New York State Regents Examination in Global History and Geography

2014 Field Test Analysis, Equating Procedure, and Scaling of Operational Test Forms

Technical Report

Prepared for the New York State Education Department
by Pearson

January 2015

Copyright

Table of Contents

Table of Contents i

List of Tables iii

List of Figures iii

Section I: Introduction 1

PURPOSE 1

Section II: Field Test Analysis 1

FILE PROCESSING AND DATA CLEANUP 2

CLASSICAL ANALYSIS 3

Item Difficulty *3*

Item Discrimination *3*

Test Reliability *4*

Scoring Reliability *5*

Inter-rater Agreement *6*

Constructed-Response Item Means and Standard Deviations *6*

Intraclass Correlation *6*

Weighted Kappa *7*

ITEM RESPONSE THEORY (IRT) AND THE CALIBRATION AND EQUATING OF THE FIELD TEST ITEMS 8

Item Calibration *9*

Item Fit Evaluation *10*

DIFFERENTIAL ITEM FUNCTIONING 11

The Mantel Chi-Square and Standardized Mean Difference *12*

Multiple Choice Items *13*

The Odds Ratio *13*

The Delta Scale *14*

DIF Classification for MC Items *14*

DIF Classification for Scaffold and Essay Items *14*

Section III: Equating Procedure 15

RANDOMLY EQUIVALENT GROUP EQUATING DESIGN 16

COMMON ITEM EQUATING DESIGN 18

Section IV: Scaling of Operational Test Forms 19

References 22

Appendix A: Classical Item Analysis 24

Appendix B: Inter-rater Consistency – Point Differences Between First and Second Reads 37

Appendix C: Additional Measures of Inter-rater Reliability and Agreement 39
Appendix D: Partial-Credit Model Item Analysis 42
Appendix E: DIF Statistics 55
Appendix F: Operational Test Maps 65
Appendix G: Scoring Tables 71

List of Tables

Table 1. Need/Resource Capacity Category Definitions 1
Table 3. Test and Scoring Reliability 5
Table 4. Criteria to Evaluate Mean-Square Fit Statistics 10
Table 5. Partial-Credit Model Item Analysis Summary 11
Table 6. DIF Classification for MC Items 14
Table 7. DIF Classification for Scaffold and Essay Items 15
Table 8. Initial Mean Abilities and Equating Constants 18

List of Figures

Figure 1. 2 × *t* Contingency Table at the k^{th} of *K* Levels. 12

Section I: Introduction

PURPOSE

The purpose of this report is to document the psychometric properties of the New York State Regents Examination in Global History and Geography. In addition, this report documents the procedures used to analyze the results of the field test and to equate and scale the operational test forms.

Section II: Field Test Analysis

In May 2014, prospective items for the New York State Regents Examination in Global History and Geography were field tested. The results of this testing were used to evaluate item quality. Only items with acceptable statistical characteristics can be selected for use on operational tests.

Representative student samples for participation in this testing were selected to mirror the demographics of the student population that is expected to take the operational test. The Need/Resource Capacity Categories in Table 1 were used as variables in the sampling plan.

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	All districts at or above the 70th percentile on the index with at least 100 students per square mile or enrollment greater than 2500
Rural	All districts at or above the 70th percentile on the index with fewer than 50 students per square mile or enrollment of fewer than 2500
Average N/RC Districts	All districts between the 20th and 70th percentiles on the index
Low N/RC Districts	All districts below the 20th percentile on the index
Charter Schools	Each charter school is a district

FILE PROCESSING AND DATA CLEANUP

The Regents examinations utilize both multiple-choice (MC) and scaffold and essay item types in order to more fully assess student ability. Multiple field test (FT) forms were given during this administration to allow for a large number of items to be field tested without placing an undue burden on the students participating in the field test; each student only took a small subset of the items being field tested. The New York State Education Department (NYSED) handled all scanning of the MC responses. Scoring of the scaffold and essay responses was performed by Measurement Incorporated (MI) under contract to the NYSED. The NYSED and MI produced separate data files which were provided to Pearson. A test map file that documented the items on each of the FT forms was also provided to Pearson by the NYSED. Finally, student data file layouts containing the position of every field within the student data files from both the NYSED and MI were also provided to Pearson by the NYSED. Upon receipt of these files, Pearson staff checked the data, test map, and layouts for consistency. Any anomalies were referred back to the NYSED for resolution. After these had been resolved and corrected as necessary, final processing of the data file took place. Merging of the NYSED and MI provided data was accomplished through uniquely assigned booklet numbers. This processing included the identification and deletion of invalid student test records through the application of a set of predefined exclusion rules[1]. Two analyses were conducted for this subject, depending on the characteristics of the items being field tested. There were a total of 28 forms used in the present field test; of these, 9 were composed exclusively of MC items while the remaining 19 consisted of an anchor test (of 30 MC items) coupled with varying numbers of scaffold and essay field test items. Among these 19 forms, identical forms were 710 & 711, 712 & 713, 714 & 715, 716 & 717, 718 & 719, 722 & 723, 724 & 725, and 726 & 727. The identical forms were combined for the equating and other analyses. The first group was equated using an equivalent groups design, and the second was equated using a common-items design. Details of both designs can be found in Section III of this document. The final field test data files for the two analyses contained 7,529 and 7,707 records, respectively.

Within the final data file used in the field test analyses, MC responses were scored according to the item keys contained in the test map; correct responses received a score of 1 while incorrect responses received a score of 0. Scaffold and essay item scores were taken directly from the student data file, with the exception that out-of-range scores were assigned scores of 0. For Item Response Theory (IRT) calibrations, blanks (i.e., missing data; not omits) were also scored as 0.

In addition to the scored data, the final data file also contained the unscored student responses and scores. Unscored data was used to calculate the percentage of students who selected the various answer choices for the MC items or the percentage of students who received each achievable score point for the scaffold and essay items.

[1] These exclusion rules flagged records without both MC and scaffold and essay components (if relevant), records with invalid or out-of-range form numbers, records without any responses, and duplicate records. These records were dropped prior to analysis.

The frequency of students leaving items blank was also calculated. The scored data were used for all other analyses.

CLASSICAL ANALYSIS

Classical Test Theory assumes that any observed test score *x* is composed of both true score *t* and error score *e*. This assumption is expressed as follows:

$$x = t + e$$

All test scores are composed of both a true and an error component. For example, the choice of test items or administration conditions might influence student responses, making a student's observed score higher or lower than the student's true ability would warrant. This error component is random and uncorrelated with (i.e., unrelated to) the student's true score. Across an infinitely large number of administrations, the mean of the error scores would be zero. Thus, the best estimate of a student's true score for any test administration (or their expected score given their [unobservable] true level of ability or true score) is that student's observed score. This expectation is expressed as follows:

$$E(x) = t$$

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 typically defined as the average of scores for a given item. For MC items, this value (commonly referred to as a p-value) ranges from 0 to 1. For scaffold and essay items, this value ranges from 0 to the maximum possible score. In order to place all item means on a common metric (ranging from 0 to 1), scaffold and essay item means were divided by the maximum points possible for the item.

Item Discrimination

Item discrimination is defined as the correlation between a score on a given test question and the overall raw test score. These correlations are Pearson correlation coefficients. For MC items, it is also known as the point-biserial correlation.

Table 2 presents a summary of the classical item analysis for each of the field test forms. The first three columns from the left 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 discrimination). Recall that for scaffold and essay items, item means were divided by the maximum number of points possible in order to place them in the same metric as the MC items. No items had difficulties that were greater than 0.90 and 12 items had correlations that were less than 0.25. In addition to the summary information provided in Table 2, further classical item statistics are provided in Appendix A.

Table 2. Classical Item Analysis Summary

Form	N-Count	No. of Items	Item Difficulty			Item Discrimination		
			<0.50	0.50 to 0.90	>0.90	<0.25	0.25 to 0.50	>0.50
701	836	30	5	25	0	0	22	8
702	848	32	8	24	0	0	21	11
703	855	32	4	28	0	1	21	10
704	829	32	6	26	0	0	18	14
705	829	32	7	25	0	0	20	12
706	821	32	8	24	0	4	20	8
707	845	32	10	22	0	3	21	8
708	833	32	13	19	0	2	26	4
709	833	32	11	21	0	2	22	8
710_711	732	1	1	0	0	0	0	1
712_713	708	1	1	0	0	0	0	1
714_715	712	1	1	0	0	0	0	1
716_717	678	1	1	0	0	0	0	1
718_719	674	1	1	0	0	0	0	1
720	360	1	1	0	0	0	0	1
721	349	1	1	0	0	0	0	1
722_723	991	11	1	10	0	0	6	5
724_725	1017	15	1	14	0	0	3	12
726_727	995	11	1	10	0	0	3	8
728	491	13	3	10	0	0	9	4

For some forms, the item counts in the “Item Difficulty” and “Item Discrimination” columns may not sum to the value in the “No. of Items” column due to DNS (Do Not Score) items. In addition, the item counts for Forms 710–728 do not include the common anchor form items.

Test Reliability

Reliability is the consistency of the results obtained from a measurement with respect to time or between 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: (1) test length and (2) homogeneity of the items. In general, the more items on the examination, the higher the reliability and the more similar the items, the higher the reliability.

Table 3 contains the internal consistency statistics for each of the field test forms under the heading “Test Reliability.” These statistics ranged from 0.85 to 0.91. It should be noted that operational tests generally are composed of more items and would be expected to have somewhat higher reliabilities than do these field test forms.

Scoring Reliability

One concern with scaffold and essay items is the reliability of the scoring process (i.e., consistency of the score assignment). Scaffold and essay items must be read by scorers who assign scores based on a comparison between the rubric and student responses. Consistency between scorers is a critical part of the reliability of the assessment. To track scorer consistency, approximately 10% of the test booklets are scored a second time (these are termed “second read scores”) and compared to the original set of scores (also known as “first read scores”).

As an overall measure of scoring reliability, the Pearson correlation coefficient between the first and second scores for all scaffold and essay items with second read scores was computed for each form. This statistic is often used as an overall indicator of scoring reliability, and it generally ranges from 0 to 1. Table 3 contains these values in the column headed “Scoring Reliability.” They ranged from 0.61 to 0.90, indicating a fair to high degree of reliability. Please note that since Forms 701–709 were composed exclusively of MC items, reliabilities for these forms could not be calculated.

Table 3. Test and Scoring Reliability

Form Number	Test Reliability	Scoring Reliability
701	0.87	N/A
702	0.88	N/A
703	0.88	N/A
704	0.89	N/A
705	0.89	N/A
706	0.85	N/A
707	0.85	N/A
708	0.86	N/A
709	0.86	N/A
710_711	0.90	0.73
712_713	0.90	0.82
714_715	0.89	0.87
716_717	0.89	0.70
718_719	0.88	0.81
720	0.90	0.90
721	0.88	0.61
722_723	0.91	0.89
724_725	0.91	0.88
726_727	0.91	0.89
728	0.90	0.81

Inter-rater Agreement

For each scaffold and essay item, the difference between the first and second reads was tracked and the number of times each possible difference between the scores occurred was tabulated. These values were then used to calculate the percentage of times each possible difference occurred. 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. Blank cells in the table indicate out-of-range differences (e.g., it is impossible for two raters to differ by more than one point in their scores on an item with a maximum possible score of one; cells in the table other than –1, 0, and 1 would therefore be blanked out).

Appendix B contains the proportion of occurrence of these differences for each scaffold and essay item. These items had rates of exact agreement in the 62–100% range. The percentage of scores that were exact or adjacent matches for these items ranged from 94–100%. Nonadjacent scores were not possible for the one-point items. Appendix C contains additional summary information regarding the first and second reads.

Constructed-Response Item Means and Standard Deviations

Appendix C also contains the mean and standard deviation of the first and second scores for each scaffold and essay item. While there were minimal differences between the standard deviation statistics, the largest difference between the item means for the first and second read scores was 0.1.

