# New York State Regents Examination in English

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

# Technical Report

Prepared for the New York State Education Department
by Pearson

**March 2012**

# Copyright

## Table of Contents

## List of Tables

# Section I: Introduction

## PURPOSE

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

# Section II: Field Test Analysis

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

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

**Table 1. Need/Resource Capacity Category Definitions**

| Need/Resource Capacity (N/RC) Category | Definition |
|---|---|
| High N/RC Districts: New York City | New York City |
| Large Cities | Buffalo, Rochester, Syracuse, Yonkers |
| Urban-Suburban | Districts at or above the 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 fewer 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 CLEANUP

Field test forms contained multiple-choice and constructed-response item types. Response data were contained in two separate files. The multiple-choice data file contained 12,523 student records, and the constructed-response data file contained 5,655 student records. The two files were combined by merging the multiple-choice records and the constructed-response records by unique test booklet number. After the exclusion rules were applied, the resulting field test data file contained 11,863 records.

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

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

## CLASSICAL ANALYSIS

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

$$x = t + e$$

In other words, error is associated with measuring a student's true score. For example, the choice of test items or the administration conditions might influence student responses, making a student's observed score higher or lower than the student's true score. The error is considered random. After repeated administrations, the mean of the error scores is virtually zero. Thus, a student's observed score is expected to equal his or her true score. This expectation is expressed as follows:

$$E(x) = t$$

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

## Item Difficulty

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

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

## Point-Biserial Correlation

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

Table 2 presents a summary of the classical item analysis for each of the field test forms. The first three columns identify the form number, the number of students who took each form, and the number of items on each field test form, respectively. The remaining columns are divided into two sections (i.e., item difficulty and point-biserial correlations). Recall that for constructed-response items, item means were divided by the maximum number of points possible in order to place them in the same metric as the multiple-choice items. For all items except three, item difficulties were equal to or below 0.90. With respect to the point-biserial correlations, none were below 0.25.

**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 |
| 601 | 512 | 8 | 1 | 6 | 1 | 0 | 5 | 3 |
| 602 | 594 | 8 | 0 | 8 | 0 | 0 | 6 | 2 |
| 603 | 598 | 8 | 0 | 6 | 2 | 0 | 4 | 4 |
| 604 | 644 | 8 | 0 | 8 | 0 | 0 | 8 | 0 |
| 605 | 531 | 8 | 0 | 8 | 0 | 0 | 7 | 1 |
| 606 | 827 | 12 | 1 | 11 | 0 | 0 | 2 | 10 |
| 607 | 827 | 12 | 0 | 12 | 0 | 0 | 2 | 10 |
| 608 | 834 | 12 | 0 | 12 | 0 | 0 | 7 | 5 |
| 609 | 797 | 12 | 1 | 11 | 0 | 0 | 2 | 10 |
| 610 | 799 | 12 | 0 | 12 | 0 | 0 | 4 | 8 |
| 611 | 501 | 7 | 0 | 7 | 0 | 0 | 4 | 3 |
| 612 | 499 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| 613 | 504 | 7 | 0 | 7 | 0 | 0 | 5 | 2 |
| 614 | 495 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| 615 | 484 | 7 | 0 | 7 | 0 | 0 | 4 | 3 |
| 616 | 500 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| 617 | 479 | 7 | 0 | 7 | 0 | 0 | 3 | 4 |
| 618 | 490 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| 619 | 473 | 7 | 0 | 7 | 0 | 0 | 4 | 3 |
| 620 | 475 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| N3 | 11863 | 10 | 1 | 9 | 0 | 0 | 5 | 5 |

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

## Test Reliability

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

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

Table 3 contains the internal consistency statistics for all of the field test forms. These statistics ranged from 0.66 to 0.85 and are based solely on the items in the individual field test forms. It is expected that these statistics associated with the operational tests would be greater because there are more items on the operational test forms.

**Table 3. Test and Scoring Reliability**

| Form Number | Test Reliability | Scoring Reliability |
|---|---|---|
| 601 | 0.78 | n/a |
| 602 | 0.80 | n/a |
| 603 | 0.76 | n/a |
| 604 | 0.76 | n/a |
| 605 | 0.81 | n/a |
| 606 | 0.83 | n/a |
| 607 | 0.85 | n/a |
| 608 | 0.81 | n/a |
| 609 | 0.85 | n/a |
| 610 | 0.85 | n/a |
| 611 | 0.80 | 0.80 |
| 612 | 0.67 | 0.76 |
| 613 | 0.79 | 0.75 |
| 614 | 0.66 | 0.86 |
| 615 | 0.79 | 0.84 |
| 616 | 0.71 | 0.84 |
| 617 | 0.82 | 0.77 |
| 618 | 0.68 | 0.86 |
| 619 | 0.81 | 0.72 |
| 620 | 0.68 | 0.84 |
| N3 | 0.70 | n/a |

Scoring Reliability

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

As an overall measure of scoring reliability, the Pearson Correlation Coefficient between the first and second scores for each of the constructed-response items was computed. This statistic is often used as an overall indicator of scoring reliability, and generally ranges from 0 to near 1. Table 3 contains the results from these analyses in

the column headed “Scoring Reliability.” The correlations ranged from 0.72 to 0.86, indicating high scoring reliability.

Inter-rater Agreement

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

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

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

| | | | Difference (First Read minus Second Read) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Form** | **Item** | **Score Points** | **−3** | **−2** | **−1** | **0** | **1** | **2** | **3** |
| 611 | 06 | 2 | 0.00 | 0.00 | 0.16 | 0.77 | 0.06 | 0.00 | 0.00 |
| 611 | 07 | 2 | 0.00 | 0.00 | 0.05 | 0.90 | 0.05 | 0.00 | 0.00 |
| 612 | Es | 6 | 0.00 | 0.02 | 0.22 | 0.60 | 0.16 | 0.00 | 0.00 |
| 613 | 06 | 2 | 0.00 | 0.00 | 0.11 | 0.78 | 0.10 | 0.00 | 0.00 |
| 613 | 07 | 2 | 0.00 | 0.00 | 0.05 | 0.80 | 0.15 | 0.00 | 0.00 |
| 614 | Es | 6 | 0.00 | 0.00 | 0.13 | 0.66 | 0.21 | 0.00 | 0.00 |
| 615 | 06 | 2 | 0.00 | 0.00 | 0.08 | 0.86 | 0.05 | 0.00 | 0.00 |
| 615 | 07 | 2 | 0.00 | 0.00 | 0.11 | 0.82 | 0.07 | 0.00 | 0.00 |
| 616 | Es | 6 | 0.00 | 0.00 | 0.17 | 0.72 | 0.12 | 0.00 | 0.00 |
| 617 | 06 | 2 | 0.00 | 0.00 | 0.07 | 0.87 | 0.07 | 0.00 | 0.00 |
| 617 | 07 | 2 | 0.00 | 0.00 | 0.08 | 0.81 | 0.11 | 0.00 | 0.00 |
| 618 | Es | 6 | 0.00 | 0.00 | 0.11 | 0.65 | 0.25 | 0.00 | 0.00 |
| 619 | 06 | 2 | 0.00 | 0.00 | 0.09 | 0.79 | 0.12 | 0.00 | 0.00 |
| 619 | 07 | 2 | 0.00 | 0.00 | 0.12 | 0.78 | 0.10 | 0.00 | 0.00 |
| 620 | Es | 6 | 0.00 | 0.01 | 0.13 | 0.71 | 0.15 | 0.00 | 0.00 |

Table 5 contains additional summary information regarding the first and second reads. In the fifth column, the percent of exact matches between the first and second scores is provided. “Adj.” is the percentage of differences with a magnitude of 1. “Total” is the sum of the two prior columns and contains values between 97.8% 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 | Intraclass Correla-tion | Wt. Kappa |
| 611 | 06 | 2 | 79 | 77.2 | 22.8 | 100.0 | 1.4 | 1.5 | 0.63 | 0.64 | 0.73 | 0.64 |
| 611 | 07 | 2 | 79 | 89.9 | 10.1 | 100.0 | 1.4 | 1.4 | 0.64 | 0.64 | 0.87 | 0.84 |
| 612 | Es | 6 | 90 | 60.0 | 37.8 | 97.8 | 3.0 | 3.1 | 0.96 | 0.99 | 0.76 | 0.59 |
| 613 | 06 | 2 | 87 | 78.2 | 21.8 | 100.0 | 1.5 | 1.5 | 0.57 | 0.57 | 0.66 | 0.61 |
| 613 | 07 | 2 | 87 | 80.5 | 19.5 | 100.0 | 1.4 | 1.3 | 0.71 | 0.69 | 0.81 | 0.73 |
| 614 | Es | 6 | 86 | 66.3 | 33.7 | 100.0 | 3.3 | 3.2 | 1.12 | 1.10 | 0.86 | 0.71 |
| 615 | 06 | 2 | 74 | 86.5 | 13.5 | 100.0 | 1.5 | 1.5 | 0.62 | 0.62 | 0.83 | 0.78 |
| 615 | 07 | 2 | 74 | 82.4 | 17.6 | 100.0 | 1.3 | 1.3 | 0.73 | 0.74 | 0.84 | 0.77 |
| 616 | Es | 6 | 95 | 71.6 | 28.4 | 100.0 | 3.3 | 3.3 | 0.94 | 0.92 | 0.84 | 0.71 |
| 617 | 06 | 2 | 75 | 86.7 | 13.3 | 100.0 | 1.5 | 1.5 | 0.55 | 0.55 | 0.78 | 0.76 |
| 617 | 07 | 2 | 75 | 81.3 | 18.7 | 100.0 | 1.5 | 1.4 | 0.64 | 0.62 | 0.76 | 0.71 |
| 618 | Es | 6 | 85 | 64.7 | 35.3 | 100.0 | 3.4 | 3.3 | 1.11 | 1.12 | 0.86 | 0.70 |
| 619 | 06 | 2 | 77 | 79.2 | 20.8 | 100.0 | 1.5 | 1.5 | 0.58 | 0.58 | 0.68 | 0.63 |
| 619 | 07 | 2 | 77 | 77.9 | 22.1 | 100.0 | 1.4 | 1.5 | 0.66 | 0.66 | 0.74 | 0.67 |
| 620 | Es | 6 | 87 | 71.3 | 27.6 | 98.9 | 3.2 | 3.2 | 0.98 | 1.03 | 0.84 | 0.71 |

* Adj. = difference of 1

## Constructed-Response Item Means and Standard Deviations

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

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

## Intraclass Correlation

The intraclass correlation was computed for each item. This correlation is an estimate of the reliability of scoring based on an average of the first and second read scores. Correlations greater than 0.60 are considered very strong because they explain more than one-third of the variance in scores. All items had intraclass correlations

greater than or equal to 0.66 (See Table 5). Consistent with other information provided in the table, these values indicate a very high level of scoring reliability.

Weighted Kappa

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

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

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

The results found in Table 5 show a high degree of consistency between the first and second reads. The Weighted Kappa statistics ranged from 0.59 to 0.84, which in all cases indicates good-to-excellent reproducibility.

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

## ITEM RESPONSE THEORY (IRT) STATISTICS

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

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

The PCM provides an overall difficulty estimate for each item. Specifically for constructed-response items when there are several points possible, individual estimates of difficulty for each of the possible score points are also calculated (i.e., step values). Each step value represents the difficulty of a student receiving a particular score point,

given that he or she has already received the prior score point. For example, if a 3-point item had step values of −1.0, 1.0, and 0.0, one could say that it is relatively easy to obtain a score of 1. However, it is much more difficult to obtain a 2, given the student has the ability to score a 1, because the difference in difficulty between a 1 and a 2 is much greater than the difference between a 0 and a 1. Also, the difference between a 2 and a 3 is not as great as the difference between a 1 and a 2. Thus, with this example, a small step is needed to go from a 0 to a 1, a large step is needed to move from a 1 to a 2, and a moderate step is needed to proceed from a 2 to a 3.

