0.41						0.41	0.51
2011_G8Sc_FT	401	CR	44	1	937	0.78	0.04	0.50	0.46						0.46	0.50
2011_G8Sc_FT	401	CR	45	1	937	0.78	0.02	0.33	0.65						0.65	0.48
2011_G8Sc_FT	401	CR	46	1	937	0.78	0.02	0.24	0.74						0.74	0.41
2011_G8Sc_FT	401	CR	47	1	937	0.78	0.02	0.12	0.86						0.86	0.37
2011_G8Sc_FT	401	CR	48	1	937	0.78	0.03	0.26	0.71						0.71	0.36
2011_G8Sc_FT	401	CR	49	1	937	0.78	0.03	0.33	0.63						0.63	0.42
2011_G8Sc_FT	401	CR	50	1	937	0.78	0.04	0.66	0.31						0.31	0.48

**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_G8Sc_FT	401	CR	51	1	937	0.78	0.07	0.65	0.28						0.28	0.45
2011_G8Sc_FT	402	MC	01	1	970	0.81	0.00		0.06	0.04	0.84	0.06			0.84	0.49
2011_G8Sc_FT	402	MC	02	1	970	0.81	0.00		0.05	0.79	0.04	0.12			0.79	0.47
2011_G8Sc_FT	402	MC	03	1	970	0.81	0.00		0.82	0.07	0.08	0.02			0.82	0.43
2011_G8Sc_FT	402	MC	04	1	970	0.81	0.00		0.02	0.07	0.04	0.86			0.86	0.44
2011_G8Sc_FT	402	MC	05	1	970	0.81	0.00		0.19	0.63	0.13	0.06			0.63	0.45
2011_G8Sc_FT	402	MC	06	1	970	0.81	0.00		0.12	0.79	0.07	0.02			0.79	0.44
2011_G8Sc_FT	402	MC	07	1	970	0.81	0.00		0.16	0.21	0.60	0.03			0.60	0.48
2011_G8Sc_FT	402	MC	08	1	970	0.81	0.00		0.24	0.08	0.07	0.61			0.61	0.49
2011_G8Sc_FT	402	MC	09	1	970	0.81	0.00		0.08	0.50	0.07	0.34			0.50	0.25
2011_G8Sc_FT	402	MC	10	1	970	0.81	0.01		0.24	0.07	0.61	0.07			0.61	0.41
2011_G8Sc_FT	402	MC	11	1	970	0.81	0.00		0.21	0.08	0.16	0.54			0.54	0.35
2011_G8Sc_FT	402	MC	12	1	970	0.81	0.00		0.07	0.12	0.36	0.44			0.44	0.35
2011_G8Sc_FT	402	MC	13	1	970	0.81	0.01		0.71	0.08	0.05	0.16			0.71	0.43
2011_G8Sc_FT	402	MC	14	1	970	0.81	0.02		0.06	0.16	0.67	0.10			0.67	0.32
2011_G8Sc_FT	402	CR	41	1	970	0.81	0.01	0.15	0.84						0.84	0.47
2011_G8Sc_FT	402	CR	42	1	970	0.81	0.01	0.25	0.74						0.74	0.41
2011_G8Sc_FT	402	CR	43	1	970	0.81	0.08	0.54	0.38						0.38	0.55
2011_G8Sc_FT	402	CR	44	1	970	0.81	0.08	0.57	0.36						0.36	0.54
2011_G8Sc_FT	402	CR	45	1	970	0.81	0.01	0.27	0.72						0.72	0.49
2011_G8Sc_FT	402	CR	46	1	970	0.81	0.06	0.27	0.67						0.67	0.54
2011_G8Sc_FT	402	CR	47	1	970	0.81	0.04	0.21	0.75						0.75	0.47
2011_G8Sc_FT	402	CR	48	1	970	0.81	0.08	0.20	0.72						0.72	0.42

**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_G8Sc_FT	402	CR	49	1	970	0.81	0.08	0.22	0.69						0.69	0.47
2011_G8Sc_FT	403	MC	01	1	949	0.75	0.00		0.75	0.10	0.05	0.09			0.75	0.44
2011_G8Sc_FT	403	MC	02	1	949	0.75	0.00		0.12	0.10	0.57	0.21			0.57	0.50
2011_G8Sc_FT	403	MC	03	1	949	0.75	0.00		0.74	0.17	0.04	0.04			0.74	0.41
2011_G8Sc_FT	403	MC	04	1	949	0.75	0.00		0.24	0.09	0.54	0.13			0.54	0.27
2011_G8Sc_FT	403	MC	05	1	949	0.75	0.01		0.48	0.34	0.12	0.05			0.48	0.31
2011_G8Sc_FT	403	MC	06	1	949	0.75	0.00		0.81	0.04	0.03	0.12			0.81	0.27
2011_G8Sc_FT	403	MC	07	1	949	0.75	0.00		0.10	0.06	0.78	0.06			0.78	0.40
2011_G8Sc_FT	403	MC	08	1	949	0.75	0.00		0.08	0.88	0.03	0.01			0.88	0.44
2011_G8Sc_FT	403	MC	09	1	949	0.75	0.00		0.25	0.09	0.14	0.52			0.52	0.35
2011_G8Sc_FT	403	MC	10	1	949	0.75	0.00		0.07	0.05	0.66	0.21			0.66	0.44
2011_G8Sc_FT	403	MC	11	1	949	0.75	0.00		0.07	0.80	0.12	0.01			0.80	0.29
2011_G8Sc_FT	403	MC	12	1	949	0.75	0.01		0.05	0.21	0.69	0.05			0.69	0.37
2011_G8Sc_FT	403	CR	41	2	949	0.75	0.01	0.06	0.12	0.81					1.74	0.40
2011_G8Sc_FT	403	CR	42													
2011_G8Sc_FT	403	CR	43	1	949	0.75	0.02	0.59	0.39						0.39	0.41
2011_G8Sc_FT	403	CR	44	1	949	0.75	0.05	0.20	0.76						0.76	0.42
2011_G8Sc_FT	403	CR	45	1	949	0.75	0.01	0.30	0.69						0.69	0.50
2011_G8Sc_FT	403	CR	46	1	949	0.75	0.03	0.36	0.61						0.61	0.47
2011_G8Sc_FT	403	CR	47	1	949	0.75	0.01	0.24	0.74						0.74	0.48
2011_G8Sc_FT	403	CR	48	1	949	0.75	0.02	0.07	0.91						0.91	0.35
2011_G8Sc_FT	403	CR	49	1	949	0.75	0.03	0.62	0.36						0.36	0.42
2011_G8Sc_FT	403	CR	50	1	949	0.75	0.05	0.65	0.31						0.31	0.37

