

New York State Regents Examination in Physics

2010 Field Test Analysis, Equating Procedure, and Scaling of Operational Test Forms

Technical Report



Prepared for the New York State Education Department
by Pearson

August 2011

Copyright

Developed and published under contract with the New York State Education Department by Pearson. Copyright © 2010 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
File Merging and Data Clean-up	2
Classical Analysis.....	2
<i>Item Difficulty</i>	3
<i>Point-Biserial Correlation</i>	3
<i>Test Reliability</i>	5
<i>Scoring Reliability</i>	6
<i>Inter-rater Agreement</i>	6
<i>Constructed-Response Item Means and Standard Deviations</i>	16
<i>Intra-class Correlation</i>	16
<i>Weighted Kappa</i>	16
Item Response Theory (IRT) Statistics	17
<i>Item Calibration</i>	18
<i>Item Fit Evaluation</i>	18
Differential Item Functioning (DIF) Statistics.....	19
Section III: Equating Procedure	21
Section IV: Scaling of Operational Test Forms	22
References	25
Appendix A: Classical Item Analysis.....	26
Appendix B: Partial Credit Model Item Analysis.....	44
Appendix C: DIF Statistics	61
Appendix D: Operational Test Maps	69
Appendix E: Scoring Tables.....	78

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	7
Table 5. First and Second Read Descriptive Statistics and Agreement	12
Table 6. Partial Credit Model Item Analysis	19
Table 7. Initial Mean Abilities and Equating Constants	22
Table 8. Classical Item Analysis	27
Table 9. Partial Credit Model Item Analysis	45
Table 10. DIF Statistics.....	62
Table 11. Operational Test Map for January 2010.....	70
Table 12. Operational Test Map for June 2010.....	74
Table 13. Scoring Table for January 2010.....	79
Table 14. Scoring Table for June 2010	80

Section I: Introduction

Purpose

The purpose of this report is to document the psychometric work on the New York State Regents Examination in Physics in 2010. Specifically, contained within this report are procedures for and results of field test analysis, equating, and scaling of operational test forms. Because of a change in vendor mid-year, the field test equating was conducted by Pearson while the scaling was conducted by the previous vendor. Information on test development can be found in the test design and development report for the New York State Regents Examination in Physics.

Section II: Field Test Analysis

In May 2010, field testing was conducted for the New York State Regents Examination in Physics 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 70 th 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 th percentile with fewer than 50 students per square mile or enrollment of less than 2500
Average N/RC Districts	All districts between the 20 th and 70 th percentiles on the index
Low N/RC Districts	All districts below the 20 th 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 Clean-up

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 15,136 student records and the constructed-response data file contained 14,690 student records. To combine the two files, the multiple-choice file served as the base file and constructed-response records were merged to the multiple-choice records using unique test booklet numbers. For multiple-choice records that did not have corresponding constructed-response records, constructed-response items were treated as non-attempted and scored as 0. After the exclusion rules were applied, the resulting field test data file contained 15,077 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 may 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 student 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. 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 one, item difficulties were below 0.90. With respect to the point-biserial correlations, most of these correlations fell between 0.25 and 0.50.

Table 2. Classical Item Analysis

			Item Difficulty			Point-Biserial		
Form	N-Count	No. of Items	<0.50	0.50 to 0.90	>0.90	<0.25	0.25 to 0.50	>0.50
841	931	20	8	12	0	0	15	5
842	928	19	11	8	0	1	11	7
843	939	20	9	11	0	1	11	8
844	934	19	6	13	0	0	14	5
845	940	19	13	6	0	2	13	4
846	951	20	8	12	0	1	13	6
847	937	19	7	12	0	2	9	8
848	951	20	4	16	0	1	12	7
849	943	19	6	12	1	0	14	5
850	954	20	10	10	0	2	15	3
851	948	16	8	4	0	0	7	5
852	961	16	7	4	0	0	6	5
853	954	15	9	3	0	0	5	7
854	938	15	4	6	0	0	3	7
855	939	17	6	9	0	0	11	4
856	929	16	9	7	0	1	9	6

* 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 'M2' are the proportions of students who received scores 0 through 2. '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 0.70 to 0.81 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
841	0.77	0.84
842	0.71	0.91
843	0.81	0.91
844	0.75	0.90
845	0.74	0.95
846	0.78	0.85
847	0.78	0.90
848	0.77	0.87
849	0.78	0.89
850	0.70	0.89
851	0.74	0.78
852	0.75	0.89
853	0.79	0.94
854	0.79	0.80
855	0.74	0.79
856	0.72	0.81

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 students' 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 0.78 to 0.95, 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 differences between the first read and second read were 0.

Table 4. Point Differences Between First and Second Reads

			Difference (First Read minus Second Read)				
Form	Item	Score Points	-2	-1	0	1	2
841	12	1	0.00	0.06	0.92	0.02	0.00
841	13	1	0.00	0.02	0.94	0.04	0.00
841	14	1	0.00	0.05	0.92	0.04	0.00
841	15	1	0.00	0.04	0.92	0.05	0.00
841	16	1	0.00	0.04	0.95	0.01	0.00
841	17	1	0.00	0.05	0.91	0.03	0.00
841	18	1	0.00	0.05	0.93	0.02	0.00
841	19	1	0.00	0.04	0.91	0.05	0.00
841	20	1	0.00	0.04	0.90	0.06	0.00
842	12	1	0.00	0.01	0.99	0.01	0.00
842	13	1	0.00	0.05	0.91	0.05	0.00
842	14	1	0.00	0.01	0.97	0.01	0.00
842	15	1	0.00	0.02	0.98	0.00	0.00
842	16	1	0.00	0.02	0.96	0.02	0.00
842	17	1	0.00	0.02	0.94	0.04	0.00
842	18	1	0.00	0.02	0.95	0.02	0.00
842	19	1	0.00	0.05	0.93	0.02	0.00
843	12	1	0.00	0.00	1.00	0.00	0.00
843	13	1	0.00	0.04	0.93	0.03	0.00
843	14	1	0.00	0.01	0.98	0.02	0.00
843	15	1	0.00	0.00	1.00	0.00	0.00
843	16	1	0.00	0.01	0.99	0.00	0.00
843	17	1	0.00	0.03	0.93	0.04	0.00
843	18	1	0.00	0.04	0.93	0.02	0.00
843	19	1	0.00	0.03	0.93	0.04	0.00
843	20	1	0.00	0.04	0.90	0.06	0.00
844	12	1	0.00	0.01	0.95	0.04	0.00
844	13	1	0.00	0.06	0.90	0.04	0.00
844	14	1	0.00	0.02	0.91	0.07	0.00
844	15	1	0.00	0.01	0.95	0.04	0.00
844	16	1	0.00	0.00	0.99	0.01	0.00

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

			Difference (First Read minus Second Read)				
Form	Item	Score Points	-2	-1	0	1	2
844	17	1	0.00	0.01	0.96	0.03	0.00
844	18	1	0.00	0.04	0.93	0.03	0.00
844	19	1	0.00	0.00	1.00	0.00	0.00
845	12	1	0.00	0.00	1.00	0.00	0.00
845	13	1	0.00	0.02	0.98	0.00	0.00
845	14	1	0.00	0.00	1.00	0.00	0.00
845	15	1	0.00	0.01	0.99	0.00	0.00
845	16	1	0.00	0.03	0.97	0.00	0.00
845	17	1	0.00	0.01	0.96	0.03	0.00
845	18	1	0.00	0.01	0.99	0.01	0.00
845	19	1	0.00	0.05	0.91	0.04	0.00
846	13	1	0.00	0.00	0.99	0.01	0.00
846	14	1	0.00	0.02	0.95	0.03	0.00
846	15	1	0.00	0.05	0.88	0.06	0.00
846	16	1	0.00	0.02	0.94	0.04	0.00
846	17	1	0.00	0.04	0.91	0.05	0.00
846	18	1	0.00	0.04	0.95	0.02	0.00
846	19	1	0.00	0.07	0.89	0.05	0.00
846	20	1	0.00	0.04	0.92	0.05	0.00
847	12	1	0.00	0.05	0.92	0.03	0.00
847	13	1	0.00	0.03	0.94	0.03	0.00
847	14	1	0.00	0.01	0.99	0.00	0.00
847	15	1	0.00	0.02	0.94	0.04	0.00
847	16	1	0.00	0.06	0.91	0.02	0.00
847	17	1	0.00	0.03	0.94	0.03	0.00
847	18	1	0.00	0.02	0.96	0.02	0.00
847	19	1	0.00	0.00	0.99	0.01	0.00
848	12	1	0.00	0.06	0.92	0.02	0.00
848	13	1	0.00	0.04	0.95	0.01	0.00
848	14	1	0.00	0.00	1.00	0.00	0.00
848	15	1	0.00	0.01	0.99	0.00	0.00

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

			Difference (First Read minus Second Read)				
Form	Item	Score Points	-2	-1	0	1	2
848	16	1	0.00	0.05	0.90	0.05	0.00
848	17	1	0.00	0.04	0.93	0.03	0.00
848	18	1	0.00	0.04	0.94	0.01	0.00
848	19	1	0.00	0.04	0.89	0.07	0.00
848	20	1	0.00	0.03	0.95	0.02	0.00
849	12	1	0.00	0.04	0.88	0.08	0.00
849	13	1	0.00	0.08	0.88	0.04	0.00
849	14	1	0.00	0.02	0.97	0.01	0.00
849	15	1	0.00	0.00	1.00	0.00	0.00
849	16	1	0.00	0.04	0.93	0.03	0.00
849	17	1	0.00	0.03	0.96	0.01	0.00
849	18	1	0.00	0.01	0.95	0.04	0.00
849	19	1	0.00	0.01	0.99	0.01	0.00
850	12	1	0.00	0.06	0.84	0.09	0.00
850	13	1	0.00	0.04	0.91	0.04	0.00
850	14	1	0.00	0.06	0.91	0.03	0.00
850	15	1	0.00	0.04	0.96	0.00	0.00
850	16	1	0.00	0.00	0.99	0.01	0.00
850	17	1	0.00	0.00	0.99	0.01	0.00
850	18	1	0.00	0.01	0.98	0.01	0.00
850	19	1	0.00	0.01	0.99	0.01	0.00
850	20	1	0.00	0.02	0.95	0.03	0.00
851	7	1	0.00	0.01	0.98	0.01	0.00
851	8	1	0.00	0.02	0.96	0.02	0.00
851	9	2	0.00	0.04	0.87	0.08	0.00
851	12	1	0.00	0.02	0.94	0.04	0.00
851	13	2	0.00	0.14	0.66	0.15	0.05
851	14	1	0.00	0.11	0.85	0.04	0.00
851	15	2	0.00	0.04	0.90	0.06	0.00
852	4	1	0.00	0.00	1.00	0.00	0.00
852	5	1	0.00	0.13	0.81	0.05	0.00

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

			Difference (First Read minus Second Read)				
Form	Item	Score Points	-2	-1	0	1	2
852	7	2	0.00	0.04	0.88	0.08	0.00
852	8	1	0.00	0.01	0.99	0.00	0.00
852	10	2	0.00	0.05	0.91	0.04	0.00
852	12	1	0.00	0.00	0.97	0.03	0.00
852	13	1	0.00	0.18	0.80	0.02	0.00
852	15	2	0.00	0.02	0.95	0.03	0.00
853	6	2	0.00	0.08	0.89	0.03	0.00
853	7	1	0.00	0.01	0.98	0.01	0.00
853	8	1	0.00	0.00	0.88	0.12	0.00
853	9	2	0.00	0.02	0.94	0.03	0.01
853	11	1	0.00	0.00	1.00	0.00	0.00
853	12	2	0.00	0.01	0.97	0.01	0.00
853	13	1	0.00	0.00	1.00	0.00	0.00
853	14	2	0.00	0.02	0.92	0.06	0.00
854	5	2	0.01	0.03	0.93	0.04	0.00
854	6	1	0.00	0.10	0.79	0.11	0.00
854	7	2	0.00	0.07	0.87	0.06	0.00
854	9	2	0.00	0.05	0.86	0.09	0.00
854	11	1	0.00	0.01	0.82	0.16	0.00
854	12	1	0.00	0.23	0.72	0.05	0.00
854	14	2	0.00	0.03	0.95	0.02	0.00
855	8	1	0.00	0.06	0.87	0.08	0.00
855	9	1	0.00	0.06	0.86	0.08	0.00
855	10	2	0.00	0.10	0.77	0.12	0.00
855	12	1	0.00	0.02	0.95	0.03	0.00
855	13	2	0.00	0.02	0.69	0.23	0.06
855	15	1	0.00	0.05	0.92	0.04	0.00
855	16	2	0.00	0.05	0.94	0.01	0.00
855	17	1	0.00	0.01	0.99	0.00	0.00

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

			Difference (First Read minus Second Read)				
Form	Item	Score Points	-2	-1	0	1	2
856	13	2	0.00	0.01	0.94	0.04	0.00
856	14	1	0.00	0.00	1.00	0.00	0.00
856	15	2	0.00	0.04	0.91	0.04	0.00
856	16	2	0.06	0.15	0.68	0.10	0.01