### Item Calibration

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

Rasch item difficulty values generally range from −3.00 to +3.00. An item with a Rasch difficulty greater than +2.0 is considered very difficult and should be examined carefully. If the item is measuring an important concept that students are having difficulty with, then the item can be useful. However, if the item is measuring a trivial concept or is written in a confusing manner, then it might not be appropriate to use on an operational test form. Likewise, any item with a Rasch difficulty less than −2.0 is considered very easy and usually provides little information regarding student achievement. The vast majority of test items should range between −2.0 and +2.0. This range represents approximately two standard deviations around the average difficulty of 0. Thus, one would expect that, based on chance, roughly 5% of the items will fall outside of that range and, therefore, these are items that should be closely examined for content.

### Item Fit Evaluation

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

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

**Table 6. Partial Credit Model Item Analysis**

| | | | Rasch | | | INFIT | | |
|---|---|---|---|---|---|---|---|---|
| **Form** | **N-Count** | **No. of Items** | **<−2.0** | **−2.0 to 2.0** | **>2.0** | **<−0.70** | **−0.70 to 1.30** | **>1.30** |
| 601 | 512 | 8 | 2 | 6 | 0 | 0 | 7 | 1 |
| 602 | 594 | 8 | 1 | 7 | 0 | 0 | 8 | 0 |
| 603 | 598 | 8 | 2 | 6 | 0 | 0 | 8 | 0 |
| 604 | 644 | 8 | 1 | 7 | 0 | 0 | 8 | 0 |
| 605 | 531 | 8 | 1 | 7 | 0 | 0 | 8 | 0 |
| 606 | 827 | 12 | 0 | 12 | 0 | 0 | 12 | 0 |
| 607 | 827 | 12 | 0 | 12 | 0 | 0 | 12 | 0 |
| 608 | 834 | 12 | 0 | 12 | 0 | 0 | 12 | 0 |
| 609 | 797 | 12 | 0 | 12 | 0 | 0 | 12 | 0 |
| 610 | 799 | 12 | 0 | 12 | 0 | 0 | 12 | 0 |
| 611 | 501 | 7 | 0 | 7 | 0 | 0 | 7 | 0 |
| 612 | 499 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 613 | 504 | 7 | 1 | 6 | 0 | 0 | 7 | 0 |
| 614 | 495 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 615 | 484 | 7 | 0 | 7 | 0 | 0 | 7 | 0 |
| 616 | 500 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 617 | 479 | 7 | 1 | 6 | 0 | 0 | 7 | 0 |
| 618 | 490 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 619 | 473 | 7 | 0 | 7 | 0 | 0 | 7 | 0 |
| 620 | 475 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| N3 | 11863 | 10 | 0 | 10 | 0 | 0 | 10 | 0 |

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

## DIFFERENTIAL ITEM FUNCTIONING (DIF) STATISTICS

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

For the multiple-choice items, the Mantel-Haenszel Delta (MHD) DIF statistics were computed (Dorans & Holland, 1992) to classify test items in three levels of DIF for each comparison: negligible DIF (A), moderate DIF (B), and large DIF (C). An item was flagged if it exhibited a B or C category of DIF, using the following rules derived from National Assessment of Educational Progress (NAEP) guidelines (Allen, Carlson, & Zalanak, 1999):

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

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

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

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

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

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

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

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

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

## Section III: Equating Procedure

The 2011 field test administration for the New York State Regents Examination in English consisted of 20 field test forms numbered 601–620 and an anchor form labeled N3. Each student participating in the field test was administered the anchor form and one of the 20 field test forms. The field 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.

Each field test form was administered with the anchor form. The field test data were arranged in an incomplete data matrix so that the anchor items were in each data line along with the unique items for each field test form. Items not appearing on the field test form are left blank and treated as not administered when item parameters are calibrated. The entire data set was then calibrated using WINSTEPS and applying the Partial Credit Model. In this calibration, the anchor items were fixed to their 2010 bank values. This places all of the item parameters on the bank scale.

Table 7 is a sample matrix equating design for three of the forms where "X" represents the presence of data and "—" represents the absence of data.

**Table 7. Incomplete Data Matrix Structure**

| Anchor | Form 601 | Form 602 | Form 603 |
|---|---|---|---|
| X | X | — | — |
| X | — | X | — |
| X | — | — | X |

An item-stability check is performed on the anchor items by examining displacement values. The displacement values indicate the difference between the bank values for the anchor items and the difficulty values for those items as if they were not fixed to the bank values. After fixing all of the items to their 2010 bank values, any item with a displacement value with a magnitude greater than 0.30 was no longer fixed, and the test form was reanalyzed. If more than one item had a displacement value with a magnitude greater than 0.30, then the item with the largest displacement was freed and the test form was reanalyzed. In a stepwise fashion, this procedure was repeated until all remaining fixed anchor items had displacements with magnitudes less than or equal to 0.30.

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

## Section IV: Scaling of Operational Test Forms

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

In order to limit wide fluctuations of raw scores that correspond to scale scores of 65 and 85 across administrations, the average Rasch item difficulty for the operational test is considered. For this examination, an average Rasch difficulty of approximately 0.451 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 June 2011 and August 2011 operational test forms. Four steps are taken in order to produce these tables and the 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 target was determined by the New York State Education Department to be January 2011. 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. One modification that is made is that for 6-point items, a constant based on existing bank values is used in place of the field test parameters. The number of points on the test form (i.e., raw score) expected across student ability levels is based on the difficulty of the items on the form. Thus, given a particular student ability level (i.e., theta), if the points are more difficult to earn on the new test than the points on the January 2011 test, the number of points expected of this student on the new test will be less than the number of points expected of this student on the baseline form.

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

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

## References


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

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

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

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

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

# Appendix A: Classical Item Analysis

**Table 8. Classical Item Analysis**

| Test | Form | Type | Item | Max | N-Count | Alpha | B | M0 | M1 | M2 | M3 | M4 | M5 | M6 | Mean | Point-Biserial |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 601 | MC | 01 | 1 | 512 | 0.78 | 0.03 | | 0.81 | 0.12 | 0.03 | 0.01 | | | 0.81 | 0.33 |
| 2011_Engl_FT | 601 | MC | 02 | 1 | 512 | 0.78 | 0.03 | | 0.01 | 0.04 | 0.92 | 0.00 | | | 0.92 | 0.29 |
| 2011_Engl_FT | 601 | MC | 03 | 1 | 512 | 0.78 | 0.03 | | 0.01 | 0.88 | 0.02 | 0.06 | | | 0.88 | 0.50 |
| 2011_Engl_FT | 601 | MC | 04 | 1 | 512 | 0.78 | 0.03 | | 0.03 | 0.12 | 0.14 | 0.68 | | | 0.68 | 0.57 |
| 2011_Engl_FT | 601 | MC | 05 | 1 | 512 | 0.78 | 0.04 | | 0.05 | 0.80 | 0.09 | 0.02 | | | 0.80 | 0.51 |
| 2011_Engl_FT | 601 | MC | 06 | 1 | 512 | 0.78 | 0.04 | | 0.68 | 0.12 | 0.10 | 0.06 | | | 0.68 | 0.54 |
| 2011_Engl_FT | 601 | MC | 07 | 1 | 512 | 0.78 | 0.04 | | 0.42 | 0.11 | 0.41 | 0.01 | | | 0.41 | 0.25 |
| 2011_Engl_FT | 601 | MC | 08 | 1 | 512 | 0.78 | 0.04 | | 0.02 | 0.14 | 0.09 | 0.72 | | | 0.72 | 0.28 |
| 2011_Engl_FT | 602 | MC | 01 | 1 | 594 | 0.80 | 0.01 | | 0.02 | 0.03 | 0.08 | 0.86 | | | 0.86 | 0.31 |
| 2011_Engl_FT | 602 | MC | 02 | 1 | 594 | 0.80 | 0.01 | | 0.86 | 0.09 | 0.03 | 0.01 | | | 0.86 | 0.39 |
| 2011_Engl_FT | 602 | MC | 03 | 1 | 594 | 0.80 | 0.01 | | 0.01 | 0.82 | 0.11 | 0.05 | | | 0.82 | 0.49 |
| 2011_Engl_FT | 602 | MC | 04 | 1 | 594 | 0.80 | 0.01 | | 0.09 | 0.06 | 0.82 | 0.02 | | | 0.82 | 0.53 |
| 2011_Engl_FT | 602 | MC | 05 | 1 | 594 | 0.80 | 0.01 | | 0.11 | 0.08 | 0.13 | 0.66 | | | 0.66 | 0.56 |
| 2011_Engl_FT | 602 | MC | 06 | 1 | 594 | 0.80 | 0.01 | | 0.90 | 0.03 | 0.04 | 0.02 | | | 0.90 | 0.48 |
| 2011_Engl_FT | 602 | MC | 07 | 1 | 594 | 0.80 | 0.02 | | 0.07 | 0.51 | 0.30 | 0.11 | | | 0.51 | 0.47 |
| 2011_Engl_FT | 602 | MC | 08 | 1 | 594 | 0.80 | 0.01 | | 0.10 | 0.07 | 0.02 | 0.79 | | | 0.79 | 0.44 |
| 2011_Engl_FT | 603 | MC | 01 | 1 | 598 | 0.76 | 0.00 | | 0.03 | 0.05 | 0.92 | 0.00 | | | 0.92 | 0.38 |
| 2011_Engl_FT | 603 | MC | 02 | 1 | 598 | 0.76 | 0.01 | | 0.93 | 0.04 | 0.01 | 0.01 | | | 0.93 | 0.40 |
| 2011_Engl_FT | 603 | MC | 03 | 1 | 598 | 0.76 | 0.01 | | 0.07 | 0.02 | 0.13 | 0.77 | | | 0.77 | 0.33 |
| 2011_Engl_FT | 603 | MC | 04 | 1 | 598 | 0.76 | 0.01 | | 0.21 | 0.73 | 0.03 | 0.02 | | | 0.73 | 0.56 |
| 2011_Engl_FT | 603 | MC | 05 | 1 | 598 | 0.76 | 0.01 | | 0.06 | 0.02 | 0.85 | 0.07 | | | 0.85 | 0.52 |
| 2011_Engl_FT | 603 | MC | 06 | 1 | 598 | 0.76 | 0.02 | | 0.75 | 0.01 | 0.17 | 0.05 | | | 0.75 | 0.51 |