**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_G8Sc_FT	403	CR	51	1	949	0.75	0.04	0.49	0.47						0.47	0.52
2011_G8Sc_FT	404	MC	01	1	994	0.85	0.00		0.02	0.08	0.19	0.71			0.71	0.28
2011_G8Sc_FT	404	MC	02	1	994	0.85	0.01		0.03	0.08	0.70	0.19			0.70	0.48
2011_G8Sc_FT	404	MC	03	1	994	0.85	0.01		0.08	0.05	0.23	0.64			0.64	0.45
2011_G8Sc_FT	404	MC	04	1	994	0.85	0.00		0.03	0.79	0.15	0.03			0.79	0.30
2011_G8Sc_FT	404	MC	05	1												
2011_G8Sc_FT	404	MC	06	1	994	0.85	0.01		0.12	0.27	0.56	0.04			0.56	0.45
2011_G8Sc_FT	404	MC	07	1	994	0.85	0.01		0.25	0.11	0.42	0.22			0.42	0.21
2011_G8Sc_FT	404	MC	08	1	994	0.85	0.01		0.36	0.15	0.41	0.07			0.41	0.38
2011_G8Sc_FT	404	MC	09	1	994	0.85	0.01		0.04	0.84	0.07	0.04			0.84	0.52
2011_G8Sc_FT	404	MC	10	1	994	0.85	0.03		0.14	0.04	0.04	0.75			0.75	0.40
2011_G8Sc_FT	404	MC	11	1	994	0.85	0.01		0.08	0.74	0.14	0.03			0.74	0.50
2011_G8Sc_FT	404	MC	12	1	994	0.85	0.01		0.05	0.04	0.22	0.69			0.69	0.50
2011_G8Sc_FT	404	MC	13	1	994	0.85	0.01		0.11	0.72	0.11	0.04			0.72	0.40
2011_G8Sc_FT	404	MC	14	1	994	0.85	0.02		0.76	0.15	0.02	0.05			0.76	0.48
2011_G8Sc_FT	404	MC	15	1	994	0.85	0.02		0.04	0.77	0.04	0.13			0.77	0.41
2011_G8Sc_FT	404	MC	16	1	994	0.85	0.02		0.12	0.67	0.03	0.16			0.67	0.52
2011_G8Sc_FT	404	MC	17	1	994	0.85	0.02		0.07	0.06	0.04	0.82			0.82	0.53
2011_G8Sc_FT	404	MC	18	1	994	0.85	0.03		0.48	0.32	0.13	0.04			0.48	0.36
2011_G8Sc_FT	404	MC	19	1	994	0.85	0.02		0.03	0.12	0.52	0.30			0.52	0.47
2011_G8Sc_FT	404	MC	20	1	994	0.85	0.03		0.69	0.06	0.16	0.06			0.69	0.53
2011_G8Sc_FT	404	MC	21	1	994	0.85	0.02		0.11	0.08	0.73	0.06			0.73	0.53
2011_G8Sc_FT	404	MC	22	1	994	0.85	0.02		0.49	0.21	0.18	0.10			0.49	0.35

**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_G8Sc_FT	404	MC	23	1	994	0.85	0.02		0.09	0.63	0.09	0.16			0.63	0.51
2011_G8Sc_FT	404	MC	24	1	994	0.85	0.03		0.87	0.05	0.03	0.03			0.87	0.54
2011_G8Sc_FT	404	MC	25	1	994	0.85	0.03		0.03	0.03	0.08	0.84			0.84	0.48
2011_G8Sc_FT	404	MC	26	1	994	0.85	0.03		0.27	0.16	0.11	0.43			0.43	0.34
2011_G8Sc_FT	404	MC	27	1	994	0.85	0.03		0.09	0.38	0.31	0.20			0.38	0.29
2011_G8Sc_FT	404	MC	28	1	994	0.85	0.03		0.15	0.05	0.05	0.72			0.72	0.54
2011_G8Sc_FT	404	MC	29	1	994	0.85	0.03		0.12	0.78	0.04	0.03			0.78	0.50
2011_G8Sc_FT	404	MC	30	1	994	0.85	0.05		0.16	0.08	0.07	0.65			0.65	0.47
2011_G8Sc_FT	405	MC	01	1	981	0.80	0.00		0.02	0.87	0.08	0.03			0.87	0.41
2011_G8Sc_FT	405	MC	02	1	981	0.80	0.00		0.77	0.11	0.08	0.04			0.77	0.44
2011_G8Sc_FT	405	MC	03	1	981	0.80	0.00		0.24	0.56	0.10	0.10			0.56	0.37
2011_G8Sc_FT	405	MC	04	1	981	0.80	0.00		0.01	0.81	0.09	0.09			0.81	0.50
2011_G8Sc_FT	405	MC	05	1	981	0.80	0.00		0.18	0.66	0.09	0.06			0.66	0.27
2011_G8Sc_FT	405	MC	06	1	981	0.80	0.00		0.29	0.01	0.66	0.03			0.66	0.51
2011_G8Sc_FT	405	MC	07	1	981	0.80	0.00		0.66	0.07	0.18	0.09			0.66	0.50
2011_G8Sc_FT	405	MC	08	1	981	0.80	0.00		0.16	0.01	0.06	0.76			0.76	0.39
2011_G8Sc_FT	405	MC	09	1	981	0.80	0.00		0.23	0.18	0.20	0.38			0.38	0.24
2011_G8Sc_FT	405	MC	10	1	981	0.80	0.00		0.13	0.39	0.28	0.19			0.39	0.41
2011_G8Sc_FT	405	MC	11	1	981	0.80	0.00		0.28	0.36	0.16	0.21			0.28	0.26
2011_G8Sc_FT	405	MC	12	1	981	0.80	0.01		0.07	0.84	0.07	0.02			0.84	0.37
2011_G8Sc_FT	405	CR	41	1	981	0.80	0.01	0.53	0.46						0.46	0.47
2011_G8Sc_FT	405	CR	42	1	981	0.80	0.01	0.63	0.36						0.36	0.37
2011_G8Sc_FT	405	CR	43	1	981	0.80	0.00	0.31	0.68						0.68	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_G8Sc_FT	405	CR	44	1	981	0.80	0.08	0.25	0.67						0.67	0.58
2011_G8Sc_FT	405	CR	45	1	981	0.80	0.04	0.59	0.37						0.37	0.49
2011_G8Sc_FT	405	CR	46	1	981	0.80	0.02	0.19	0.78						0.78	0.37
2011_G8Sc_FT	405	CR	47	1	981	0.80	0.03	0.34	0.63						0.63	0.51
2011_G8Sc_FT	405	CR	48	1	981	0.80	0.07	0.36	0.56						0.56	0.54
2011_G8Sc_FT	405	CR	49	1	981	0.80	0.12	0.53	0.35						0.35	0.52
2011_G8Sc_FT	405	CR	50	1	981	0.80	0.04	0.24	0.72						0.72	0.54
2011_G8Sc_FT	405	CR	51	1	981	0.80	0.04	0.24	0.72						0.72	0.45
2011_G8Sc_FT	406	MC	01	1	968	0.80	0.00		0.10	0.05	0.20	0.64			0.64	0.55
2011_G8Sc_FT	406	MC	02	1	968	0.80	0.00		0.17	0.03	0.11	0.69			0.69	0.46
2011_G8Sc_FT	406	MC	03	1	968	0.80	0.00		0.27	0.09	0.60	0.04			0.60	0.43
2011_G8Sc_FT	406	MC	04	1	968	0.80	0.00		0.17	0.09	0.67	0.06			0.67	0.43
2011_G8Sc_FT	406	MC	05	1												
2011_G8Sc_FT	406	MC	06	1	968	0.80	0.00		0.56	0.37	0.03	0.04			0.56	0.33
2011_G8Sc_FT	406	MC	07	1	968	0.80	0.00		0.44	0.13	0.35	0.08			0.44	0.43
2011_G8Sc_FT	406	MC	08	1	968	0.80	0.00		0.10	0.13	0.70	0.07			0.70	0.41
2011_G8Sc_FT	406	MC	09	1	968	0.80	0.00		0.03	0.29	0.57	0.11			0.57	0.25
2011_G8Sc_FT	406	MC	10	1	968	0.80	0.00		0.05	0.14	0.10	0.71			0.71	0.43
2011_G8Sc_FT	406	MC	11	1	968	0.80	0.00		0.79	0.11	0.06	0.04			0.79	0.36
2011_G8Sc_FT	406	MC	12	1	968	0.80	0.02		0.13	0.44	0.11	0.29			0.44	0.33
2011_G8Sc_FT	406	CR	41	1	968	0.80	0.03	0.28	0.69						0.69	0.43
2011_G8Sc_FT	406	CR	42													
2011_G8Sc_FT	406	CR	43	1	968	0.80	0.08	0.49	0.43						0.43	0.56