Table 5 contains additional summary information regarding the first and second reads. In the fourth column the percent of exact matches between the first and second scores is provided. 'Adj.' is the percentage of differences with a magnitude of one. 'Total' is the sum of the two prior columns and contains values between 92.8% 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	Intra-Class Correlation	Wt Kappa
841	12	1	133	91.7	8.3	100.0	0.7	0.8	0.44	0.42	0.78	0.78
841	13	1	133	94.0	6.0	100.0	0.8	0.8	0.37	0.38	0.78	0.78
841	14	1	132	91.7	8.3	100.0	0.3	0.3	0.46	0.46	0.80	0.80
841	15	1	132	91.7	8.3	100.0	0.5	0.5	0.50	0.50	0.83	0.83
841	16	1	129	95.3	4.7	100.0	0.2	0.3	0.43	0.45	0.88	0.88
841	17	1	129	91.5	8.5	100.0	0.8	0.8	0.43	0.42	0.77	0.76
841	18	1	129	93.0	7.0	100.0	0.3	0.3	0.45	0.46	0.83	0.83
841	19	1	129	90.7	9.3	100.0	0.6	0.6	0.49	0.49	0.80	0.80
841	20	1	136	90.4	9.6	100.0	0.8	0.8	0.42	0.43	0.74	0.74
842	12	1	135	98.5	1.5	100.0	0.6	0.6	0.49	0.49	0.97	0.97
842	13	1	133	91.0	9.0	100.0	0.4	0.4	0.48	0.48	0.81	0.81
842	14	1	135	97.0	3.0	100.0	0.5	0.5	0.50	0.50	0.94	0.94
842	15	1	131	98.5	1.5	100.0	0.4	0.4	0.48	0.48	0.97	0.97
842	16	1	128	96.1	3.9	100.0	0.5	0.5	0.50	0.50	0.92	0.92
842	17	1	128	93.8	6.2	100.0	0.6	0.5	0.50	0.50	0.87	0.87
842	18	1	128	95.3	4.7	100.0	0.4	0.4	0.49	0.49	0.90	0.90
842	19	1	128	93.0	7.0	100.0	0.1	0.1	0.23	0.27	0.44	0.43
843	12	1	131	100.0	0.0	100.0	0.7	0.7	0.44	0.44	1.00	1.00
843	13	1	133	93.2	6.8	100.0	0.5	0.5	0.50	0.50	0.86	0.86
843	14	1	133	97.7	2.3	100.0	0.7	0.6	0.48	0.48	0.95	0.95
843	15	1	131	100.0	0.0	100.0	0.2	0.2	0.43	0.43	1.00	1.00
843	16	1	131	99.2	0.8	100.0	0.7	0.7	0.47	0.47	0.98	0.98
843	17	1	135	93.3	6.7	100.0	0.4	0.4	0.48	0.48	0.86	0.86
843	18	1	135	93.3	6.7	100.0	0.6	0.6	0.49	0.49	0.86	0.86
843	19	1	135	93.3	6.7	100.0	0.4	0.3	0.48	0.48	0.85	0.85
843	20	1	135	89.6	10.4	100.0	0.5	0.4	0.50	0.50	0.79	0.79
844	12	1	134	94.8	5.2	100.0	0.5	0.5	0.50	0.50	0.90	0.90
844	13	1	134	89.6	10.4	100.0	0.6	0.6	0.50	0.50	0.79	0.79
844	14	1	138	90.6	9.4	100.0	0.4	0.4	0.50	0.49	0.81	0.81
844	15	1	138	94.9	5.1	100.0	0.6	0.6	0.49	0.49	0.90	0.89

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	Intra-Class Correlation	Wt Kappa
844	16	1	137	99.3	0.7	100.0	0.6	0.6	0.49	0.49	0.98	0.98
844	17	1	131	96.2	3.8	100.0	0.3	0.3	0.47	0.47	0.91	0.91
844	18	1	131	93.1	6.9	100.0	0.3	0.3	0.45	0.45	0.83	0.83
844	19	1	136	100.0	0.0	100.0	0.6	0.6	0.50	0.50	1.00	1.00
845	12	1	136	100.0	0.0	100.0	0.3	0.3	0.46	0.46	1.00	1.00
845	13	1	133	98.5	1.5	100.0	0.3	0.3	0.46	0.46	0.96	0.96
845	14	1	135	100.0	0.0	100.0	0.2	0.2	0.38	0.38	1.00	1.00
845	15	1	137	99.3	0.7	100.0	0.5	0.5	0.50	0.50	0.99	0.99
845	16	1	137	97.1	2.9	100.0	0.3	0.3	0.45	0.47	0.93	0.93
845	17	1	137	96.4	3.6	100.0	0.5	0.5	0.50	0.50	0.93	0.93
845	18	1	137	98.5	1.5	100.0	0.5	0.5	0.50	0.50	0.97	0.97
845	19	1	137	91.2	8.8	100.0	0.5	0.5	0.50	0.50	0.82	0.82
846	13	1	130	99.2	0.8	100.0	0.8	0.8	0.38	0.39	0.97	0.97
846	14	1	130	94.6	5.4	100.0	0.4	0.4	0.49	0.49	0.89	0.89
846	15	1	130	88.5	11.5	100.0	0.5	0.5	0.50	0.50	0.77	0.77
846	16	1	139	94.2	5.8	100.0	0.5	0.4	0.50	0.50	0.88	0.88
846	17	1	139	90.6	9.4	100.0	0.7	0.7	0.45	0.45	0.77	0.77
846	18	1	131	94.7	5.3	100.0	0.5	0.5	0.50	0.50	0.89	0.89
846	19	1	131	88.5	11.5	100.0	0.2	0.2	0.39	0.41	0.65	0.65
846	20	1	131	91.6	8.4	100.0	0.4	0.4	0.49	0.49	0.82	0.82
847	12	1	131	91.6	8.4	100.0	0.4	0.4	0.49	0.49	0.83	0.82
847	13	1	131	93.9	6.1	100.0	0.6	0.6	0.50	0.50	0.88	0.88
847	14	1	125	99.2	0.8	100.0	0.6	0.6	0.49	0.49	0.98	0.98
847	15	1	125	93.6	6.4	100.0	0.4	0.4	0.49	0.49	0.87	0.87
847	16	1	125	91.2	8.8	100.0	0.6	0.7	0.49	0.47	0.81	0.81
847	17	1	125	93.6	6.4	100.0	0.5	0.5	0.50	0.50	0.87	0.87
847	18	1	125	96.0	4.0	100.0	0.6	0.6	0.50	0.50	0.92	0.92
847	19	1	125	99.2	0.8	100.0	0.2	0.2	0.41	0.41	0.98	0.98
848	12	1	141	92.2	7.8	100.0	0.6	0.7	0.48	0.47	0.83	0.83
848	13	1	141	95.0	5.0	100.0	0.4	0.4	0.49	0.50	0.90	0.90

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	Intra-Class Correlation	Wt Kappa
848	14	1	139	100.0	0.0	100.0	0.6	0.6	0.49	0.49	1.00	1.00
848	15	1	135	99.3	0.7	100.0	0.8	0.8	0.42	0.42	0.98	0.98
848	16	1	138	89.9	10.1	100.0	0.9	0.9	0.35	0.35	0.59	0.59
848	17	1	138	92.8	7.2	100.0	0.8	0.8	0.42	0.41	0.79	0.79
848	18	1	138	94.2	5.8	100.0	0.4	0.5	0.50	0.50	0.88	0.88
848	19	1	138	89.1	10.9	100.0	0.6	0.6	0.49	0.49	0.77	0.77
848	20	1	139	95.0	5.0	100.0	0.7	0.7	0.46	0.45	0.88	0.88
849	12	1	140	87.9	12.1	100.0	0.4	0.4	0.50	0.49	0.75	0.75
849	13	1	140	87.9	12.1	100.0	0.5	0.5	0.50	0.50	0.76	0.76
849	14	1	135	97.0	3.0	100.0	0.2	0.2	0.40	0.41	0.91	0.91
849	15	1	136	100.0	0.0	100.0	0.6	0.6	0.49	0.49	1.00	1.00
849	16	1	136	93.4	6.6	100.0	0.3	0.3	0.46	0.46	0.84	0.84
849	17	1	136	96.3	3.7	100.0	0.5	0.6	0.50	0.50	0.93	0.93
849	18	1	136	94.9	5.1	100.0	0.5	0.5	0.50	0.50	0.90	0.90
849	19	1	138	98.6	1.4	100.0	0.6	0.6	0.49	0.49	0.97	0.97
850	12	1	141	84.4	15.6	100.0	0.6	0.6	0.49	0.49	0.67	0.67
850	13	1	141	91.5	8.5	100.0	0.7	0.7	0.48	0.48	0.81	0.81
850	14	1	141	90.8	9.2	100.0	0.6	0.6	0.49	0.48	0.81	0.80
850	15	1	141	96.5	3.5	100.0	0.6	0.7	0.48	0.47	0.92	0.92
850	16	1	136	98.5	1.5	100.0	0.3	0.3	0.45	0.45	0.96	0.96
850	17	1	136	99.3	0.7	100.0	0.4	0.3	0.48	0.48	0.98	0.98
850	18	1	136	97.8	2.2	100.0	0.3	0.3	0.45	0.46	0.95	0.95
850	19	1	138	98.6	1.4	100.0	0.0	0.0	0.19	0.19	0.79	0.79
850	20	1	138	94.9	5.1	100.0	0.2	0.2	0.37	0.36	0.81	0.81
851	07	1	142	97.9	2.1	100.0	0.8	0.7	0.43	0.44	0.94	0.94
851	08	1	147	95.9	4.1	100.0	0.6	0.6	0.48	0.48	0.91	0.91
851	09	2	143	87.4	12.6	100.0	0.7	0.7	0.68	0.65	0.86	0.82
851	12	1	157	94.3	5.7	100.0	0.6	0.6	0.50	0.50	0.88	0.88
851	13	2	257	66.1	28.8	94.9	0.7	0.6	0.81	0.72	0.59	0.51
851	14	1	260	85.0	15.0	100.0	0.3	0.4	0.47	0.49	0.69	0.68

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	Intra-Class Correlation	Wt Kappa
851	15	2	139	89.9	10.1	100.0	0.7	0.6	0.81	0.79	0.92	0.88
852	04	1	143	100.0	0.0	100.0	0.4	0.4	0.50	0.50	1.00	1.00
852	05	1	257	81.3	18.7	100.0	0.4	0.5	0.49	0.50	0.63	0.63
852	07	2	138	87.7	12.3	100.0	1.1	1.0	0.77	0.76	0.90	0.85
852	08	1	138	99.3	0.7	100.0	0.4	0.4	0.48	0.48	0.98	0.98
852	10	2	142	90.8	9.2	100.0	1.0	1.0	0.82	0.82	0.93	0.90
852	12	1	156	97.4	2.6	100.0	0.6	0.6	0.49	0.49	0.95	0.95
852	13	1	248	80.2	19.8	100.0	0.2	0.3	0.38	0.47	0.54	0.50
852	15	2	139	95.0	5.0	100.0	0.9	0.9	0.83	0.84	0.96	0.94
853	06	2	138	89.1	10.9	100.0	1.0	1.0	0.84	0.83	0.92	0.88
853	07	1	138	97.8	2.2	100.0	0.2	0.2	0.39	0.39	0.93	0.93
853	08	1	235	87.7	12.3	100.0	0.5	0.4	0.50	0.48	0.77	0.75
853	09	2	137	94.2	5.1	99.3	0.3	0.3	0.63	0.65	0.90	0.86
853	11	1	140	100.0	0.0	100.0	0.7	0.7	0.46	0.46	1.00	1.00
853	12	2	140	97.1	2.9	100.0	0.8	0.8	0.82	0.80	0.98	0.97
853	13	1	141	100.0	0.0	100.0	0.4	0.4	0.48	0.48	1.00	1.00
853	14	2	141	92.2	7.8	100.0	0.8	0.8	0.89	0.87	0.95	0.92
854	05	2	136	92.6	6.6	99.3	0.9	0.9	0.83	0.80	0.93	0.91
854	06	1	335	79.1	20.9	100.0	0.5	0.5	0.50	0.50	0.58	0.58
854	07	2	135	87.4	12.6	100.0	1.1	1.1	0.78	0.76	0.89	0.85
854	09	2	134	85.8	14.2	100.0	1.2	1.2	0.77	0.75	0.88	0.83
854	11	1	284	82.4	17.6	100.0	0.9	0.7	0.33	0.45	0.52	0.47
854	12	1	310	71.9	28.1	100.0	0.2	0.4	0.41	0.49	0.41	0.37
854	14	2	138	94.9	5.1	100.0	0.8	0.8	0.79	0.79	0.96	0.94
855	08	1	145	86.9	13.1	100.0	0.9	0.9	0.32	0.35	0.41	0.41
855	09	1	145	86.2	13.8	100.0	0.8	0.8	0.37	0.40	0.53	0.53
855	10	2	146	77.4	22.6	100.0	0.9	0.9	0.72	0.74	0.78	0.71
855	12	1	169	94.7	5.3	100.0	0.8	0.8	0.41	0.42	0.85	0.85
855	13	2	248	68.5	25.4	94.0	1.1	0.8	0.79	0.73	0.66	0.57
855	15	1	155	91.6	8.4	100.0	0.5	0.5	0.50	0.50	0.83	0.83

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	Intra-Class Correlation	Wt Kappa
855	16	2	138	94.2	5.8	100.0	0.8	0.9	0.90	0.92	0.97	0.94
855	17	1	146	98.6	1.4	100.0	0.2	0.2	0.42	0.43	0.96	0.96
856	13	2	135	94.1	5.9	100.0	1.0	1.0	0.86	0.86	0.96	0.94
856	14	1	134	100.0	0.0	100.0	0.5	0.5	0.50	0.50	1.00	1.00
856	15	2	134	91.0	9.0	100.0	1.1	1.1	0.82	0.84	0.94	0.90
856	16	2	250	68.0	24.8	92.8	0.7	0.8	0.69	0.79	0.54	0.51

* Adj. = difference of one

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 scores was 0.3, while there were minimal differences among standard deviation statistics for most items.

Intra-class Correlation

The intra-class 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 reads. Correlations greater than 0.60 are considered very strong because they explain more than one-third of the variance in scores. All but 10 items had intra-class correlations greater than or equal to 0.63 (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. For all but one item, the Weighted Kappa statistics ranged from 0.41 to 1.0, 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 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 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 they have 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. Different samples of students 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 may 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. 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. Most of the items fell within the moderate -2.0 to +2.0 difficulty range and only five items had an INFIT statistic outside the typical range.

Table 6. 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
841	931	20	0	20	0	0	20	0
842	928	19	2	16	1	0	18	1
843	939	20	1	18	1	0	20	0
844	934	19	1	18	0	0	19	0
845	940	19	0	19	0	0	19	0
846	951	20	0	20	0	0	20	0
847	937	19	1	18	0	0	18	1
848	951	20	0	20	0	0	20	0
849	943	19	3	16	0	0	19	0
850	954	20	0	19	1	0	20	0
851	948	16	0	12	0	0	12	0
852	961	16	0	11	0	0	11	0
853	954	15	0	11	1	0	11	1
854	938	15	1	9	0	0	9	1
855	939	17	2	13	0	0	15	0
856	929	16	1	15	0	0	15	1

* A complete listing of all Partial Credit Model item analysis results is presented in Appendix B.