**Table 8. Classical Item Analysis *(continued)***

| Test | Form | Type | Item | Max | N-Count | Alpha | B | M0 | M1 | M2 | M3 | M4 | M5 | M6 | Mean | Point-Biserial |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 603 | MC | 07 | 1 | 598 | 0.76 | 0.02 | | 0.03 | 0.04 | 0.08 | 0.83 | | | 0.83 | 0.51 |
| 2011_Engl_FT | 603 | MC | 08 | 1 | 598 | 0.76 | 0.01 | | 0.09 | 0.76 | 0.10 | 0.05 | | | 0.76 | 0.33 |
| 2011_Engl_FT | 604 | MC | 01 | 1 | 644 | 0.76 | 0.01 | | 0.11 | 0.71 | 0.15 | 0.03 | | | 0.71 | 0.40 |
| 2011_Engl_FT | 604 | MC | 02 | 1 | 644 | 0.76 | 0.00 | | 0.83 | 0.12 | 0.03 | 0.02 | | | 0.83 | 0.41 |
| 2011_Engl_FT | 604 | MC | 03 | 1 | 644 | 0.76 | 0.01 | | 0.77 | 0.10 | 0.05 | 0.08 | | | 0.77 | 0.42 |
| 2011_Engl_FT | 604 | MC | 04 | 1 | 644 | 0.76 | 0.00 | | 0.05 | 0.03 | 0.79 | 0.12 | | | 0.79 | 0.36 |
| 2011_Engl_FT | 604 | MC | 05 | 1 | 644 | 0.76 | 0.00 | | 0.08 | 0.02 | 0.01 | 0.88 | | | 0.88 | 0.45 |
| 2011_Engl_FT | 604 | MC | 06 | 1 | 644 | 0.76 | 0.01 | | 0.09 | 0.04 | 0.16 | 0.70 | | | 0.70 | 0.41 |
| 2011_Engl_FT | 604 | MC | 07 | 1 | 644 | 0.76 | 0.01 | | 0.42 | 0.52 | 0.03 | 0.02 | | | 0.52 | 0.32 |
| 2011_Engl_FT | 604 | MC | 08 | 1 | 644 | 0.76 | 0.00 | | 0.14 | 0.09 | 0.72 | 0.05 | | | 0.72 | 0.37 |
| 2011_Engl_FT | 605 | MC | 01 | 1 | 531 | 0.81 | 0.01 | | 0.19 | 0.06 | 0.71 | 0.03 | | | 0.71 | 0.45 |
| 2011_Engl_FT | 605 | MC | 02 | 1 | 531 | 0.81 | 0.01 | | 0.05 | 0.02 | 0.29 | 0.63 | | | 0.63 | 0.48 |
| 2011_Engl_FT | 605 | MC | 03 | 1 | 531 | 0.81 | 0.01 | | 0.02 | 0.90 | 0.04 | 0.02 | | | 0.90 | 0.37 |
| 2011_Engl_FT | 605 | MC | 04 | 1 | 531 | 0.81 | 0.01 | | 0.70 | 0.09 | 0.10 | 0.10 | | | 0.70 | 0.45 |
| 2011_Engl_FT | 605 | MC | 05 | 1 | 531 | 0.81 | 0.01 | | 0.05 | 0.05 | 0.14 | 0.75 | | | 0.75 | 0.44 |
| 2011_Engl_FT | 605 | MC | 06 | 1 | 531 | 0.81 | 0.01 | | 0.10 | 0.07 | 0.77 | 0.05 | | | 0.77 | 0.50 |
| 2011_Engl_FT | 605 | MC | 07 | 1 | 531 | 0.81 | 0.01 | | 0.80 | 0.06 | 0.02 | 0.10 | | | 0.80 | 0.61 |
| 2011_Engl_FT | 605 | MC | 08 | 1 | 531 | 0.81 | 0.01 | | 0.10 | 0.73 | 0.10 | 0.06 | | | 0.73 | 0.36 |
| 2011_Engl_FT | 606 | MC | 11 | 1 | 827 | 0.83 | 0.01 | | 0.65 | 0.14 | 0.13 | 0.08 | | | 0.65 | 0.51 |
| 2011_Engl_FT | 606 | MC | 12 | 1 | 827 | 0.83 | 0.01 | | 0.09 | 0.09 | 0.78 | 0.03 | | | 0.78 | 0.54 |
| 2011_Engl_FT | 606 | MC | 13 | 1 | 827 | 0.83 | 0.01 | | 0.07 | 0.78 | 0.06 | 0.07 | | | 0.78 | 0.54 |
| 2011_Engl_FT | 606 | MC | 14 | 1 | 827 | 0.83 | 0.01 | | 0.86 | 0.04 | 0.06 | 0.02 | | | 0.86 | 0.57 |
| 2011_Engl_FT | 606 | MC | 15 | 1 | 827 | 0.83 | 0.02 | | 0.13 | 0.08 | 0.71 | 0.06 | | | 0.71 | 0.51 |

**Table 8. Classical Item Analysis *(continued)***

| Test | Form | Type | Item | Max | N-Count | Alpha | B | M0 | M1 | M2 | M3 | M4 | M5 | M6 | Mean | Point-Biserial |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 606 | MC | 16 | 1 | 827 | 0.83 | 0.02 | | 0.04 | 0.16 | 0.06 | 0.71 | | | 0.71 | 0.51 |
| 2011_Engl_FT | 606 | MC | 17 | 1 | 827 | 0.83 | 0.03 | | 0.21 | 0.59 | 0.12 | 0.06 | | | 0.59 | 0.49 |
| 2011_Engl_FT | 606 | MC | 18 | 1 | 827 | 0.83 | 0.03 | | 0.75 | 0.07 | 0.08 | 0.07 | | | 0.75 | 0.54 |
| 2011_Engl_FT | 606 | MC | 19 | 1 | 827 | 0.83 | 0.04 | | 0.05 | 0.15 | 0.15 | 0.62 | | | 0.62 | 0.54 |
| 2011_Engl_FT | 606 | MC | 20 | 1 | 827 | 0.83 | 0.04 | | 0.14 | 0.18 | 0.49 | 0.15 | | | 0.49 | 0.59 |
| 2011_Engl_FT | 606 | MC | 21 | 1 | 827 | 0.83 | 0.04 | | 0.14 | 0.23 | 0.52 | 0.07 | | | 0.52 | 0.42 |
| 2011_Engl_FT | 606 | MC | 22 | 1 | 827 | 0.83 | 0.04 | | 0.06 | 0.81 | 0.06 | 0.05 | | | 0.81 | 0.53 |
| 2011_Engl_FT | 607 | MC | 11 | 1 | 827 | 0.85 | 0.01 | | 0.14 | 0.04 | 0.13 | 0.68 | | | 0.68 | 0.50 |
| 2011_Engl_FT | 607 | MC | 12 | 1 | 827 | 0.85 | 0.01 | | 0.11 | 0.55 | 0.14 | 0.18 | | | 0.55 | 0.35 |
| 2011_Engl_FT | 607 | MC | 13 | 1 | 827 | 0.85 | 0.01 | | 0.06 | 0.06 | 0.84 | 0.03 | | | 0.84 | 0.58 |
| 2011_Engl_FT | 607 | MC | 14 | 1 | 827 | 0.85 | 0.02 | | 0.05 | 0.03 | 0.89 | 0.01 | | | 0.89 | 0.56 |
| 2011_Engl_FT | 607 | MC | 15 | 1 | 827 | 0.85 | 0.02 | | 0.80 | 0.05 | 0.04 | 0.08 | | | 0.80 | 0.56 |
| 2011_Engl_FT | 607 | MC | 16 | 1 | 827 | 0.85 | 0.02 | | 0.07 | 0.25 | 0.11 | 0.54 | | | 0.54 | 0.54 |
| 2011_Engl_FT | 607 | MC | 17 | 1 | 827 | 0.85 | 0.03 | | 0.10 | 0.10 | 0.72 | 0.05 | | | 0.72 | 0.55 |
| 2011_Engl_FT | 607 | MC | 18 | 1 | 827 | 0.85 | 0.03 | | 0.76 | 0.08 | 0.04 | 0.09 | | | 0.76 | 0.58 |
| 2011_Engl_FT | 607 | MC | 19 | 1 | 827 | 0.85 | 0.03 | | 0.10 | 0.08 | 0.07 | 0.72 | | | 0.72 | 0.56 |
| 2011_Engl_FT | 607 | MC | 20 | 1 | 827 | 0.85 | 0.04 | | 0.23 | 0.63 | 0.04 | 0.07 | | | 0.63 | 0.60 |
| 2011_Engl_FT | 607 | MC | 21 | 1 | 827 | 0.85 | 0.05 | | 0.05 | 0.10 | 0.06 | 0.74 | | | 0.74 | 0.61 |
| 2011_Engl_FT | 607 | MC | 22 | 1 | 827 | 0.85 | 0.05 | | 0.05 | 0.21 | 0.64 | 0.06 | | | 0.64 | 0.60 |
| 2011_Engl_FT | 608 | MC | 11 | 1 | 834 | 0.81 | 0.01 | | 0.64 | 0.04 | 0.28 | 0.03 | | | 0.64 | 0.40 |
| 2011_Engl_FT | 608 | MC | 12 | 1 | 834 | 0.81 | 0.02 | | 0.01 | 0.03 | 0.13 | 0.82 | | | 0.82 | 0.38 |
| 2011_Engl_FT | 608 | MC | 13 | 1 | 834 | 0.81 | 0.02 | | 0.11 | 0.04 | 0.68 | 0.15 | | | 0.68 | 0.48 |
| 2011_Engl_FT | 608 | MC | 14 | 1 | 834 | 0.81 | 0.02 | | 0.10 | 0.74 | 0.04 | 0.10 | | | 0.74 | 0.56 |

**Table 8. Classical Item Analysis *(continued)***

| Test | Form | Type | Item | Max | N-Count | Alpha | B | M0 | M1 | M2 | M3 | M4 | M5 | M6 | Mean | Point-Biserial |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 608 | MC | 15 | 1 | 834 | 0.81 | 0.02 | | 0.68 | 0.08 | 0.06 | 0.15 | | | 0.68 | 0.52 |
| 2011_Engl_FT | 608 | MC | 16 | 1 | 834 | 0.81 | 0.02 | | 0.14 | 0.26 | 0.51 | 0.06 | | | 0.51 | 0.37 |
| 2011_Engl_FT | 608 | MC | 17 | 1 | 834 | 0.81 | 0.03 | | 0.63 | 0.26 | 0.03 | 0.05 | | | 0.63 | 0.58 |
| 2011_Engl_FT | 608 | MC | 18 | 1 | 834 | 0.81 | 0.03 | | 0.26 | 0.06 | 0.13 | 0.53 | | | 0.53 | 0.41 |
| 2011_Engl_FT | 608 | MC | 19 | 1 | 834 | 0.81 | 0.03 | | 0.07 | 0.04 | 0.83 | 0.03 | | | 0.83 | 0.48 |
| 2011_Engl_FT | 608 | MC | 20 | 1 | 834 | 0.81 | 0.03 | | 0.08 | 0.07 | 0.05 | 0.77 | | | 0.77 | 0.56 |
| 2011_Engl_FT | 608 | MC | 21 | 1 | 834 | 0.81 | 0.04 | | 0.06 | 0.71 | 0.07 | 0.12 | | | 0.71 | 0.56 |
| 2011_Engl_FT | 608 | MC | 22 | 1 | 834 | 0.81 | 0.04 | | 0.71 | 0.05 | 0.05 | 0.15 | | | 0.71 | 0.45 |
| 2011_Engl_FT | 609 | MC | 11 | 1 | 797 | 0.85 | 0.01 | | 0.05 | 0.09 | 0.10 | 0.75 | | | 0.75 | 0.61 |
| 2011_Engl_FT | 609 | MC | 12 | 1 | 797 | 0.85 | 0.01 | | 0.11 | 0.04 | 0.78 | 0.06 | | | 0.78 | 0.61 |
| 2011_Engl_FT | 609 | MC | 13 | 1 | 797 | 0.85 | 0.01 | | 0.07 | 0.82 | 0.08 | 0.02 | | | 0.82 | 0.55 |
| 2011_Engl_FT | 609 | MC | 14 | 1 | 797 | 0.85 | 0.01 | | 0.66 | 0.11 | 0.16 | 0.05 | | | 0.66 | 0.52 |
| 2011_Engl_FT | 609 | MC | 15 | 1 | 797 | 0.85 | 0.02 | | 0.06 | 0.16 | 0.69 | 0.07 | | | 0.69 | 0.48 |
| 2011_Engl_FT | 609 | MC | 16 | 1 | 797 | 0.85 | 0.02 | | 0.47 | 0.11 | 0.04 | 0.38 | | | 0.47 | 0.48 |
| 2011_Engl_FT | 609 | MC | 17 | 1 | 797 | 0.85 | 0.02 | | 0.10 | 0.81 | 0.05 | 0.02 | | | 0.81 | 0.54 |
| 2011_Engl_FT | 609 | MC | 18 | 1 | 797 | 0.85 | 0.03 | | 0.75 | 0.08 | 0.11 | 0.04 | | | 0.75 | 0.56 |
| 2011_Engl_FT | 609 | MC | 19 | 1 | 797 | 0.85 | 0.03 | | 0.07 | 0.12 | 0.76 | 0.03 | | | 0.76 | 0.54 |
| 2011_Engl_FT | 609 | MC | 20 | 1 | 797 | 0.85 | 0.03 | | 0.07 | 0.05 | 0.08 | 0.78 | | | 0.78 | 0.65 |
| 2011_Engl_FT | 609 | MC | 21 | 1 | 797 | 0.85 | 0.04 | | 0.22 | 0.12 | 0.56 | 0.06 | | | 0.56 | 0.54 |
| 2011_Engl_FT | 609 | MC | 22 | 1 | 797 | 0.85 | 0.04 | | 0.60 | 0.18 | 0.13 | 0.05 | | | 0.60 | 0.56 |
| 2011_Engl_FT | 610 | MC | 11 | 1 | 799 | 0.85 | 0.01 | | 0.04 | 0.03 | 0.84 | 0.08 | | | 0.84 | 0.50 |
| 2011_Engl_FT | 610 | MC | 12 | 1 | 799 | 0.85 | 0.01 | | 0.08 | 0.05 | 0.03 | 0.83 | | | 0.83 | 0.51 |
| 2011_Engl_FT | 610 | MC | 13 | 1 | 799 | 0.85 | 0.01 | | 0.10 | 0.79 | 0.05 | 0.05 | | | 0.79 | 0.60 |