**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_G8Sc_FT	406	CR	44	1	968	0.80	0.02	0.35	0.63						0.63	0.50
2011_G8Sc_FT	406	CR	45	1	968	0.80	0.02	0.36	0.62						0.62	0.42
2011_G8Sc_FT	406	CR	46	1	968	0.80	0.02	0.18	0.80						0.80	0.39
2011_G8Sc_FT	406	CR	47	1	968	0.80	0.05	0.38	0.57						0.57	0.53
2011_G8Sc_FT	406	CR	48	1	968	0.80	0.03	0.35	0.62						0.62	0.52
2011_G8Sc_FT	406	CR	49	1	968	0.80	0.07	0.60	0.33						0.33	0.43
2011_G8Sc_FT	406	CR	50	1	968	0.80	0.07	0.49	0.45						0.45	0.41
2011_G8Sc_FT	406	CR	51	1	968	0.80	0.23	0.49	0.28						0.28	0.50
2011_G8Sc_FT	406	CR	52	1	968	0.80	0.17	0.42	0.41						0.41	0.54
2011_G8Sc_FT	406	CR	53	1	968	0.80	0.14	0.69	0.17						0.17	0.32
2011_G8Sc_FT	407	MC	01	1	953	0.79	0.00		0.02	0.02	0.94	0.02			0.94	0.32
2011_G8Sc_FT	407	MC	02	1	953	0.79	0.00		0.11	0.20	0.59	0.09			0.59	0.49
2011_G8Sc_FT	407	MC	03	1	953	0.79	0.01		0.05	0.12	0.27	0.55			0.55	0.46
2011_G8Sc_FT	407	MC	04	1	953	0.79	0.00		0.07	0.18	0.11	0.65			0.65	0.50
2011_G8Sc_FT	407	MC	05	1	953	0.79	0.01		0.09	0.11	0.16	0.63			0.63	0.35
2011_G8Sc_FT	407	MC	06	1	953	0.79	0.00		0.07	0.63	0.25	0.06			0.63	0.51
2011_G8Sc_FT	407	MC	07	1	953	0.79	0.00		0.78	0.07	0.01	0.14			0.78	0.30
2011_G8Sc_FT	407	MC	08	1	953	0.79	0.00		0.32	0.06	0.45	0.17			0.45	0.35
2011_G8Sc_FT	407	MC	09	1	953	0.79	0.00		0.15	0.61	0.10	0.13			0.61	0.32
2011_G8Sc_FT	407	MC	10	1	953	0.79	0.00		0.73	0.12	0.08	0.07			0.73	0.51
2011_G8Sc_FT	407	MC	11	1	953	0.79	0.00		0.08	0.05	0.76	0.11			0.76	0.38
2011_G8Sc_FT	407	MC	12	1	953	0.79	0.00		0.07	0.02	0.09	0.82			0.82	0.33
2011_G8Sc_FT	407	MC	13	1	953	0.79	0.02		0.70	0.05	0.04	0.19			0.70	0.42

**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_G8Sc_FT	407	CR	41	1	953	0.79	0.00	0.08	0.91						0.91	0.34
2011_G8Sc_FT	407	CR	42	1	953	0.79	0.02	0.40	0.58						0.58	0.34
2011_G8Sc_FT	407	CR	43	1	953	0.79	0.01	0.15	0.83						0.83	0.43
2011_G8Sc_FT	407	CR	44	1	953	0.79	0.02	0.39	0.58						0.58	0.46
2011_G8Sc_FT	407	CR	45	1	953	0.79	0.01	0.17	0.82						0.82	0.47
2011_G8Sc_FT	407	CR	46	1	953	0.79	0.01	0.42	0.57						0.57	0.41
2011_G8Sc_FT	407	CR	47	1	953	0.79	0.02	0.21	0.77						0.77	0.36
2011_G8Sc_FT	407	CR	48	1	953	0.79	0.05	0.55	0.40						0.40	0.33
2011_G8Sc_FT	407	CR	49	1	953	0.79	0.05	0.51	0.44						0.44	0.35
2011_G8Sc_FT	407	CR	50	1	953	0.79	0.09	0.57	0.34						0.34	0.56
2011_G8Sc_FT	407	CR	51	1	953	0.79	0.08	0.28	0.64						0.64	0.43
2011_G8Sc_FT	407	CR	52	1	953	0.79	0.07	0.19	0.73						0.73	0.43
2011_G8Sc_FT	408	MC	01	1	956	0.81	0.00		0.04	0.51	0.40	0.04			0.51	0.39
2011_G8Sc_FT	408	MC	02	1	956	0.81	0.00		0.06	0.04	0.09	0.81			0.81	0.42
2011_G8Sc_FT	408	MC	03	1	956	0.81	0.00		0.06	0.72	0.06	0.15			0.72	0.40
2011_G8Sc_FT	408	MC	04	1	956	0.81	0.00		0.10	0.71	0.10	0.08			0.71	0.36
2011_G8Sc_FT	408	MC	05	1	956	0.81	0.00		0.10	0.83	0.03	0.04			0.83	0.42
2011_G8Sc_FT	408	MC	06	1	956	0.81	0.00		0.45	0.19	0.21	0.13			0.45	0.33
2011_G8Sc_FT	408	MC	07	1	956	0.81	0.00		0.53	0.06	0.30	0.11			0.53	0.30
2011_G8Sc_FT	408	MC	08	1												
2011_G8Sc_FT	408	MC	09	1	956	0.81	0.00		0.14	0.16	0.53	0.17			0.53	0.49
2011_G8Sc_FT	408	MC	10	1	956	0.81	0.00		0.17	0.60	0.11	0.12			0.60	0.39
2011_G8Sc_FT	408	MC	11	1	956	0.81	0.00		0.37	0.58	0.03	0.01			0.58	0.35



**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_G8Sc_FT	408	MC	12	1	956	0.81	0.01		0.02	0.82	0.13	0.01			0.82	0.41
2011_G8Sc_FT	408	CR	41	1	956	0.81	0.01	0.15	0.83						0.83	0.32
2011_G8Sc_FT	408	CR	42	1	956	0.81	0.01	0.32	0.67						0.67	0.57
2011_G8Sc_FT	408	CR	43	1	956	0.81	0.01	0.28	0.71						0.71	0.53
2011_G8Sc_FT	408	CR	44	1	956	0.81	0.01	0.21	0.78						0.78	0.27
2011_G8Sc_FT	408	CR	45	1	956	0.81	0.02	0.28	0.69						0.69	0.53
2011_G8Sc_FT	408	CR	46	1	956	0.81	0.04	0.68	0.28						0.28	0.46
2011_G8Sc_FT	408	CR	47	1	956	0.81	0.01	0.37	0.62						0.62	0.44
2011_G8Sc_FT	408	CR	48	1	956	0.81	0.07	0.11	0.82						0.82	0.39
2011_G8Sc_FT	408	CR	49	1	956	0.81	0.08	0.56	0.36						0.36	0.58
2011_G8Sc_FT	408	CR	50	1	956	0.81	0.07	0.10	0.83						0.83	0.48
2011_G8Sc_FT	408	CR	51	1	956	0.81	0.08	0.52	0.40						0.40	0.47
2011_G8Sc_FT	408	CR	52	1	956	0.81	0.08	0.18	0.74						0.74	0.50
2011_G8Sc_FT	408	CR	53	1	956	0.81	0.11	0.46	0.43						0.43	0.64
2011_G8Sc_FT	409	MC	01	1	976	0.84	0.00		0.78	0.05	0.09	0.09			0.78	0.47
2011_G8Sc_FT	409	MC	02	1	976	0.84	0.00		0.10	0.22	0.61	0.07			0.61	0.42
2011_G8Sc_FT	409	MC	03	1	976	0.84	0.00		0.33	0.58	0.05	0.04			0.58	0.43
2011_G8Sc_FT	409	MC	04	1	976	0.84	0.00		0.45	0.20	0.19	0.16			0.45	0.46
2011_G8Sc_FT	409	MC	05	1	976	0.84	0.00		0.79	0.15	0.04	0.02			0.79	0.44
2011_G8Sc_FT	409	MC	06	1	976	0.84	0.00		0.04	0.15	0.06	0.75			0.75	0.26
2011_G8Sc_FT	409	MC	07	1	976	0.84	0.00		0.06	0.86	0.03	0.05			0.86	0.38
2011_G8Sc_FT	409	MC	08	1	976	0.84	0.01		0.07	0.55	0.29	0.08			0.55	0.53
2011_G8Sc_FT	409	MC	09	1	976	0.84	0.00		0.33	0.06	0.07	0.54			0.33	0.32