** 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 titled 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 2010 field test administration for the New York State Regents Examination in Physics consisted of 15 field test forms numbered 841–855 and one anchor form labeled 856. The field test forms contained multiple-choice and constructed-response items. All students participating in the field test were administered one of the 16 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 2009 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 non-extreme students. This average ability estimate was used to equate the remaining field test forms as well as updating the item parameters for the anchor form.

As part of the anchor item equating, an item-stability check was performed. After fixing all of the items to their 2009 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 a strong level of stability in the items used on the anchor form.

The equated mean ability estimate for form 856 was -0.25. This value served as the target mean ability for 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 856). 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
841	0.12	-0.36
842	-0.10	-0.15
843	-0.08	-0.16
844	0.25	-0.47
845	-0.39	0.13
846	0.15	-0.38
847	0.33	-0.55
848	0.55	-0.76
849	0.40	-0.61
850	-0.14	-0.11
851	-0.41	0.15
852	-0.17	-0.07
853	-0.46	0.20
854	-0.01	-0.21
855	0.58	-0.77

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 2010 operational

test maps with content information regarding each item included on the January 2010 and June 2010 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 -0.319 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 and June 2010 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 2004. 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 June 2004 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 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, 213–20.
- Cronbach, L.J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Dorans, N.J. and Holland, P.W. 1992. DIF Detection and Description: Mantel–Haenszel and Standardization. In *Differential Item Functioning: Theory and Practice*, edited by P.W. Holland and H. Wainer, 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
2010_Phys_FT	841	MC	01	1	931	0.77	0.01		0.55	0.25	0.10	0.09			0.55	0.42
2010_Phys_FT	841	MC	02	1	931	0.77	0.00		0.30	0.60	0.02	0.07			0.60	0.31
2010_Phys_FT	841	MC	03	1	931	0.77	0.00		0.65	0.14	0.12	0.09			0.65	0.44
2010_Phys_FT	841	MC	04	1	931	0.77	0.03		0.08	0.13	0.42	0.33			0.42	0.28
2010_Phys_FT	841	MC	05	1	931	0.77	0.03		0.15	0.58	0.14	0.10			0.58	0.44
2010_Phys_FT	841	MC	06	1	931	0.77	0.02		0.49	0.03	0.20	0.27			0.49	0.48
2010_Phys_FT	841	MC	07	1	931	0.77	0.00		0.08	0.23	0.43	0.26			0.43	0.28
2010_Phys_FT	841	MC	08	1	931	0.77	0.01		0.25	0.03	0.05	0.65			0.65	0.30
2010_Phys_FT	841	MC	09	1	931	0.77	0.04		0.06	0.29	0.53	0.09			0.53	0.32
2010_Phys_FT	841	MC	10	1	931	0.77	0.01		0.08	0.45	0.20	0.26			0.45	0.34
2010_Phys_FT	841	MC	11	1	931	0.77	0.01		0.35	0.59	0.04	0.01			0.59	0.37
2010_Phys_FT	841	CR	12	1	931	0.77	0.08	0.23	0.69						0.69	0.45
2010_Phys_FT	841	CR	13	1	931	0.77	0.08	0.15	0.76						0.76	0.47
2010_Phys_FT	841	CR	14	1	931	0.77	0.15	0.55	0.31						0.31	0.53
2010_Phys_FT	841	CR	15	1	931	0.77	0.15	0.42	0.43						0.43	0.55
2010_Phys_FT	841	CR	16	1	931	0.77	0.11	0.69	0.20						0.20	0.60
2010_Phys_FT	841	CR	17	1	931	0.77	0.11	0.23	0.66						0.66	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
2010_Phys_FT	841	CR	18	1	931	0.77	0.27	0.48	0.25						0.25	0.58
2010_Phys_FT	841	CR	19	1	931	0.77	0.28	0.20	0.52						0.52	0.60
2010_Phys_FT	841	CR	20	1	931	0.77	0.13	0.20	0.67						0.67	0.44
2010_Phys_FT	842	MC	01	1	928	0.71	0.01		0.19	0.51	0.18	0.10			0.51	0.37
2010_Phys_FT	842	MC	02	1	928	0.71	0.00		0.01	0.01	0.15	0.82			0.82	0.32
2010_Phys_FT	842	MC	03	1	928	0.71	0.00		0.65	0.14	0.19	0.01			0.65	0.28
2010_Phys_FT	842	MC	04	1	928	0.71	0.00		0.02	0.65	0.30	0.02			0.65	0.37
2010_Phys_FT	842	MC	05	1	928	0.71	0.01		0.45	0.49	0.03	0.02			0.49	0.41
2010_Phys_FT	842	MC	06	1	928	0.71	0.00		0.82	0.02	0.13	0.02			0.82	0.31
2010_Phys_FT	842	MC	07	1	928	0.71	0.01		0.08	0.36	0.39	0.16			0.39	0.28
2010_Phys_FT	842	MC	08	1	928	0.71	0.07		0.31	0.18	0.22	0.22			0.31	0.11
2010_Phys_FT	842	MC	09	1	928	0.71	0.00		0.22	0.67	0.08	0.03			0.67	0.38
2010_Phys_FT	842	MC	10	1	928	0.71	0.01		0.13	0.22	0.36	0.28			0.36	0.35
2010_Phys_FT	842	MC	11	1	928	0.71	0.02		0.59	0.27	0.07	0.04			0.59	0.43
2010_Phys_FT	842	CR	12	1	928	0.71	0.10	0.31	0.60						0.60	0.51
2010_Phys_FT	842	CR	13	1	928	0.71	0.11	0.57	0.32						0.32	0.54
2010_Phys_FT	842	CR	14	1	928	0.71	0.15	0.38	0.47						0.47	0.57
2010_Phys_FT	842	CR	15	1	928	0.71	0.15	0.52	0.32						0.32	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
2010_Phys_FT	842	CR	16	1	928	0.71	0.26	0.34	0.40						0.40	0.52
2010_Phys_FT	842	CR	17	1	928	0.71	0.26	0.28	0.46						0.46	0.54
2010_Phys_FT	842	CR	18	1	928	0.71	0.30	0.36	0.34						0.34	0.53
2010_Phys_FT	842	CR	19	1	928	0.71	0.35	0.59	0.06						0.06	0.26
2010_Phys_FT	843	MC	01	1	939	0.81	0.00		0.88	0.04	0.06	0.02			0.88	0.35
2010_Phys_FT	843	MC	02	1	939	0.81	0.01		0.47	0.17	0.19	0.16			0.47	0.34
2010_Phys_FT	843	MC	03	1	939	0.81	0.01		0.28	0.28	0.17	0.26			0.26	0.38
2010_Phys_FT	843	MC	04	1	939	0.81	0.01		0.19	0.28	0.13	0.40			0.40	0.46
2010_Phys_FT	843	MC	05	1	939	0.81	0.01		0.75	0.12	0.07	0.06			0.75	0.40
2010_Phys_FT	843	MC	06	1	939	0.81	0.00		0.14	0.02	0.06	0.78			0.78	0.35
2010_Phys_FT	843	MC	07	1	939	0.81	0.01		0.18	0.63	0.10	0.08			0.63	0.36
2010_Phys_FT	843	MC	08	1	939	0.81	0.02		0.36	0.09	0.31	0.23			0.23	0.26
2010_Phys_FT	843	MC	09	1	939	0.81	0.04		0.26	0.50	0.15	0.06			0.50	0.40
2010_Phys_FT	843	MC	10	1	939	0.81	0.01		0.22	0.06	0.09	0.61			0.61	0.30
2010_Phys_FT	843	MC	11	1	939	0.81	0.01		0.49	0.12	0.26	0.11			0.12	0.08
2010_Phys_FT	843	CR	12	1	939	0.81	0.12	0.23	0.64						0.64	0.60
2010_Phys_FT	843	CR	13	1	939	0.81	0.17	0.33	0.50						0.50	0.62
2010_Phys_FT	843	CR	14	1	939	0.81	0.17	0.20	0.63						0.63	0.64

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
2010_Phys_FT	843	CR	15	1	939	0.81	0.13	0.65	0.22						0.22	0.43
2010_Phys_FT	843	CR	16	1	939	0.81	0.13	0.25	0.62						0.62	0.56
2010_Phys_FT	843	CR	17	1	939	0.81	0.24	0.47	0.30						0.30	0.66
2010_Phys_FT	843	CR	18	1	939	0.81	0.24	0.26	0.50						0.50	0.72
2010_Phys_FT	843	CR	19	1	939	0.81	0.27	0.44	0.30						0.30	0.63
2010_Phys_FT	843	CR	20	1	939	0.81	0.27	0.32	0.41						0.41	0.65
2010_Phys_FT	844	MC	01	1	934	0.75	0.00		0.21	0.10	0.01	0.68			0.68	0.43
2010_Phys_FT	844	MC	02	1	934	0.75	0.00		0.12	0.60	0.20	0.08			0.60	0.31
2010_Phys_FT	844	MC	03	1	934	0.75	0.02		0.13	0.69	0.09	0.07			0.69	0.44
2010_Phys_FT	844	MC	04	1	934	0.75	0.00		0.07	0.06	0.08	0.79			0.79	0.26
2010_Phys_FT	844	MC	05	1	934	0.75	0.02		0.34	0.27	0.34	0.04			0.34	0.27
2010_Phys_FT	844	MC	06	1	934	0.75	0.00		0.14	0.28	0.54	0.04			0.54	0.29
2010_Phys_FT	844	MC	07	1	934	0.75	0.00		0.03	0.86	0.07	0.03			0.86	0.43
2010_Phys_FT	844	MC	08	1	934	0.75	0.01		0.12	0.68	0.13	0.07			0.68	0.43
2010_Phys_FT	844	MC	09	1	934	0.75	0.03		0.07	0.16	0.68	0.07			0.68	0.40
2010_Phys_FT	844	MC	10	1	934	0.75	0.01		0.03	0.44	0.09	0.42			0.42	0.35
2010_Phys_FT	844	MC	11	1	934	0.75	0.03		0.17	0.21	0.43	0.17			0.43	0.31
2010_Phys_FT	844	CR	12	1	934	0.75	0.06	0.38	0.56						0.56	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
2010_Phys_FT	844	CR	13	1	934	0.75	0.06	0.39	0.55						0.55	0.45
2010_Phys_FT	844	CR	14	1	934	0.75	0.12	0.48	0.40						0.40	0.56
2010_Phys_FT	844	CR	15	1	934	0.75	0.12	0.33	0.55						0.55	0.55
2010_Phys_FT	844	CR	16	1	934	0.75	0.15	0.33	0.52						0.52	0.57
2010_Phys_FT	844	CR	17	1	934	0.75	0.26	0.46	0.28						0.28	0.56
2010_Phys_FT	844	CR	18	1	934	0.75	0.26	0.50	0.24						0.24	0.59
2010_Phys_FT	844	CR	19	1	934	0.75	0.15	0.27	0.58						0.58	0.45
2010_Phys_FT	845	MC	01	1	940	0.74	0.04		0.07	0.54	0.32	0.03			0.32	0.23
2010_Phys_FT	845	MC	02	1	940	0.74	0.00		0.10	0.50	0.38	0.01			0.38	0.46
2010_Phys_FT	845	MC	03	1	940	0.74	0.01		0.17	0.25	0.38	0.20			0.38	0.17
2010_Phys_FT	845	MC	04	1	940	0.74	0.01		0.30	0.46	0.12	0.10			0.30	0.28
2010_Phys_FT	845	MC	05	1	940	0.74	0.02		0.08	0.08	0.63	0.20			0.63	0.44
2010_Phys_FT	845	MC	06	1	940	0.74	0.00		0.21	0.58	0.14	0.07			0.58	0.31
2010_Phys_FT	845	MC	07	1	940	0.74	0.02		0.09	0.38	0.41	0.10			0.41	0.33
2010_Phys_FT	845	MC	08	1	940	0.74	0.00		0.29	0.62	0.07	0.01			0.29	0.43
2010_Phys_FT	845	MC	09	1	940	0.74	0.01		0.06	0.52	0.21	0.20			0.52	0.41
2010_Phys_FT	845	MC	10	1	940	0.74	0.01		0.79	0.16	0.02	0.01			0.79	0.27
2010_Phys_FT	845	MC	11	1	940	0.74	0.07		0.23	0.16	0.37	0.17			0.37	0.26