Intraclass Correlation

In addition, Appendix C contains the intraclass correlations for the items. These correlations are calculated using a formulation given by Shrout and Fleiss (1979). Specifically, they described six different models based on various configurations of judges and targets (in this case, papers that are being scored). For this assessment, the purpose of the statistic is to describe the reliability of single ratings, and each paper is scored by two judges who are randomly assigned from the larger pool of judges, and who score multiple papers. This description fits their "Case 1." Further, they distinguish between situations where the score assigned to the paper is that of a single rater versus that when the score is the mean of *k* raters. Since the students' operational scores are those from single (i.e., the first) raters, the proper intraclass correlation in this instance is termed by Shrout and Fleiss as "ICC(1,1)." It will be referred to herein simply as the "intraclass correlation" (ICC).

While the ICC is a bona fide correlation coefficient, it differs from a regular correlation coefficient in that its value remains the same regardless of how the raters are ordered. A regular Pearson correlation coefficient would change values if half of the second raters were switched to the first position, while the ICC would maintain a consistent value. Because the papers were randomly assigned to the judges, ordering is arbitrary, and thus the ICC is a more appropriate measure of reliability than the Pearson correlation coefficient in this situation. The ICC ranges from zero (the scores given by the two judges are unrelated) to one (the scores from the two judges match perfectly);

negative values are possible, but rare, and have essentially the same meaning as values of zero. It should also be noted that the ICC can be affected by low degrees of variance in the scores being related, similar to the way that regular Pearson correlation coefficients are affected. ICCs for items where almost every examinee achieved the same score point (e.g., an extremely easy dichotomous item where almost every examinee was able to answer it correctly) may have a low or negative ICC even though almost all ratings by the judges matched exactly.

McGraw and Wong (1996, Table 4, p. 35) state that the ICC can be interpreted as "the degree of absolute agreement among measurements made on randomly selected objects. It estimates the correlation of any two measurements." Since it is a correlation coefficient, its square indicates the percent of variance in the scores that is accounted for by the relationship between the two sets of scores (i.e., the two measurements). In this case, these scores are those of the pair of judges. ICC values greater than 0.60 indicate that at least 36% (0.60^2) of the variation in the scores given by the raters is accounted for by variations in the responses to the items that are being scored (e.g., variations in the ability being measured) rather than by variations caused by a combination of differences in the severity of the judges, interactions between judge severity and the items, and random error (e.g., variations exterior to the ability being measured). It is generally preferred that items have ICCs at this level or higher. Only one item had ICCs below 0.60. Consistent with other information provided in the table, these values indicate a high to very high level of scoring reliability for almost all of the items in the field test.

Weighted Kappa

Weighted Kappa (Cohen, 1968) was also calculated for each item based on the first and second reads and is included in Appendix C as well. This statistic is an estimate of the agreement of the score classifications over and above that which would be expected to occur by chance. Similar to the ICC, its value can range between zero (the scores given by the judges agree as often as would be expected by chance) and one (scores given by the judges agree perfectly). In addition, negative values are possible, but rare, and have the same interpretation as zero values. One set of guidelines for the evaluation of this statistic is (Fleiss, 1981):

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

The results show good reproducibility for most of the items field tested. Forty-one items displayed excellent reproducibility and only one item had marginal reproducibility. The scoring reliability analyses offer evidence that the scoring of the scaffold and essay items was performed in a reliable manner.

ITEM RESPONSE THEORY (IRT) AND THE CALIBRATION AND EQUATING OF THE FIELD TEST ITEMS

While classical test theory-based statistical measures are useful for assessing the suitability of items for operational use (i.e., use as part of an assessment used to measure student ability and thus having real-world consequences for students, teachers, schools, and administrators), their values are dependent on both the psychometric properties of the items and the ability distributions of the samples upon which they are based. In other words, classical test theory-based statistics are *sample-dependent statistics*.

In contrast, Item Response Theory (IRT) based statistics are not dependent on the sample over which they are estimated—they are invariant across different samples (Hambleton, Swaminathan, & Rogers, 1991; Lord, 1980). This invariance allows student ability to be estimated on a common metric even if different sets of items are used (as with different test forms over different test administrations).

The process of estimating IRT-based item parameters is referred to as "item calibration," and the placing of these parameters on a common metric or scale is termed "equating." While one reason for the field testing of items is to allow for their suitability for use in the operational measurement of student ability to be assessed, the data resulting from field testing is also used to place items on the scale of the operational test (i.e., they are equated to the operational metric). Once items are on this common metric, any form composed of items from this pool can be scaled (the process through which scale score equivalents for each achievable raw score are derived) and the resulting scale scores will be directly comparable to those from other administrations, even though the underlying test forms are composed of different sets of items.

There are several variations of IRT differing mainly in the way item behavior is modeled. The New York State Regents Examinations use the Rasch family of IRT statistics (Rasch, 1980; Masters, 1982) to calibrate, scale, and equate all subjects.

The most basic expression of the Rasch model is in the item characteristic curve. It conceptualizes the probability of a correct response to an item as a function of the ability level and the item's difficulty. The probability of a correct response is bounded by "1" (certainty of a correct response) and "0" (certainty of an incorrect response). The ability scale is theoretically unbounded. In practice, the ability scale ranges from approximately −4 to +4 logits. The relationship between examinee ability θ, item difficulty D_i, and probability of answering the item correctly P_i is shown in the equation below:

$$P_i(\theta) = \frac{\exp(\theta - D_i)}{1 + \exp(\theta - D_i)}$$

Examinee ability (θ) and item difficulty (D_i) are on the same scale. This is useful for certain purposes. An examinee with an ability level equal to the item difficulty will have a 50% chance of answering the item correctly; if his or her ability level is higher than the

item difficulty, then the probability of answering the item correctly is commensurately higher, and the converse is also true.

The Rasch Partial Credit Model (PCM) (Masters, 1982) is a direct extension of the dichotomous one-parameter IRT model above. For an item involving *m* score categories, the general expression for the probability of achieving a score of *x* on the item is given by

$$P_x(\theta) = \frac{\exp[\sum_{k=0}^{x}(\theta - D_k)]}{\sum_{h=0}^{m}\exp[\sum_{k=0}^{h}(\theta - D_k)]}$$

where

$$D_0 \equiv 0.0$$

In the above equation, P_x is the probability of achieving a score of *x* given an ability of θ; *m* is the number of achievable score points minus one (note that the subscript *k* runs from 0 to *m*); and D_k is the step parameter for step *k*. The steps are numbered from 0 to the number of achievable score points minus one, and step 0 (D_0) is defined as being equal to zero. Note that a four-point item, for example, usually has five achievable score points (0, 1, 2, 3, and 4), thus the step numbers usually mirror the achievable point values.

According to this model, the probability of an examinee scoring in a particular category (step) is the sum of the logit (log-odds) differences between θ and *Dk* of all the completed steps, divided by the sum of the differences of all the steps of an item. Thissen and Steinberg (1986) refer to this model as a divide-by-total model. The parameters estimated by this model are *mi* −1 threshold (difficulty) estimates, and they represent the points on the ability continuum where the probability of the examinee achieving score mi exceeds that of *mi*-1. The mean of these threshold estimates is used as an overall summary of the polytomous item's difficulty.

If the number of achievable score points is one (i.e., the item is dichotomous), then the PCM reduces to the basic Rasch IRT model for dichotomous items. This means that dichotomous and polytomous items are being scaled using a common model and therefore can be calibrated, equated, and scaled together. It should be noted that the Rasch model assumes that all items have equal levels of discrimination and that there is no guessing on MC items. However items that violate these assumptions to a large degree are usually flagged for item-model misfit.

Item Calibration

When interpreting IRT item parameters, it is important to remember that they do not have an absolute scale. Their scale (in terms of mean and standard deviation) is purely arbitrary. It is conventional to set the mean of the item difficulties to zero when an assessment is scaled for the first time. Rasch IRT scales the theta measures in terms of *logits*, or "log-odds units." The length of a logit varies from test to test, but generally the standard deviation of the item difficulties of a test scaled for the first time will be somewhere in the area of 0.6–0.8. While the item difficulties are invariant with respect to

one another, the absolute level of difficulty represented by their mean is dependent on the overall difficulty of the group of items with which it was tested. In addition, there is no basis for assuming that the difficulty values are normally distributed around their mean—their distribution depends solely upon the intrinsic difficulties of the items themselves. Thus, if a particularly difficult set of items (relative to the set of items originally calibrated) was field tested, their overall mean would most probably be greater than zero, and their standard deviation would be considerably less than one. In addition, they would most probably not be normally distributed.

Rasch item difficulties generally range from −3.0 to 3.0, although very easy or difficult items can fall outside of this range. Items should not be discounted solely on the basis of their difficulty. A particular topic may require either a difficult or an easy item. Items are usually most useful if their difficulty is close to a cut score, as items provide the highest level of information at the ability level equal to their difficulty. Items with difficulties farther away from the cuts provide less information about students with abilities close to the cut scores (and, hence, are more susceptible to misclassification), but are still useful. In general, items should be selected for use based on their content, with their Rasch difficulty being only a secondary consideration.

Item Fit Evaluation

The INFIT statistic is used to assess how well items fit the Rasch model. Rasch theory models the probability of a student being able to answer an item correctly as a function of the student's level of ability and the item's difficulty, as stated previously. The Rasch model also assumes that discriminations of items do not differ, and that the items are not susceptible to guessing. If these assumptions do not hold (an item has an extremely high or low level of discrimination), then the item's behavior will not be well modeled by Rasch IRT. Guidelines for interpretation of the INFIT statistic are taken from Linacre (2005) and can be found in Table 4 below.

Table 4. Criteria to Evaluate Mean-Square Fit Statistics

INFIT	Interpretation
>2.0	Distorts or degrades the measurement system
1.5–2.0	Unproductive for construction of measurement, but not degrading
0.5–1.5	Productive for measurement
<0.5	Unproductive for measurement, but not degrading. May produce misleadingly good reliabilities and separations

INFIT is an information-weighted fit statistic, which is more sensitive to unexpected behavior affecting responses to items near the student's measure (or ability) level. In general, values near 1.0 indicate little distortion of the measurement system, while values less than 1.0 indicate observations are too predictable (redundancy, model overfit). Values greater than 1.0 indicate unpredictability (unmodeled noise, model underfit).

Table 5 contains a summary of the analysis for each of the field test forms. The first column from the left 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 final columns show the frequency of items at three levels of difficulty (easier items with a Rasch difficulty <−2.0, moderate items with a Rasch difficulty between −2.0 and 2.0, and more difficult items with a Rasch difficulty >2.0), and frequencies of item misfits as classified in the preceding table. Most of the items fell within the moderate −2.0 to +2.0 difficulty range, and there was no item with an INFIT statistic outside the range most productive for measurement. Item level results of the analysis can be found in Appendix D.

Table 5. Partial-Credit Model Item Analysis Summary

Form	N-Count	No. of Items	Rasch			INFIT			
			<−2.0	−2.0 to 2.0	>2.0	<0.5	0.5 to 1.5	1.5 to 2.0	>2.0
701	836	30	0	30	0	0	30	0	0
702	848	32	0	32	0	0	32	0	0
703	855	32	0	32	0	0	32	0	0
704	829	32	0	32	0	0	32	0	0
705	829	32	0	32	0	0	32	0	0
706	821	32	0	32	0	0	32	0	0
707	845	32	0	32	0	0	32	0	0
708	833	32	0	31	1	0	32	0	0
709	833	32	0	31	1	0	32	0	0
710_711	732	31	0	30	1	0	31	0	0
712_713	708	31	0	30	1	0	31	0	0
714_715	712	31	0	31	0	0	31	0	0
716_717	678	31	0	31	0	0	31	0	0
718_719	674	31	0	30	1	0	31	0	0
720	360	31	0	30	1	0	31	0	0
721	349	31	0	30	1	0	31	0	0
722_723	991	41	0	40	1	0	41	0	0
724_725	1017	45	0	45	0	0	45	0	0
726_727	995	41	0	41	0	0	41	0	0
728	491	43	0	43	0	0	43	0	0

For some forms, the item counts in the “Rasch” and “INFIT” columns may not sum to the value in the “No. of Items” column due to DNS (Do Not Score) items.