**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_G8Sc_FT	409	MC	10	1	976	0.84	0.00		0.17	0.11	0.11	0.60			0.60	0.46
2011_G8Sc_FT	409	MC	11	1	976	0.84	0.00		0.22	0.19	0.45	0.14			0.45	0.34
2011_G8Sc_FT	409	MC	12	1	976	0.84	0.01		0.08	0.81	0.09	0.01			0.81	0.40
2011_G8Sc_FT	409	CR	41	1	976	0.84	0.04	0.53	0.43						0.43	0.48
2011_G8Sc_FT	409	CR	42	1	976	0.84	0.03	0.46	0.51						0.51	0.51
2011_G8Sc_FT	409	CR	43	1	976	0.84	0.01	0.40	0.59						0.59	0.53
2011_G8Sc_FT	409	CR	44	1	976	0.84	0.06	0.29	0.65						0.65	0.46
2011_G8Sc_FT	409	CR	45	1	976	0.84	0.01	0.27	0.72						0.72	0.55
2011_G8Sc_FT	409	CR	46	1	976	0.84	0.02	0.21	0.78						0.78	0.53
2011_G8Sc_FT	409	CR	47	1	976	0.84	0.05	0.43	0.52						0.52	0.54
2011_G8Sc_FT	409	CR	48	1	976	0.84	0.03	0.24	0.73						0.73	0.52
2011_G8Sc_FT	409	CR	49	1	976	0.84	0.06	0.45	0.49						0.49	0.45
2011_G8Sc_FT	409	CR	50	1	976	0.84	0.09	0.41	0.49						0.49	0.48
2011_G8Sc_FT	409	CR	51	1	976	0.84	0.05	0.31	0.64						0.64	0.44
2011_G8Sc_FT	409	CR	52	1	976	0.84	0.11	0.73	0.16						0.16	0.42
2011_G8Sc_FT	409	CR	53	1	976	0.84	0.08	0.16	0.76						0.76	0.46

## **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_G8Sc_FT	401	MC	01	1	937	-2.4202							0.94
2011_G8Sc_FT	401	MC	02	1	937	-0.7847							1.11
2011_G8Sc_FT	401	MC	03	1	937	-0.8232							0.93
2011_G8Sc_FT	401	MC	04	1	937	-2.5600							0.84
2011_G8Sc_FT	401	MC	05	1	937	-1.3679							1.00
2011_G8Sc_FT	401	MC	06	1	937	-1.7715							0.91
2011_G8Sc_FT	401	MC	07	1	937	-0.0045							1.05
2011_G8Sc_FT	401	MC	08	1	937	-0.0333							1.04
2011_G8Sc_FT	401	MC	09	1	937	-0.8077							0.99
2011_G8Sc_FT	401	MC	10	1	937	0.0639							0.99
2011_G8Sc_FT	401	MC	11	1	937	1.5215							1.01
2011_G8Sc_FT	401	MC	12	1	937	1.6298							1.07
2011_G8Sc_FT	401	CR	41	2	937	-0.0841	-0.3602	0.3602					1.08
2011_G8Sc_FT	401	CR	42	1	937	1.2580							1.01
2011_G8Sc_FT	401	CR	43	1	937	1.2849							0.92
2011_G8Sc_FT	401	CR	44	1	937	1.0303							0.94
2011_G8Sc_FT	401	CR	45	1	937	0.0977							0.96
2011_G8Sc_FT	401	CR	46	1	937	-0.4256							1.01
2011_G8Sc_FT	401	CR	47	1	937	-1.3265							0.97
2011_G8Sc_FT	401	CR	48	1	937	-0.2173							1.08
2011_G8Sc_FT	401	CR	49	1	937	0.1921							1.03
2011_G8Sc_FT	401	CR	50	1	937	1.8316							0.93
2011_G8Sc_FT	401	CR	51	1	937	2.0077							0.97

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

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_G8Sc_FT	402	MC	01	1	970	-1.1386							0.89
2011_G8Sc_FT	402	MC	02	1	970	-0.7117							0.95
2011_G8Sc_FT	402	MC	03	1	970	-0.9861							0.96
2011_G8Sc_FT	402	MC	04	1	970	-1.3441							0.93
2011_G8Sc_FT	402	MC	05	1	970	0.2205							1.01
2011_G8Sc_FT	402	MC	06	1	970	-0.7624							0.96
2011_G8Sc_FT	402	MC	07	1	970	0.3530							0.98
2011_G8Sc_FT	402	MC	08	1	970	0.3056							0.96
2011_G8Sc_FT	402	MC	09	1	970	0.8610							1.25
2011_G8Sc_FT	402	MC	10	1	970	0.2897							1.06
2011_G8Sc_FT	402	MC	11	1	970	0.6423							1.14
2011_G8Sc_FT	402	MC	12	1	970	1.1679							1.13
2011_G8Sc_FT	402	MC	13	1	970	-0.2184							1.01
2011_G8Sc_FT	402	MC	14	1	970	-0.0156							1.16
2011_G8Sc_FT	402	CR	41	1	970	-1.0950							0.92
2011_G8Sc_FT	402	CR	42	1	970	-0.4281							1.02
2011_G8Sc_FT	402	CR	43	1	970	1.4633							0.85
2011_G8Sc_FT	402	CR	44	1	970	1.5891							0.85
2011_G8Sc_FT	402	CR	45	1	970	-0.3030							0.95
2011_G8Sc_FT	402	CR	46	1	970	-0.0156							0.89
2011_G8Sc_FT	402	CR	47	1	970	-0.4409							0.96
2011_G8Sc_FT	402	CR	48	1	970	-0.2604							1.03
2011_G8Sc_FT	402	CR	49	1	970	-0.1359							0.98
2011_G8Sc_FT	403	MC	01	1	949	-0.4747							0.96

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

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_G8Sc_FT	403	MC	02	1	949	0.5003							0.92
2011_G8Sc_FT	403	MC	03	1	949	-0.4051							0.99
2011_G8Sc_FT	403	MC	04	1	949	0.6259							1.17
2011_G8Sc_FT	403	MC	05	1	949	0.9104							1.12
2011_G8Sc_FT	403	MC	06	1	949	-0.8500							1.08
2011_G8Sc_FT	403	MC	07	1	949	-0.6338							0.98
2011_G8Sc_FT	403	MC	08	1	949	-1.4272							0.90
2011_G8Sc_FT	403	MC	09	1	949	0.7408							1.08
2011_G8Sc_FT	403	MC	10	1	949	0.0523							0.98
2011_G8Sc_FT	403	MC	11	1	949	-0.7391							1.06
2011_G8Sc_FT	403	MC	12	1	949	-0.0859							1.04
2011_G8Sc_FT	403	CR	41	2	949	-0.9460	0.2895	-0.2895					1.07
2011_G8Sc_FT	403	CR	42										
2011_G8Sc_FT	403	CR	43	1	949	1.3534							1.03
2011_G8Sc_FT	403	CR	44	1	949	-0.4875							0.96
2011_G8Sc_FT	403	CR	45	1	949	-0.0915							0.91
2011_G8Sc_FT	403	CR	46	1	949	0.3218							0.95
2011_G8Sc_FT	403	CR	47	1	949	-0.4176							0.91
2011_G8Sc_FT	403	CR	48	1	949	-1.7730							0.93
2011_G8Sc_FT	403	CR	49	1	949	1.5294							0.99
2011_G8Sc_FT	403	CR	50	1	949	1.7993							1.05
2011_G8Sc_FT	403	CR	51	1	949	0.9905							0.90
2011_G8Sc_FT	404	MC	01	1	994	-0.4000							1.25
2011_G8Sc_FT	404	MC	02	1	994	-0.0700							0.93