**Table 8. Classical Item Analysis *(continued)***

| Test | Form | Type | Item | Max | N-Count | Alpha | B | M0 | M1 | M2 | M3 | M4 | M5 | M6 | Mean | Point-Biserial |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 610 | MC | 14 | 1 | 799 | 0.85 | 0.02 | | 0.72 | 0.06 | 0.07 | 0.14 | | | 0.72 | 0.44 |
| 2011_Engl_FT | 610 | MC | 15 | 1 | 799 | 0.85 | 0.02 | | 0.20 | 0.08 | 0.66 | 0.05 | | | 0.66 | 0.62 |
| 2011_Engl_FT | 610 | MC | 16 | 1 | 799 | 0.85 | 0.02 | | 0.32 | 0.51 | 0.08 | 0.07 | | | 0.51 | 0.46 |
| 2011_Engl_FT | 610 | MC | 17 | 1 | 799 | 0.85 | 0.02 | | 0.79 | 0.04 | 0.13 | 0.02 | | | 0.79 | 0.64 |
| 2011_Engl_FT | 610 | MC | 18 | 1 | 799 | 0.85 | 0.02 | | 0.06 | 0.06 | 0.04 | 0.82 | | | 0.82 | 0.54 |
| 2011_Engl_FT | 610 | MC | 19 | 1 | 799 | 0.85 | 0.02 | | 0.08 | 0.07 | 0.73 | 0.11 | | | 0.73 | 0.57 |
| 2011_Engl_FT | 610 | MC | 20 | 1 | 799 | 0.85 | 0.02 | | 0.04 | 0.87 | 0.03 | 0.05 | | | 0.87 | 0.54 |
| 2011_Engl_FT | 610 | MC | 21 | 1 | 799 | 0.85 | 0.03 | | 0.68 | 0.08 | 0.18 | 0.03 | | | 0.68 | 0.55 |
| 2011_Engl_FT | 610 | MC | 22 | 1 | 799 | 0.85 | 0.04 | | 0.14 | 0.10 | 0.70 | 0.03 | | | 0.70 | 0.50 |
| 2011_Engl_FT | 611 | MC | 01 | 1 | 501 | 0.80 | 0.00 | | 0.13 | 0.76 | 0.06 | 0.05 | | | 0.76 | 0.44 |
| 2011_Engl_FT | 611 | MC | 02 | 1 | 501 | 0.80 | 0.01 | | 0.08 | 0.20 | 0.62 | 0.09 | | | 0.62 | 0.42 |
| 2011_Engl_FT | 611 | MC | 03 | 1 | 501 | 0.80 | 0.00 | | 0.07 | 0.05 | 0.04 | 0.83 | | | 0.83 | 0.53 |
| 2011_Engl_FT | 611 | MC | 04 | 1 | 501 | 0.80 | 0.00 | | 0.07 | 0.08 | 0.09 | 0.76 | | | 0.76 | 0.27 |
| 2011_Engl_FT | 611 | MC | 05 | 1 | 501 | 0.80 | 0.00 | | 0.75 | 0.04 | 0.09 | 0.12 | | | 0.75 | 0.41 |
| 2011_Engl_FT | 611 | CR | 06 | 2 | 501 | 0.80 | 0.01 | 0.09 | 0.39 | 0.51 | | | | | 1.41 | 0.59 |
| 2011_Engl_FT | 611 | CR | 07 | 2 | 501 | 0.80 | 0.06 | 0.11 | 0.37 | 0.47 | | | | | 1.30 | 0.63 |
| 2011_Engl_FT | 612 | CR | Es | 6 | 499 | 0.67 | 0.00 | 0.01 | 0.07 | 0.17 | 0.40 | 0.27 | 0.07 | 0.01 | 3.08 | 0.68 |
| 2011_Engl_FT | 613 | MC | 01 | 1 | 504 | 0.79 | 0.00 | | 0.05 | 0.03 | 0.83 | 0.09 | | | 0.83 | 0.36 |
| 2011_Engl_FT | 613 | MC | 02 | 1 | 504 | 0.79 | 0.01 | | 0.78 | 0.12 | 0.03 | 0.06 | | | 0.78 | 0.42 |
| 2011_Engl_FT | 613 | MC | 03 | 1 | 504 | 0.79 | 0.00 | | 0.07 | 0.10 | 0.05 | 0.77 | | | 0.77 | 0.41 |
| 2011_Engl_FT | 613 | MC | 04 | 1 | 504 | 0.79 | 0.01 | | 0.53 | 0.17 | 0.20 | 0.09 | | | 0.53 | 0.42 |
| 2011_Engl_FT | 613 | MC | 05 | 1 | 504 | 0.79 | 0.00 | | 0.03 | 0.90 | 0.03 | 0.03 | | | 0.90 | 0.42 |
| 2011_Engl_FT | 613 | CR | 06 | 2 | 504 | 0.79 | 0.01 | 0.10 | 0.44 | 0.46 | | | | | 1.36 | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 613 | CR | 07 | 2 | 504 | 0.79 | 0.07 | 0.12 | 0.35 | 0.47 | | | | | 1.28 | 0.62 |
| 2011_Engl_FT | 614 | CR | Es | 6 | 495 | 0.66 | 0.00 | 0.02 | 0.08 | 0.10 | 0.39 | 0.31 | 0.07 | 0.02 | 3.18 | 0.74 |
| 2011_Engl_FT | 615 | MC | 01 | 1 | 484 | 0.79 | 0.00 | | 0.09 | 0.84 | 0.02 | 0.05 | | | 0.84 | 0.40 |
| 2011_Engl_FT | 615 | MC | 02 | 1 | 484 | 0.79 | 0.00 | | 0.04 | 0.05 | 0.05 | 0.85 | | | 0.85 | 0.44 |
| 2011_Engl_FT | 615 | MC | 03 | 1 | 484 | 0.79 | 0.00 | | 0.69 | 0.06 | 0.05 | 0.21 | | | 0.69 | 0.51 |
| 2011_Engl_FT | 615 | MC | 04 | 1 | 484 | 0.79 | 0.00 | | 0.08 | 0.07 | 0.79 | 0.07 | | | 0.79 | 0.39 |
| 2011_Engl_FT | 615 | MC | 05 | 1 | 484 | 0.79 | 0.00 | | 0.04 | 0.04 | 0.87 | 0.05 | | | 0.87 | 0.39 |
| 2011_Engl_FT | 615 | CR | 06 | 2 | 484 | 0.79 | 0.02 | 0.12 | 0.36 | 0.50 | | | | | 1.36 | 0.60 |
| 2011_Engl_FT | 615 | CR | 07 | 2 | 484 | 0.79 | 0.07 | 0.12 | 0.37 | 0.43 | | | | | 1.24 | 0.59 |
| 2011_Engl_FT | 616 | CR | Es | 6 | 500 | 0.71 | 0.00 | 0.01 | 0.07 | 0.11 | 0.33 | 0.37 | 0.10 | 0.00 | 3.30 | 0.74 |
| 2011_Engl_FT | 617 | MC | 01 | 1 | 479 | 0.82 | 0.01 | | 0.03 | 0.19 | 0.76 | 0.01 | | | 0.76 | 0.51 |
| 2011_Engl_FT | 617 | MC | 02 | 1 | 479 | 0.82 | 0.01 | | 0.23 | 0.60 | 0.02 | 0.14 | | | 0.60 | 0.48 |
| 2011_Engl_FT | 617 | MC | 03 | 1 | 479 | 0.82 | 0.01 | | 0.03 | 0.03 | 0.04 | 0.90 | | | 0.90 | 0.45 |
| 2011_Engl_FT | 617 | MC | 04 | 1 | 479 | 0.82 | 0.00 | | 0.79 | 0.12 | 0.04 | 0.05 | | | 0.79 | 0.35 |
| 2011_Engl_FT | 617 | MC | 05 | 1 | 479 | 0.82 | 0.01 | | 0.05 | 0.81 | 0.09 | 0.04 | | | 0.81 | 0.53 |
| 2011_Engl_FT | 617 | CR | 06 | 2 | 479 | 0.82 | 0.01 | 0.06 | 0.48 | 0.44 | | | | | 1.36 | 0.58 |
| 2011_Engl_FT | 617 | CR | 07 | 2 | 479 | 0.82 | 0.06 | 0.09 | 0.41 | 0.44 | | | | | 1.29 | 0.65 |
| 2011_Engl_FT | 618 | CR | Es | 6 | 490 | 0.68 | 0.00 | 0.02 | 0.08 | 0.10 | 0.36 | 0.33 | 0.10 | 0.02 | 3.28 | 0.71 |
| 2011_Engl_FT | 619 | MC | 01 | 1 | 473 | 0.81 | 0.00 | | 0.83 | 0.13 | 0.01 | 0.02 | | | 0.83 | 0.49 |
| 2011_Engl_FT | 619 | MC | 02 | 1 | 473 | 0.81 | 0.00 | | 0.08 | 0.87 | 0.02 | 0.03 | | | 0.87 | 0.51 |
| 2011_Engl_FT | 619 | MC | 03 | 1 | 473 | 0.81 | 0.00 | | 0.17 | 0.01 | 0.02 | 0.80 | | | 0.80 | 0.49 |
| 2011_Engl_FT | 619 | MC | 04 | 1 | 473 | 0.81 | 0.00 | | 0.15 | 0.01 | 0.80 | 0.03 | | | 0.80 | 0.49 |