DIFFERENTIAL ITEM FUNCTIONING

Differential Item Functioning (DIF) occurs when members of a particular group have a different probability of success than members of another group who have the same level of ability for reasons unrelated to the academic skill or construct being measured.

For example, items testing English grammar skills may be more difficult for LEP students as opposed to non-LEP students, but such differences are likely due to the fact that the item measures an academic skill related to English language proficiency. Such items would not be considered to be functioning differentially.

The Mantel Chi-Square and Standardized Mean Difference

The Mantel χ^2 is a conditional mean comparison of the ordered response categories for reference and focal groups combined over values of the matching variable score. "Ordered" means that a response earning a score of "1" on an item is better than a response earning a score of "0," or "2" is better than "1,"and so on. "Conditional," on the other hand, refers to the comparison of members from the two groups who received the same score on the matching variable, that is, the total test score in our analysis.

Group	Item Score				Total
	y_1	y_2	…	y_T	
Reference	n_{R1k}	n_{R2k}	…	n_{Rtk}	n_{R+k}
Focal	n_{F1k}	n_{F2k}	…	n_{Ftk}	n_{F+k}
Total	n_{+1k}	n_{+2k}	…	n_{+tk}	n_{++k}

Figure 1. 2 × *t* Contingency Table at the k^{th} of *K* Levels.

Figure 1 (from Zwick, Donoghue, & Grima, 1993) shows a 2 × *t* contingency table at the k^{th} of *K* levels, where *t* represents the number of response categories and *k* represents the number of levels of the matching variable. The values $y_1, y_2, \ldots, y_T$ represent the *t* scores that can be gained on the item. The values n_{Ftk} and n_{Rtk} represent the numbers of focal and reference groups who are at the k^{th} level of the matching variable and gain an item score of y_t. The "+" indicates the total number over a particular index (Zwick et al., 1993). The Mantel statistic is defined as the following formula:

$$Mantel\chi^2 = \frac{\left(\sum_k F_k - \sum_k E(F_k)\right)^2}{\sum_k Var(F_k)}$$

in which F_k represents the sum of scores for the focal group at the k^{th} level of the matching variable and is defined as follows:

$$F_k = \sum_t y_t n_{Ftk}$$

The expectation of F_k under the null hypothesis is

$$E(F_k) = \frac{n_{F+k}}{n_{++k}} \sum_t y_t n_{Ftk}$$

The variance of F_k under the null hypothesis is as follows:

$$Var(F_k) = \frac{n_{R+k} n_{F+k}}{n_{++k}^2 (n_{++k} - 1)} \left[(n_{++k} \sum_t y_t^2 n_{+tk}) - (\sum_t y_t n_{+tk})^2 \right]$$

Under *H0*, the Mantel statistic has a chi-square distribution with one degree of freedom. In DIF applications, rejecting *H0* suggests that the students of the reference and focal groups who are similar in overall test performance tend to differ in their mean performance on the item. For dichotomous items, the statistic is identical to the Mantel-Haenszel (MH) (1959) statistic without the continuity correction (Zwick et al., 1993).

A summary statistic to accompany the Mantel approach is the standardized mean difference (SMD) between the reference and focal groups proposed by Dorans and Schmitt (1991). This statistic compares the means of the reference and focal groups, adjusting for differences in the distribution of the reference and focal group members across the values of the matching variable. The SMD has the following form:

$$SMD = \sum_k p_{Fk} m_{Fk} - \sum_k p_{Fk} m_{Rk}$$

in which

$$p_{Fk} = \frac{n_{F+k}}{n_{F++}}$$

is the proportion of the focal group members who are at the k^{th} level of the matching variable;

$$m_{Fk} = \frac{1}{n_{F+k} \sum_t y_t n_{Ftk}}$$

is the mean item score of the focal group members at the k^{th} level; and m_{Rk} is the analogous value for the reference group. As can be seen from the equation above, the SMD is the difference between the unweighted item mean of the focal group and the weighted item mean of the reference group. The weights for the reference group are applied to make the weighted number of the reference-group students the same as in the focal group within the same level of ability. A negative SMD value implies that the focal group has a lower mean item score than the reference group, conditional on the matching variable.

Multiple Choice Items

For the MC items, the *MH* odds ratio (converted to the ETS delta scale [D]) is used to classify items into one of three categories of DIF.

The Odds Ratio

The odds of a correct response (proportion passing divided by proportion failing) are *P/Q* or *P/(1–P)*. The *odds ratio* is the odds of a correct response of the reference group divided by the odds of a correct response of the focal group. For a given item, the odds ratio is defined as follows:

$$\alpha_{M-H} = \frac{P_r / Q_r}{P_f / Q_f}$$

and, the corresponding null hypothesis is that the odds of getting the item correct are equal for the two groups. Thus, the odds ratio is equal to 1:

$$\alpha_{M-H} = \frac{P_r / Q_r}{P_f / Q_f} = 1$$

The Delta Scale

To make the odds ratio symmetrical around zero with its range being in the interval −∞ to +∞, the odds ratio is transformed into a log odds ratio according to this equation:

$$\beta_{MH} = \ln(\alpha_{MH})$$

This simple natural logarithm transformation of the odds ratio is symmetrical around zero. This DIF measure is a signed index; a positive value signifies DIF in favor of the reference group, a negative value indicates DIF in favor of the focal group, and zero has the interpretation of equal odds of success on the item. β_{MH} also has the advantage of a linear relationship to other interval scale metrics (Camilli & Shepard, 1994). β_{MH} is placed on the ETS delta scale (D) using the following equation:

$$D = -2.35\beta_{MH}$$

DIF Classification for MC Items

Table 6 depicts DIF classifications for MC items. Classification depends on the delta (D) value and the significance of its difference from zero ($p < 0.05$). The criteria are derived from those used by the National Assessment of Educational Progress (Allen, Carlson, & Zelenak, 1999) in the development of their assessments.

Table 6. DIF Classification for MC Items

Category	Description	Criterion
A	No DIF	D not significantly different from zero or |D| < 1.0
B	Moderate DIF	1.0 ≤ |D| < 1.5 or not otherwise A or C
C	High DIF	D is significantly different from zero and |D| ≥ 1.5

DIF Classification for Scaffold and Essay Items

The SMD is divided by the total group item standard deviation to obtain an effect-size value for the SMD (ES_{SMD}). The value of ES_{SMD} and the significance of the Mantel χ^2 statistic ($p < 0.05$) are then used to determine the DIF category of the item as depicted in Table 7.

Table 7. DIF Classification for Scaffold and Essay Items

Category	Description	Criterion
AA	No DIF	Non-significant Mantel χ^2 or $\lvert ES_{SMD} \rvert \leq 0.17$
BB	Moderate DIF	Significant Mantel χ^2 and $0.17 < \lvert ES_{SMD} \rvert \leq 0.25$
CC	High DIF	Significant Mantel χ^2 and $0.25 < \lvert ES_{SMD} \rvert$

Reliable DIF results are dependent on the number of examinees in both the focal and reference groups. Clauser and Mazor (1998) state that a minimum of 200 to 250 examinees per group are sufficient to provide reliable results. Some testing organizations require as many as 300 to 400 examinees per group (Zwick, 2012) in some applications. For the field testing of the Regents examinations, the sample sizes were such that only comparisons based on gender (e.g., males vs. females) were possible. Even for gender, sample sizes were only moderately large, and so the results should be interpreted with caution.

The DIF statistics for gender are shown in Appendix E. MC items in DIF categories "B" and "C" and scaffold and essay items in categories "BB" and "CC" were flagged. These flags are shown in the "DIF Category" column ("A" and "AA" category items will have blank cells here). The "Favored Group" column indicates which gender is favored for items that are flagged.

Section III: Equating Procedure

Students particpating in the 2014 field test administration for the New York State Regents Examination in Global History and Geography received one of 28 test forms (numbered 701–728). It should be noted that among these test forms, some forms were identical. These were 710 & 711, 712 & 713, 714 & 715, 716 & 717, 718 & 719, 722 & 723, 724 & 725, and 726 & 727. The identical forms were combined for the equating and other analyses. Form 701 was the anchor form for the equating and was an intact form that had been administered in the prior year. The items on Form 701 were also embedded within Forms 710–728 as an internal anchor set. Because the form had been previously administered, its items had known parameters on the operational scale. The remaining test forms were composed of items that had not been administered to New York State students. Test forms were spiraled within classrooms, so that students had an equal chance of receiving any of the 28 forms, depending solely on their ordinal position within the classroom. In essence, students were randomly assigned to test forms, forming randomly equivalent groups taking each of the forms. Appendices A and D (with the classical and Rasch IRT item level statistics) may be consulted to determine the characteristics of the items (e.g., item type and maximum number of points possible) that made up each form. All forms were calibrated using Winsteps, version 3.60 (Linacre, 2005).

RANDOMLY EQUIVALENT GROUP EQUATING DESIGN

The equating analyses for Forms 701–709 were based on the assumption that the groups taking the different forms had equivalent ability distributions and means. Given the random assignment of forms to examinees, this was a reasonable assumption. The initial step in the analyses was to calibrate all forms, both the anchor form and the remaining field test forms.

The anchor form calibration began with all anchor item difficulty parameters fixed to their known values from the previous year. Because it is possible for item parameters to "drift" (shift their difficulties relative to one another), a stability check was integrated into the analysis.

Winsteps provides an item level statistic, termed "displacement." Linacre (2011, p 545) describes this statistic as:

> …the size of the change in the parameter estimate that would be observed in the next estimation iteration if this parameter was free (unanchored) and all other parameter estimates were anchored at their current values. For a parameter (item or student) that is anchored in the main estimation, (the displacement value) indicates the size of disagreement between an estimate based on the current data and the anchor value.

This statistic was used to identify items with difficulties that had shifted, relative to the difficulties of the other items on the form. After the intial calibration run, the Winsteps displacement values for all anchor form items were examined for absolute values greater than 0.30. If present, the item with the largest absolute displacement value was removed from anchored status but remained on the test form. Its difficulty value was subsequently reestimated relative to the difficulties of the remaining anchored items. The Winsteps calibration was then rerun with the reduced anchor set, after which the displacement values were again checked for absolute values in excess of 0.30. If another was found, it was also removed from anchored status and the calibration rerun. This iterative procedure continued until all anchored items had displacements of 0.30 or less. No item was identified as having drifted for the 2014 analyses.

After a stable anchor item set had been identified, the mean of the ability estimates of the students who took the anchor form was computed[2]. This mean ability was then used as the target ability for the forms with the field test items. Because the groups taking the different forms were randomly equivalent and thus had the same mean ability, adjustment of the parameters of the field test items on any form to values that produced an ability distribution for students who had taken the form with a mean equal to the target ability from the anchor form would result in the parameters for the field test

[2] Because under Rasch IRT the ability of students with extreme scores (either zero or perfect scores) cannot be exactly computed (they are equal to $-\infty$ and $+\infty$, respectively), they were excluded from this and all other analyses for both the anchor and other field test forms.

items on that form being equated to the scale of the anchor form, which was also the operational scale.

The equated mean ability estimate for Form 701 was 0.41. This value became the target mean ability estimate for the field test forms.

At this point in the analyses, the calibration of the anchor form was complete. The next step was the initial calibration of the field test forms. This was a “free” calibration, meaning that the item parameters were not constrained in any way. This initial calibration produced a set of Rasch difficulty parameters for the items on each form. Also produced as a part of the Winsteps calibration was a set of person ability estimates for each form.