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

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_G8Sc_FT	404	MC	03	1	994	0.0800							1.01
2011_G8Sc_FT	404	MC	04	1	994	-0.4700							1.04
2011_G8Sc_FT	404	MC	05	1									
2011_G8Sc_FT	404	MC	06	1	994	0.6600							0.98
2011_G8Sc_FT	404	MC	07	1	994	1.1600							1.23
2011_G8Sc_FT	404	MC	08	1	994	1.1100							1.02
2011_G8Sc_FT	404	MC	09	1	994	-1.1300							0.85
2011_G8Sc_FT	404	MC	10	1	994	-0.5300							1.05
2011_G8Sc_FT	404	MC	11	1	994	-0.3000							0.89
2011_G8Sc_FT	404	MC	12	1	994	-0.2900							1.01
2011_G8Sc_FT	404	MC	13	1	994	-0.2200							1.02
2011_G8Sc_FT	404	MC	14	1	994	-0.3000							0.86
2011_G8Sc_FT	404	MC	15	1	994	-0.7400							1.09
2011_G8Sc_FT	404	MC	16	1	994	0.0600							0.90
2011_G8Sc_FT	404	MC	17	1	994	-1.1200							0.95
2011_G8Sc_FT	404	MC	18	1	994	0.8400							1.08
2011_G8Sc_FT	404	MC	19	1	994	0.6500							0.98
2011_G8Sc_FT	404	MC	20	1	994	-0.3000							0.98
2011_G8Sc_FT	404	MC	21	1	994	-0.1700							0.86
2011_G8Sc_FT	404	MC	22	1	994	0.7800							1.11
2011_G8Sc_FT	404	MC	23	1	994	0.1700							0.93
2011_G8Sc_FT	404	MC	24	1	994	-1.4800							0.88
2011_G8Sc_FT	404	MC	25	1	994	-1.0200							0.87
2011_G8Sc_FT	404	MC	26	1	994	1.1500							1.11

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

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_G8Sc_FT	404	MC	27	1	994	1.5100							1.15
2011_G8Sc_FT	404	MC	28	1	994	0.0000							0.82
2011_G8Sc_FT	404	MC	29	1	994	-0.5600							0.87
2011_G8Sc_FT	404	MC	30	1	994	0.2300							0.95
2011_G8Sc_FT	405	MC	01	1	981	-1.3690							0.95
2011_G8Sc_FT	405	MC	02	1	981	-0.6149							0.98
2011_G8Sc_FT	405	MC	03	1	981	0.5272							1.10
2011_G8Sc_FT	405	MC	04	1	981	-0.8871							0.89
2011_G8Sc_FT	405	MC	05	1	981	0.0300							1.22
2011_G8Sc_FT	405	MC	06	1	981	0.0190							0.93
2011_G8Sc_FT	405	MC	07	1	981	0.0573							0.94
2011_G8Sc_FT	405	MC	08	1	981	-0.5548							1.05
2011_G8Sc_FT	405	MC	09	1	981	1.4420							1.23
2011_G8Sc_FT	405	MC	10	1	981	1.3642							1.04
2011_G8Sc_FT	405	MC	11	1	981	1.9913							1.13
2011_G8Sc_FT	405	MC	12	1	981	-1.0957							1.01
2011_G8Sc_FT	405	CR	41	1	981	1.0208							0.99
2011_G8Sc_FT	405	CR	42	1	981	1.5314							1.10
2011_G8Sc_FT	405	CR	43	1	981	-0.0814							0.94
2011_G8Sc_FT	405	CR	44	1	981	-0.0253							0.85
2011_G8Sc_FT	405	CR	45	1	981	1.4839							0.93
2011_G8Sc_FT	405	CR	46	1	981	-0.7115							1.04
2011_G8Sc_FT	405	CR	47	1	981	0.1758							0.94
2011_G8Sc_FT	405	CR	48	1	981	0.5272							0.91



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

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_G8Sc_FT	405	CR	49	1	981	1.5955							0.89
2011_G8Sc_FT	405	CR	50	1	981	-0.2728							0.88
2011_G8Sc_FT	405	CR	51	1	981	-0.2968							0.98
2011_G8Sc_FT	406	MC	01	1	968	0.1277							0.88
2011_G8Sc_FT	406	MC	02	1	968	-0.1371							0.97
2011_G8Sc_FT	406	MC	03	1	968	0.3581							1.02
2011_G8Sc_FT	406	MC	04	1	968	-0.0136							1.01
2011_G8Sc_FT	406	MC	05	1									
2011_G8Sc_FT	406	MC	06	1	968	0.5211							1.14
2011_G8Sc_FT	406	MC	07	1	968	1.1274							1.02
2011_G8Sc_FT	406	MC	08	1	968	-0.1715							1.01
2011_G8Sc_FT	406	MC	09	1	968	0.5110							1.24
2011_G8Sc_FT	406	MC	10	1	968	-0.2238							0.98
2011_G8Sc_FT	406	MC	11	1	968	-0.7224							1.03
2011_G8Sc_FT	406	MC	12	1	968	1.1223							1.14
2011_G8Sc_FT	406	CR	41	1	968	-0.1257							0.98
2011_G8Sc_FT	406	CR	42										
2011_G8Sc_FT	406	CR	43	1	968	1.1833							0.87
2011_G8Sc_FT	406	CR	44	1	968	0.2020							0.93
2011_G8Sc_FT	406	CR	45	1	968	0.2388							1.03
2011_G8Sc_FT	406	CR	46	1	968	-0.8091							0.96
2011_G8Sc_FT	406	CR	47	1	968	0.5009							0.91
2011_G8Sc_FT	406	CR	48	1	968	0.2493							0.92
2011_G8Sc_FT	406	CR	49	1	968	1.6875							1.01