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
2010_Phys_FT	845	CR	12	1	940	0.74	0.12	0.56	0.32						0.32	0.50
2010_Phys_FT	845	CR	13	1	940	0.74	0.12	0.55	0.33						0.33	0.40
2010_Phys_FT	845	CR	14	1	940	0.74	0.29	0.51	0.20						0.20	0.46
2010_Phys_FT	845	CR	15	1	940	0.74	0.14	0.31	0.55						0.55	0.45
2010_Phys_FT	845	CR	16	1	940	0.74	0.21	0.46	0.33						0.33	0.58
2010_Phys_FT	845	CR	17	1	940	0.74	0.21	0.25	0.54						0.54	0.60
2010_Phys_FT	845	CR	18	1	940	0.74	0.30	0.32	0.38						0.38	0.63
2010_Phys_FT	845	CR	19	1	940	0.74	0.31	0.25	0.44						0.44	0.66
2010_Phys_FT	846	MC	01	1	951	0.78	0.01		0.54	0.18	0.04	0.23			0.54	0.36
2010_Phys_FT	846	MC	02	1	951	0.78	0.00		0.01	0.30	0.66	0.03			0.66	0.40
2010_Phys_FT	846	MC	03	1	951	0.78	0.01		0.09	0.14	0.11	0.65			0.65	0.43
2010_Phys_FT	846	MC	04	1	951	0.78	0.00		0.01	0.01	0.61	0.37			0.37	0.39
2010_Phys_FT	846	MC	05	1	951	0.78	0.01		0.05	0.18	0.22	0.54			0.54	0.20
2010_Phys_FT	846	MC	06	1	951	0.78	0.01		0.02	0.55	0.40	0.03			0.55	0.31
2010_Phys_FT	846	MC	07	1	951	0.78	0.02		0.35	0.46	0.10	0.07			0.46	0.34
2010_Phys_FT	846	MC	08	1	951	0.78	0.01		0.19	0.04	0.56	0.21			0.56	0.31
2010_Phys_FT	846	MC	09	1	951	0.78	0.03		0.07	0.07	0.69	0.14			0.69	0.41
2010_Phys_FT	846	MC	10	1	951	0.78	0.01		0.05	0.73	0.15	0.06			0.73	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
2010_Phys_FT	846	MC	11	1	951	0.78	0.03		0.16	0.17	0.61	0.03			0.61	0.50
2010_Phys_FT	846	MC	12	1	951	0.78	0.06		0.22	0.18	0.44	0.09			0.44	0.44
2010_Phys_FT	846	CR	13	1	951	0.78	0.09	0.15	0.76						0.76	0.50
2010_Phys_FT	846	CR	14	1	951	0.78	0.12	0.51	0.36						0.36	0.53
2010_Phys_FT	846	CR	15	1	951	0.78	0.12	0.35	0.53						0.53	0.55
2010_Phys_FT	846	CR	16	1	951	0.78	0.16	0.37	0.47						0.47	0.50
2010_Phys_FT	846	CR	17	1	951	0.78	0.16	0.14	0.70						0.70	0.53
2010_Phys_FT	846	CR	18	1	951	0.78	0.28	0.28	0.44						0.44	0.59
2010_Phys_FT	846	CR	19	1	951	0.78	0.38	0.44	0.18						0.18	0.54
2010_Phys_FT	846	CR	20	1	951	0.78	0.37	0.28	0.35						0.35	0.62
2010_Phys_FT	847	MC	01	1	937	0.78	0.00		0.09	0.40	0.10	0.40			0.40	0.28
2010_Phys_FT	847	MC	02	1	937	0.78	0.00		0.12	0.09	0.09	0.70			0.70	0.41
2010_Phys_FT	847	MC	03	1	937	0.78	0.00		0.88	0.06	0.02	0.04			0.88	0.30
2010_Phys_FT	847	MC	04	1	937	0.78	0.02		0.06	0.72	0.14	0.06			0.72	0.40
2010_Phys_FT	847	MC	05	1	937	0.78	0.01		0.81	0.04	0.09	0.05			0.81	0.48
2010_Phys_FT	847	MC	06	1	937	0.78	0.01		0.43	0.16	0.19	0.21			0.43	0.24
2010_Phys_FT	847	MC	07	1	937	0.78	0.00		0.15	0.39	0.32	0.14			0.39	0.33
2010_Phys_FT	847	MC	08	1	937	0.78	0.00		0.60	0.35	0.04	0.00			0.60	0.27

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
2010_Phys_FT	847	MC	09	1	937	0.78	0.00		0.07	0.66	0.20	0.07			0.66	0.34
2010_Phys_FT	847	MC	10	1	937	0.78	0.01		0.03	0.16	0.18	0.62			0.62	0.43
2010_Phys_FT	847	MC	11	1	937	0.78	0.01		0.58	0.02	0.30	0.09			0.58	0.15
2010_Phys_FT	847	CR	12	1	937	0.78	0.15	0.48	0.37						0.37	0.59
2010_Phys_FT	847	CR	13	1	937	0.78	0.15	0.28	0.56						0.56	0.61
2010_Phys_FT	847	CR	14	1	937	0.78	0.13	0.29	0.58						0.58	0.58
2010_Phys_FT	847	CR	15	1	937	0.78	0.17	0.46	0.37						0.37	0.64
2010_Phys_FT	847	CR	16	1	937	0.78	0.17	0.22	0.61						0.61	0.68
2010_Phys_FT	847	CR	17	1	937	0.78	0.24	0.28	0.48						0.48	0.65
2010_Phys_FT	847	CR	18	1	937	0.78	0.24	0.16	0.60						0.60	0.66
2010_Phys_FT	847	CR	19	1	937	0.78	0.32	0.45	0.23						0.23	0.54
2010_Phys_FT	848	MC	01	1	951	0.77	0.00		0.04	0.21	0.72	0.02			0.72	0.40
2010_Phys_FT	848	MC	02	1	951	0.77	0.01		0.10	0.77	0.09	0.04			0.77	0.47
2010_Phys_FT	848	MC	03	1	951	0.77	0.01		0.11	0.68	0.13	0.08			0.68	0.52
2010_Phys_FT	848	MC	04	1	951	0.77	0.01		0.16	0.19	0.57	0.07			0.57	0.24
2010_Phys_FT	848	MC	05	1	951	0.77	0.00		0.14	0.09	0.51	0.25			0.51	0.39
2010_Phys_FT	848	MC	06	1	951	0.77	0.03		0.51	0.12	0.29	0.05			0.51	0.36
2010_Phys_FT	848	MC	07	1	951	0.77	0.01		0.11	0.23	0.03	0.62			0.62	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
2010_Phys_FT	848	MC	08	1	951	0.77	0.01		0.49	0.30	0.19	0.00			0.49	0.30
2010_Phys_FT	848	MC	09	1	951	0.77	0.01		0.38	0.14	0.45	0.02			0.38	0.31
2010_Phys_FT	848	MC	10	1	951	0.77	0.01		0.13	0.81	0.03	0.02			0.81	0.39
2010_Phys_FT	848	MC	11	1	951	0.77	0.03		0.13	0.17	0.10	0.57			0.57	0.39
2010_Phys_FT	848	CR	12	1	951	0.77	0.11	0.25	0.64						0.64	0.51
2010_Phys_FT	848	CR	13	1	951	0.77	0.12	0.48	0.40						0.40	0.34
2010_Phys_FT	848	CR	14	1	951	0.77	0.09	0.34	0.57						0.57	0.57
2010_Phys_FT	848	CR	15	1	951	0.77	0.09	0.18	0.73						0.73	0.54
2010_Phys_FT	848	CR	16	1	951	0.77	0.08	0.14	0.79						0.79	0.46
2010_Phys_FT	848	CR	17	1	951	0.77	0.08	0.22	0.70						0.70	0.40
2010_Phys_FT	848	CR	18	1	951	0.77	0.17	0.43	0.40						0.40	0.52
2010_Phys_FT	848	CR	19	1	951	0.77	0.17	0.22	0.60						0.60	0.59
2010_Phys_FT	848	CR	20	1	951	0.77	0.19	0.13	0.68						0.68	0.58
2010_Phys_FT	849	MC	01	1	943	0.78	0.00		0.54	0.20	0.18	0.08			0.54	0.33
2010_Phys_FT	849	MC	02	1	943	0.78	0.01		0.18	0.28	0.43	0.11			0.43	0.32
2010_Phys_FT	849	MC	03	1	943	0.78	0.01		0.04	0.05	0.84	0.07			0.84	0.27
2010_Phys_FT	849	MC	04	1	943	0.78	0.00		0.02	0.92	0.01	0.04			0.92	0.41
2010_Phys_FT	849	MC	05	1	943	0.78	0.01		0.56	0.22	0.17	0.03			0.56	0.34

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
2010_Phys_FT	849	MC	06	1	943	0.78	0.00		0.84	0.03	0.04	0.09			0.84	0.49
2010_Phys_FT	849	MC	07	1	943	0.78	0.01		0.10	0.06	0.03	0.80			0.80	0.48
2010_Phys_FT	849	MC	08	1	943	0.78	0.00		0.10	0.57	0.30	0.03			0.57	0.41
2010_Phys_FT	849	MC	09	1	943	0.78	0.01		0.05	0.30	0.58	0.05			0.58	0.41
2010_Phys_FT	849	MC	10	1	943	0.78	0.01		0.09	0.13	0.20	0.57			0.57	0.46
2010_Phys_FT	849	MC	11	1	943	0.78	0.01		0.03	0.63	0.30	0.03			0.63	0.42
2010_Phys_FT	849	CR	12	1	943	0.78	0.13	0.47	0.40						0.40	0.53
2010_Phys_FT	849	CR	13	1	943	0.78	0.14	0.40	0.46						0.46	0.55
2010_Phys_FT	849	CR	14	1	943	0.78	0.33	0.48	0.19						0.19	0.41
2010_Phys_FT	849	CR	15	1	943	0.78	0.12	0.31	0.58						0.58	0.57
2010_Phys_FT	849	CR	16	1	943	0.78	0.21	0.48	0.31						0.31	0.48
2010_Phys_FT	849	CR	17	1	943	0.78	0.21	0.25	0.54						0.54	0.63
2010_Phys_FT	849	CR	18	1	943	0.78	0.23	0.31	0.46						0.46	0.61
2010_Phys_FT	849	CR	19	1	943	0.78	0.24	0.22	0.54						0.54	0.50
2010_Phys_FT	850	MC	01	1	954	0.70	0.00		0.09	0.17	0.71	0.03			0.71	0.39
2010_Phys_FT	850	MC	02	1	954	0.70	0.00		0.60	0.27	0.03	0.10			0.60	0.28
2010_Phys_FT	850	MC	03	1	954	0.70	0.00		0.04	0.82	0.06	0.08			0.82	0.38

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
2010_Phys_FT	850	MC	04	1	954	0.70	0.02		0.04	0.05	0.40	0.49			0.49	0.25
2010_Phys_FT	850	MC	05	1	954	0.70	0.02		0.68	0.13	0.07	0.10			0.68	0.40
2010_Phys_FT	850	MC	06	1	954	0.70	0.01		0.21	0.55	0.07	0.17			0.55	0.33
2010_Phys_FT	850	MC	07	1	954	0.70	0.01		0.11	0.18	0.65	0.03			0.65	0.43
2010_Phys_FT	850	MC	08	1	954	0.70	0.02		0.45	0.06	0.39	0.07			0.39	0.36
2010_Phys_FT	850	MC	09	1	954	0.70	0.02		0.35	0.04	0.46	0.14			0.35	0.16
2010_Phys_FT	850	MC	10	1	954	0.70	0.02		0.20	0.25	0.04	0.49			0.49	0.30
2010_Phys_FT	850	MC	11	1	954	0.70	0.04		0.10	0.39	0.41	0.06			0.41	0.20
2010_Phys_FT	850	CR	12	1	954	0.70	0.07	0.32	0.61						0.61	0.50
2010_Phys_FT	850	CR	13	1	954	0.70	0.07	0.24	0.68						0.68	0.44
2010_Phys_FT	850	CR	14	1	954	0.70	0.10	0.29	0.61						0.61	0.58
2010_Phys_FT	850	CR	15	1	954	0.70	0.10	0.32	0.58						0.58	0.60
2010_Phys_FT	850	CR	16	1	954	0.70	0.18	0.56	0.25						0.25	0.44
2010_Phys_FT	850	CR	17	1	954	0.70	0.28	0.42	0.30						0.30	0.51
2010_Phys_FT	850	CR	18	1	954	0.70	0.34	0.38	0.27						0.27	0.47
2010_Phys_FT	850	CR	19	1	954	0.70	0.50	0.45	0.05						0.05	0.35
2010_Phys_FT	850	CR	20	1	954	0.70	0.50	0.33	0.17						0.17	0.45
2010_Phys_FT	851	MC	01	1	948	0.74	0.01		0.15	0.56	0.06	0.23			0.56	0.33

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
2010_Phys_FT	851	MC	02	1	948	0.74	0.06		0.04	0.23	0.46	0.22			0.22	0.42
2010_Phys_FT	851	MC	03	1	948	0.74	0.00		0.12	0.45	0.05	0.38			0.38	0.46
2010_Phys_FT	851	MC	04	1	948	0.74	0.03		0.05	0.21	0.55	0.17			0.55	0.42
2010_Phys_FT	851	MC	05	1	948	0.74	0.00		0.49	0.03	0.28	0.19			0.28	0.28
2010_Phys_FT	851	MC	06	1												
2010_Phys_FT	851	CR	07	1	948	0.74	0.04	0.30	0.66						0.66	0.36
2010_Phys_FT	851	CR	08	1	948	0.74	0.06	0.37	0.57						0.57	0.50
2010_Phys_FT	851	CR	09	2	948	0.74	0.17	0.31	0.39	0.13					0.65	0.57
2010_Phys_FT	851	MC	10	1												
2010_Phys_FT	851	MC	11	1												
2010_Phys_FT	851	CR	12	1	948	0.74	0.21	0.34	0.45						0.45	0.59
2010_Phys_FT	851	CR	13	2	948	0.74	0.27	0.30	0.23	0.20					0.62	0.72
2010_Phys_FT	851	CR	14	1	948	0.74	0.27	0.39	0.35						0.35	0.64
2010_Phys_FT	851	CR	15	2	948	0.74	0.27	0.25	0.27	0.21					0.69	0.68
2010_Phys_FT	851	MC	16	1												
2010_Phys_FT	852	MC	01	1	961	0.75	0.01		0.56	0.19	0.18	0.07			0.56	0.47
2010_Phys_FT	852	MC	02	1	961	0.75	0.02		0.11	0.55	0.23	0.08			0.55	0.47
2010_Phys_FT	852	MC	03	1	961	0.75	0.01		0.07	0.19	0.13	0.61			0.61	0.36

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
2010_Phys_FT	852	CR	04	1	961	0.75	0.13	0.45	0.42						0.42	0.45
2010_Phys_FT	852	CR	05	1	961	0.75	0.20	0.38	0.41						0.41	0.54
2010_Phys_FT	852	MC	06	1												
2010_Phys_FT	852	CR	07	2	961	0.75	0.12	0.24	0.34	0.30					0.93	0.73
2010_Phys_FT	852	CR	08	1	961	0.75	0.15	0.54	0.30						0.30	0.40
2010_Phys_FT	852	MC	09	1												
2010_Phys_FT	852	CR	10	2	961	0.75	0.15	0.20	0.37	0.27					0.91	0.69
2010_Phys_FT	852	MC	11	1												
2010_Phys_FT	852	CR	12	1	961	0.75	0.16	0.21	0.63						0.63	0.59
2010_Phys_FT	852	CR	13	1	961	0.75	0.13	0.70	0.18						0.18	0.35
2010_Phys_FT	852	MC	14	1												
2010_Phys_FT	852	CR	15	2	961	0.75	0.20	0.21	0.25	0.34					0.92	0.71
2010_Phys_FT	852	MC	16	1												
2010_Phys_FT	853	MC	01	1	954	0.79	0.04		0.08	0.36	0.47	0.06			0.47	0.45
2010_Phys_FT	853	MC	02	1	954	0.79	0.01		0.53	0.06	0.22	0.17			0.53	0.50
2010_Phys_FT	853	MC	03	1	954	0.79	0.01		0.10	0.24	0.58	0.08			0.58	0.25
2010_Phys_FT	853	MC	04	1	954	0.79	0.06		0.12	0.33	0.34	0.14			0.34	0.34
2010_Phys_FT	853	MC	05	1												