The next step was the computation of an equating constant for each form. Under Rasch IRT, if all of the difficulty parameters on a form have a constant added to them, the ability estimates for examinees will also be changed from their previous values by the amount represented by that constant. Therefore, to adjust the item difficulty parameters such that the mean of the ability distribution is set equal to the target mean ability from the anchor form, an equating constant was calculated for each field test form by subtracting the field test form mean ability from the target mean ability. This value was then added to the Rasch difficulty parameter of all items on the field test form. These adjusted values were then used as anchors for a second Winsteps calibration of the field test form. The mean of the student ability values from this second calibration was computed and compared to the target mean. If the anchored field test mean ability differed from the target mean ability by 0.005 or more, then an additional equating constant was computed using the difference between the mean ability from the field test form anchored run and the target mean ability, and another anchored run was completed. This process continued until all adjusted field test form mean abilities were within the 0.005 tolerance limit around the targeted mean ability. The final equating constant for any field test form was the sum of the constants from each anchored round for that form. At this point, with the adjusted mean abilities for the field test forms all equal (within the specified limits) to the target abilities, all of the adjusted field test item parameters and the anchor item parameters were on the common operational scale, and thus could be used in any subsequent operational administration. The initial mean abilities and final equating constants for the field test forms can be found in Table 8.

Table 8. Initial Mean Abilities and Equating Constants

Form Number	Mean Ability	Constant
702	0.45	−0.03
703	0.54	−0.12
704	0.47	−0.05
705	0.70	−0.28
706	0.42	−0.01
707	0.30	0.11
708	0.10	0.30
709	0.25	0.16

COMMON ITEM EQUATING DESIGN

An equating design utilizing common items was done for Forms 710–728. This design does not require the assumption of randomly equivalent groups as does the design used to equate Forms 702–709, but that assumption is tenable and should be so noted. All that is required is a set of representative common items. In this case, the 30 items from Form 701 formed this anchor set and were administered to all students along with their assigned set of field test items.

Using this design, the field test forms can either be calibrated individually or together. For this assessment, the forms were calibrated separately. The data file for such an approach contains one record per student . Within the data records, each distinct column corresponds to an item that was administered on the form. If the student was presented with the item, then his or her item score appears in that column.

The difficulty parameters for the anchor items were fixed at their “known” parameters from their most recent use (in this case, the 2013 Global History and Geography field test administration), and the parameters of the field test items were estimated relative to those of the anchor items. Because it is possible for item parameters to “drift” (shift their difficulties relative to one another over time), a stability check was integrated into the analysis.

The Winsteps displacement statistic was used to identify items with difficulties that had shifted, relative to the difficulties of the other items on the form. After the intial calibration run, the Winsteps displacement values for all anchor form items were examined for absolute values greater than 0.30. If present, the item with the largest absolute displacement value was removed from anchored status but remained on the test form. Its difficulty value was subsequently reestimated relative to the difficulties of the remaining anchored items. The Winsteps calibration was then rerun with the reduced anchor set, after which the displacement values were again checked for absolute values in excess of 0.30. If another was found, it was also removed from anchored status and the calibration rerun. This iterative procedure continued until all

anchored items had displacements of 0.30 or less. When the iterative procedure was finished, the parameters resulting from the final run were then in the operational metric, and the calibrations were complete. Depending on the form, between one and four items were identified as having drifted for the 2014 analyses.

Section IV: Scaling of Operational Test Forms

Operational test items were selected based on content coverage, content accuracy, and statistical quality. The sets of items on each operational test conformed to the coverage determined by content experts working from the learning standards established by the New York State Education Department and explicated in the test blueprint. Each item's classical and Rasch statistics were used to assess item quality. Items were selected to vary in difficulty to measure students' abilities accurately across the ability continuum. Appendix F contains the 2014 operational test maps for the January, June, and August administrations.

All Regents examinations have two cut scores, which are set at the scale scores of 65 and 85. One of the primary considerations during test construction was to select items so as to minimize changes in the raw scores corresponding to these two scale scores. Maintaining a consistent mean Rasch difficulty level from administration to administration facilitates this. For this assessment, the target value for the mean Rasch difficulty was set at 0.515. It should be noted that the raw scores corresponding to the scale score cut scores may still fluctuate even if the mean Rasch difficulty level is maintained at the target value due to differences in the distributions of the Rasch difficulty values among the items from administration to administration.

The relationship between raw and scale scores is explicated in the scoring tables for each administration. These tables can be found in Appendix G and cover the January, June, and August administrations. These tables are the end product of the following scaling procedure.

All Regents examinations are equated back to a base scale that is held constant from year to year. Specifically, they are equated to the base scale through the use of a calibrated item pool. The Rasch difficulties from the items' initial administration in a previous year's field test are used to equate the scale for the current administration to the base administration. For this exam, the base administration was the June 2004 administration. Scale scores from the 2014 administrations are on the same scale and can be directly compared to scale scores on all previous administrations back to and including the June 2004 administration.

When the base administration was concluded, the initial raw score-to-scale score relationship was established. Four raw scores were fixed at specific scale scores. Scale scores of 0 and 100 were fixed to correspond to the minimum and maximum possible raw scores. In addition, a standard setting had been held to determine the passing and passing with distinction cut scores in the raw score metric. The scale score points of 65

and 85 were set to correspond to those raw score cuts. A third degree polynomial is required in order to fit a line exactly to four arbitrary points (e.g., the raw scores corresponding to the four critical scale scores of 0, 65, 85, and 100). The general form of this best–fitting line is:

$$SS = m_3*RS^3 + m_2*RS^2 + m_1*RS + m_0$$

where *SS* is the scaled score, *RS* is the raw score, and m_0 through m_3 are the transformation constants that convert the raw score into the scale score (please note that m_0 will always be equal to zero in this application since a raw score of zero corresponds to a scale score of zero). The above relationship and the values of m_1 to m_3 specific to this subject were then used to determine the scale scores corresponding to the remainder of the raw scores on the exam. This initial relationship between the raw and scale scores became the base scale.

The Rasch difficulty parameters for the items on the base form were then used to derive a raw score-to-Rasch student ability (theta score) relationship. This allowed the relationship between the Rasch theta score and the scale score to be known, mediated through their common relationship with the raw scores.

In succeeding years, each test form was selected from the pool of items that had been tested in previous years' field tests, each of which had known Rasch item difficulty parameter(s). These known parameters were used to construct the relationship between the raw and Rasch theta scores for that particular form[3]. Because the Rasch difficulty parameters are all on a common scale, the Rasch theta scores were also on a common scale with previously administered forms. The remaining step in the scaling process was to find the scale score equivalent for the Rasch theta score corresponding to each raw score point on the new form using the theta to scale score relationship established in the base year. This was done via linear interpolation.

This process results in a relationship between the raw scores on the form and the overall scale scores. The scale scores corresponding to each raw score are then rounded to the nearest integer for reporting on the conversion chart (posted at the close of each administration). The only exceptions are for the minimum and maximum raw scores and the raw scores that correspond to the scaled cut scores of 65 and 85.

The minimum (zero) and maximum possible raw scores are assigned scale scores of 0 and 100, respectively. In the event that there are raw scores less than the

[3] All Regents examinations are pre-equated, meaning that the parameters used to derive the relationship between the raw and scale scores are estimated prior to the construction and administration of the operational form. These field tests are administered to as small a sample of students as possible in order to minimize the impact on student instructional time across the state. The small N-Counts associated with such administrations are sufficient for reasonably accurate estimation of most items' parameters; however, for the five-point items, their parameters can be unstable when estimated across as small a sample as is typically used. Therefore, a set of constants is used for these items' parameters on operational exams. These constants were set by the NYSED and are based on the values in the bank for all similar items.

maximum with scale scores that round to 100, their scale scores are set equal to 99. A similar process is followed with the minimum score; if any raw scores other than zero have scale scores that round to zero, their scale scores are instead set equal to one.

With regard to the cuts, if two or more scale scores round to either 65 or 85, the lowest raw score's scale score is set equal to a 65 or 85 and the scale scores corresponding to the higher raw scores are set to 66 or 86 as appropriate. If no scale score rounds to either of these two critical cuts, then the raw score with the largest scale score that is less than the cut is set equal to the cut. The overarching principle when two raw scores both round to either scale score cut is that the lower of the raw scores is always assigned to be equal to the cut so that students are never penalized for this ambiguity.

References

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

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice, 17*(1), 31–44.

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., & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach* (ETS Research Report No. 91-49). Princeton, NJ: Educational Testing Service.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions.* 2nd ed. New York: John Wiley.

Hambleton, R. K, Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage Publications, Inc.

Linacre, J. M. (2005). WINSTEPS Rasch measurement computer program and manual (PDF file) v 3.60. Chicago: Winsteps.com

Linacre, J. M. (2011). *A user's guide to WINSTEPS MINISTEP Rasch-model computer programs: Program manual 3.73.0* (PDF file). Chicago: Winsteps.com

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients, *Psychological Methods, 1*(1), 30–46.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51,* 567–577.

Zwick, R. (2012). *A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement* (ETS Research Report No. 12-08). Princeton, NJ: Educational Testing Service.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30*, 233–251.

Appendix A: Classical Item Analysis

In the following table, “Max” is the maximum number of possible points. “N-Count” refers to the number of student records in the analysis. “Alpha” contains Cronbach's Coefficient α (since this is a test [form] level statistic, it has the same value for all items within each form). For MC 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 scaffold and essay items, “B” represents the proportion of students who left the item blank, and “M0” through “M5” are the proportions of students who received scores of 0 through 5. “Mean” is the average of the scores received by the students. The final (right) column contains the Point-Biserial correlation for each item. There may be some instances of items with missing statistics; this occurs when an item was not scored. The anchor form items for Forms 716–725 are not included (these items’ statistics are represented by their use in Form 701).

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	701	MC	01	1	836	0.87	0.01		0.06	0.04	0.09	0.80		0.80	0.49
2014_GLHG	701	MC	02	1	836	0.87	0.00		0.70	0.12	0.09	0.08		0.70	0.45
2014_GLHG	701	MC	03	1	836	0.87	0.01		0.13	0.07	0.17	0.62		0.62	0.45
2014_GLHG	701	MC	04	1	836	0.87	0.00		0.23	0.22	0.42	0.12		0.42	0.45
2014_GLHG	701	MC	05	1	836	0.87	0.01		0.51	0.21	0.13	0.14		0.51	0.39
2014_GLHG	701	MC	06	1	836	0.87	0.00		0.10	0.62	0.23	0.04		0.62	0.39
2014_GLHG	701	MC	07	1	836	0.87	0.01		0.16	0.13	0.61	0.09		0.61	0.44
2014_GLHG	701	MC	08	1	836	0.87	0.01		0.19	0.12	0.13	0.55		0.55	0.32
2014_GLHG	701	MC	09	1	836	0.87	0.01		0.05	0.08	0.83	0.03		0.83	0.44
2014_GLHG	701	MC	10	1	836	0.87	0.01		0.17	0.69	0.07	0.06		0.69	0.49
2014_GLHG	701	MC	11	1	836	0.87	0.01		0.08	0.16	0.18	0.57		0.57	0.47
2014_GLHG	701	MC	12	1	836	0.87	0.01		0.80	0.08	0.04	0.07		0.80	0.50
2014_GLHG	701	MC	13	1	836	0.87	0.01		0.08	0.52	0.20	0.18		0.52	0.52
2014_GLHG	701	MC	14	1	836	0.87	0.01		0.49	0.15	0.15	0.19		0.49	0.37
2014_GLHG	701	MC	15	1	836	0.87	0.01		0.14	0.09	0.69	0.07		0.69	0.51
2014_GLHG	701	MC	16	1	836	0.87	0.01		0.58	0.20	0.14	0.07		0.58	0.38
2014_GLHG	701	MC	17	1	836	0.87	0.01		0.15	0.57	0.20	0.06		0.57	0.45