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

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_G8Sc_FT	406	CR	50	1	968	1.0971							1.05
2011_G8Sc_FT	406	CR	51	1	968	1.9735							0.90
2011_G8Sc_FT	406	CR	52	1	968	1.2807							0.88
2011_G8Sc_FT	406	CR	53	1	968	2.7174							1.06
2011_G8Sc_FT	407	MC	01	1	953	-2.2072							0.93
2011_G8Sc_FT	407	MC	02	1	953	0.3965							0.93
2011_G8Sc_FT	407	MC	03	1	953	0.5948							0.97
2011_G8Sc_FT	407	MC	04	1	953	0.1018							0.92
2011_G8Sc_FT	407	MC	05	1	953	0.1770							1.08
2011_G8Sc_FT	407	MC	06	1	953	0.2194							0.92
2011_G8Sc_FT	407	MC	07	1	953	-0.6463							1.09
2011_G8Sc_FT	407	MC	08	1	953	1.0614							1.07
2011_G8Sc_FT	407	MC	09	1	953	0.2773							1.12
2011_G8Sc_FT	407	MC	10	1	953	-0.3278							0.90
2011_G8Sc_FT	407	MC	11	1	953	-0.5445							1.01
2011_G8Sc_FT	407	MC	12	1	953	-0.9357							1.02
2011_G8Sc_FT	407	MC	13	1	953	-0.1495							0.99
2011_G8Sc_FT	407	CR	41	1	953	-1.8532							0.96
2011_G8Sc_FT	407	CR	42	1	953	0.4579							1.11
2011_G8Sc_FT	407	CR	43	1	953	-1.0339							0.93
2011_G8Sc_FT	407	CR	44	1	953	0.4273							0.97
2011_G8Sc_FT	407	CR	45	1	953	-0.9120							0.89
2011_G8Sc_FT	407	CR	46	1	953	0.5038							1.04
2011_G8Sc_FT	407	CR	47	1	953	-0.5982							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_G8Sc_FT	407	CR	48	1	953	1.3176							1.12
2011_G8Sc_FT	407	CR	49	1	953	1.0969							1.09
2011_G8Sc_FT	407	CR	50	1	953	1.6233							0.82
2011_G8Sc_FT	407	CR	51	1	953	0.1716							1.00
2011_G8Sc_FT	407	CR	52	1	953	-0.3524							0.98
2011_G8Sc_FT	408	MC	01	1	956	0.7916							1.09
2011_G8Sc_FT	408	MC	02	1	956	-0.8709							0.97
2011_G8Sc_FT	408	MC	03	1	956	-0.3160							1.04
2011_G8Sc_FT	408	MC	04	1	956	-0.2125							1.08
2011_G8Sc_FT	408	MC	05	1	956	-1.0480							0.97
2011_G8Sc_FT	408	MC	06	1	956	1.0765							1.17
2011_G8Sc_FT	408	MC	07	1	956	0.6882							1.20
2011_G8Sc_FT	408	MC	08	1									
2011_G8Sc_FT	408	MC	09	1	956	0.6882							0.97
2011_G8Sc_FT	408	MC	10	1	956	0.3267							1.08
2011_G8Sc_FT	408	MC	11	1	956	0.4486							1.14
2011_G8Sc_FT	408	MC	12	1	956	-0.9818							0.97
2011_G8Sc_FT	408	CR	41	1	956	-1.0734							1.05
2011_G8Sc_FT	408	CR	42	1	956	-0.0197							0.86
2011_G8Sc_FT	408	CR	43	1	956	-0.2305							0.89
2011_G8Sc_FT	408	CR	44	1	956	-0.6860							1.15
2011_G8Sc_FT	408	CR	45	1	956	-0.1354							0.89
2011_G8Sc_FT	408	CR	46	1	956	1.9919							0.94
2011_G8Sc_FT	408	CR	47	1	956	0.2516							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_G8Sc_FT	408	CR	48	1	956	-0.9494							1.00
2011_G8Sc_FT	408	CR	49	1	956	1.5602							0.83
2011_G8Sc_FT	408	CR	50	1	956	-1.0063							0.89
2011_G8Sc_FT	408	CR	51	1	956	1.3248							0.99
2011_G8Sc_FT	408	CR	52	1	956	-0.3976							0.92
2011_G8Sc_FT	408	CR	53	1	956	1.2183							0.78
2011_G8Sc_FT	409	MC	01	1	976	-0.6867							0.94
2011_G8Sc_FT	409	MC	02	1	976	0.3087							1.07
2011_G8Sc_FT	409	MC	03	1	976	0.4613							1.06
2011_G8Sc_FT	409	MC	04	1	976	1.1021							1.01
2011_G8Sc_FT	409	MC	05	1	976	-0.7501							0.96
2011_G8Sc_FT	409	MC	06	1	976	-0.5249							1.18
2011_G8Sc_FT	409	MC	07	1	976	-1.3115							0.99
2011_G8Sc_FT	409	MC	08	1	976	0.5809							0.93
2011_G8Sc_FT	409	MC	09	1	976	1.7163							1.18
2011_G8Sc_FT	409	MC	10	1	976	0.3299							1.02
2011_G8Sc_FT	409	MC	11	1	976	1.1230							1.18
2011_G8Sc_FT	409	MC	12	1	976	-0.9130							1.00
2011_G8Sc_FT	409	CR	41	1	976	1.2174							0.99
2011_G8Sc_FT	409	CR	42	1	976	0.7819							0.96
2011_G8Sc_FT	409	CR	43	1	976	0.3774							0.94
2011_G8Sc_FT	409	CR	44	1	976	0.0816							1.00
2011_G8Sc_FT	409	CR	45	1	976	-0.3229							0.87
2011_G8Sc_FT	409	CR	46	1	976	-0.6728							0.86

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

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2011_G8Sc_FT	409	CR	47	1	976	0.7356							0.93
2011_G8Sc_FT	409	CR	48	1	976	-0.3784							0.91
2011_G8Sc_FT	409	CR	49	1	976	0.8951							1.05
2011_G8Sc_FT	409	CR	50	1	976	0.8848							1.00
2011_G8Sc_FT	409	CR	51	1	976	0.1310							1.04
2011_G8Sc_FT	409	CR	52	1	976	2.9143							0.96
2011_G8Sc_FT	409	CR	53	1	976	-0.5711							0.95

## **Appendix C: DIF Statistics**

**Table 10. DIF Statistics**

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
401	01	MC	-0.72	0.81	-0.06		
401	02	MC	-0.47	1.28	-0.08		
401	03	MC	0.67	2.25	0.08		
401	04	MC	0.82	0.78	0.05		
401	05	MC	0.13	0.06	0.02		
401	06	MC	0.27	0.20	0.03		
401	07	MC	-0.93	6.45	-0.15		
401	08	MC	-0.75	4.25	-0.14		
401	09	MC	-0.60	1.88	-0.10		
401	10	MC	-0.34	0.83	-0.06		
401	11	MC	0.76	4.23	0.12		
401	12	MC	-0.08	0.05	0.00		
401	41	CR		2.90	0.10		
401	42	CR		2.06	0.10		
401	43	CR		0.00	-0.01		
401	44	CR		2.24	0.09		
401	45	CR		0.60	-0.03		
401	46	CR		0.18	0.02		
401	47	CR		1.81	0.08		
401	48	CR		9.01	-0.19	BB	M
401	49	CR		3.22	-0.11		
401	50	CR		5.40	0.13		
401	51	CR		0.94	0.06		
402	01	MC	-0.52	1.10	-0.07		

**Table 10. DIF Statistics (continued)**

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
402	02	MC	-0.85	3.81	-0.10		
402	03	MC	-0.77	2.82	-0.10		
402	04	MC	0.50	0.97	0.08		
402	05	MC	0.22	0.38	0.04		
402	06	MC	-0.64	2.21	-0.08		
402	07	MC	-1.05	8.40	-0.18	B	M
402	08	MC	-0.97	6.98	-0.15		
402	09	MC	-0.05	0.03	-0.01		
402	10	MC	-1.05	8.93	-0.20	B	M
402	11	MC	0.62	3.47	0.12		
402	12	MC	-0.01	0.00	0.01		
402	13	MC	0.37	0.97	0.07		
402	14	MC	-0.14	0.16	-0.02		
402	41	CR		4.40	0.13		
402	42	CR		22.05	0.30	CC	F
402	43	CR		0.02	0.00		
402	44	CR		0.92	-0.05		
402	45	CR		0.01	-0.02		
402	46	CR		2.69	0.10		
402	47	CR		3.15	0.09		
402	48	CR		4.04	0.10		
402	49	CR		0.03	-0.01		
403	01	MC	-0.76	3.43	-0.11		
403	02	MC	-0.45	1.51	-0.07		

**Table 10. DIF Statistics (continued)**

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
403	03	MC	-0.65	2.72	-0.10		
403	04	MC	0.98	8.86	0.21		
403	05	MC	0.47	1.97	0.09		
403	06	MC	-0.85	4.04	-0.15		
403	07	MC	-0.67	2.58	-0.11		
403	08	MC	-1.08	3.68	-0.13		
403	09	MC	-0.08	0.06	-0.02		
403	10	MC	-0.78	4.72	-0.13		
403	11	MC	-0.70	2.97	-0.10		
403	12	MC	-0.45	1.50	-0.08		
403	41	CR		9.90	0.20	BB	F
403	43	CR		4.21	0.13		
403	44	CR		0.08	0.02		
403	45	CR		1.87	0.09		
403	46	CR		6.56	0.15		
403	47	CR		1.06	-0.08		
403	48	CR		0.03	0.02		
403	49	CR		1.61	0.08		
403	50	CR		0.00	0.00		
403	51	CR		1.54	-0.07		
404	01	MC	-0.01	0.00	0.00		
404	02	MC	-0.08	0.04	-0.01		
404	03	MC	0.20	0.32	0.04		
404	04	MC	0.33	0.70	0.06		
404	06	MC	-0.39	1.24	-0.06		