**Table 8. Classical Item Analysis *(continued)***

| Test | Form | Type | Item | Max | N-Count | Alpha | B | M0 | M1 | M2 | M3 | M4 | M5 | M6 | Mean | Point-Biserial |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 619 | MC | 05 | 1 | 473 | 0.81 | 0.00 | | 0.75 | 0.19 | 0.05 | 0.01 | | | 0.75 | 0.41 |
| 2011_Engl_FT | 619 | CR | 06 | 2 | 473 | 0.81 | 0.01 | 0.04 | 0.44 | 0.51 | | | | | 1.46 | 0.58 |
| 2011_Engl_FT | 619 | CR | 07 | 2 | 473 | 0.81 | 0.07 | 0.09 | 0.37 | 0.47 | | | | | 1.31 | 0.61 |
| 2011_Engl_FT | 620 | CR | Es | 6 | 475 | 0.68 | 0.00 | 0.01 | 0.06 | 0.14 | 0.35 | 0.36 | 0.07 | 0.00 | 3.21 | 0.71 |
| 2011_Engl_FT | N3 | MC | 01 | 1 | 11,863 | 0.70 | 0.01 | | 0.16 | 0.03 | 0.04 | 0.77 | | | 0.77 | 0.47 |
| 2011_Engl_FT | N3 | MC | 02 | 1 | 11,863 | 0.70 | 0.02 | | 0.25 | 0.24 | 0.42 | 0.07 | | | 0.42 | 0.42 |
| 2011_Engl_FT | N3 | MC | 03 | 1 | 11,863 | 0.70 | 0.01 | | 0.05 | 0.83 | 0.03 | 0.07 | | | 0.83 | 0.55 |
| 2011_Engl_FT | N3 | MC | 04 | 1 | 11,863 | 0.70 | 0.02 | | 0.75 | 0.11 | 0.05 | 0.07 | | | 0.75 | 0.48 |
| 2011_Engl_FT | N3 | MC | 05 | 1 | 11,863 | 0.70 | 0.02 | | 0.07 | 0.68 | 0.18 | 0.06 | | | 0.68 | 0.50 |
| 2011_Engl_FT | N3 | MC | 06 | 1 | 11,863 | 0.70 | 0.03 | | 0.08 | 0.16 | 0.23 | 0.50 | | | 0.50 | 0.55 |
| 2011_Engl_FT | N3 | MC | 07 | 1 | 11,863 | 0.70 | 0.02 | | 0.64 | 0.13 | 0.12 | 0.09 | | | 0.64 | 0.60 |
| 2011_Engl_FT | N3 | MC | 08 | 1 | 11,863 | 0.70 | 0.03 | | 0.25 | 0.10 | 0.12 | 0.51 | | | 0.51 | 0.46 |
| 2011_Engl_FT | N3 | MC | 09 | 1 | 11,863 | 0.70 | 0.03 | | 0.20 | 0.63 | 0.04 | 0.11 | | | 0.63 | 0.60 |
| 2011_Engl_FT | N3 | MC | 10 | 1 | 11,863 | 0.70 | 0.03 | | 0.16 | 0.08 | 0.58 | 0.16 | | | 0.58 | 0.61 |

# Appendix B: Partial Credit Model Item Analysis

## Table 9. Partial Credit Model Item Analysis

| Test | Form | Type | Item | Max | N-Count | RID | S1 | S2 | S3 | S4 | S5 | S6 | INFIT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 601 | MC | 01 | 1 | 512 | −1.3957 | | | | | | | 1.09 |
| 2011_Engl_FT | 601 | MC | 02 | 1 | 512 | −2.5001 | | | | | | | 0.99 |
| 2011_Engl_FT | 601 | MC | 03 | 1 | 512 | −2.0655 | | | | | | | 0.83 |
| 2011_Engl_FT | 601 | MC | 04 | 1 | 512 | −0.4783 | | | | | | | 0.88 |
| 2011_Engl_FT | 601 | MC | 05 | 1 | 512 | −1.2874 | | | | | | | 0.89 |
| 2011_Engl_FT | 601 | MC | 06 | 1 | 512 | −0.4551 | | | | | | | 0.93 |
| 2011_Engl_FT | 601 | MC | 07 | 1 | 512 | 0.9825 | | | | | | | 1.33 |
| 2011_Engl_FT | 601 | MC | 08 | 1 | 512 | −0.7301 | | | | | | | 1.22 |
| 2011_Engl_FT | 602 | MC | 01 | 1 | 594 | −1.6756 | | | | | | | 1.11 |
| 2011_Engl_FT | 602 | MC | 02 | 1 | 594 | −1.7437 | | | | | | | 1.00 |
| 2011_Engl_FT | 602 | MC | 03 | 1 | 594 | −1.3227 | | | | | | | 0.95 |
| 2011_Engl_FT | 602 | MC | 04 | 1 | 594 | −1.3801 | | | | | | | 0.88 |
| 2011_Engl_FT | 602 | MC | 05 | 1 | 594 | −0.2568 | | | | | | | 0.94 |
| 2011_Engl_FT | 602 | MC | 06 | 1 | 594 | −2.1722 | | | | | | | 0.83 |
| 2011_Engl_FT | 602 | MC | 07 | 1 | 594 | 0.6458 | | | | | | | 1.06 |
| 2011_Engl_FT | 602 | MC | 08 | 1 | 594 | −1.1321 | | | | | | | 1.02 |
| 2011_Engl_FT | 603 | MC | 01 | 1 | 598 | −2.5502 | | | | | | | 0.99 |
| 2011_Engl_FT | 603 | MC | 02 | 1 | 598 | −2.8731 | | | | | | | 0.95 |
| 2011_Engl_FT | 603 | MC | 03 | 1 | 598 | −1.1090 | | | | | | | 1.15 |
| 2011_Engl_FT | 603 | MC | 04 | 1 | 598 | −0.8033 | | | | | | | 0.88 |
| 2011_Engl_FT | 603 | MC | 05 | 1 | 598 | −1.7581 | | | | | | | 0.87 |
| 2011_Engl_FT | 603 | MC | 06 | 1 | 598 | −0.9689 | | | | | | | 0.94 |
| 2011_Engl_FT | 603 | MC | 07 | 1 | 598 | −1.5871 | | | | | | | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 603 | MC | 08 | 1 | 598 | −1.0147 | | | | | | | 1.14 |
| 2011_Engl_FT | 604 | MC | 01 | 1 | 644 | −0.8665 | | | | | | | 1.06 |
| 2011_Engl_FT | 604 | MC | 02 | 1 | 644 | −1.6938 | | | | | | | 0.97 |
| 2011_Engl_FT | 604 | MC | 03 | 1 | 644 | −1.2256 | | | | | | | 1.00 |
| 2011_Engl_FT | 604 | MC | 04 | 1 | 644 | −1.3676 | | | | | | | 1.06 |
| 2011_Engl_FT | 604 | MC | 05 | 1 | 644 | −2.1829 | | | | | | | 0.88 |
| 2011_Engl_FT | 604 | MC | 06 | 1 | 644 | −0.8011 | | | | | | | 1.04 |
| 2011_Engl_FT | 604 | MC | 07 | 1 | 644 | 0.1571 | | | | | | | 1.22 |
| 2011_Engl_FT | 604 | MC | 08 | 1 | 644 | −0.9236 | | | | | | | 1.10 |
| 2011_Engl_FT | 605 | MC | 01 | 1 | 531 | −0.8957 | | | | | | | 1.06 |
| 2011_Engl_FT | 605 | MC | 02 | 1 | 531 | −0.4163 | | | | | | | 1.04 |
| 2011_Engl_FT | 605 | MC | 03 | 1 | 531 | −2.5956 | | | | | | | 0.97 |
| 2011_Engl_FT | 605 | MC | 04 | 1 | 531 | −0.8477 | | | | | | | 1.06 |
| 2011_Engl_FT | 605 | MC | 05 | 1 | 531 | −1.1331 | | | | | | | 1.04 |
| 2011_Engl_FT | 605 | MC | 06 | 1 | 531 | −1.3067 | | | | | | | 0.96 |
| 2011_Engl_FT | 605 | MC | 07 | 1 | 531 | −1.5671 | | | | | | | 0.78 |
| 2011_Engl_FT | 605 | MC | 08 | 1 | 531 | −1.0185 | | | | | | | 1.15 |
| 2011_Engl_FT | 606 | MC | 11 | 1 | 827 | 0.0565 | | | | | | | 0.97 |
| 2011_Engl_FT | 606 | MC | 12 | 1 | 827 | −0.7837 | | | | | | | 0.89 |
| 2011_Engl_FT | 606 | MC | 13 | 1 | 827 | −0.8192 | | | | | | | 0.87 |
| 2011_Engl_FT | 606 | MC | 14 | 1 | 827 | −1.4661 | | | | | | | 0.79 |
| 2011_Engl_FT | 606 | MC | 15 | 1 | 827 | −0.3127 | | | | | | | 0.95 |
| 2011_Engl_FT | 606 | MC | 16 | 1 | 827 | −0.3431 | | | | | | | 0.95 |
| 2011_Engl_FT | 606 | MC | 17 | 1 | 827 | 0.4030 | | | | | | | 1.01 |

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

| Test | Form | Type | Item | Max | N-Count | RID | S1 | S2 | S3 | S4 | S5 | S6 | INFIT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 606 | MC | 18 | 1 | 827 | −0.5963 | | | | | | | 0.90 |
| 2011_Engl_FT | 606 | MC | 19 | 1 | 827 | 0.2215 | | | | | | | 0.94 |
| 2011_Engl_FT | 606 | MC | 20 | 1 | 827 | 0.9292 | | | | | | | 0.86 |
| 2011_Engl_FT | 606 | MC | 21 | 1 | 827 | 0.7783 | | | | | | | 1.10 |
| 2011_Engl_FT | 606 | MC | 22 | 1 | 827 | −0.9760 | | | | | | | 0.88 |
| 2011_Engl_FT | 607 | MC | 11 | 1 | 827 | −0.1350 | | | | | | | 1.02 |
| 2011_Engl_FT | 607 | MC | 12 | 1 | 827 | 0.6251 | | | | | | | 1.26 |
| 2011_Engl_FT | 607 | MC | 13 | 1 | 827 | −1.2764 | | | | | | | 0.82 |
| 2011_Engl_FT | 607 | MC | 14 | 1 | 827 | −1.8367 | | | | | | | 0.78 |
| 2011_Engl_FT | 607 | MC | 15 | 1 | 827 | −0.9854 | | | | | | | 0.87 |
| 2011_Engl_FT | 607 | MC | 16 | 1 | 827 | 0.6792 | | | | | | | 0.95 |
| 2011_Engl_FT | 607 | MC | 17 | 1 | 827 | −0.3738 | | | | | | | 0.93 |
| 2011_Engl_FT | 607 | MC | 18 | 1 | 827 | −0.6638 | | | | | | | 0.87 |
| 2011_Engl_FT | 607 | MC | 19 | 1 | 827 | −0.4057 | | | | | | | 0.92 |
| 2011_Engl_FT | 607 | MC | 20 | 1 | 827 | 0.1773 | | | | | | | 0.87 |
| 2011_Engl_FT | 607 | MC | 21 | 1 | 827 | −0.5281 | | | | | | | 0.84 |
| 2011_Engl_FT | 607 | MC | 22 | 1 | 827 | 0.1205 | | | | | | | 0.88 |
| 2011_Engl_FT | 608 | MC | 11 | 1 | 834 | 0.1045 | | | | | | | 1.08 |
| 2011_Engl_FT | 608 | MC | 12 | 1 | 834 | −1.0402 | | | | | | | 1.00 |
| 2011_Engl_FT | 608 | MC | 13 | 1 | 834 | −0.1488 | | | | | | | 0.98 |
| 2011_Engl_FT | 608 | MC | 14 | 1 | 834 | −0.5278 | | | | | | | 0.87 |
| 2011_Engl_FT | 608 | MC | 15 | 1 | 834 | −0.1559 | | | | | | | 0.93 |
| 2011_Engl_FT | 608 | MC | 16 | 1 | 834 | 0.7877 | | | | | | | 1.13 |
| 2011_Engl_FT | 608 | MC | 17 | 1 | 834 | 0.1708 | | | | | | | 0.86 |