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	701	MC	18	1	836	0.87	0.02		0.12	0.16	0.10	0.60		0.60	0.57
2014_GLHG	701	MC	19	1	836	0.87	0.02		0.14	0.14	0.11	0.59		0.59	0.55
2014_GLHG	701	MC	20	1	836	0.87	0.02		0.18	0.48	0.12	0.20		0.48	0.31
2014_GLHG	701	MC	21	1	836	0.87	0.02		0.61	0.13	0.11	0.14		0.61	0.48
2014_GLHG	701	MC	22	1	836	0.87	0.02		0.18	0.06	0.68	0.06		0.68	0.51
2014_GLHG	701	MC	23	1	836	0.87	0.03		0.62	0.19	0.09	0.08		0.62	0.55
2014_GLHG	701	MC	24	1	836	0.87	0.03		0.16	0.55	0.11	0.16		0.55	0.37
2014_GLHG	701	MC	25	1	836	0.87	0.03		0.12	0.56	0.11	0.17		0.56	0.43
2014_GLHG	701	MC	26	1	836	0.87	0.03		0.23	0.10	0.56	0.07		0.56	0.38
2014_GLHG	701	MC	27	1	836	0.87	0.04		0.26	0.18	0.42	0.11		0.42	0.46
2014_GLHG	701	MC	28	1	836	0.87	0.04		0.12	0.11	0.14	0.60		0.60	0.61
2014_GLHG	701	MC	29	1	836	0.87	0.03		0.17	0.41	0.23	0.15		0.41	0.40
2014_GLHG	701	MC	30	1	836	0.87	0.04		0.21	0.12	0.55	0.08		0.55	0.58
2014_GLHG	702	MC	01	1	848	0.88	0.00		0.07	0.16	0.53	0.24		0.53	0.46
2014_GLHG	702	MC	02	1	848	0.88	0.00		0.07	0.84	0.06	0.03		0.84	0.43
2014_GLHG	702	MC	03	1	848	0.88	0.01		0.17	0.39	0.18	0.25		0.39	0.46
2014_GLHG	702	MC	04	1	848	0.88	0.01		0.57	0.12	0.09	0.21		0.57	0.32
2014_GLHG	702	MC	05	1	848	0.88	0.00		0.09	0.05	0.79	0.07		0.79	0.51
2014_GLHG	702	MC	06	1	848	0.88	0.00		0.83	0.05	0.08	0.03		0.83	0.51
2014_GLHG	702	MC	07	1	848	0.88	0.00		0.05	0.31	0.11	0.52		0.52	0.47
2014_GLHG	702	MC	08	1	848	0.88	0.01		0.24	0.49	0.06	0.19		0.49	0.33
2014_GLHG	702	MC	09	1	848	0.88	0.00		0.06	0.64	0.09	0.20		0.64	0.40
2014_GLHG	702	MC	10	1	848	0.88	0.01		0.54	0.14	0.18	0.12		0.54	0.35
2014_GLHG	702	MC	11	1	848	0.88	0.01		0.11	0.09	0.12	0.68		0.68	0.55
2014_GLHG	702	MC	12	1	848	0.88	0.01		0.71	0.14	0.04	0.09		0.71	0.56
2014_GLHG	702	MC	13	1	848	0.88	0.01		0.24	0.18	0.12	0.45		0.45	0.47
2014_GLHG	702	MC	14	1	848	0.88	0.01		0.50	0.05	0.08	0.35		0.50	0.40
2014_GLHG	702	MC	15	1	848	0.88	0.02		0.06	0.08	0.73	0.11		0.73	0.46

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	702	MC	16	1	848	0.88	0.02		0.17	0.09	0.11	0.60		0.60	0.57
2014_GLHG	702	MC	17	1	848	0.88	0.02		0.10	0.59	0.11	0.17		0.59	0.37
2014_GLHG	702	MC	18	1	848	0.88	0.02		0.22	0.21	0.08	0.48		0.48	0.43
2014_GLHG	702	MC	19	1	848	0.88	0.01		0.15	0.10	0.68	0.05		0.68	0.57
2014_GLHG	702	MC	20	1	848	0.88	0.02		0.27	0.43	0.16	0.13		0.43	0.45
2014_GLHG	702	MC	21	1	848	0.88	0.02		0.20	0.19	0.07	0.52		0.52	0.47
2014_GLHG	702	MC	22	1	848	0.88	0.03		0.19	0.43	0.16	0.20		0.43	0.43
2014_GLHG	702	MC	23	1	848	0.88	0.03		0.07	0.12	0.21	0.56		0.56	0.56
2014_GLHG	702	MC	24	1	848	0.88	0.03		0.11	0.18	0.58	0.10		0.58	0.49
2014_GLHG	702	MC	25	1	848	0.88	0.04		0.13	0.67	0.11	0.05		0.67	0.47
2014_GLHG	702	MC	26	1	848	0.88	0.04		0.65	0.09	0.16	0.05		0.65	0.53
2014_GLHG	702	MC	27	1	848	0.88	0.04		0.70	0.08	0.10	0.06		0.70	0.58
2014_GLHG	702	MC	28	1	848	0.88	0.05		0.14	0.13	0.17	0.52		0.52	0.54
2014_GLHG	702	MC	29	1	848	0.88	0.04		0.54	0.09	0.06	0.27		0.54	0.53
2014_GLHG	702	MC	30	1	848	0.88	0.04		0.09	0.56	0.10	0.22		0.56	0.46
2014_GLHG	702	MC	31	1	848	0.88	0.05		0.14	0.30	0.40	0.10		0.40	0.38
2014_GLHG	702	MC	32	1	848	0.88	0.05		0.49	0.19	0.10	0.17		0.49	0.42
2014_GLHG	703	MC	01	1	855	0.88	0.00		0.04	0.04	0.07	0.84		0.84	0.38
2014_GLHG	703	MC	02	1	855	0.88	0.01		0.68	0.10	0.10	0.11		0.68	0.42
2014_GLHG	703	MC	03	1	855	0.88	0.01		0.07	0.60	0.19	0.14		0.60	0.21
2014_GLHG	703	MC	04	1	855	0.88	0.01		0.69	0.07	0.08	0.15		0.69	0.34
2014_GLHG	703	MC	05	1	855	0.88	0.01		0.18	0.06	0.67	0.08		0.67	0.54
2014_GLHG	703	MC	06	1	855	0.88	0.01		0.09	0.17	0.52	0.21		0.52	0.48
2014_GLHG	703	MC	07	1	855	0.88	0.00		0.08	0.72	0.12	0.08		0.72	0.50
2014_GLHG	703	MC	08	1	855	0.88	0.01		0.64	0.16	0.14	0.05		0.64	0.47
2014_GLHG	703	MC	09	1	855	0.88	0.00		0.16	0.23	0.09	0.52		0.52	0.36
2014_GLHG	703	MC	10	1	855	0.88	0.00		0.35	0.30	0.18	0.17		0.35	0.31
2014_GLHG	703	MC	11	1	855	0.88	0.00		0.14	0.22	0.53	0.11		0.53	0.49

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	703	MC	12	1	855	0.88	0.01		0.15	0.12	0.13	0.59		0.59	0.57
2014_GLHG	703	MC	13	1	855	0.88	0.01		0.08	0.08	0.71	0.12		0.71	0.56
2014_GLHG	703	MC	14	1	855	0.88	0.01		0.12	0.27	0.53	0.07		0.53	0.36
2014_GLHG	703	MC	15	1	855	0.88	0.01		0.20	0.10	0.66	0.03		0.66	0.50
2014_GLHG	703	MC	16	1	855	0.88	0.01		0.26	0.13	0.10	0.50		0.50	0.51
2014_GLHG	703	MC	17	1	855	0.88	0.01		0.16	0.41	0.14	0.28		0.41	0.39
2014_GLHG	703	MC	18	1	855	0.88	0.02		0.07	0.15	0.11	0.66		0.66	0.45
2014_GLHG	703	MC	19	1	855	0.88	0.02		0.06	0.60	0.20	0.13		0.60	0.42
2014_GLHG	703	MC	20	1	855	0.88	0.02		0.07	0.07	0.09	0.74		0.74	0.61
2014_GLHG	703	MC	21	1	855	0.88	0.02		0.58	0.17	0.07	0.15		0.58	0.42
2014_GLHG	703	MC	22	1	855	0.88	0.03		0.21	0.11	0.19	0.46		0.46	0.38
2014_GLHG	703	MC	23	1	855	0.88	0.03		0.11	0.61	0.08	0.17		0.61	0.48
2014_GLHG	703	MC	24	1	855	0.88	0.02		0.10	0.11	0.67	0.09		0.67	0.58
2014_GLHG	703	MC	25	1	855	0.88	0.03		0.15	0.11	0.63	0.07		0.63	0.57
2014_GLHG	703	MC	26	1	855	0.88	0.04		0.56	0.15	0.14	0.12		0.56	0.54
2014_GLHG	703	MC	27	1	855	0.88	0.04		0.45	0.16	0.13	0.22		0.45	0.35
2014_GLHG	703	MC	28	1	855	0.88	0.04		0.11	0.68	0.11	0.06		0.68	0.60
2014_GLHG	703	MC	29	1	855	0.88	0.04		0.06	0.54	0.23	0.13		0.54	0.46
2014_GLHG	703	MC	30	1	855	0.88	0.04		0.68	0.11	0.08	0.09		0.68	0.46
2014_GLHG	703	MC	31	1	855	0.88	0.05		0.61	0.11	0.15	0.08		0.61	0.56
2014_GLHG	703	MC	32	1	855	0.88	0.05		0.13	0.12	0.60	0.10		0.60	0.48
2014_GLHG	704	MC	01	1	829	0.89	0.00		0.52	0.22	0.13	0.12		0.52	0.38
2014_GLHG	704	MC	02	1	829	0.89	0.00		0.16	0.11	0.68	0.04		0.68	0.43
2014_GLHG	704	MC	03	1	829	0.89	0.01		0.55	0.08	0.29	0.07		0.55	0.43
2014_GLHG	704	MC	04	1	829	0.89	0.01		0.09	0.67	0.08	0.15		0.67	0.33
2014_GLHG	704	MC	05	1	829	0.89	0.01		0.36	0.04	0.05	0.54		0.54	0.54
2014_GLHG	704	MC	06	1	829	0.89	0.01		0.58	0.09	0.18	0.14		0.58	0.47
2014_GLHG	704	MC	07	1	829	0.89	0.01		0.06	0.83	0.07	0.04		0.83	0.48