**Table 10. DIF Statistics (continued)**

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
404	07	MC	0.95	8.58	0.19		
404	08	MC	0.13	0.15	0.01		
404	09	MC	-0.56	1.18	-0.07		
404	10	MC	0.29	0.53	0.04		
404	11	MC	-0.01	0.00	0.01		
404	12	MC	-0.09	0.05	-0.01		
404	13	MC	-0.18	0.24	-0.03		
404	14	MC	-1.06	6.30	-0.15	B	M
404	15	MC	-0.30	0.56	-0.05		
404	16	MC	0.44	1.33	0.07		
404	17	MC	-0.69	2.04	-0.08		
404	18	MC	0.01	0.00	0.01		
404	19	MC	1.15	10.94	0.20	B	F
404	20	MC	0.01	0.00	0.00		
404	21	MC	-1.09	7.26	-0.16	B	M
404	22	MC	-0.37	1.23	-0.07		
404	23	MC	-0.63	2.88	-0.10		
404	24	MC	0.62	1.18	0.06		
404	25	MC	-0.19	0.16	-0.03		
404	26	MC	-0.24	0.56	-0.05		
404	27	MC	-0.01	0.00	0.01		
404	28	MC	-0.69	2.78	-0.10		
404	29	MC	0.81	3.60	0.12		
404	30	MC	0.82	5.28	0.12		
405	01	MC	-0.85	2.71	-0.11		



**Table 10. DIF Statistics (continued)**

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
405	02	MC	0.53	1.65	0.07		
405	03	MC	0.14	0.17	0.03		
405	04	MC	-0.66	2.00	-0.09		
405	05	MC	-0.03	0.01	-0.02		
405	06	MC	-0.18	0.21	-0.02		
405	07	MC	-1.26	11.25	-0.20	B	M
405	08	MC	-0.34	0.74	-0.05		
405	09	MC	0.54	2.74	0.10		
405	10	MC	-0.06	0.03	-0.02		
405	11	MC	-0.38	1.07	-0.07		
405	12	MC	0.89	3.68	0.10		
405	41	CR		1.70	0.08		
405	42	CR		1.17	0.07		
405	43	CR		13.17	-0.21	BB	M
405	44	CR		0.09	-0.01		
405	45	CR		0.02	0.01		
405	46	CR		2.61	0.10		
405	47	CR		0.63	0.05		
405	48	CR		1.67	-0.07		
405	49	CR		0.08	0.00		
405	50	CR		1.87	0.07		
405	51	CR		6.17	0.16		
406	01	MC	-0.07	0.03	-0.02		
406	02	MC	-0.05	0.02	-0.02		
406	03	MC	0.26	0.59	0.05		

**Table 10. DIF Statistics (continued)**

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
406	04	MC	0.26	0.53	0.04		
406	06	MC	-0.49	2.13	-0.10		
406	07	MC	-0.34	0.97	-0.05		
406	08	MC	-0.86	5.21	-0.13		
406	09	MC	0.32	0.98	0.07		
406	10	MC	0.75	3.90	0.13		
406	11	MC	0.12	0.08	0.02		
406	12	MC	-0.34	1.02	-0.06		
406	41	CR		2.28	0.11		
406	43	CR		1.43	0.05		
406	44	CR		0.06	-0.01		
406	45	CR		1.55	-0.07		
406	46	CR		0.92	0.07		
406	47	CR		0.63	0.04		
406	48	CR		1.71	-0.08		
406	49	CR		0.14	0.03		
406	50	CR		4.09	0.10		
406	51	CR		5.74	-0.13		
406	52	CR		1.55	0.07		
406	53	CR		9.69	-0.20	BB	M
407	01	MC	0.23	0.11	0.03		
407	02	MC	-0.91	5.90	-0.15		
407	03	MC	-0.10	0.08	-0.01		
407	04	MC	0.43	1.28	0.06		
407	05	MC	-0.09	0.07	-0.02		

**Table 10. DIF Statistics (continued)**

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
407	06	MC	-0.54	2.08	-0.10		
407	07	MC	-1.23	9.30	-0.22	B	M
407	08	MC	-0.47	1.96	-0.10		
407	09	MC	-0.45	1.77	-0.09		
407	10	MC	0.15	0.14	0.03		
407	11	MC	0.20	0.25	0.02		
407	12	MC	0.18	0.17	0.03		
407	13	MC	0.01	0.00	0.00		
407	41	CR		0.96	0.04		
407	42	CR		3.33	-0.10		
407	43	CR		0.31	0.04		
407	44	CR		5.18	-0.14		
407	45	CR		0.47	0.03		
407	46	CR		4.38	-0.13		
407	47	CR		1.35	0.10		
407	48	CR		5.04	0.16		
407	49	CR		0.00	0.00		
407	50	CR		8.19	0.16		
407	51	CR		13.16	0.22	BB	F
407	52	CR		9.44	0.22	BB	F
408	01	MC	0.76	4.95	0.15		
408	02	MC	0.02	0.00	0.00		
408	03	MC	0.11	0.09	0.02		
408	04	MC	-0.50	1.89	-0.07		
408	05	MC	-0.05	0.01	0.00		

**Table 10. DIF Statistics (continued)**

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
408	06	MC	-0.43	1.67	-0.09		
408	07	MC	0.53	2.50	0.09		
408	09	MC	-0.40	1.27	-0.08		
408	10	MC	0.26	0.55	0.05		
408	11	MC	-0.94	7.65	-0.17		
408	12	MC	0.02	0.00	0.01		
408	41	CR		8.47	0.19	BB	F
408	42	CR		0.07	0.01		
408	43	CR		0.08	-0.03		
408	44	CR		1.98	-0.09		
408	45	CR		0.92	0.05		
408	46	CR		0.10	0.02		
408	47	CR		0.02	-0.02		
408	48	CR		0.07	-0.02		
408	49	CR		5.26	0.12		
408	50	CR		1.10	0.07		
408	51	CR		2.05	-0.07		
408	52	CR		1.79	0.07		
408	53	CR		14.22	-0.18	BB	M
409	01	MC	0.31	0.53	0.05		
409	02	MC	0.41	1.35	0.07		
409	03	MC	-0.24	0.47	-0.04		
409	04	MC	0.03	0.01	-0.01		
409	05	MC	0.94	4.74	0.14		
409	06	MC	1.26	11.29	0.22	B	F

**Table 10. DIF Statistics (continued)**

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
409	07	MC	-0.10	0.04	-0.01		
409	08	MC	0.22	0.36	0.04		
409	09	MC	-0.15	0.19	-0.02		
409	10	MC	-1.20	11.03	-0.20	B	M
409	11	MC	-1.53	20.58	-0.28	C	M
409	12	MC	0.03	0.01	0.00		
409	41	CR		1.66	0.07		
409	42	CR		7.69	0.15		
409	43	CR		0.84	0.06		
409	44	CR		1.01	-0.06		

**Table 10. DIF Statistics (continued)**

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
409	45	CR		0.45	-0.04		
409	46	CR		0.05	-0.01		
409	47	CR		1.70	0.07		
409	48	CR		0.04	0.02		
409	49	CR		3.77	0.11		
409	50	CR		2.05	-0.09		
409	51	CR		3.62	-0.12		
409	52	CR		9.17	-0.17	BB	M
409	53	CR		0.83	0.06		