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
2010_Phys_FT	853	CR	06	2	954	0.79	0.12	0.27	0.31	0.30					0.92	0.69
2010_Phys_FT	853	CR	07	1	954	0.79	0.20	0.68	0.12						0.12	0.38
2010_Phys_FT	853	CR	08	1	954	0.79	0.27	0.29	0.45						0.45	0.62
2010_Phys_FT	853	CR	09	2	954	0.79	0.46	0.33	0.10	0.10					0.31	0.59
2010_Phys_FT	853	MC	10	1												
2010_Phys_FT	853	CR	11	1	954	0.79	0.14	0.22	0.64						0.64	0.51
2010_Phys_FT	853	CR	12	2	954	0.79	0.25	0.23	0.29	0.23					0.74	0.71
2010_Phys_FT	853	CR	13	1	954	0.79	0.30	0.32	0.38						0.38	0.66
2010_Phys_FT	853	CR	14	2	954	0.79	0.33	0.18	0.20	0.29					0.78	0.73
2010_Phys_FT	853	MC	15	1												
2010_Phys_FT	854	MC	01	1	938	0.79	0.00		0.28	0.12	0.09	0.51			0.51	0.46
2010_Phys_FT	854	MC	02	1	938	0.79	0.02		0.17	0.08	0.17	0.57			0.57	0.56
2010_Phys_FT	854	MC	03	1	938	0.79	0.05		0.17	0.38	0.30	0.09			0.38	0.32
2010_Phys_FT	854	MC	04	1												
2010_Phys_FT	854	CR	05	2	938	0.79	0.11	0.35	0.27	0.27					0.81	0.72
2010_Phys_FT	854	CR	06	1	938	0.79	0.15	0.34	0.51						0.51	0.62
2010_Phys_FT	854	CR	07	2	938	0.79	0.13	0.18	0.37	0.32					1.00	0.76
2010_Phys_FT	854	MC	08	1												

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
2010_Phys_FT	854	CR	09	2	938	0.79	0.13	0.17	0.35	0.36					1.06	0.70
2010_Phys_FT	854	MC	10	1												
2010_Phys_FT	854	CR	11	1	938	0.79	0.12	0.04	0.83						0.83	0.52
2010_Phys_FT	854	CR	12	1	938	0.79	0.14	0.67	0.20						0.20	0.42
2010_Phys_FT	854	MC	13	1												
2010_Phys_FT	854	CR	14	2	938	0.79	0.22	0.21	0.33	0.25					0.82	0.73
2010_Phys_FT	854	MC	15	1												
2010_Phys_FT	855	MC	01	1	939	0.74	0.00		0.06	0.82	0.09	0.04			0.82	0.29
2010_Phys_FT	855	MC	02	1	939	0.74	0.01		0.10	0.66	0.06	0.16			0.66	0.44
2010_Phys_FT	855	MC	03	1	939	0.74	0.00		0.06	0.08	0.10	0.75			0.75	0.35
2010_Phys_FT	855	MC	04	1	939	0.74	0.01		0.57	0.09	0.26	0.06			0.57	0.36
2010_Phys_FT	855	MC	05	1	939	0.74	0.00		0.04	0.08	0.75	0.13			0.75	0.35
2010_Phys_FT	855	MC	06	1	939	0.74	0.01		0.04	0.07	0.44	0.45			0.44	0.38
2010_Phys_FT	855	MC	07	1	939	0.74	0.02		0.26	0.16	0.11	0.45			0.45	0.39
2010_Phys_FT	855	CR	08	1	939	0.74	0.03	0.11	0.86						0.86	0.32
2010_Phys_FT	855	CR	09	1	939	0.74	0.03	0.14	0.83						0.83	0.46
2010_Phys_FT	855	CR	10	2	939	0.74	0.09	0.25	0.45	0.21					0.87	0.63
2010_Phys_FT	855	MC	11	1												

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
2010_Phys_FT	855	CR	12	1	939	0.74	0.08	0.15	0.77						0.77	0.56
2010_Phys_FT	855	CR	13	2	939	0.74	0.10	0.17	0.38	0.34					1.07	0.70
2010_Phys_FT	855	MC	14	1												
2010_Phys_FT	855	CR	15	1	939	0.74	0.09	0.45	0.46						0.46	0.42
2010_Phys_FT	855	CR	16	2	939	0.74	0.19	0.33	0.13	0.35					0.83	0.71
2010_Phys_FT	855	CR	17	1	939	0.74	0.20	0.52	0.28						0.28	0.45
2010_Phys_FT	856	MC	01	1	929	0.72	0.00		0.02	0.49	0.48	0.01			0.48	0.17
2010_Phys_FT	856	MC	02	1	929	0.72	0.00		0.05	0.09	0.21	0.64			0.64	0.53
2010_Phys_FT	856	MC	03	1	929	0.72	0.02		0.47	0.12	0.32	0.06			0.47	0.31
2010_Phys_FT	856	MC	04	1	929	0.72	0.01		0.35	0.16	0.40	0.08			0.35	0.53
2010_Phys_FT	856	MC	05	1	929	0.72	0.00		0.03	0.08	0.02	0.87			0.87	0.30
2010_Phys_FT	856	MC	06	1	929	0.72	0.00		0.07	0.24	0.13	0.55			0.55	0.37
2010_Phys_FT	856	MC	07	1	929	0.72	0.01		0.19	0.44	0.35	0.01			0.44	0.26
2010_Phys_FT	856	MC	08	1	929	0.72	0.00		0.65	0.05	0.13	0.17			0.65	0.40
2010_Phys_FT	856	MC	09	1	929	0.72	0.00		0.06	0.04	0.06	0.83			0.83	0.29
2010_Phys_FT	856	MC	10	1	929	0.72	0.00		0.02	0.06	0.46	0.46			0.46	0.25
2010_Phys_FT	856	MC	11	1	929	0.72	0.01		0.34	0.28	0.23	0.14			0.34	0.34

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
2010_Phys_FT	856	MC	12	1	929	0.72	0.02		0.21	0.14	0.12	0.51			0.51	0.46
2010_Phys_FT	856	CR	13	2	929	0.72	0.14	0.25	0.27	0.34					0.95	0.68
2010_Phys_FT	856	CR	14	1	929	0.72	0.16	0.37	0.46						0.46	0.65
2010_Phys_FT	856	CR	15	2	929	0.72	0.15	0.18	0.30	0.37					1.03	0.68
2010_Phys_FT	856	CR	16	2	929	0.72	0.17	0.26	0.43	0.13					0.70	0.58

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
2010_Phys_FT	841	MC	01	1	931	-0.4941							1.02
2010_Phys_FT	841	MC	02	1	931	-0.7477							1.13
2010_Phys_FT	841	MC	03	1	931	-0.9889							0.98
2010_Phys_FT	841	MC	04	1	931	0.1067							1.19
2010_Phys_FT	841	MC	05	1	931	-0.6589							1.00
2010_Phys_FT	841	MC	06	1	931	-0.1982							0.96
2010_Phys_FT	841	MC	07	1	931	0.1015							1.18
2010_Phys_FT	841	MC	08	1	931	-1.0054							1.13
2010_Phys_FT	841	MC	09	1	931	-0.4021							1.14
2010_Phys_FT	841	MC	10	1	931	-0.0339							1.12
2010_Phys_FT	841	MC	11	1	931	-0.7006							1.07
2010_Phys_FT	841	CR	12	1	931	-1.2032							0.94
2010_Phys_FT	841	CR	13	1	931	-1.6271							0.88
2010_Phys_FT	841	CR	14	1	931	0.7277							0.91
2010_Phys_FT	841	CR	15	1	931	0.0962							0.89
2010_Phys_FT	841	CR	16	1	931	1.3720							0.76
2010_Phys_FT	841	CR	17	1	931	-1.0553							0.94

Table 9. Partial Credit Model Item Analysis (continued)

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2010_Phys_FT	841	CR	18	1	931	1.0426							0.80
2010_Phys_FT	841	CR	19	1	931	-0.3562							0.83
2010_Phys_FT	841	CR	20	1	931	-1.0946							0.96
2010_Phys_FT	842	MC	01	1	928	-0.3078							1.06
2010_Phys_FT	842	MC	02	1	928	-2.0034							1.02
2010_Phys_FT	842	MC	03	1	928	-0.9549							1.14
2010_Phys_FT	842	MC	04	1	928	-0.9387							1.04
2010_Phys_FT	842	MC	05	1	928	-0.1880							1.02
2010_Phys_FT	842	MC	06	1	928	-2.0034							1.02
2010_Phys_FT	842	MC	07	1	928	0.2591							1.15
2010_Phys_FT	842	MC	08	1	928	0.6745							1.31
2010_Phys_FT	842	MC	09	1	928	-1.0424							1.03
2010_Phys_FT	842	MC	10	1	928	0.4236							1.06
2010_Phys_FT	842	MC	11	1	928	-0.6812							0.99
2010_Phys_FT	842	CR	12	1	928	-0.7070							0.90
2010_Phys_FT	842	CR	13	1	928	0.6116							0.87
2010_Phys_FT	842	CR	14	1	928	-0.1229							0.85
2010_Phys_FT	842	CR	15	1	928	0.6003							0.88
2010_Phys_FT	842	CR	16	1	928	0.2019							0.90

Table 9. Partial Credit Model Item Analysis (continued)

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2010_Phys_FT	842	CR	17	1	928	-0.0628							0.88
2010_Phys_FT	842	CR	18	1	928	0.5386							0.88
2010_Phys_FT	842	CR	19	1	928	2.8586							0.99
2010_Phys_FT	843	MC	01	1	939	-2.7200							0.99
2010_Phys_FT	843	MC	02	1	939	-0.1090							1.23
2010_Phys_FT	843	MC	03	1	939	1.0916							1.07
2010_Phys_FT	843	MC	04	1	939	0.2935							1.04
2010_Phys_FT	843	MC	05	1	939	-1.6154							1.05
2010_Phys_FT	843	MC	06	1	939	-1.7874							1.09
2010_Phys_FT	843	MC	07	1	939	-0.8865							1.17
2010_Phys_FT	843	MC	08	1	939	1.2743							1.21
2010_Phys_FT	843	MC	09	1	939	-0.2176							1.15
2010_Phys_FT	843	MC	10	1	939	-0.8067							1.27
2010_Phys_FT	843	MC	11	1	939	2.1467							1.26
2010_Phys_FT	843	CR	12	1	939	-0.9791							0.83
2010_Phys_FT	843	CR	13	1	939	-0.2448							0.83
2010_Phys_FT	843	CR	14	1	939	-0.9095							0.78
2010_Phys_FT	843	CR	15	1	939	1.3406							1.01
2010_Phys_FT	843	CR	16	1	939	-0.8579							0.90

Table 9. Partial Credit Model Item Analysis (continued)

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2010_Phys_FT	843	CR	17	1	939	0.8558							0.73
2010_Phys_FT	843	CR	18	1	939	-0.2231							0.70
2010_Phys_FT	843	CR	19	1	939	0.8495							0.78
2010_Phys_FT	843	CR	20	1	939	0.2151							0.78
2010_Phys_FT	844	MC	01	1	934	-1.0958							0.99
2010_Phys_FT	844	MC	02	1	934	-0.7215							1.14
2010_Phys_FT	844	MC	03	1	934	-1.1644							0.98
2010_Phys_FT	844	MC	04	1	934	-1.7733							1.09
2010_Phys_FT	844	MC	05	1	934	0.5582							1.18
2010_Phys_FT	844	MC	06	1	934	-0.4128							1.18
2010_Phys_FT	844	MC	07	1	934	-2.2773							0.88
2010_Phys_FT	844	MC	08	1	934	-1.1015							0.99
2010_Phys_FT	844	MC	09	1	934	-1.1071							1.02
2010_Phys_FT	844	MC	10	1	934	0.1429							1.10
2010_Phys_FT	844	MC	11	1	934	0.0910							1.15
2010_Phys_FT	844	CR	12	1	934	-0.5250							0.98
2010_Phys_FT	844	CR	13	1	934	-0.4586							0.99
2010_Phys_FT	844	CR	14	1	934	0.2583							0.87
2010_Phys_FT	844	CR	15	1	934	-0.4739							0.88

Table 9. Partial Credit Model Item Analysis (continued)