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	704	MC	08	1	829	0.89	0.01		0.15	0.07	0.69	0.08		0.69	0.54
2014_GLHG	704	MC	09	1	829	0.89	0.01		0.16	0.55	0.14	0.12		0.55	0.30
2014_GLHG	704	MC	10	1	829	0.89	0.01		0.51	0.08	0.12	0.27		0.51	0.47
2014_GLHG	704	MC	11	1	829	0.89	0.01		0.06	0.10	0.64	0.20		0.64	0.56
2014_GLHG	704	MC	12	1	829	0.89	0.01		0.03	0.06	0.08	0.81		0.81	0.56
2014_GLHG	704	MC	13	1	829	0.89	0.02		0.10	0.69	0.09	0.09		0.69	0.49
2014_GLHG	704	MC	14	1	829	0.89	0.02		0.12	0.17	0.12	0.57		0.57	0.49
2014_GLHG	704	MC	15	1	829	0.89	0.02		0.25	0.09	0.16	0.48		0.48	0.33
2014_GLHG	704	MC	16	1	829	0.89	0.02		0.71	0.08	0.06	0.12		0.71	0.49
2014_GLHG	704	MC	17	1	829	0.89	0.02		0.19	0.17	0.06	0.57		0.57	0.53
2014_GLHG	704	MC	18	1	829	0.89	0.03		0.12	0.71	0.08	0.06		0.71	0.57
2014_GLHG	704	MC	19	1	829	0.89	0.03		0.36	0.11	0.30	0.20		0.36	0.40
2014_GLHG	704	MC	20	1	829	0.89	0.03		0.07	0.11	0.47	0.31		0.47	0.46
2014_GLHG	704	MC	21	1	829	0.89	0.04		0.29	0.09	0.19	0.40		0.40	0.42
2014_GLHG	704	MC	22	1	829	0.89	0.05		0.20	0.10	0.14	0.51		0.51	0.58
2014_GLHG	704	MC	23	1	829	0.89	0.05		0.11	0.62	0.12	0.09		0.62	0.46
2014_GLHG	704	MC	24	1	829	0.89	0.05		0.13	0.61	0.07	0.14		0.61	0.52
2014_GLHG	704	MC	25	1	829	0.89	0.05		0.68	0.08	0.12	0.07		0.68	0.60
2014_GLHG	704	MC	26	1	829	0.89	0.05		0.05	0.08	0.66	0.17		0.66	0.40
2014_GLHG	704	MC	27	1	829	0.89	0.07		0.25	0.10	0.50	0.07		0.50	0.52
2014_GLHG	704	MC	28	1	829	0.89	0.07		0.52	0.15	0.15	0.10		0.52	0.58
2014_GLHG	704	MC	29	1	829	0.89	0.07		0.11	0.13	0.17	0.51		0.51	0.52
2014_GLHG	704	MC	30	1	829	0.89	0.07		0.23	0.18	0.46	0.07		0.46	0.53
2014_GLHG	704	MC	31	1	829	0.89	0.07		0.15	0.11	0.57	0.09		0.57	0.57
2014_GLHG	704	MC	32	1	829	0.89	0.09		0.11	0.48	0.21	0.11		0.48	0.38
2014_GLHG	705	MC	01	1	829	0.89	0.01		0.10	0.17	0.05	0.68		0.68	0.40
2014_GLHG	705	MC	02	1	829	0.89	0.01		0.03	0.77	0.10	0.09		0.77	0.42
2014_GLHG	705	MC	03	1	829	0.89	0.01		0.06	0.83	0.05	0.05		0.83	0.49

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	705	MC	04	1	829	0.89	0.01		0.29	0.19	0.17	0.34		0.34	0.31
2014_GLHG	705	MC	05	1	829	0.89	0.00		0.04	0.13	0.79	0.03		0.79	0.43
2014_GLHG	705	MC	06	1	829	0.89	0.01		0.64	0.12	0.17	0.07		0.64	0.49
2014_GLHG	705	MC	07	1	829	0.89	0.01		0.62	0.08	0.14	0.15		0.62	0.48
2014_GLHG	705	MC	08	1	829	0.89	0.01		0.05	0.08	0.15	0.70		0.70	0.54
2014_GLHG	705	MC	09	1	829	0.89	0.01		0.11	0.07	0.76	0.05		0.76	0.47
2014_GLHG	705	MC	10	1	829	0.89	0.01		0.23	0.56	0.15	0.05		0.56	0.46
2014_GLHG	705	MC	11	1	829	0.89	0.01		0.26	0.61	0.08	0.03		0.61	0.46
2014_GLHG	705	MC	12	1	829	0.89	0.01		0.74	0.09	0.11	0.04		0.74	0.51
2014_GLHG	705	MC	13	1	829	0.89	0.02		0.18	0.48	0.10	0.23		0.48	0.45
2014_GLHG	705	MC	14	1	829	0.89	0.01		0.04	0.05	0.03	0.86		0.86	0.48
2014_GLHG	705	MC	15	1	829	0.89	0.02		0.66	0.10	0.19	0.03		0.66	0.57
2014_GLHG	705	MC	16	1	829	0.89	0.02		0.05	0.04	0.79	0.10		0.79	0.51
2014_GLHG	705	MC	17	1	829	0.89	0.03		0.69	0.17	0.08	0.02		0.69	0.45
2014_GLHG	705	MC	18	1	829	0.89	0.03		0.21	0.52	0.14	0.09		0.52	0.34
2014_GLHG	705	MC	19	1	829	0.89	0.03		0.34	0.14	0.41	0.08		0.41	0.29
2014_GLHG	705	MC	20	1	829	0.89	0.03		0.17	0.45	0.13	0.23		0.45	0.33
2014_GLHG	705	MC	21	1	829	0.89	0.03		0.11	0.12	0.10	0.63		0.63	0.56
2014_GLHG	705	MC	22	1	829	0.89	0.03		0.51	0.23	0.14	0.08		0.51	0.41
2014_GLHG	705	MC	23	1	829	0.89	0.04		0.48	0.14	0.13	0.21		0.48	0.37
2014_GLHG	705	MC	24	1	829	0.89	0.04		0.07	0.08	0.74	0.08		0.74	0.55
2014_GLHG	705	MC	25	1	829	0.89	0.05		0.49	0.20	0.13	0.14		0.49	0.56
2014_GLHG	705	MC	26	1	829	0.89	0.05		0.14	0.13	0.12	0.55		0.55	0.61
2014_GLHG	705	MC	27	1	829	0.89	0.06		0.11	0.45	0.30	0.08		0.45	0.48
2014_GLHG	705	MC	28	1	829	0.89	0.06		0.10	0.09	0.69	0.06		0.69	0.65
2014_GLHG	705	MC	29	1	829	0.89	0.06		0.64	0.10	0.13	0.07		0.64	0.52
2014_GLHG	705	MC	30	1	829	0.89	0.06		0.08	0.12	0.61	0.12		0.61	0.52
2014_GLHG	705	MC	31	1	829	0.89	0.07		0.07	0.72	0.10	0.04		0.72	0.50

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	705	MC	32	1	829	0.89	0.08		0.62	0.19	0.07	0.04		0.62	0.53
2014_GLHG	706	MC	01	1	821	0.85	0.00		0.87	0.06	0.02	0.05		0.87	0.22
2014_GLHG	706	MC	02	1	821	0.85	0.00		0.13	0.63	0.14	0.09		0.63	0.39
2014_GLHG	706	MC	03	1	821	0.85	0.00		0.41	0.41	0.08	0.10		0.41	0.39
2014_GLHG	706	MC	04	1	821	0.85	0.00		0.06	0.06	0.81	0.08		0.81	0.44
2014_GLHG	706	MC	05	1	821	0.85	0.00		0.77	0.09	0.10	0.03		0.77	0.45
2014_GLHG	706	MC	06	1	821	0.85	0.00		0.17	0.04	0.12	0.67		0.67	0.49
2014_GLHG	706	MC	07	1	821	0.85	0.01		0.15	0.16	0.11	0.56		0.56	0.38
2014_GLHG	706	MC	08	1	821	0.85	0.01		0.08	0.16	0.59	0.17		0.59	0.59
2014_GLHG	706	MC	09	1	821	0.85	0.01		0.28	0.55	0.09	0.07		0.55	0.31
2014_GLHG	706	MC	10	1	821	0.85	0.01		0.12	0.14	0.58	0.15		0.58	0.52
2014_GLHG	706	MC	11	1	821	0.85	0.01		0.18	0.35	0.21	0.25		0.35	0.23
2014_GLHG	706	MC	12	1	821	0.85	0.01		0.17	0.15	0.14	0.53		0.53	0.27
2014_GLHG	706	MC	13	1	821	0.85	0.02		0.13	0.29	0.53	0.04		0.53	0.41
2014_GLHG	706	MC	14	1	821	0.85	0.02		0.07	0.14	0.73	0.04		0.73	0.50
2014_GLHG	706	MC	15	1	821	0.85	0.02		0.22	0.42	0.22	0.11		0.42	0.35
2014_GLHG	706	MC	16	1	821	0.85	0.02		0.17	0.07	0.70	0.04		0.70	0.49
2014_GLHG	706	MC	17	1	821	0.85	0.03		0.58	0.09	0.23	0.07		0.58	0.38
2014_GLHG	706	MC	18	1	821	0.85	0.02		0.03	0.24	0.22	0.48		0.48	0.22
2014_GLHG	706	MC	19	1	821	0.85	0.03		0.66	0.11	0.17	0.04		0.66	0.44
2014_GLHG	706	MC	20	1	821	0.85	0.02		0.39	0.43	0.04	0.11		0.43	0.37
2014_GLHG	706	MC	21	1	821	0.85	0.03		0.18	0.20	0.22	0.38		0.38	0.24
2014_GLHG	706	MC	22	1	821	0.85	0.03		0.17	0.58	0.13	0.08		0.58	0.50
2014_GLHG	706	MC	23	1	821	0.85	0.04		0.67	0.10	0.07	0.13		0.67	0.59
2014_GLHG	706	MC	24	1	821	0.85	0.03		0.05	0.07	0.77	0.07		0.77	0.53
2014_GLHG	706	MC	25	1	821	0.85	0.04		0.53	0.19	0.07	0.18		0.53	0.35
2014_GLHG	706	MC	26	1	821	0.85	0.04		0.08	0.65	0.07	0.16		0.65	0.52
2014_GLHG	706	MC	27	1	821	0.85	0.04		0.57	0.11	0.15	0.12		0.57	0.38

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	706	MC	28	1	821	0.85	0.05		0.13	0.60	0.15	0.07		0.60	0.52
2014_GLHG	706	MC	29	1	821	0.85	0.05		0.39	0.11	0.38	0.07		0.39	0.43
2014_GLHG	706	MC	30	1	821	0.85	0.06		0.06	0.13	0.10	0.65		0.65	0.60
2014_GLHG	706	MC	31	1	821	0.85	0.06		0.22	0.32	0.10	0.31		0.31	0.35
2014_GLHG	706	MC	32	1	821	0.85	0.05		0.62	0.17	0.07	0.09		0.62	0.56
2014_GLHG	707	MC	01	1	845	0.85	0.00		0.11	0.24	0.59	0.06		0.59	0.42
2014_GLHG	707	MC	02	1	845	0.85	0.02		0.24	0.21	0.13	0.40		0.40	0.39
2014_GLHG	707	MC	03	1	845	0.85	0.01		0.42	0.15	0.22	0.20		0.42	0.17
2014_GLHG	707	MC	04	1	845	0.85	0.01		0.31	0.08	0.55	0.05		0.55	0.36
2014_GLHG	707	MC	05	1	845	0.85	0.00		0.02	0.08	0.24	0.66		0.66	0.42
2014_GLHG	707	MC	06	1	845	0.85	0.01		0.60	0.15	0.18	0.07		0.60	0.36
2014_GLHG	707	MC	07	1	845	0.85	0.01		0.13	0.75	0.06	0.05		0.75	0.23
2014_GLHG	707	MC	08	1	845	0.85	0.01		0.55	0.16	0.21	0.07		0.55	0.29
2014_GLHG	707	MC	09	1	845	0.85	0.01		0.05	0.83	0.04	0.06		0.83	0.35
2014_GLHG	707	MC	10	1	845	0.85	0.01		0.57	0.20	0.11	0.11		0.57	0.48
2014_GLHG	707	MC	11	1	845	0.85	0.01		0.07	0.08	0.72	0.11		0.72	0.52
2014_GLHG	707	MC	12	1	845	0.85	0.01		0.63	0.19	0.09	0.07		0.63	0.59
2014_GLHG	707	MC	13	1	845	0.85	0.02		0.07	0.15	0.64	0.12		0.64	0.40
2014_GLHG	707	MC	14	1	845	0.85	0.02		0.12	0.39	0.22	0.24		0.39	0.43
2014_GLHG	707	MC	15	1	845	0.85	0.02		0.56	0.10	0.11	0.21		0.56	0.38
2014_GLHG	707	MC	16	1	845	0.85	0.03		0.22	0.51	0.14	0.10		0.51	0.53
2014_GLHG	707	MC	17	1	845	0.85	0.02		0.45	0.15	0.14	0.23		0.45	0.40
2014_GLHG	707	MC	18	1	845	0.85	0.03		0.07	0.28	0.08	0.55		0.55	0.59
2014_GLHG	707	MC	19	1	845	0.85	0.03		0.19	0.09	0.63	0.06		0.63	0.46
2014_GLHG	707	MC	20	1	845	0.85	0.03		0.12	0.11	0.10	0.64		0.64	0.50
2014_GLHG	707	MC	21	1	845	0.85	0.03		0.33	0.36	0.24	0.04		0.36	0.33
2014_GLHG	707	MC	22	1	845	0.85	0.03		0.07	0.09	0.11	0.70		0.70	0.46
2014_GLHG	707	MC	23	1	845	0.85	0.03		0.16	0.61	0.12	0.07		0.61	0.52