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

| Test | Form | Type | Item | Max | N-Count | RID | S1 | S2 | S3 | S4 | S5 | S6 | INFIT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 608 | MC | 18 | 1 | 834 | 0.7003 | | | | | | | 1.08 |
| 2011_Engl_FT | 608 | MC | 19 | 1 | 834 | −1.2014 | | | | | | | 0.89 |
| 2011_Engl_FT | 608 | MC | 20 | 1 | 834 | −0.6901 | | | | | | | 0.86 |
| 2011_Engl_FT | 608 | MC | 21 | 1 | 834 | −0.3293 | | | | | | | 0.87 |
| 2011_Engl_FT | 608 | MC | 22 | 1 | 834 | −0.2851 | | | | | | | 1.01 |
| 2011_Engl_FT | 609 | MC | 11 | 1 | 797 | −0.6759 | | | | | | | 0.84 |
| 2011_Engl_FT | 609 | MC | 12 | 1 | 797 | −0.8506 | | | | | | | 0.83 |
| 2011_Engl_FT | 609 | MC | 13 | 1 | 797 | −1.1633 | | | | | | | 0.88 |
| 2011_Engl_FT | 609 | MC | 14 | 1 | 797 | −0.0871 | | | | | | | 0.98 |
| 2011_Engl_FT | 609 | MC | 15 | 1 | 797 | −0.2734 | | | | | | | 1.04 |
| 2011_Engl_FT | 609 | MC | 16 | 1 | 797 | 1.0275 | | | | | | | 1.01 |
| 2011_Engl_FT | 609 | MC | 17 | 1 | 797 | −1.0787 | | | | | | | 0.90 |
| 2011_Engl_FT | 609 | MC | 18 | 1 | 797 | −0.6670 | | | | | | | 0.90 |
| 2011_Engl_FT | 609 | MC | 19 | 1 | 797 | −0.7209 | | | | | | | 0.93 |
| 2011_Engl_FT | 609 | MC | 20 | 1 | 797 | −0.8697 | | | | | | | 0.77 |
| 2011_Engl_FT | 609 | MC | 21 | 1 | 797 | 0.4789 | | | | | | | 0.95 |
| 2011_Engl_FT | 609 | MC | 22 | 1 | 797 | 0.2521 | | | | | | | 0.93 |
| 2011_Engl_FT | 610 | MC | 11 | 1 | 799 | −1.2232 | | | | | | | 0.92 |
| 2011_Engl_FT | 610 | MC | 12 | 1 | 799 | −1.1668 | | | | | | | 0.91 |
| 2011_Engl_FT | 610 | MC | 13 | 1 | 799 | −0.8150 | | | | | | | 0.85 |
| 2011_Engl_FT | 610 | MC | 14 | 1 | 799 | −0.3101 | | | | | | | 1.07 |
| 2011_Engl_FT | 610 | MC | 15 | 1 | 799 | 0.0517 | | | | | | | 0.84 |
| 2011_Engl_FT | 610 | MC | 16 | 1 | 799 | 0.9131 | | | | | | | 1.06 |
| 2011_Engl_FT | 610 | MC | 17 | 1 | 799 | −0.8344 | | | | | | | 0.77 |

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

| Test | Form | Type | Item | Max | N-Count | RID | S1 | S2 | S3 | S4 | S5 | S6 | INFIT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 610 | MC | 18 | 1 | 799 | −1.0793 | | | | | | | 0.87 |
| 2011_Engl_FT | 610 | MC | 19 | 1 | 799 | −0.3764 | | | | | | | 0.90 |
| 2011_Engl_FT | 610 | MC | 20 | 1 | 799 | −1.5440 | | | | | | | 0.84 |
| 2011_Engl_FT | 610 | MC | 21 | 1 | 799 | −0.0864 | | | | | | | 0.93 |
| 2011_Engl_FT | 610 | MC | 22 | 1 | 799 | −0.1808 | | | | | | | 1.00 |
| 2011_Engl_FT | 611 | MC | 01 | 1 | 501 | −0.8826 | | | | | | | 1.01 |
| 2011_Engl_FT | 611 | MC | 02 | 1 | 501 | −0.0726 | | | | | | | 1.10 |
| 2011_Engl_FT | 611 | MC | 03 | 1 | 501 | −1.4840 | | | | | | | 0.84 |
| 2011_Engl_FT | 611 | MC | 04 | 1 | 501 | −0.9241 | | | | | | | 1.24 |
| 2011_Engl_FT | 611 | MC | 05 | 1 | 501 | −0.8281 | | | | | | | 1.07 |
| 2011_Engl_FT | 611 | CR | 06 | 2 | 501 | −0.6689 | −1.0968 | 1.0968 | | | | | 0.99 |
| 2011_Engl_FT | 611 | CR | 07 | 2 | 501 | −0.2382 | −0.8131 | 0.8131 | | | | | 0.97 |
| 2011_Engl_FT | 612 | CR | Es | 6 | 499 | 1.3700 | −4.2770 | −2.6566 | −1.8370 | 0.1874 | 2.4155 | 6.1677 | 1.10 |
| 2011_Engl_FT | 613 | MC | 01 | 1 | 504 | −1.4465 | | | | | | | 1.06 |
| 2011_Engl_FT | 613 | MC | 02 | 1 | 504 | −1.1193 | | | | | | | 1.01 |
| 2011_Engl_FT | 613 | MC | 03 | 1 | 504 | −1.0476 | | | | | | | 1.03 |
| 2011_Engl_FT | 613 | MC | 04 | 1 | 504 | 0.3871 | | | | | | | 1.09 |
| 2011_Engl_FT | 613 | MC | 05 | 1 | 504 | −2.1714 | | | | | | | 0.91 |
| 2011_Engl_FT | 613 | CR | 06 | 2 | 504 | −0.5859 | −1.2334 | 1.2334 | | | | | 0.90 |
| 2011_Engl_FT | 613 | CR | 07 | 2 | 504 | −0.2136 | −0.6686 | 0.6686 | | | | | 0.98 |
| 2011_Engl_FT | 614 | CR | Es | 6 | 495 | 0.9898 | −3.1134 | −1.6682 | −2.1208 | 0.2324 | 2.4868 | 4.1831 | 0.92 |
| 2011_Engl_FT | 615 | MC | 01 | 1 | 484 | −1.4785 | | | | | | | 0.99 |
| 2011_Engl_FT | 615 | MC | 02 | 1 | 484 | −1.5367 | | | | | | | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | 615 | MC | 03 | 1 | 484 | −0.3468 | | | | | | | 0.97 |
| 2011_Engl_FT | 615 | MC | 04 | 1 | 484 | −1.0177 | | | | | | | 1.05 |
| 2011_Engl_FT | 615 | MC | 05 | 1 | 484 | −1.7229 | | | | | | | 0.99 |
| 2011_Engl_FT | 615 | CR | 06 | 2 | 484 | −0.3272 | −0.8200 | 0.8200 | | | | | 1.02 |
| 2011_Engl_FT | 615 | CR | 07 | 2 | 484 | 0.0302 | −0.7768 | 0.7768 | | | | | 1.06 |
| 2011_Engl_FT | 616 | CR | Es | 6 | 500 | −0.2420 | −2.9895 | −0.8987 | −0.8314 | 1.0329 | 3.6868 | | 0.98 |
| 2011_Engl_FT | 617 | MC | 01 | 1 | 479 | −1.0487 | | | | | | | 0.95 |
| 2011_Engl_FT | 617 | MC | 02 | 1 | 479 | −0.0417 | | | | | | | 1.06 |
| 2011_Engl_FT | 617 | MC | 03 | 1 | 479 | −2.3157 | | | | | | | 0.90 |
| 2011_Engl_FT | 617 | MC | 04 | 1 | 479 | −1.2348 | | | | | | | 1.13 |
| 2011_Engl_FT | 617 | MC | 05 | 1 | 479 | −1.4007 | | | | | | | 0.91 |
| 2011_Engl_FT | 617 | CR | 06 | 2 | 479 | −0.8121 | −1.6135 | 1.6135 | | | | | 1.02 |
| 2011_Engl_FT | 617 | CR | 07 | 2 | 479 | −0.3887 | −1.0844 | 1.0844 | | | | | 0.95 |
| 2011_Engl_FT | 618 | CR | Es | 6 | 490 | 0.8645 | −3.4262 | −1.6010 | −1.9161 | 0.1952 | 2.3115 | 4.4366 | 1.07 |
| 2011_Engl_FT | 619 | MC | 01 | 1 | 473 | −1.4901 | | | | | | | 0.91 |
| 2011_Engl_FT | 619 | MC | 02 | 1 | 473 | −1.8372 | | | | | | | 0.86 |
| 2011_Engl_FT | 619 | MC | 03 | 1 | 473 | −1.2073 | | | | | | | 0.94 |
| 2011_Engl_FT | 619 | MC | 04 | 1 | 473 | −1.2409 | | | | | | | 0.94 |
| 2011_Engl_FT | 619 | MC | 05 | 1 | 473 | −0.8338 | | | | | | | 1.09 |
| 2011_Engl_FT | 619 | CR | 06 | 2 | 473 | −1.1098 | −1.6466 | 1.6466 | | | | | 0.99 |
| 2011_Engl_FT | 619 | CR | 07 | 2 | 473 | −0.2968 | −0.8853 | 0.8853 | | | | | 1.05 |
| 2011_Engl_FT | 620 | CR | Es | 6 | 475 | 1.1609 | −3.9428 | −2.4144 | −1.8761 | −0.1337 | 2.6774 | 5.6895 | 1.00 |
| 2011_Engl_FT | N3 | MC | 01 | 1 | 11,863 | −0.7100 | | | | | | | 0.99 |

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

| Test | Form | Type | Item | Max | N-Count | RID | S1 | S2 | S3 | S4 | S5 | S6 | INFIT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011_Engl_FT | N3 | MC | 02 | 1 | 11,863 | 1.0400 | | | | | | | 1.20 |
| 2011_Engl_FT | N3 | MC | 03 | 1 | 11,863 | −1.3300 | | | | | | | 0.87 |
| 2011_Engl_FT | N3 | MC | 04 | 1 | 11,863 | −0.7469 | | | | | | | 1.04 |
| 2011_Engl_FT | N3 | MC | 05 | 1 | 11,863 | −0.1900 | | | | | | | 1.06 |
| 2011_Engl_FT | N3 | MC | 06 | 1 | 11,863 | 0.7900 | | | | | | | 1.00 |
| 2011_Engl_FT | N3 | MC | 07 | 1 | 11,863 | −0.2100 | | | | | | | 0.95 |
| 2011_Engl_FT | N3 | MC | 08 | 1 | 11,863 | 0.6900 | | | | | | | 1.15 |
| 2011_Engl_FT | N3 | MC | 09 | 1 | 11,863 | −0.1200 | | | | | | | 0.96 |
| 2011_Engl_FT | N3 | MC | 10 | 1 | 11,863 | 0.2900 | | | | | | | 0.90 |

# Appendix C: DIF Statistics

## Table 10. DIF Statistics

| Form | Item | Item Type | MH Delta | MH Chi-Sq | Effect Size | DIF Category | Favored Group |
|---|---|---|---|---|---|---|---|
| 601 | 01 | MC | −0.02 | 0.00 | 0.00 | | |
| 601 | 02 | MC | −1.20 | 1.88 | −0.09 | | |
| 601 | 03 | MC | 0.48 | 0.35 | 0.04 | | |
| 601 | 04 | MC | −0.28 | 0.26 | −0.04 | | |
| 601 | 05 | MC | −0.12 | 0.04 | −0.02 | | |
| 601 | 06 | MC | 0.12 | 0.06 | 0.01 | | |
| 601 | 07 | MC | 0.30 | 0.41 | 0.06 | | |
| 601 | 08 | MC | −0.59 | 1.28 | −0.11 | | |
| 601 | 09 | MC | 0.12 | 1.07 | 0.02 | | |
| 601 | 10 | MC | 0.05 | 0.31 | 0.01 | | |
| 601 | 11 | MC | 0.88 | 42.32 | 0.11 | | |
| 601 | 12 | MC | 0.68 | 38.49 | 0.11 | | |
| 601 | 13 | MC | 0.18 | 3.08 | 0.03 | | |
| 601 | 14 | MC | −1.48 | 20.98 | −0.25 | B | M |
| 601 | 15 | MC | −0.21 | 4.06 | −0.03 | | |
| 601 | 16 | MC | 0.46 | 24.66 | 0.09 | | |
| 601 | 17 | MC | 1.21 | 37.51 | 0.20 | B | F |
| 601 | 18 | MC | −0.15 | 2.03 | −0.02 | | |
| 602 | 01 | MC | 1.42 | 5.21 | 0.18 | B | F |
| 602 | 02 | MC | 0.67 | 1.07 | 0.08 | | |
| 602 | 03 | MC | −1.85 | 10.22 | −0.22 | C | M |
| 602 | 04 | MC | −0.69 | 1.24 | −0.08 | | |
| 602 | 05 | MC | 0.03 | 0.00 | 0.00 | | |
| 602 | 06 | MC | −1.57 | 3.40 | −0.11 | | |