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2010_Phys_FT	844	CR	16	1	934	-0.3215							0.86
2010_Phys_FT	844	CR	17	1	934	0.9087							0.84
2010_Phys_FT	844	CR	18	1	934	1.1047							0.78
2010_Phys_FT	844	CR	19	1	934	-0.6279							0.99
2010_Phys_FT	845	MC	01	1	940	0.6308							1.18
2010_Phys_FT	845	MC	02	1	940	0.3173							0.96
2010_Phys_FT	845	MC	03	1	940	0.3435							1.27
2010_Phys_FT	845	MC	04	1	940	0.7628							1.13
2010_Phys_FT	845	MC	05	1	940	-0.8370							0.97
2010_Phys_FT	845	MC	06	1	940	-0.6059							1.12
2010_Phys_FT	845	MC	07	1	940	0.1985							1.11
2010_Phys_FT	845	MC	08	1	940	0.7923							0.98
2010_Phys_FT	845	MC	09	1	940	-0.3461							1.02
2010_Phys_FT	845	MC	10	1	940	-1.7623							1.06
2010_Phys_FT	845	MC	11	1	940	0.3698							1.17
2010_Phys_FT	845	CR	12	1	940	0.6591							0.92
2010_Phys_FT	845	CR	13	1	940	0.5695							1.02
2010_Phys_FT	845	CR	14	1	940	1.3824							0.92
2010_Phys_FT	845	CR	15	1	940	-0.4753							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
2010_Phys_FT	845	CR	16	1	940	0.5972							0.84
2010_Phys_FT	845	CR	17	1	940	-0.4255							0.80
2010_Phys_FT	845	CR	18	1	940	0.3278							0.79
2010_Phys_FT	845	CR	19	1	940	0.0364							0.75
2010_Phys_FT	846	MC	01	1	951	-0.4281							1.11
2010_Phys_FT	846	MC	02	1	951	-1.0160							1.03
2010_Phys_FT	846	MC	03	1	951	-0.9671							1.00
2010_Phys_FT	846	MC	04	1	951	0.3992							1.06
2010_Phys_FT	846	MC	05	1	951	-0.4533							1.29
2010_Phys_FT	846	MC	01	1	951	-0.4281							1.11
2010_Phys_FT	846	MC	06	1	951	-0.4634							1.17
2010_Phys_FT	846	MC	08	1	951	-0.5191							1.15
2010_Phys_FT	846	MC	09	1	951	-1.1950							1.01
2010_Phys_FT	846	MC	10	1	951	-1.4018							1.03
2010_Phys_FT	846	MC	11	1	951	-0.7917							0.93
2010_Phys_FT	846	MC	12	1	951	0.0470							1.01
2010_Phys_FT	846	CR	13	1	951	-1.5721							0.87
2010_Phys_FT	846	CR	14	1	951	0.4210							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
2010_Phys_FT	846	CR	15	1	951	-0.3827							0.89
2010_Phys_FT	846	CR	16	1	951	-0.1006							0.95
2010_Phys_FT	846	CR	17	1	951	-1.2411							0.87
2010_Phys_FT	846	CR	18	1	951	0.0265							0.84
2010_Phys_FT	846	CR	19	1	951	1.5383							0.84
2010_Phys_FT	846	CR	20	1	951	0.4869							0.80
2010_Phys_FT	847	MC	01	1	937	0.2484							1.23
2010_Phys_FT	847	MC	02	1	937	-1.2823							1.03
2010_Phys_FT	847	MC	03	1	937	-2.5651							1.03
2010_Phys_FT	847	MC	04	1	937	-1.3496							1.04
2010_Phys_FT	847	MC	05	1	937	-1.9807							0.88
2010_Phys_FT	847	MC	06	1	937	0.1186							1.27
2010_Phys_FT	847	MC	07	1	937	0.3087							1.17
2010_Phys_FT	847	MC	08	1	937	-0.7537							1.24
2010_Phys_FT	847	MC	09	1	937	-1.0474							1.14
2010_Phys_FT	847	MC	10	1	937	-0.8463							1.03
2010_Phys_FT	847	MC	11	1	937	-0.6089							1.39
2010_Phys_FT	847	CR	12	1	937	0.4200							0.84
2010_Phys_FT	847	CR	13	1	937	-0.5347							0.83

Table 9. Partial Credit Model Item Analysis (continued)

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2010_Phys_FT	847	CR	14	1	937	-0.6143							0.86
2010_Phys_FT	847	CR	15	1	937	0.3919							0.78
2010_Phys_FT	847	CR	16	1	937	-0.7808							0.73
2010_Phys_FT	847	CR	17	1	937	-0.1305							0.79
2010_Phys_FT	847	CR	18	1	937	-0.7429							0.76
2010_Phys_FT	847	CR	19	1	937	1.2723							0.84
2010_Phys_FT	848	MC	01	1	951	-1.3423							1.03
2010_Phys_FT	848	MC	02	1	951	-1.6405							0.96
2010_Phys_FT	848	MC	03	1	951	-1.0783							0.92
2010_Phys_FT	848	MC	04	1	951	-0.5272							1.22
2010_Phys_FT	848	MC	05	1	951	-0.2517							1.05
2010_Phys_FT	848	MC	06	1	951	-0.2617							1.09
2010_Phys_FT	848	MC	07	1	951	-0.8113							1.07
2010_Phys_FT	848	MC	08	1	951	-0.1821							1.16
2010_Phys_FT	848	MC	09	1	951	0.3900							1.11
2010_Phys_FT	848	MC	10	1	951	-1.8728							1.00
2010_Phys_FT	848	MC	11	1	951	-0.5475							1.05
2010_Phys_FT	848	CR	12	1	951	-0.8859							0.92
2010_Phys_FT	848	CR	13	1	951	0.2448							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
2010_Phys_FT	848	CR	14	1	951	-0.5576							0.86
2010_Phys_FT	848	CR	15	1	951	-1.3856							0.87
2010_Phys_FT	848	CR	16	1	951	-1.7535							0.93
2010_Phys_FT	848	CR	17	1	951	-1.2105							1.03
2010_Phys_FT	848	CR	18	1	951	0.2808							0.89
2010_Phys_FT	848	CR	19	1	951	-0.6961							0.83
2010_Phys_FT	848	CR	20	1	951	-1.0783							0.84
2010_Phys_FT	849	MC	01	1	943	-0.4273							1.19
2010_Phys_FT	849	MC	02	1	943	0.1554							1.19
2010_Phys_FT	849	MC	03	1	943	-2.2064							1.13
2010_Phys_FT	849	MC	04	1	943	-3.1734							0.85
2010_Phys_FT	849	MC	05	1	943	-0.5118							1.17
2010_Phys_FT	849	MC	06	1	943	-2.1973							0.86
2010_Phys_FT	849	MC	07	1	943	-1.8849							0.91
2010_Phys_FT	849	MC	08	1	943	-0.5383							1.09
2010_Phys_FT	849	MC	09	1	943	-0.6290							1.08
2010_Phys_FT	849	MC	10	1	943	-0.5595							1.03
2010_Phys_FT	849	MC	11	1	943	-0.8469							1.08
2010_Phys_FT	849	CR	12	1	943	0.2789							0.92

Table 9. Partial Credit Model Item Analysis (continued)

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2010_Phys_FT	849	CR	13	1	943	-0.0296							0.90
2010_Phys_FT	849	CR	14	1	943	1.5574							0.98
2010_Phys_FT	849	CR	15	1	943	-0.6022							0.88
2010_Phys_FT	849	CR	16	1	943	0.7733							0.99
2010_Phys_FT	849	CR	17	1	943	-0.4116							0.80
2010_Phys_FT	849	CR	18	1	943	-0.0086							0.82
2010_Phys_FT	849	CR	19	1	943	-0.4063							0.97
2010_Phys_FT	850	MC	01	1	954	-1.2639							0.99
2010_Phys_FT	850	MC	02	1	954	-0.7273							1.12
2010_Phys_FT	850	MC	03	1	954	-1.9505							0.96
2010_Phys_FT	850	MC	04	1	954	-0.1979							1.16
2010_Phys_FT	850	MC	05	1	954	-1.1072							1.00
2010_Phys_FT	850	MC	06	1	954	-0.4623							1.08
2010_Phys_FT	850	MC	07	1	954	-0.9626							0.97
2010_Phys_FT	850	MC	08	1	954	0.2800							1.04
2010_Phys_FT	850	MC	09	1	954	0.4643							1.24
2010_Phys_FT	850	MC	10	1	954	-0.2171							1.12
2010_Phys_FT	850	MC	11	1	954	0.1705							1.20
2010_Phys_FT	850	CR	12	1	954	-0.7423							0.90

Table 9. Partial Credit Model Item Analysis (continued)

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2010_Phys_FT	850	CR	13	1	954	-1.1236							0.95
2010_Phys_FT	850	CR	14	1	954	-0.7724							0.83
2010_Phys_FT	850	CR	15	1	954	-0.6035							0.81
2010_Phys_FT	850	CR	16	1	954	0.9896							0.91
2010_Phys_FT	850	CR	17	1	954	0.7185							0.87
2010_Phys_FT	850	CR	18	1	954	0.8624							0.91
2010_Phys_FT	850	CR	19	1	954	2.9294							0.88
2010_Phys_FT	850	CR	20	1	954	1.5381							0.87
2010_Phys_FT	851	MC	01	1	948	-0.5659							1.19
2010_Phys_FT	851	MC	02	1	948	1.2512							1.02
2010_Phys_FT	851	MC	03	1	948	0.3228							1.03
2010_Phys_FT	851	MC	04	1	948	-0.5250							1.08
2010_Phys_FT	851	MC	05	1	948	0.8472							1.23
2010_Phys_FT	851	MC	06	1									
2010_Phys_FT	851	CR	07	1	948	-1.0817							1.10
2010_Phys_FT	851	CR	08	1	948	-0.6122							0.96
2010_Phys_FT	851	CR	09	2	948	0.6676	-0.9880	0.9880					1.04
2010_Phys_FT	851	MC	10	1									
2010_Phys_FT	851	MC	11	1									

Table 9. Partial Credit Model Item Analysis (continued)

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2010_Phys_FT	851	CR	12	1	948	-0.0270							0.86
2010_Phys_FT	851	CR	13	2	948	0.5688	-0.1121	0.1121					0.81
2010_Phys_FT	851	CR	14	1	948	0.4918							0.80
2010_Phys_FT	851	CR	15	2	948	0.4206	-0.2954	0.2954					0.90
2010_Phys_FT	851	MC	16	1									
2010_Phys_FT	852	MC	01	1	961	-0.5755							1.09
2010_Phys_FT	852	MC	02	1	961	-0.5597							1.08
2010_Phys_FT	852	MC	03	1	961	-0.8715							1.24
2010_Phys_FT	852	CR	04	1	961	0.1087							1.08
2010_Phys_FT	852	CR	05	1	961	0.1296							0.95
2010_Phys_FT	852	MC	06	1									
2010_Phys_FT	852	CR	07	2	961	-0.1314	-0.5602	0.5602					0.82
2010_Phys_FT	852	CR	08	1	961	0.7211							1.12
2010_Phys_FT	852	MC	09	1									
2010_Phys_FT	852	CR	10	2	961	-0.0824	-0.7196	0.7196					0.89
2010_Phys_FT	852	MC	11	1									
2010_Phys_FT	852	CR	12	1	961	-0.9437							0.89
2010_Phys_FT	852	CR	13	1	961	1.5281							1.08
2010_Phys_FT	852	MC	14	1									

Table 9. Partial Credit Model Item Analysis (continued)

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2010_Phys_FT	852	CR	15	2	961	-0.1176	-0.0864	0.0864					0.89
2010_Phys_FT	852	MC	16	1									
2010_Phys_FT	853	MC	01	1	954	-0.1572							1.12
2010_Phys_FT	853	MC	02	1	954	-0.4664							1.02
2010_Phys_FT	853	MC	03	1	954	-0.6989							1.39
2010_Phys_FT	853	MC	04	1	954	0.5071							1.23
2010_Phys_FT	853	MC	05	1									
2010_Phys_FT	853	CR	06	2	954	-0.1188	-0.4652	0.4652					0.94
2010_Phys_FT	853	CR	07	1	954	2.1418							1.03
2010_Phys_FT	853	CR	08	1	954	-0.0451							0.85
2010_Phys_FT	853	CR	09	2	954	1.4258	0.4280	-0.4280					0.93
2010_Phys_FT	853	MC	10	1									
2010_Phys_FT	853	CR	11	1	954	-1.0302							0.96
2010_Phys_FT	853	CR	12	2	954	0.3140	-0.4055	0.4055					0.87
2010_Phys_FT	853	CR	13	1	954	0.3032							0.79
2010_Phys_FT	853	CR	14	2	954	0.1807	0.1241	-0.1241					0.85
2010_Phys_FT	853	MC	15	1									
2010_Phys_FT	854	MC	01	1	938	-0.3227							1.18
2010_Phys_FT	854	MC	02	1	938	-0.6237							1.02

Table 9. Partial Credit Model Item Analysis (continued)

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	S6	INFIT
2010_Phys_FT	854	MC	03	1	938	0.3569							1.35
2010_Phys_FT	854	MC	04	1									
2010_Phys_FT	854	CR	05	2	938	0.1970	-0.3388	0.3388					0.90
2010_Phys_FT	854	CR	06	1	938	-0.3171							0.91
2010_Phys_FT	854	CR	07	2	938	-0.2838	-0.8186	0.8186					0.79
2010_Phys_FT	854	MC	08	1									
2010_Phys_FT	854	CR	09	2	938	-0.4355	-0.7385	0.7385					0.97
2010_Phys_FT	854	MC	10	1									
2010_Phys_FT	854	CR	11	1	938	-2.4375							0.89
2010_Phys_FT	854	CR	12	1	938	1.5361							1.09
2010_Phys_FT	854	MC	13	1									
2010_Phys_FT	854	CR	14	2	938	0.1983	-0.6544	0.6544					0.89
2010_Phys_FT	854	MC	15	1									
2010_Phys_FT	855	MC	01	1	939	-1.9774							1.13
2010_Phys_FT	855	MC	02	1	939	-0.9892							1.02
2010_Phys_FT	855	MC	03	1	939	-1.5092							1.10
2010_Phys_FT	855	MC	04	1	939	-0.5541							1.13
2010_Phys_FT	855	MC	05	1	939	-1.5291							1.10
2010_Phys_FT	855	MC	06	1	939	0.0828							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
2010_Phys_FT	855	MC	07	1	939	0.0362							1.10
2010_Phys_FT	855	CR	08	1	939	-2.2794							1.05
2010_Phys_FT	855	CR	09	1	939	-2.0356							0.90
2010_Phys_FT	855	CR	10	2	939	0.1579	-1.0211	1.0211					0.92
2010_Phys_FT	855	MC	11	1									
2010_Phys_FT	855	CR	12	1	939	-1.6452							0.82
2010_Phys_FT	855	CR	13	2	939	-0.3705	-0.6775	0.6775					0.81
2010_Phys_FT	855	MC	14	1									
2010_Phys_FT	855	CR	15	1	939	-0.0256							1.06
2010_Phys_FT	855	CR	16	2	939	0.1158	0.8084	-0.8084					0.81
2010_Phys_FT	855	CR	17	1	939	0.9521							0.98
2010_Phys_FT	856	MC	01	1	929	0.1300							1.31
2010_Phys_FT	856	MC	02	1	929	-0.7700							0.86
2010_Phys_FT	856	MC	03	1	929	-0.1200							1.12
2010_Phys_FT	856	MC	04	1	929	0.5600							0.90
2010_Phys_FT	856	MC	05	1	929	-2.1000							0.87
2010_Phys_FT	856	MC	06	1	929	-0.5000							1.05
2010_Phys_FT	856	MC	07	1	929	-0.0300							1.18
2010_Phys_FT	856	MC	08	1	929	-0.9600							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
2010_Phys_FT	856	MC	09	1	929	-1.9000							0.98
2010_Phys_FT	856	MC	10	1	929	0.0700							1.20
2010_Phys_FT	856	MC	11	1	929	0.6400							1.12
2010_Phys_FT	856	MC	12	1	929	-0.2400							0.97
2010_Phys_FT	856	CR	13	2	929	-0.1600	-0.0800	0.0800					0.83
2010_Phys_FT	856	CR	14	1	929	-0.1700							0.76
2010_Phys_FT	856	CR	15	2	929	-0.4900	-0.0300	0.0300					0.80
2010_Phys_FT	856	CR	16	2	929	0.4500	-1.0700	1.0700					0.92