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	707	MC	24	1	845	0.85	0.04		0.41	0.13	0.29	0.12		0.41	0.26
2014_GLHG	707	MC	25	1	845	0.85	0.05		0.14	0.17	0.60	0.05		0.60	0.54
2014_GLHG	707	MC	26	1	845	0.85	0.05		0.13	0.08	0.63	0.11		0.63	0.54
2014_GLHG	707	MC	27	1	845	0.85	0.05		0.09	0.08	0.38	0.40		0.40	0.43
2014_GLHG	707	MC	28	1	845	0.85	0.05		0.26	0.46	0.06	0.17		0.46	0.17
2014_GLHG	707	MC	29	1	845	0.85	0.05		0.16	0.21	0.48	0.10		0.48	0.41
2014_GLHG	707	MC	30	1	845	0.85	0.07		0.14	0.55	0.08	0.15		0.55	0.43
2014_GLHG	707	MC	31	1	845	0.85	0.06		0.18	0.23	0.37	0.17		0.37	0.38
2014_GLHG	707	MC	32	1	845	0.85	0.06		0.12	0.61	0.14	0.08		0.61	0.56
2014_GLHG	708	MC	01	1	833	0.86	0.01		0.70	0.07	0.14	0.08		0.70	0.44
2014_GLHG	708	MC	02	1	833	0.86	0.00		0.83	0.08	0.06	0.03		0.83	0.41
2014_GLHG	708	MC	03	1	833	0.86	0.00		0.03	0.81	0.13	0.03		0.81	0.31
2014_GLHG	708	MC	04	1	833	0.86	0.01		0.05	0.13	0.67	0.14		0.67	0.41
2014_GLHG	708	MC	05	1	833	0.86	0.01		0.14	0.57	0.21	0.08		0.57	0.45
2014_GLHG	708	MC	06	1	833	0.86	0.00		0.15	0.42	0.03	0.39		0.39	0.41
2014_GLHG	708	MC	07	1	833	0.86	0.00		0.08	0.54	0.13	0.24		0.54	0.40
2014_GLHG	708	MC	08	1	833	0.86	0.01		0.55	0.17	0.15	0.12		0.55	0.56
2014_GLHG	708	MC	09	1	833	0.86	0.01		0.31	0.20	0.21	0.26		0.21	0.30
2014_GLHG	708	MC	10	1	833	0.86	0.01		0.11	0.22	0.50	0.17		0.50	0.47
2014_GLHG	708	MC	11	1	833	0.86	0.02		0.63	0.11	0.17	0.08		0.63	0.43
2014_GLHG	708	MC	12	1	833	0.86	0.02		0.24	0.45	0.19	0.10		0.10	0.15
2014_GLHG	708	MC	13	1	833	0.86	0.02		0.15	0.14	0.54	0.16		0.54	0.50
2014_GLHG	708	MC	14	1	833	0.86	0.02		0.63	0.11	0.12	0.12		0.63	0.50
2014_GLHG	708	MC	15	1	833	0.86	0.02		0.32	0.44	0.13	0.08		0.44	0.48
2014_GLHG	708	MC	16	1	833	0.86	0.03		0.12	0.51	0.19	0.16		0.51	0.33
2014_GLHG	708	MC	17	1	833	0.86	0.02		0.22	0.19	0.10	0.47		0.47	0.32
2014_GLHG	708	MC	18	1	833	0.86	0.03		0.47	0.22	0.13	0.15		0.47	0.43
2014_GLHG	708	MC	19	1	833	0.86	0.03		0.45	0.15	0.16	0.20		0.45	0.52

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	708	MC	20	1	833	0.86	0.03		0.43	0.12	0.32	0.10		0.32	0.20
2014_GLHG	708	MC	21	1	833	0.86	0.04		0.21	0.15	0.22	0.38		0.38	0.46
2014_GLHG	708	MC	22	1	833	0.86	0.04		0.12	0.29	0.07	0.48		0.48	0.49
2014_GLHG	708	MC	23	1	833	0.86	0.04		0.10	0.60	0.08	0.17		0.60	0.54
2014_GLHG	708	MC	24	1	833	0.86	0.05		0.22	0.12	0.43	0.17		0.43	0.32
2014_GLHG	708	MC	25	1	833	0.86	0.06		0.50	0.09	0.23	0.13		0.50	0.42
2014_GLHG	708	MC	26	1	833	0.86	0.06		0.10	0.13	0.54	0.17		0.54	0.44
2014_GLHG	708	MC	27	1	833	0.86	0.07		0.20	0.52	0.14	0.06		0.52	0.47
2014_GLHG	708	MC	28	1	833	0.86	0.06		0.14	0.11	0.60	0.08		0.60	0.50
2014_GLHG	708	MC	29	1	833	0.86	0.06		0.07	0.09	0.73	0.05		0.73	0.55
2014_GLHG	708	MC	30	1	833	0.86	0.07		0.62	0.12	0.09	0.10		0.62	0.48
2014_GLHG	708	MC	31	1	833	0.86	0.08		0.18	0.12	0.19	0.43		0.43	0.48
2014_GLHG	708	MC	32	1	833	0.86	0.09		0.20	0.48	0.15	0.09		0.48	0.48
2014_GLHG	709	MC	01	1	833	0.86	0.00		0.45	0.19	0.20	0.15		0.45	0.35
2014_GLHG	709	MC	02	1	833	0.86	0.00		0.29	0.30	0.23	0.17		0.17	0.09
2014_GLHG	709	MC	03	1	833	0.86	0.00		0.22	0.08	0.62	0.08		0.62	0.42
2014_GLHG	709	MC	04	1	833	0.86	0.00		0.09	0.05	0.03	0.83		0.83	0.49
2014_GLHG	709	MC	05	1	833	0.86	0.00		0.17	0.75	0.05	0.03		0.75	0.47
2014_GLHG	709	MC	06	1	833	0.86	0.01		0.31	0.19	0.24	0.25		0.31	0.35
2014_GLHG	709	MC	07	1	833	0.86	0.01		0.25	0.12	0.50	0.11		0.50	0.42
2014_GLHG	709	MC	08	1	833	0.86	0.01		0.13	0.18	0.08	0.60		0.60	0.50
2014_GLHG	709	MC	09	1	833	0.86	0.00		0.06	0.03	0.19	0.72		0.72	0.44
2014_GLHG	709	MC	10	1	833	0.86	0.01		0.10	0.40	0.30	0.19		0.30	0.39
2014_GLHG	709	MC	11	1	833	0.86	0.01		0.11	0.17	0.07	0.64		0.64	0.44
2014_GLHG	709	MC	12	1	833	0.86	0.01		0.70	0.12	0.13	0.04		0.70	0.51
2014_GLHG	709	MC	13	1	833	0.86	0.01		0.11	0.62	0.11	0.15		0.62	0.43
2014_GLHG	709	MC	14	1	833	0.86	0.01		0.66	0.13	0.12	0.08		0.66	0.53
2014_GLHG	709	MC	15	1	833	0.86	0.02		0.12	0.08	0.56	0.21		0.56	0.31

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	709	MC	16	1	833	0.86	0.02		0.08	0.05	0.73	0.12		0.73	0.51
2014_GLHG	709	MC	17	1	833	0.86	0.03		0.41	0.28	0.13	0.15		0.41	0.43
2014_GLHG	709	MC	18	1	833	0.86	0.03		0.06	0.69	0.14	0.08		0.69	0.57
2014_GLHG	709	MC	19	1	833	0.86	0.03		0.11	0.52	0.25	0.09		0.52	0.45
2014_GLHG	709	MC	20	1	833	0.86	0.03		0.69	0.10	0.09	0.09		0.69	0.59
2014_GLHG	709	MC	21	1	833	0.86	0.03		0.33	0.17	0.10	0.37		0.37	0.41
2014_GLHG	709	MC	22	1	833	0.86	0.03		0.16	0.57	0.18	0.06		0.57	0.42
2014_GLHG	709	MC	23	1	833	0.86	0.03		0.23	0.08	0.48	0.17		0.48	0.39
2014_GLHG	709	MC	24	1	833	0.86	0.03		0.16	0.08	0.16	0.57		0.57	0.49
2014_GLHG	709	MC	25	1	833	0.86	0.03		0.06	0.60	0.17	0.14		0.60	0.59
2014_GLHG	709	MC	26	1	833	0.86	0.04		0.32	0.21	0.37	0.07		0.37	0.34
2014_GLHG	709	MC	27	1	833	0.86	0.04		0.66	0.13	0.10	0.08		0.66	0.48
2014_GLHG	709	MC	28	1	833	0.86	0.04		0.49	0.16	0.12	0.19		0.49	0.29
2014_GLHG	709	MC	29	1	833	0.86	0.04		0.11	0.63	0.13	0.09		0.63	0.57
2014_GLHG	709	MC	30	1	833	0.86	0.04		0.14	0.06	0.55	0.21		0.55	0.37
2014_GLHG	709	MC	31	1	833	0.86	0.09		0.46	0.10	0.33	0.03		0.33	0.20
2014_GLHG	709	MC	32	1	833	0.86	0.09		0.20	0.11	0.18	0.42		0.42	0.59
2014_GLHG	710_711	ESS	THM	5	732	0.90	0.21	0.29	0.35	0.13	0.02	0.00	0.00	0.68	0.54
2014_GLHG	712_713	ESS	THM	5	708	0.90	0.22	0.16	0.34	0.22	0.06	0.00	0.00	0.97	0.54
2014_GLHG	714_715	ESS	THM	5	712	0.89	0.20	0.15	0.39	0.20	0.05	0.00	0.00	0.96	0.54
2014_GLHG	716_717	ESS	THM	5	678	0.89	0.24	0.21	0.37	0.14	0.03	0.00	0.00	0.76	0.55
2014_GLHG	718_719	ESS	THM	5	674	0.88	0.22	0.27	0.36	0.14	0.02	0.00	0.00	0.69	0.57
2014_GLHG	720	ESS	THM	5	360	0.90	0.24	0.19	0.37	0.16	0.04	0.00	0.00	0.82	0.52
2014_GLHG	721	ESS	THM	5	349	0.88	0.23	0.36	0.29	0.11	0.01	0.00	0.00	0.53	0.60
2014_GLHG	722_723	SCF	01	1	991	0.91	0.08	0.08	0.84					0.84	0.40
2014_GLHG	722_723	SCF	02	1	991	0.91	0.12	0.38	0.50					0.50	0.48
2014_GLHG	722_723	SCF	03	2	991	0.91	0.12	0.05	0.16	0.67				1.51	0.56
2014_GLHG	722_723	SCF	04	2	991	0.91	0.13	0.05	0.17	0.65				1.46	0.56