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

## **Appendix D: Operational Test Map**

**Table 11. Operational Test Map for June 2011**

Position	Item Type	Max Points	Weight	Standard	Mean	Point-Biserial	Rasch	S1	S2	S3	S4	S5	S6
1	MC	1	1	1	0.87	0.39	-1.40						
2	MC	1	1	1	0.81	0.27	-1.02						
3	MC	1	1	1	0.91	0.29	-1.82						
4	MC	1	1	4	0.66	0.41	0.00						
5	MC	1	1	4	0.71	0.43	-0.26						
6	MC	1	1	1	0.73	0.40	-0.35						
7	MC	1	1	4	0.50	0.29	0.80						
8	MC	1	1	4	0.83	0.33	-1.01						
9	MC	1	1	4	0.42	0.26	1.22						
10	MC	1	1	4	0.80	0.33	-0.86						
11	MC	1	1	4	0.95	0.30	-2.48						
12	MC	1	1	4	0.65	0.35	0.03						
13	MC	1	1	4	0.88	0.45	-1.56						
14	MC	1	1	4	0.87	0.43	-1.43						
15	MC	1	1	4	0.43	0.36	1.13						
16	MC	1	1	4	0.50	0.53	0.78						
17	MC	1	1	4	0.42	0.45	1.21						
18	MC	1	1	4	0.75	0.28	-0.50						
19	MC	1	1	4	0.66	0.49	0.02						
20	MC	1	1	4	0.55	0.28	0.56						
21	MC	1	1	4	0.74	0.41	-0.44						
22	MC	1	1	4	0.78	0.47	-0.64						
23	MC	1	1	4	0.48	0.36	0.76						

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

Position	Item Type	Max Points	Weight	Standard	Mean	Point-Biserial	Rasch	S1	S2	S3	S4	S5	S6
24	MC	1	1	4	0.76	0.46	-0.52						
25	MC	1	1	4	0.32	0.40	1.71						
26	MC	1	1	4	0.83	0.36	-1.18						
27	MC	1	1	4	0.83	0.34	-1.06						
28	MC	1	1	4	0.80	0.33	-0.89						
29	MC	1	1	4	0.56	0.38	0.50						
30	MC	1	1	4	0.71	0.44	-0.27						
31	MC	1	1	4	0.44	0.48	1.09						
32	MC	1	1	4	0.44	0.41	1.09						
33	MC	1	1	4	0.81	0.44	-0.86						
34	MC	1	1	4	0.69	0.27	-0.16						
35	MC	1	1	4	0.44	0.37	1.09						
36	MC	1	1	4	0.53	0.35	0.68						
37	MC	1	1	4	0.92	0.25	-2.03						
38	MC	1	1	4	0.49	0.47	0.87						
39	MC	1	1	4	0.75	0.45	-0.49						
40	MC	1	1	4	0.52	0.46	0.67						
41	MC	1	1	4	0.51	0.23	0.76						
42	MC	1	1	4	0.41	0.44	1.26						
43	MC	1	1	4	0.47	0.35	0.94						
44	MC	1	1	4	0.92	0.19	-2.06						
45	MC	1	1	4	0.78	0.44	-0.81						
46	CR	1	1	1	0.59	0.40	0.37						

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

Position	Item Type	Max Points	Weight	Standard	Mean	Point-Biserial	Rasch	S1	S2	S3	S4	S5	S6
47	CR	1	1	1	0.52	0.52	0.69						
48	CR	1	1	1	0.74	0.40	-0.47						
49	CR	1	1	4	0.55	0.51	0.54						
50	CR	1	1	1	0.61	0.49	0.25						
51	CR	1	1	1	0.50	0.50	0.83						
52	CR	1	1	4	0.70	0.52	-0.19						
53	CR	1	1	1	0.93	0.37	-2.22						
54	CR	1	1	1	0.61	0.49	0.24						
55	CR	1	1	4	0.19	0.44	2.53						
56	CR	1	1	1	0.85	0.34	-1.18						
57	CR	1	1	6	0.72	0.46	-0.31						
58	CR	1	1	6	0.87	0.39	-1.39						
59	CR	1	1	6	0.64	0.52	0.14						
60	CR	2	1	4	1.75	0.52	-0.92	0.62	-0.62				
61	CR	2	1	4	1.29	0.65	0.25	1.17	-1.17				
62	CR	1	1	4	0.40	0.47	1.29						
63	CR	1	1	4	0.60	0.50	0.31						
64	CR	1	1	4	0.81	0.45	-0.88						
65	CR	1	1	4	0.36	0.35	1.51						
66	CR	1	1	4	0.28	0.48	1.93						
67	CR	1	1	4	0.72	0.44	-0.33						
68	CR	1	1	4	0.31	0.54	1.76						
69	CR	1	1	4	0.41	0.51	1.21						

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

Position	Item Type	Max Points	Weight	Standard	Mean	Point-Biserial	Rasch	S1	S2	S3	S4	S5	S6
70	CR	1	1	4	0.59	0.50	0.36						
71	CR	1	1	4	0.60	0.60	0.32						
72	CR	1	1	4	0.67	0.49	-0.06						
73	CR	1	1	4	0.49	0.58	0.84						
74	CR	1	1	4	0.60	0.54	0.32						
75	CR	1	1	4	0.82	0.43	-0.96						
76	CR	1	1	4	0.26	0.46	2.06						
77	CR	1	1	4	0.26	0.46	2.04						
78	CR	1	1	4	0.44	0.41	1.08						
79	CR	1	1	4	0.53	0.51	0.63						
80	CR	1	1	4	0.34	0.39	1.61						
81	CR	1	1	4	0.73	0.40	-0.38						
82	CR	1	1	4	0.40	0.50	1.29						
83	CR	1	1	4	0.59	0.50	0.35						



## **Appendix E: Scoring Table**

**Table 12. 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	-6.093	0.000		23	-1.107	32.699		46	0.300	61.350		69	1.850	84.689
1	-4.867	1.852		24	-1.037	34.401		47	0.358	62.358		70	1.941	85.623
2	-4.143	3.817		25	-0.968	35.523		48	0.416	63.355		71	2.036	86.553
3	-3.707	4.995		26	-0.901	37.048		49	0.475	64.348		72	2.137	87.478
4	-3.389	6.185		27	-0.835	38.631		50	0.533	65.346		73	2.243	88.398
5	-3.136	7.750		28	-0.770	39.697		51	0.593	66.338		74	2.356	89.315
6	-2.924	8.937		29	-0.706	40.771		52	0.653	67.320		75	2.478	90.227
7	-2.741	10.857		30	-0.643	41.825		53	0.713	68.623		76	2.609	91.133
8	-2.579	12.041		31	-0.580	43.745		54	0.774	70.291		77	2.753	92.036
9	-2.432	13.426		32	-0.519	44.951		55	0.836	71.279		78	2.913	92.922
10	-2.299	14.958		33	-0.458	45.999		56	0.899	72.255		79	3.093	93.802
11	-2.175	16.238		34	-0.398	47.289		57	0.962	73.227		80	3.302	94.673
12	-2.060	17.744		35	-0.338	48.892		58	1.027	74.192		81	3.551	95.537
13	-1.953	19.342		36	-0.279	50.126		59	1.093	75.169		82	3.865	96.407
14	-1.851	20.537		37	-0.220	51.167		60	1.160	76.139		83	4.297	97.478
15	-1.755	22.070		38	-0.162	52.198		61	1.228	77.100		84	5.017	99.246
16	-1.663	23.673		39	-0.104	53.235		62	1.298	78.059		85	6.239	100.000
17	-1.575	24.769		40	-0.046	54.273		63	1.370	78.965				
18	-1.490	26.052		41	0.012	55.566		64	1.444	79.035				
19	-1.409	27.689		42	0.070	57.306		65	1.520	79.953				
20	-1.330	29.049		43	0.127	58.316		66	1.598	81.713				
21	-1.254	30.437		44	0.185	59.326		67	1.679	82.811				
22	-1.180	31.810		45	0.242	60.337		68	1.763	83.751				