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

| Form | Item | Item Type | MH Delta | MH Chi-Sq | Effect Size | DIF Category | Favored Group |
|---|---|---|---|---|---|---|---|
| 602 | 07 | MC | 0.82 | 3.15 | 0.12 | | |
| 602 | 08 | MC | 0.31 | 0.31 | 0.04 | | |
| 603 | 01 | MC | 0.97 | 1.30 | 0.06 | | |
| 603 | 02 | MC | 0.40 | 0.19 | 0.02 | | |
| 603 | 03 | MC | −0.08 | 0.02 | −0.02 | | |
| 603 | 04 | MC | −2.50 | 20.26 | −0.30 | C | M |
| 603 | 05 | MC | 0.15 | 0.06 | 0.05 | | |
| 603 | 06 | MC | −0.07 | 0.02 | −0.01 | | |
| 603 | 07 | MC | 0.01 | 0.00 | −0.01 | | |
| 603 | 08 | MC | −0.16 | 0.11 | 0.00 | | |
| 604 | 01 | MC | −0.36 | 0.62 | −0.07 | | |
| 604 | 02 | MC | 0.58 | 1.12 | 0.07 | | |
| 604 | 03 | MC | −0.74 | 2.20 | −0.08 | | |
| 604 | 04 | MC | −0.47 | 0.84 | −0.07 | | |
| 604 | 05 | MC | −1.27 | 3.30 | −0.12 | | |
| 604 | 06 | MC | 0.74 | 2.65 | 0.12 | | |
| 604 | 07 | MC | −0.06 | 0.02 | −0.02 | | |
| 604 | 08 | MC | −0.45 | 0.96 | −0.08 | | |
| 605 | 01 | MC | −1.04 | 3.89 | −0.15 | B | M |
| 605 | 02 | MC | −0.54 | 1.24 | −0.10 | | |
| 605 | 03 | MC | −0.70 | 0.78 | −0.06 | | |
| 605 | 04 | MC | −0.49 | 0.91 | −0.05 | | |
| 605 | 05 | MC | −0.14 | 0.07 | 0.00 | | |
| 605 | 06 | MC | 0.74 | 1.64 | 0.09 | | |

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

| Form | Item | Item Type | MH Delta | MH Chi-Sq | Effect Size | DIF Category | Favored Group |
|---|---|---|---|---|---|---|---|
| 605 | 07 | MC | −2.42 | 10.96 | −0.20 | C | M |
| 605 | 08 | MC | −0.29 | 0.32 | −0.03 | | |
| 606 | 11 | MC | 0.61 | 2.29 | 0.09 | | |
| 606 | 12 | MC | 0.22 | 0.20 | 0.01 | | |
| 606 | 13 | MC | −0.16 | 0.10 | −0.02 | | |
| 606 | 14 | MC | 0.14 | 0.05 | 0.02 | | |
| 606 | 15 | MC | −0.21 | 0.23 | −0.03 | | |
| 606 | 16 | MC | 1.77 | 16.54 | 0.23 | C | F |
| 606 | 17 | MC | 0.54 | 1.94 | 0.10 | | |
| 606 | 18 | MC | −0.39 | 0.69 | −0.05 | | |
| 606 | 19 | MC | −0.01 | 0.00 | −0.03 | | |
| 606 | 20 | MC | −2.73 | 38.31 | −0.35 | C | M |
| 606 | 21 | MC | −0.75 | 3.94 | −0.11 | | |
| 606 | 22 | MC | 0.47 | 0.87 | 0.04 | | |
| 607 | 11 | MC | −0.10 | 0.05 | −0.01 | | |
| 607 | 12 | MC | 0.12 | 0.11 | 0.04 | | |
| 607 | 13 | MC | 0.63 | 1.04 | 0.05 | | |
| 607 | 14 | MC | 1.62 | 4.41 | 0.10 | | |
| 607 | 15 | MC | 0.35 | 0.43 | 0.03 | | |
| 607 | 16 | MC | 0.30 | 0.56 | 0.04 | | |
| 607 | 17 | MC | −1.56 | 10.95 | −0.20 | C | M |
| 607 | 18 | MC | 0.62 | 1.54 | 0.06 | | |
| 607 | 19 | MC | −0.17 | 0.14 | −0.01 | | |
| 607 | 20 | MC | −0.05 | 0.02 | −0.01 | | |
| 607 | 21 | MC | 0.75 | 2.29 | 0.09 | | |

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

| Form | Item | Item Type | MH Delta | MH Chi-Sq | Effect Size | DIF Category | Favored Group |
|---|---|---|---|---|---|---|---|
| 607 | 22 | MC | −1.34 | 8.64 | −0.16 | B | M |
| 608 | 11 | MC | 0.27 | 0.51 | 0.05 | | |
| 608 | 12 | MC | −0.62 | 1.72 | −0.09 | | |
| 608 | 13 | MC | 0.36 | 0.79 | 0.05 | | |
| 608 | 14 | MC | 0.64 | 1.92 | 0.10 | | |
| 608 | 15 | MC | 0.59 | 2.00 | 0.08 | | |
| 608 | 16 | MC | −0.11 | 0.10 | −0.02 | | |
| 608 | 17 | MC | −1.62 | 14.20 | −0.22 | C | M |
| 608 | 18 | MC | −0.91 | 6.22 | −0.17 | | |
| 608 | 19 | MC | −0.22 | 0.17 | −0.03 | | |
| 608 | 20 | MC | −0.47 | 0.97 | −0.06 | | |
| 608 | 21 | MC | −0.09 | 0.04 | −0.02 | | |
| 608 | 22 | MC | −0.73 | 3.12 | −0.12 | | |
| 609 | 11 | MC | 0.44 | 0.74 | 0.04 | | |
| 609 | 12 | MC | 0.25 | 0.23 | 0.04 | | |
| 609 | 13 | MC | 0.16 | 0.09 | 0.02 | | |
| 609 | 14 | MC | 0.28 | 0.46 | 0.04 | | |
| 609 | 15 | MC | −1.38 | 9.94 | −0.21 | B | M |
| 609 | 16 | MC | 0.43 | 1.14 | 0.08 | | |
| 609 | 17 | MC | 0.10 | 0.04 | 0.01 | | |
| 609 | 18 | MC | −0.70 | 2.04 | −0.09 | | |
| 609 | 19 | MC | 0.09 | 0.04 | 0.01 | | |
| 609 | 20 | MC | 0.08 | 0.02 | 0.01 | | |
| 609 | 21 | MC | −0.44 | 1.19 | −0.07 | | |
| 609 | 22 | MC | −0.31 | 0.53 | −0.05 | | |

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

| Form | Item | Item Type | MH Delta | MH Chi-Sq | Effect Size | DIF Category | Favored Group |
|---|---|---|---|---|---|---|---|
| 610 | 11 | MC | 1.04 | 4.03 | 0.15 | B | F |
| 610 | 12 | MC | 0.33 | 0.36 | 0.03 | | |
| 610 | 13 | MC | −1.43 | 6.89 | −0.16 | B | M |
| 610 | 14 | MC | 0.08 | 0.04 | 0.03 | | |
| 610 | 15 | MC | −0.79 | 2.77 | −0.09 | | |
| 610 | 16 | MC | −0.94 | 5.47 | −0.15 | | |
| 610 | 17 | MC | 0.09 | 0.03 | 0.01 | | |
| 610 | 18 | MC | 0.92 | 2.96 | 0.10 | | |
| 610 | 19 | MC | −0.64 | 1.84 | −0.10 | | |
| 610 | 20 | MC | 0.26 | 0.17 | 0.02 | | |
| 610 | 21 | MC | −0.45 | 1.02 | −0.07 | | |
| 610 | 22 | MC | 1.00 | 5.76 | 0.12 | B | F |
| 611 | 01 | MC | 0.53 | 1.00 | 0.10 | | |
| 611 | 02 | MC | −2.96 | 31.95 | −0.45 | C | M |
| 611 | 03 | MC | −0.99 | 1.85 | −0.11 | | |
| 611 | 04 | MC | −0.27 | 0.27 | −0.04 | | |
| 611 | 05 | MC | −2.17 | 13.61 | −0.28 | C | M |
| 611 | 06 | CR | | 10.87 | 0.23 | BB | F |
| 611 | 07 | CR | | 0.04 | 0.00 | | |
| 612 | Es | CR | | 8.10 | 0.26 | | |
| 613 | 01 | MC | −0.87 | 2.00 | −0.15 | | |
| 613 | 02 | MC | 0.19 | 0.11 | 0.03 | | |
| 613 | 03 | MC | −0.58 | 1.05 | −0.08 | | |
| 613 | 04 | MC | −0.43 | 0.81 | −0.07 | | |
| 613 | 05 | MC | 0.03 | 0.00 | 0.00 | | |

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

| Form | Item | Item Type | MH Delta | MH Chi-Sq | Effect Size | DIF Category | Favored Group |
|---|---|---|---|---|---|---|---|
| 613 | 06 | CR | | 0.61 | −0.07 | | |
| 613 | 07 | CR | | 1.06 | 0.08 | | |
| 614 | Es | CR | | 8.18 | 0.17 | | |
| 615 | 01 | MC | 0.72 | 1.05 | 0.07 | | |
| 615 | 02 | MC | −0.63 | 0.76 | −0.08 | | |
| 615 | 03 | MC | −0.10 | 0.03 | −0.03 | | |
| 615 | 04 | MC | 1.38 | 5.36 | 0.17 | B | F |
| 615 | 05 | MC | 0.19 | 0.07 | 0.02 | | |
| 615 | 06 | CR | | 11.39 | 0.26 | CC | F |
| 615 | 07 | CR | | 4.64 | 0.16 | | |
| 616 | Es | CR | | 0.71 | 0.08 | | |
| 617 | 01 | MC | 0.54 | 0.81 | 0.07 | | |
| 617 | 02 | MC | −0.28 | 0.28 | −0.05 | | |
| 617 | 03 | MC | 3.17 | 11.73 | 0.24 | C | F |
| 617 | 04 | MC | −0.08 | 0.02 | −0.02 | | |
| 617 | 05 | MC | 0.37 | 0.33 | 0.06 | | |
| 617 | 06 | CR | | 6.51 | 0.19 | BB | F |
| 617 | 07 | CR | | 0.40 | −0.06 | | |
| 618 | Es | CR | | 4.53 | 0.21 | | |
| 619 | 01 | MC | 0.62 | 0.85 | 0.07 | | |
| 619 | 02 | MC | −0.05 | 0.00 | −0.02 | | |
| 619 | 03 | MC | −0.04 | 0.00 | 0.00 | | |
| 619 | 04 | MC | −0.50 | 0.60 | −0.06 | | |
| 619 | 05 | MC | −0.98 | 3.02 | −0.14 | | |
| 619 | 06 | CR | | 11.40 | 0.26 | CC | F |

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

| Form | Item | Item Type | MH Delta | MH Chi-Sq | Effect Size | DIF Category | Favored Group |
|---|---|---|---|---|---|---|---|
| 619 | 07 | CR | | 8.21 | 0.20 | BB | F |
| 620 | Es | CR | | 5.25 | 0.15 | | |