Appendix C: DIF Statistics

Table 10. DIF Statistics

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
841	1	MC	-0.10	0.08	-0.03		
841	2	MC	-1.78	27.23	-0.35	C	M
841	3	MC	-0.26	0.49	-0.06		
841	4	MC	-0.54	2.65	-0.13		
841	5	MC	0.40	1.34	0.06		
841	6	MC	-0.88	5.98	-0.15		
841	7	MC	0.55	2.67	0.10		
841	8	MC	-0.54	2.40	-0.10		
841	9	MC	-0.06	0.03	-0.01		
841	10	MC	-0.07	0.04	-0.03		
841	11	MC	-1.35	15.55	-0.25	B	M
841	12	OE		14.52	0.25	BB	F
841	13	OE		1.56	0.09		
841	14	OE		9.39	0.17	BB	F
841	15	OE		5.68	0.13		
841	16	OE		1.75	-0.06		
841	17	OE		16.46	0.28	CC	F
841	18	OE		0.01	0.01		
841	19	OE		4.09	0.12		
841	20	OE		0.23	-0.01		
842	1	MC	0.53	2.54	0.11		
842	2	MC	0.26	0.34	0.03		

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
842	3	MC	0.19	0.29	0.04		
842	4	MC	-0.86	5.99	-0.15		
842	5	MC	-0.21	0.36	-0.04		
842	6	MC	-1.05	5.79	-0.15	B	M
842	7	MC	-1.10	10.60	-0.21	B	M
842	8	MC	1.38	16.04	0.26	B	F
842	9	MC	-1.16	10.31	-0.21	B	M
842	10	MC	-0.09	0.07	-0.02		
842	11	MC	-0.64	3.29	-0.11		
842	12	OE		0.19	-0.03		
842	13	OE		8.50	-0.16		
842	14	OE		0.00	0.00		
842	15	OE		0.07	0.02		
842	16	OE		17.85	0.23	BB	F
842	17	OE		12.25	0.20	BB	F
842	18	OE		4.14	0.12		
842	19	OE		1.37	0.08		
843	1	MC	-0.63	1.44	-0.06		
843	2	MC	-1.45	18.66	-0.26	B	M
843	3	MC	0.91	5.31	0.12		
843	4	MC	0.05	0.02	0.01		
843	5	MC	-0.61	2.36	-0.10		

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
843	6	MC	0.07	0.03	0.02		
843	7	MC	0.72	4.26	0.10		
843	8	MC	0.31	0.61	0.06		
843	9	MC	-0.87	6.54	-0.15		
843	10	MC	-0.66	3.83	-0.14		
843	11	MC	-0.36	0.55	-0.05		
843	12	OE		0.16	-0.01		
843	13	OE		3.30	0.10		
843	14	OE		1.72	0.07		
843	15	OE		4.78	-0.11		
843	16	OE		1.52	0.06		
843	17	OE		2.98	0.06		
843	18	OE		0.54	0.03		
843	19	OE		3.59	0.10		
843	20	OE		5.39	0.11		
844	1	MC	-1.71	20.91	-0.27	C	M
844	2	MC	0.00	0.00	-0.01		
844	3	MC	-0.31	0.68	-0.04		
844	4	MC	-0.88	4.56	-0.13		
844	5	MC	0.63	3.18	0.12		
844	6	MC	-0.88	7.24	-0.17		
844	7	MC	-1.46	7.48	-0.15	B	M
844	8	MC	-0.52	1.98	-0.10		

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
844	9	MC	0.47	1.62	0.07		
844	10	MC	-0.42	1.57	-0.10		
844	11	MC	-0.52	2.43	-0.10		
844	12	OE		21.03	0.29	CC	F
844	13	OE		5.56	0.14		
844	14	OE		8.72	0.17		
844	15	OE		8.03	0.16		
844	16	OE		0.00	-0.01		
844	17	OE		3.92	0.12		
844	18	OE		3.72	-0.11		
844	19	OE		0.20	0.04		
845	1	MC	-1.29	13.90	-0.23	B	M
845	2	MC	-0.45	1.59	-0.08		
845	3	MC	-0.55	2.81	-0.11		
845	4	MC	-0.22	0.38	-0.04		
845	5	MC	1.51	17.57	0.25	C	F
845	6	MC	-0.37	1.25	-0.07		
845	7	MC	-1.27	14.53	-0.24	B	M
845	8	MC	-1.37	13.16	-0.22	B	M
845	9	MC	-0.25	0.56	-0.06		
845	10	MC	0.04	0.01	0.01		
845	11	MC	-0.32	0.93	-0.06		
845	12	OE		0.59	0.04		

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
845	13	OE		2.23	-0.09		
845	14	OE		1.55	0.07		
845	15	OE		0.78	0.04		
845	16	OE		11.81	0.18	BB	F
845	17	OE		6.33	0.13		
845	18	OE		22.24	0.24	BB	F
845	19	OE		16.50	0.21	BB	F
846	1	MC	-0.91	7.41	-0.17		
846	2	MC	-1.93	27.63	-0.32	C	M
846	3	MC	-0.25	0.46	-0.04		
846	4	MC	-1.40	15.90	-0.23	B	M
846	5	MC	-0.23	0.56	-0.05		
846	6	MC	0.08	0.06	0.01		
846	7	MC	0.81	6.08	0.16		
846	8	MC	-0.36	1.18	-0.07		
846	9	MC	0.51	1.91	0.09		
846	10	MC	-1.40	13.32	-0.23	B	M
846	11	MC	-0.10	0.07	-0.02		
846	12	MC	0.48	1.86	0.07		
846	13	OE		0.02	-0.02		
846	14	OE		10.33	0.18	BB	F
846	15	OE		11.14	0.18	BB	F
846	16	OE		8.18	0.17		

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
846	17	OE		1.44	0.06		
846	18	OE		0.73	0.05		
846	19	OE		4.17	0.10		
846	20	OE		0.82	0.05		
847	1	MC	-1.29	14.57	-0.24	B	M
847	2	MC	-0.23	0.35	-0.04		
847	3	MC	-0.41	0.64	-0.04		
847	4	MC	-0.76	4.02	-0.12		
847	5	MC	1.70	12.48	0.20	C	F
847	6	MC	0.37	1.34	0.05		
847	7	MC	0.14	0.17	0.03		
847	8	MC	-0.68	4.28	-0.13		
847	9	MC	0.79	5.00	0.14		
847	10	MC	-0.70	3.79	-0.11		
847	11	MC	-0.11	0.12	-0.03		
847	12	OE		4.59	0.12		
847	13	OE		0.22	0.02		
847	14	OE		5.38	-0.12		
847	15	OE		1.39	0.07		
847	16	OE		0.08	0.02		
847	17	OE		9.96	0.17		
847	18	OE		4.05	0.10		
847	19	OE		0.89	-0.06		

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
848	1	MC	0.66	2.95	0.11		
848	2	MC	0.71	2.76	0.10		
848	3	MC	-0.67	2.93	-0.10		
848	4	MC	0.02	0.00	0.00		
848	5	MC	-0.99	8.63	-0.19		
848	6	MC	0.05	0.03	-0.01		
848	7	MC	-0.16	0.20	-0.03		
848	8	MC	-0.01	0.00	0.02		
848	9	MC	-1.14	11.26	-0.22	B	M
848	10	MC	0.22	0.27	0.01		
848	11	MC	-0.57	2.88	-0.11		
848	12	OE		4.12	-0.11		
848	13	OE		11.43	-0.20	BB	M
848	14	OE		0.09	0.02		
848	15	OE		1.76	0.08		
848	16	OE		10.94	0.19	BB	F
848	17	OE		8.87	0.19	BB	F
848	18	OE		11.05	0.19	BB	F
848	19	OE		2.26	0.07		
848	20	OE		0.82	0.05		
849	1	MC	-0.93	7.72	-0.17		
849	2	MC	-0.56	2.80	-0.09		
849	3	MC	0.48	1.17	0.08		

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
849	4	MC	1.55	4.53	0.12		
849	5	MC	0.81	5.85	0.15		
849	6	MC	1.92	12.29	0.19	C	F
849	7	MC	-0.87	3.71	-0.11		
849	8	MC	-3.11	77.51	-0.57	C	M
849	9	MC	-1.65	22.27	-0.29	C	M
849	10	MC	-0.46	1.74	-0.10		
849	11	MC	-0.50	2.02	-0.06		
849	12	OE		6.79	0.15		
849	13	OE		3.69	0.11		
849	14	OE		2.41	-0.08		
849	15	OE		1.93	0.06		
849	16	OE		9.66	0.18	BB	F
849	17	OE		23.36	0.23	BB	F
849	18	OE		10.40	0.15		
849	19	OE		3.77	0.12		
850	1	MC	-0.45	1.45	-0.07		
850	2	MC	0.68	4.25	0.12		
850	3	MC	-0.68	2.34	-0.08		
850	4	MC	-0.71	4.99	-0.14		
850	5	MC	1.16	9.91	0.19	B	F
850	6	MC	0.14	0.18	0.03		
850	7	MC	-0.26	0.54	-0.03		

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
850	8	MC	-0.70	4.14	-0.12		
850	9	MC	-0.08	0.06	-0.01		
850	10	MC	-0.62	3.61	-0.13		
850	11	MC	0.36	1.28	0.05		
850	12	OE		0.15	-0.02		
850	13	OE		2.55	0.09		
850	14	OE		6.04	0.13		
850	15	OE		2.90	0.09		
850	16	OE		0.04	0.01		
850	17	OE		0.08	-0.02		
850	18	OE		6.47	-0.14		
850	19	OE		0.70	0.04		
850	20	OE		0.03	0.01		
851	1	MC	-0.29	0.76	-0.05		
851	2	MC	0.54	1.67	0.08		
851	3	MC	-0.12	0.11	-0.01		
851	4	MC	-0.11	0.11	-0.03		
851	5	MC	-0.40	1.25	-0.07		
851	6	MC					
851	7	OE		5.60	-0.15		
851	8	OE		5.69	-0.14		
851	9	OE		0.14	0.03		
851	10	MC					

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
851	11	MC					
851	12	OE		7.84	0.15		
851	13	OE		15.36	0.19	BB	F
851	14	OE		0.29	0.02		
851	15	OE		13.36	0.18	BB	F
851	16	MC					
852	1	MC	-0.08	0.06	-0.02		
852	2	MC	-0.30	0.76	-0.04		
852	3	MC	0.13	0.15	0.04		
852	4	OE		0.00	0.01		
852	5	OE		0.00	0.00		
852	6	MC					
852	7	OE		15.01	0.18	BB	F
852	8	OE		3.86	0.11		
852	9	MC					
852	10	OE		6.79	0.11		
852	11	MC					
852	12	OE		1.77	-0.06		
852	13	OE		8.97	-0.18	BB	M
852	14	MC					
852	15	OE		0.28	0.01		
852	16	MC					
853	1	MC	-0.57	2.75	-0.11		

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
853	2	MC	-0.59	2.74	-0.09		
853	3	MC	0.14	0.18	0.03		
853	4	MC	-0.53	2.35	-0.10		
853	5	MC					
853	6	OE		16.89	0.21	BB	F
853	7	OE		1.39	-0.08		
853	8	OE		6.44	-0.13		
853	9	OE		1.39	-0.06		
853	10	MC					
853	11	OE		10.11	0.17		
853	12	OE		10.35	0.16		
853	13	OE		1.05	-0.05		
853	14	OE		2.27	0.07		
853	15	MC					
854	1	MC	1.01	8.50	0.18	B	F
854	2	MC	-0.68	3.39	-0.10		
854	3	MC	-0.82	5.71	-0.15		
854	4	MC					
854	5	OE		1.36	0.05		
854	6	OE		8.73	-0.17		
854	7	OE		5.60	0.11		
854	8	MC					
854	9	OE		11.12	0.17		

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
854	10	MC					
854	11	OE		0.14	-0.02		
854	12	OE		22.79	-0.29	CC	M
854	13	MC					
854	14	OE		3.96	0.09		
854	15	MC					
855	1	MC	0.19	0.18	0.02		
855	2	MC	0.27	0.56	0.04		
855	3	MC	0.37	0.91	0.07		
855	4	MC	-0.18	0.28	-0.02		
855	5	MC	-0.95	5.96	-0.17		
855	6	MC	0.27	0.60	0.05		
855	7	MC	0.88	6.38	0.17		
855	8	OE		1.22	0.08		
855	9	OE		0.00	-0.01		
855	10	OE		5.68	0.14		
855	11	MC					
855	12	OE		0.08	0.02		
855	13	OE		19.65	0.22	BB	F
855	14	MC					
855	15	OE		11.97	-0.23	BB	M
855	16	OE		0.15	0.01		
855	17	OE		17.41	-0.26	CC	M

Table 10. DIF Statistics (continued)

Form	Item	Item Type	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
856	1	MC	0.12	0.15	0.03		
856	2	MC	0.84	4.82	0.13		
856	3	MC	0.29	0.77	0.04		
856	4	MC	-0.96	5.99	-0.13		
856	5	MC	0.89	3.12	0.11		
856	6	MC	-0.10	0.09	-0.02		
856	7	MC	-0.40	1.50	-0.07		
856	8	MC	1.88	25.94	0.32	C	F