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

# Appendix D: Operational Test Maps

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

| Position | Item Type | Max Points | Weight | Standard | Mean | Point-Biserial | Rasch | S1 | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | MC | 1 | 1 | 1 | 0.68 | 0.34 | −0.56 | | | | | | |
| 2 | MC | 1 | 1 | CPI | 0.93 | 0.35 | −3.03 | | | | | | |
| 3 | MC | 1 | 1 | 3 | 0.78 | 0.33 | −1.20 | | | | | | |
| 4 | MC | 1 | 1 | CPI | 0.91 | 0.45 | −2.68 | | | | | | |
| 5 | MC | 1 | 1 | 3 | 0.85 | 0.51 | −1.82 | | | | | | |
| 6 | MC | 1 | 1 | 3 | 0.76 | 0.42 | −1.09 | | | | | | |
| 7 | MC | 1 | 1 | 2 | 0.62 | 0.46 | −0.21 | | | | | | |
| 8 | MC | 1 | 1 | 2 | 0.52 | 0.47 | 0.37 | | | | | | |
| 9 | MC | 1 | 1 | 3 | 0.74 | 0.56 | −0.75 | | | | | | |
| 10 | MC | 1 | 1 | 2 | 0.76 | 0.59 | −0.88 | | | | | | |
| 11 | MC | 1 | 1 | 1 | 0.57 | 0.46 | 0.28 | | | | | | |
| 12 | MC | 1 | 1 | 3 | 0.65 | 0.50 | −0.19 | | | | | | |
| 13 | MC | 1 | 1 | 3 | 0.67 | 0.50 | −0.32 | | | | | | |
| 14 | MC | 1 | 1 | 3 | 0.68 | 0.49 | −0.37 | | | | | | |
| 15 | MC | 1 | 1 | 1 | 0.75 | 0.58 | −0.80 | | | | | | |
| 16 | MC | 1 | 1 | 3 | 0.77 | 0.56 | −0.95 | | | | | | |
| 17 | MC | 1 | 1 | 1 | 0.71 | 0.54 | −0.58 | | | | | | |
| 18 | MC | 1 | 1 | CPI | 0.66 | 0.53 | −0.25 | | | | | | |
| 19 | MC | 1 | 1 | 2 | 0.79 | 0.60 | −1.10 | | | | | | |
| 20 | MC | 1 | 1 | 3 | 0.76 | 0.57 | −0.90 | | | | | | |
| 21 | MC | 1 | 1 | 2 | 0.94 | 0.41 | −2.96 | | | | | | |
| 22 | MC | 1 | 1 | 1 | 0.78 | 0.38 | −1.20 | | | | | | |
| 23 | MC | 1 | 1 | 3 | 0.84 | 0.50 | −1.67 | | | | | | |

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

| Position | Item Type | Max Points | Weight | Standard | Mean | Point-Biserial | Rasch | S1 | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | MC | 1 | 1 | CPI | 0.50 | 0.45 | 0.52 | | | | | | |
| 25 | MC | 1 | 1 | 3 | 0.80 | 0.36 | −1.33 | | | | | | |
| 26 | CR | 2 | 3 | CPI, 1, 2, 3 | 1.19 | 0.58 | −0.02 | −0.94 | 0.94 | | | | |
| 27 | CR | 2 | 3 | CPI, 1, 2 | 1.09 | 0.64 | 0.27 | −0.80 | 0.80 | | | | |
| 28 | Essay | 6 | 3 | CPI, 1, 2, 3 | 3.07 | 0.66 | 1.67 | −1.73 | −3.00 | −2.44 | −1.03 | 1.68 | 6.51 |

**Table 12. Operational Test Map for August 2011**

| Position | Item Type | Max Points | Weight | Standard | Mean | Point-Biserial | Rasch | S1 | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | MC | 1 | 1 | S3 | 0.81 | 0.50 | −1.52 | | | | | | |
| 2 | MC | 1 | 1 | S3 | 0.78 | 0.45 | −1.25 | | | | | | |
| 3 | MC | 1 | 1 | S3 | 0.62 | 0.45 | −0.20 | | | | | | |
| 4 | MC | 1 | 1 | S2 | 0.85 | 0.53 | −1.83 | | | | | | |
| 5 | MC | 1 | 1 | S2 | 0.86 | 0.45 | −1.98 | | | | | | |
| 6 | MC | 1 | 1 | S2 | 0.66 | 0.47 | −0.40 | | | | | | |
| 7 | MC | 1 | 1 | CPI | 0.89 | 0.43 | −2.36 | | | | | | |
| 8 | MC | 1 | 1 | S1 | 0.93 | 0.5 | −3.08 | | | | | | |
| 9 | MC | 1 | 1 | S3 | 0.73 | 0.62 | −0.61 | | | | | | |
| 10 | MC | 1 | 1 | S1 | 0.73 | 0.52 | −0.58 | | | | | | |
| 11 | MC | 1 | 1 | S2 | 0.72 | 0.51 | −0.55 | | | | | | |
| 12 | MC | 1 | 1 | S1 | 0.78 | 0.57 | −0.96 | | | | | | |
| 13 | MC | 1 | 1 | S3 | 0.83 | 0.61 | −1.33 | | | | | | |
| 14 | MC | 1 | 1 | S3 | 0.63 | 0.53 | −0.01 | | | | | | |
| 15 | MC | 1 | 1 | S2 | 0.64 | 0.48 | −0.04 | | | | | | |
| 16 | MC | 1 | 1 | S3 | 0.71 | 0.56 | −0.51 | | | | | | |
| 17 | MC | 1 | 1 | CPI | 0.53 | 0.58 | 0.55 | | | | | | |
| 18 | MC | 1 | 1 | CPI | 0.61 | 0.51 | 0.12 | | | | | | |
| 19 | MC | 1 | 1 | S1 | 0.64 | 0.54 | −0.08 | | | | | | |
| 20 | MC | 1 | 1 | S3 | 0.68 | 0.55 | −0.32 | | | | | | |
| 21 | MC | 1 | 1 | S3 | 0.9 | 0.44 | −2.48 | | | | | | |
| 22 | MC | 1 | 1 | S3 | 0.92 | 0.4 | −2.73 | | | | | | |

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

| Position | Item Type | Max Points | Weight | Standard | Mean | Point-Biserial | Rasch | S1 | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | MC | 1 | 1 | S2 | 0.9 | 0.45 | −2.39 | | | | | | |
| 24 | MC | 1 | 1 | S3 | 0.83 | 0.44 | −1.66 | | | | | | |
| 25 | MC | 1 | 1 | S2 | 0.59 | 0.39 | −0.03 | | | | | | |
| 26 | CR | 2 | 3 | CP1,S1,S2, S3 | 1.26 | 0.59 | −0.27 | −0.86 | 0.86 | | | | |
| 27 | CR | 2 | 3 | CPI,S1,S2 | 1.04 | 0.62 | 0.38 | −1.05 | 1.05 | | | | |
| 28 | Essay | 6 | 3 | CPI,S1,S2, S3 | 3.11 | 0.69 | 0.43 | −0.38 | −1.60 | −1.47 | −0.02 | 3.47 | |

# Appendix E: Scoring Tables

## Table 13. Scoring Table for June 2011

| Raw Score | Ability | Scale Score | | Raw Score | Ability | Scale Score | | Raw Score | Ability | Scale Score |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | −5.970 | 0.000 | | 23 | −0.144 | 40.727 | | 46 | 1.567 | 82.275 |
| 1 | −4.709 | 1.438 | | 24 | −0.039 | 42.518 | | 47 | 1.663 | 84.141 |
| 2 | −3.937 | 3.096 | | 25 | 0.064 | 44.309 | | 48 | 1.774 | 86.020 |
| 3 | −3.456 | 4.665 | | 26 | 0.162 | 46.099 | | 49 | 1.907 | 87.914 |
| 4 | −3.096 | 6.332 | | 27 | 0.257 | 47.887 | | 50 | 2.069 | 89.828 |
| 5 | −2.802 | 8.064 | | 28 | 0.347 | 49.675 | | 51 | 2.277 | 91.766 |
| 6 | −2.553 | 9.840 | | 29 | 0.433 | 51.465 | | 52 | 2.554 | 93.735 |
| 7 | −2.333 | 11.644 | | 30 | 0.513 | 53.254 | | 53 | 2.955 | 95.741 |
| 8 | −2.135 | 13.466 | | 31 | 0.589 | 55.045 | | 54 | 3.649 | 97.793 |
| 9 | −1.955 | 15.297 | | 32 | 0.661 | 56.836 | | 55 | 4.858 | 99.809 |
| 10 | −1.787 | 17.133 | | 33 | 0.728 | 58.628 | | | | |
| 11 | −1.631 | 18.970 | | 34 | 0.793 | 60.423 | | | | |
| 12 | −1.483 | 20.806 | | 35 | 0.855 | 62.220 | | | | |
| 13 | −1.342 | 22.638 | | 36 | 0.915 | 64.020 | | | | |
| 14 | −1.207 | 24.466 | | 37 | 0.974 | 65.823 | | | | |
| 15 | −1.076 | 26.290 | | 38 | 1.032 | 67.629 | | | | |
| 16 | −0.950 | 28.108 | | 39 | 1.090 | 69.440 | | | | |
| 17 | −0.827 | 29.922 | | 40 | 1.149 | 71.255 | | | | |
| 18 | −0.708 | 31.732 | | 41 | 1.209 | 73.076 | | | | |
| 19 | −0.590 | 33.538 | | 42 | 1.271 | 74.902 | | | | |
| 20 | −0.476 | 35.340 | | 43 | 1.336 | 76.733 | | | | |
| 21 | −0.363 | 37.138 | | 44 | 1.406 | 78.572 | | | | |
| 22 | −0.252 | 38.933 | | 45 | 1.483 | 80.419 | | | | |

## Table 14. Scoring Table for August 2011

| Raw Score | Ability | Scale Score | | Raw Score | Ability | Scale Score | | Raw Score | Ability | Scale Score |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | −6.100 | 0.000 | | 23 | −0.212 | 39.593 | | 46 | 1.562 | 82.171 |
| 1 | −4.848 | 1.228 | | 24 | −0.101 | 41.460 | | 47 | 1.659 | 84.066 |
| 2 | −4.085 | 2.722 | | 25 | 0.008 | 43.326 | | 48 | 1.771 | 85.975 |
| 3 | −3.611 | 4.030 | | 26 | 0.113 | 45.187 | | 49 | 1.905 | 87.898 |
| 4 | −3.254 | 5.494 | | 27 | 0.213 | 47.044 | | 50 | 2.071 | 89.841 |
| 5 | −2.963 | 7.064 | | 28 | 0.309 | 48.897 | | 51 | 2.282 | 91.807 |
| 6 | −2.713 | 8.672 | | 29 | 0.399 | 50.747 | | 52 | 2.563 | 93.796 |
| 7 | −2.492 | 10.325 | | 30 | 0.485 | 52.595 | | 53 | 2.970 | 95.812 |
| 8 | −2.291 | 12.021 | | 31 | 0.564 | 54.440 | | 54 | 3.673 | 97.856 |
| 9 | −2.106 | 13.755 | | 32 | 0.639 | 56.280 | | 55 | 4.888 | 99.855 |
| 10 | −1.934 | 15.519 | | 33 | 0.710 | 58.119 | | | | |
| 11 | −1.772 | 17.311 | | 34 | 0.777 | 59.954 | | | | |
| 12 | −1.618 | 19.123 | | 35 | 0.841 | 61.790 | | | | |
| 13 | −1.471 | 20.951 | | 36 | 0.902 | 63.626 | | | | |
| 14 | −1.330 | 22.793 | | 37 | 0.962 | 65.462 | | | | |
| 15 | −1.194 | 24.645 | | 38 | 1.021 | 67.301 | | | | |
| 16 | −1.061 | 26.505 | | 39 | 1.080 | 69.142 | | | | |
| 17 | −0.932 | 28.368 | | 40 | 1.140 | 70.986 | | | | |
| 18 | −0.806 | 30.237 | | 41 | 1.201 | 72.834 | | | | |
| 19 | −0.683 | 32.108 | | 42 | 1.263 | 74.687 | | | | |
| 20 | −0.562 | 33.980 | | 43 | 1.330 | 76.547 | | | | |
| 21 | −0.443 | 35.852 | | 44 | 1.400 | 78.413 | | | | |
| 22 | −0.327 | 37.723 | | 45 | 1.477 | 80.287 | | | | |