*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
856	9	MC	-1.71	14.16	-0.25	C	M
856	10	MC	-2.40	54.23	-0.50	C	M
856	11	MC	0.43	1.54	0.09		
856	12	MC	-0.35	0.98	-0.06		
856	13	OE		8.29	0.14		
856	14	OE		0.60	0.05		
856	15	OE		0.81	0.05		
856	16	OE		0.59	-0.04		

Appendix D: Operational Test Maps

Table 11. Operational Test Map for January 2010

Position	Item Type	Max Points	Mean	Point-Biserial	Rasch	S1	S2	S3	S4
1	MC	1	0.41	0.36	0.17				
2	MC	1	0.50	0.37	-0.24				
3	MC	1	0.70	0.43	-1.24				
4	MC	1	0.66	0.37	-0.98				
5	MC	1	0.45	0.32	0.01				
6	MC	1	0.38	0.28	0.34				
7	MC	1	0.64	0.30	-0.91				
8	MC	1	0.68	0.32	-1.07				
9	MC	1	0.58	0.45	-0.64				
10	MC	1	0.49	0.31	-0.20				
11	MC	1	0.68	0.28	-1.05				
12	MC	1	0.66	0.38	-0.99				
13	MC	1	0.50	0.27	-0.23				
14	MC	1	0.59	-0.01	-0.62				
15	MC	1	0.37	0.27	0.39				
16	MC	1	0.52	0.45	-0.32				
17	MC	1	0.26	0.14	1.06				
18	MC	1	0.64	0.41	-0.91				
19	MC	1	0.61	0.48	-0.72				
20	MC	1	0.35	0.33	0.57				
21	MC	1	0.55	0.25	-0.46				

Table 11. Operational Test Map for January 2010 (continued)

Position	Item Type	Max Points	Mean	Point-Biserial	Rasch	S1	S2	S3	S4
22	MC	1	0.42	0.42	0.15				
23	MC	1	0.50	0.35	-0.24				
24	MC	1	0.43	0.32	0.11				
25	MC	1	0.63	0.35	-0.81				
26	MC	1	0.82	0.29	-1.93				
27	MC	1	0.63	0.47	-0.82				
28	MC	1	0.82	0.38	-1.94				
29	MC	1	0.59	0.30	-0.65				
30	MC	1	0.52	0.29	-0.31				
31	MC	1	0.43	0.33	0.10				
32	MC	1	0.37	0.42	0.40				
33	MC	1	0.58	0.33	-0.61				
34	MC	1	0.57	0.40	-0.58				
35	MC	1	0.39	0.29	0.26				
36	MC	1	0.64	0.30	-0.91				
37	MC	1	0.47	0.36	-0.11				
38	MC	1	0.35	0.54	0.48				
39	MC	1	0.70	0.36	-1.22				
40	MC	1	0.43	0.38	0.10				
41	MC	1	0.44	0.20	0.17				
42	MC	1	0.41	0.38	0.16				
43	MC	1	0.43	0.50	0.07				

Table 11. Operational Test Map for January 2010 (continued)

Position	Item Type	Max Points	Mean	Point-Biserial	Rasch	S1	S2	S3	S4
44	MC	1	0.50	0.38	-0.22				
45	MC	1	0.43	0.52	0.12				
46	MC	1	0.70	0.46	-1.19				
47	MC	1	0.57	0.29	-0.50				
48	MC	1	0.45	0.29	-0.03				
49	MC	1	0.30	0.35	0.72				
50	MC	1	0.53	0.20	-0.32				
51	CR	1	0.32	0.56	0.63				
52	CR	1	0.12	0.39	2.03				
53	CR	2	1.31	0.57	-0.88	-0.25	0.25		
54	CR	2	1.00	0.65	-0.25	0.82	-0.82		
55	CR	1	0.44	0.38	0.06				
56	CR	2	0.97	0.61	-0.18	0.17	-0.17		
57	CR	1	0.53	0.47	-0.40				
58	CR	1	0.69	0.50	-1.12				
59	CR	1	0.62	0.36	-0.70				
60	CR	1	0.71	0.28	-1.16				
61	CR	1	0.94	0.22	-3.15				
62	CR	1	0.10	0.33	2.25				
63	CR	1	0.24	0.41	1.28				
64	CR	1	0.80	0.34	-1.74				
65	CR	1	0.33	0.36	0.73				

Table 11. Operational Test Map for January 2010 (continued)

Position	Item Type	Max Points	Mean	Point-Biserial	Rasch	S1	S2	S3	S4
66	CR	1	0.72	0.33	-1.26				
67	CR	1	0.42	0.46	0.30				
68	CR	1	0.60	0.46	-0.61				
69	CR	1	0.57	0.30	-0.56				
70	CR	1	0.64	0.40	-0.90				
71	CR	2	0.58	0.59	0.74	-0.53	0.53		
72	CR	1	0.36	0.55	0.43				
73	CR	1	0.48	0.49	-0.13				
74	CR	2	0.99	0.64	-0.23	0.20	-0.20		
75	CR	2	0.79	0.68	0.16	0.47	-0.47		
76	CR	1	0.51	0.53	-0.30				
77	CR	1	0.16	0.42	1.80				
78	CR	1	0.36	0.48	0.55				
79	CR	1	0.50	0.41	-0.16				

Table 12. Operational Test Map for June 2010

Position	Item Type	Max Points	Mean	Point-Biserial	Rasch	S1	S2	S3	S4
1	MC	1	0.40	0.35	0.22				
2	MC	1	0.72	0.38	-1.28				
3	MC	1	0.63	0.44	-0.88				
4	MC	1	0.59	0.45	-0.69				
5	MC	1	0.56	0.30	-0.52				
6	MC	1	0.36	0.45	0.41				
7	MC	1	0.48	0.34	-0.14				
8	MC	1	0.47	0.42	-0.07				
9	MC	1	0.49	0.29	-0.19				
10	MC	1	0.45	0.30	-0.01				
11	MC	1	0.60	0.36	-0.70				
12	MC	1	0.55	0.38	-0.45				
13	MC	1	0.48	0.40	-0.12				
14	MC	1	0.73	0.28	-1.35				
15	MC	1	0.30	0.18	0.74				
16	MC	1	0.85	0.27	-2.15				
17	MC	1	0.51	0.45	-0.31				
18	MC	1	0.74	0.42	-1.46				
19	MC	1	0.80	0.16	-1.80				
20	MC	1	0.71	0.22	-1.13				
21	MC	1	0.48	0.40	-0.16				

Table 12. Operational Test Map for June 2010 (continued)

Position	Item Type	Max Points	Mean	Point-Biserial	Rasch	S1	S2	S3	S4
22	MC	1	0.43	0.22	0.08				
23	MC	1	0.46	0.27	0.08				
24	MC	1	0.56	0.22	-0.50				
25	MC	1	0.65	0.35	-0.92				
26	MC	1	0.36	0.24	0.43				
27	MC	1	0.63	0.45	-0.85				
28	MC	1	0.68	0.39	-1.12				
29	MC	1	0.80	0.38	-1.86				
30	MC	1	0.59	0.37	-0.66				
31	MC	1	0.45	0.31	-0.02				
32	MC	1	0.35	0.36	0.49				
33	MC	1	0.38	0.42	0.40				
34	MC	1	0.81	0.30	-1.83				
35	MC	1	0.61	0.35	-0.75				
36	MC	1	0.50	0.22	-0.25				
37	MC	1	0.38	0.14	0.43				
38	MC	1	0.47	0.44	-0.08				
39	MC	1	0.43	0.44	0.08				
40	MC	1	0.59	0.43	-0.62				
41	MC	1	0.77	0.36	-1.59				
42	MC	1	0.31	0.36	0.67				
43	MC	1	0.50	0.32	-0.22				

Table 12. Operational Test Map for June 2010 (continued)

Position	Item Type	Max Points	Mean	Point-Biserial	Rasch	S1	S2	S3	S4
44	MC	1	0.35	0.33	0.50				
45	MC	1	0.63	0.39	-0.74				
46	MC	1	0.28	0.12	0.82				
47	MC	1	0.52	0.44	-0.33				
48	MC	1	0.50	0.45	-0.26				
49	MC	1	0.24	0.37	1.11				
50	MC	1	0.54	0.32	-0.44				
51	CR	1	0.77	0.46	-1.62				
52	CR	1	0.26	0.55	1.00				
53	CR	2	0.41	0.68	1.02	0.60	-0.60		
54	CR	1	0.30	0.40	0.81				
55	CR	2	1.31	0.52	-0.92	-0.56	0.56		
56	CR	2	0.68	0.53	0.71	-1.28	1.28		
57	CR	1	0.78	0.41	-1.53				
58	CR	2	0.84	0.51	0.23	-0.45	0.45		
59	CR	1	0.29	0.48	0.82				
60	CR	2	0.93	0.61	-0.08	-0.19	0.19		
61	CR	1	0.82	0.45	-1.96				
62	CR	1	0.63	0.38	-0.84				
63	CR	1	0.58	0.47	-0.61				
64	CR	2	0.89	0.58	0.01	-0.32	0.32		
65	CR	2	1.03	0.64	-0.30	0.80	-0.80		

Table 12. Operational Test Map for June 2010 (continued)

Position	Item Type	Max Points	Mean	Point-Biserial	Rasch	S1	S2	S3	S4
66	CR	1	0.32	0.56	0.66				
67	CR	1	0.33	0.61	0.63				
68	CR	2	0.85	0.66	0.01	1.28	-1.28		
69	CR	1	0.37	0.48	0.36				
70	CR	2	0.88	0.67	-0.04	1.26	-1.26		
71	CR	1	0.49	0.51	-0.18				
72	CR	1	0.47	0.53	-0.09				
73	CR	1	0.36	0.57	0.44				
74	CR	2	0.46	0.68	0.90	0.29	-0.29		
75	CR	1	0.29	0.59	0.82				

Appendix E: Scoring Tables

Table 13. Scoring Table for January 2010

Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score
0	-5.711	0.370		22	-1.383	35.518		44	-0.156	60.840		66	1.120	82.401
1	-4.985	2.167		23	-1.316	36.826		45	-0.105	61.860		67	1.195	83.377
2	-4.259	4.209		24	-1.251	38.134		46	-0.054	62.863		68	1.274	84.309
3	-3.823	6.094		25	-1.188	39.429		47	-0.003	63.880		69	1.356	85.306
4	-3.506	7.978		26	-1.127	40.646		48	0.049	64.902		70	1.443	86.280
5	-3.255	9.778		27	-1.067	41.840		49	0.101	65.922		71	1.534	87.202
6	-3.047	11.556		28	-1.008	43.048		50	0.153	66.941		72	1.630	88.182
7	-2.867	13.326		29	-0.950	44.248		51	0.206	67.865		73	1.732	89.125
8	-2.709	15.012		30	-0.893	45.457		52	0.259	68.881		74	1.842	90.015
9	-2.567	16.651		31	-0.837	46.657		53	0.313	69.917		75	1.960	90.972
10	-2.438	18.289		32	-0.782	47.780		54	0.368	70.837		76	2.089	91.919
11	-2.320	19.839		33	-0.728	48.944		55	0.424	71.865		77	2.231	92.853
12	-2.210	21.351		34	-0.674	50.089		56	0.480	72.789		78	2.389	93.777
13	-2.107	22.865		35	-0.621	51.189		57	0.537	73.796		79	2.568	94.675
14	-2.011	24.369		36	-0.568	52.304		58	0.596	74.740		80	2.776	95.559
15	-1.919	25.883		37	-0.516	53.404		59	0.655	75.667		81	3.026	96.486
16	-1.833	27.305		38	-0.464	54.520		60	0.716	76.665		82	3.341	97.367
17	-1.750	28.720		39	-0.412	55.565		61	0.779	77.622		83	3.774	98.188
18	-1.671	30.110		40	-0.361	56.641		62	0.843	78.611		84	4.497	99.227
19	-1.595	31.508		41	-0.310	57.680		63	0.909	79.558		85	5.220	100.338
20	-1.522	32.897		42	-0.258	58.802		64	0.977	80.548				
21	-1.451	34.229		43	-0.207	59.824		65	1.047	81.454				

Table 14. Scoring Table for June 2010

Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score		Raw Score	Ability	Scale Score
0	-5.631	0.566		22	-1.332	36.511		44	-0.104	61.880		66	1.127	82.497
1	-4.914	2.351		23	-1.265	37.850		45	-0.054	62.863		67	1.198	83.412
2	-4.197	4.447		24	-1.199	39.202		46	-0.004	63.860		68	1.272	84.284
3	-3.768	6.400		25	-1.136	40.469		47	0.047	64.863		69	1.349	85.224
4	-3.456	8.307		26	-1.073	41.720		48	0.097	65.843		70	1.430	86.150
5	-3.209	10.141		27	-1.012	42.966		49	0.148	66.843		71	1.515	87.004
6	-3.003	11.960		28	-0.953	44.186		50	0.199	67.744		72	1.606	87.949
7	-2.825	13.762		29	-0.895	45.414		51	0.251	68.727		73	1.701	88.871
8	-2.668	15.459		30	-0.837	46.657		52	0.303	69.727		74	1.804	89.704
9	-2.527	17.128		31	-0.781	47.800		53	0.355	70.617		75	1.915	90.618
10	-2.398	18.813		32	-0.726	48.989		54	0.408	71.578		76	2.036	91.539
11	-2.280	20.370		33	-0.671	50.151		55	0.462	72.487		77	2.169	92.449
12	-2.169	21.935		34	-0.617	51.272		56	0.516	73.432		78	2.318	93.361
13	-2.066	23.491		35	-0.564	52.388		57	0.572	74.359		79	2.487	94.270
14	-1.969	25.050		36	-0.511	53.510		58	0.628	75.236		80	2.685	95.182
15	-1.877	26.580		37	-0.459	54.627		59	0.685	76.167		81	2.924	96.121
16	-1.789	28.046		38	-0.408	55.643		60	0.743	77.070		82	3.227	97.100
17	-1.705	29.508		39	-0.357	56.727		61	0.803	78.003		83	3.649	98.013
18	-1.625	30.948		40	-0.306	57.760		62	0.864	78.906		84	4.359	99.015
19	-1.548	32.399		41	-0.255	58.868		63	0.927	79.830		85	5.069	100.106
20	-1.474	33.801		42	-0.205	59.863		64	0.991	80.730				
21	-1.402	35.154		43	-0.154	60.880		65	1.058	81.593				