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	722_723	SCF	5a	1	991	0.91	0.13	0.14	0.73					0.73	0.47
2014_GLHG	722_723	SCF	5b	1	991	0.91	0.14	0.12	0.73					0.73	0.52
2014_GLHG	722_723	SCF	06	1	991	0.91	0.14	0.06	0.80					0.80	0.50
2014_GLHG	722_723	SCF	07	1	991	0.91	0.17	0.32	0.50					0.50	0.49
2014_GLHG	722_723	SCF	08	1	991	0.91	0.18	0.14	0.68					0.68	0.59
2014_GLHG	722_723	SCF	09	1	991	0.91	0.19	0.24	0.57					0.57	0.46
2014_GLHG	722_723	ESS	DBQ	5	991	0.91	0.07	0.31	0.29	0.27	0.07	0.00	0.00	1.03	0.56
2014_GLHG	724_725	SCF	1a	1	1017	0.91	0.08	0.17	0.75					0.75	0.44
2014_GLHG	724_725	SCF	1b	1	1017	0.91	0.11	0.27	0.62					0.62	0.55
2014_GLHG	724_725	SCF	02	1	1017	0.91	0.09	0.35	0.56					0.56	0.49
2014_GLHG	724_725	SCF	3a	1	1017	0.91	0.10	0.11	0.79					0.79	0.54
2014_GLHG	724_725	SCF	3b	1	1017	0.91	0.10	0.10	0.80					0.80	0.58
2014_GLHG	724_725	SCF	4a	1	1017	0.91	0.10	0.28	0.62					0.62	0.43
2014_GLHG	724_725	SCF	4b	1	1017	0.91	0.12	0.06	0.83					0.83	0.57
2014_GLHG	724_725	SCF	5a	1	1017	0.91	0.14	0.22	0.63					0.63	0.56
2014_GLHG	724_725	SCF	5b	1	1017	0.91	0.15	0.26	0.59					0.59	0.55
2014_GLHG	724_725	SCF	06	1	1017	0.91	0.15	0.14	0.71					0.71	0.52
2014_GLHG	724_725	SCF	7a	1	1017	0.91	0.15	0.10	0.76					0.76	0.54
2014_GLHG	724_725	SCF	7b	1	1017	0.91	0.15	0.05	0.80					0.80	0.60
2014_GLHG	724_725	SCF	8a	1	1017	0.91	0.17	0.08	0.75					0.75	0.59
2014_GLHG	724_725	SCF	8b	1	1017	0.91	0.18	0.15	0.67					0.67	0.60
2014_GLHG	724_725	ESS	DBQ	5	1017	0.91	0.06	0.29	0.27	0.28	0.09	0.00	0.00	1.10	0.55
2014_GLHG	726_727	SCF	01	2	995	0.91	0.08	0.10	0.32	0.49				1.30	0.58
2014_GLHG	726_727	SCF	02	1	995	0.91	0.10	0.34	0.56					0.56	0.50
2014_GLHG	726_727	SCF	3a	1	995	0.91	0.12	0.14	0.73					0.73	0.47
2014_GLHG	726_727	SCF	3b	1	995	0.91	0.13	0.10	0.76					0.76	0.50
2014_GLHG	726_727	SCF	04	1	995	0.91	0.13	0.18	0.69					0.69	0.55
2014_GLHG	726_727	SCF	05	1	995	0.91	0.16	0.15	0.69					0.69	0.57

Test	Form	Type	Item	Max	N-Count	Alpha	B	M0	M1	M2	M3	M4	M5	Mean	Point-Biserial
2014_GLHG	726_727	SCF	06	1	995	0.91	0.15	0.09	0.76					0.76	0.58
2014_GLHG	726_727	SCF	07	1	995	0.91	0.17	0.19	0.64					0.64	0.61
2014_GLHG	726_727	SCF	08	1	995	0.91	0.19	0.20	0.61					0.61	0.53
2014_GLHG	726_727	SCF	09	2	995	0.91	0.20	0.09	0.19	0.52				1.23	0.59
2014_GLHG	726_727	ESS	DBQ	5	995	0.91	0.06	0.35	0.29	0.26	0.04	0.00	0.00	0.92	0.59
2014_GLHG	728	SCF	01	1	491	0.90	0.09	0.07	0.84					0.84	0.45
2014_GLHG	728	SCF	02	1	491	0.90	0.09	0.24	0.67					0.67	0.39
2014_GLHG	728	SCF	03	2	491	0.90	0.11	0.06	0.32	0.51				1.34	0.51
2014_GLHG	728	SCF	04	1	491	0.90	0.12	0.17	0.72					0.72	0.49
2014_GLHG	728	SCF	5a	1	491	0.90	0.14	0.08	0.78					0.78	0.52
2014_GLHG	728	SCF	5b	1	491	0.90	0.16	0.18	0.67					0.67	0.56
2014_GLHG	728	SCF	06	1	491	0.90	0.15	0.14	0.70					0.70	0.49
2014_GLHG	728	SCF	7a	1	491	0.90	0.16	0.44	0.40					0.40	0.41
2014_GLHG	728	SCF	7b	1	491	0.90	0.18	0.47	0.34					0.34	0.48
2014_GLHG	728	SCF	08	1	491	0.90	0.20	0.27	0.53					0.53	0.43
2014_GLHG	728	SCF	9a	1	491	0.90	0.22	0.25	0.53					0.53	0.41
2014_GLHG	728	SCF	9b	1	491	0.90	0.23	0.20	0.57					0.57	0.46
2014_GLHG	728	ESS	DBQ	5	491	0.90	0.07	0.35	0.28	0.24	0.06	0.00	0.00	0.93	0.59

Appendix B: Inter-rater Consistency – Point Differences Between First and Second Reads

The first three columns from the left contain the form ID, item sequence number, and number of score points for each item. The remaining columns contain the percentage of times each possible difference between the first and second raters' scores occurred. Blank cells indicate out-of-range differences (e.g., differences greater than the maximum possible given the point value of that particular item).

Form	Item	Score Points	Difference (First Read Minus Second Read)										
			−5	−4	−3	−2	−1	0	1	2	3	4	5
710_711	31	5	0%	0%	0%	0%	15%	73%	11%	1%	0%	0%	0%
712_713	31	5	0%	0%	0%	1%	12%	75%	12%	0%	0%	0%	0%
714_715	31	5	0%	0%	0%	0%	13%	83%	5%	0%	0%	0%	0%
716_717	31	5	0%	0%	0%	0%	19%	67%	12%	2%	0%	0%	0%
718_719	31	5	0%	0%	0%	0%	12%	76%	12%	0%	0%	0%	0%
720	31	5	0%	0%	0%	0%	6%	82%	12%	0%	0%	0%	0%
721	31	5	0%	0%	0%	0%	12%	76%	9%	3%	0%	0%	0%
722_723	31	1					1%	95%	4%				
722_723	32	1					4%	85%	10%				
722_723	33	2				0%	4%	90%	5%	1%			
722_723	34	2				1%	5%	88%	7%	0%			
722_723	35	1					5%	92%	3%				
722_723	36	1					1%	97%	2%				
722_723	37	1					2%	98%	1%				
722_723	38	1					5%	92%	3%				
722_723	39	1					2%	97%	1%				
722_723	40	1					5%	93%	2%				
722_723	41	5	0%	0%	0%	1%	10%	76%	14%	0%	0%	0%	0%
724_725	31	1					2%	95%	3%				
724_725	32	1					1%	98%	1%				
724_725	33	1					2%	98%	0%				
724_725	34	1					3%	94%	3%				
724_725	35	1					1%	96%	3%				
724_725	36	1					9%	85%	6%				
724_725	37	1					2%	97%	1%				
724_725	38	1					3%	86%	11%				
724_725	39	1					3%	95%	2%				
724_725	40	1					3%	95%	2%				
724_725	41	1					1%	98%	1%				
724_725	42	1					0%	100%	0%				

Form	Item	Score Points	Difference (First Read Minus Second Read)										
			−5	−4	−3	−2	−1	0	1	2	3	4	5
724_725	43	1					2%	98%	1%				
724_725	44	1					3%	94%	3%				
724_725	45	5	0%	0%	0%	0%	15%	70%	14%	1%	0%	0%	0%
726_727	31	2				0%	4%	91%	5%	0%			
726_727	32	1					3%	90%	7%				
726_727	33	1					1%	97%	2%				
726_727	34	1					2%	97%	1%				
726_727	35	1					5%	92%	3%				
726_727	36	1					3%	93%	4%				
726_727	37	1					3%	95%	2%				
726_727	38	1					4%	90%	6%				
726_727	39	1					3%	93%	4%				
726_727	40	2				1%	12%	80%	7%	0%			
726_727	41	5	0%	0%	0%	0%	9%	82%	9%	0%	0%	0%	0%
728	31	1					2%	98%	0%				
728	32	1					7%	88%	5%				
728	33	2				3%	17%	62%	16%	2%			
728	34	1					7%	90%	3%				
728	35	1					1%	97%	2%				
728	36	1					0%	100%	0%				
728	37	1					1%	98%	1%				
728	38	1					13%	82%	5%				
728	39	1					5%	93%	1%				
728	40	1					8%	85%	8%				
728	41	1					8%	89%	3%				
728	42	1					3%	88%	9%				
728	43	5	0%	0%	0%	0%	9%	80%	11%	0%	0%	0%	0%

Appendix C: Additional Measures of Inter-rater Reliability and Agreement

The first four columns from the left contain the form ID, item sequence number, number of score points, and the total count of items receiving a first and second read. In the fifth column the percent of exact matches between the first and second scores is provided. The following column ("Adj.") is the percentage of the first and second scores with a difference of −1 or 1. "Total" is the sum of Exact and Adjacent matches (e.g., the two prior columns).

Form	Item	Score Points	Total N-Count	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intraclass Corr.	Weighted Kappa
				Exact	Adj.	Total	First Read	Second Read	First Read	Second Read		
710_711	31	5	125	72.8%	26.4%	99.2%	0.8	0.8	0.74	0.76	0.74	0.64
712_713	31	5	115	74.8%	24.3%	99.1%	1.1	1.1	0.84	0.90	0.82	0.72
714_715	31	5	126	82.5%	17.5%	100.0%	1.0	1.1	0.79	0.80	0.86	0.80
716_717	31	5	109	67.0%	31.2%	98.2%	1.0	1.1	0.74	0.84	0.69	0.58
718_719	31	5	117	76.1%	23.9%	100.0%	0.9	0.9	0.79	0.80	0.81	0.71
720	31	5	65	81.5%	18.5%	100.0%	1.0	0.9	1.01	0.92	0.90	0.82
721	31	5	67	76.1%	20.9%	97.0%	0.6	0.5	0.70	0.59	0.61	0.59
722_723	31	1	171	94.7%	5.3%	100.0%	0.9	0.9	0.30	0.34	0.74	0.74
722_723	32	1	182	85.2%	14.8%	100.0%	0.6	0.5	0.49	0.50	0.70	0.70
722_723	33	2	169	89.9%	9.5%	99.4%	1.6	1.5	0.74	0.73	0.89	0.84
722_723	34	2	178	87.6%	11.8%	99.4%	1.5	1.5	0.75	0.74	0.87	0.82
722_723	35	1	176	92.0%	8.0%	100.0%	0.8	0.8	0.43	0.41	0.77	0.77
722_723	36	1	179	97.2%	2.8%	100.0%	0.8	0.8	0.43	0.43	0.92	0.92
722_723	37	1	177	97.7%	2.3%	100.0%	0.8	0.8	0.38	0.37	0.92	0.92
722_723	38	1	181	91.7%	8.3%	100.0%	0.5	0.5	0.50	0.50	0.84	0.83
722_723	39	1	176	97.2%	2.8%	100.0%	0.7	0.7	0.44	0.44	0.93	0.93
722_723	40	1	174	92.5%	7.5%	100.0%	0.6	0.7	0.49	0.48	0.84	0.84
722_723	41	5	183	76.0%	23.5%	99.5%	1.4	1.4	0.86	0.87	0.83	0.74
724_725	31	1	180	95.0%	5.0%	100.0%	0.8	0.8	0.41	0.42	0.85	0.85

Form	Item	Score Points	Total N-Count	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intraclass Corr.	Weighted Kappa
				Exact	Adj.	Total	First Read	Second Read	First Read	Second Read		
724_725	32	1	179	98.3%	1.7%	100.0%	0.6	0.6	0.49	0.49	0.96	0.96
724_725	33	1	181	98.3%	1.7%	100.0%	0.5	0.5	0.50	0.50	0.97	0.97
724_725	34	1	178	94.4%	5.6%	100.0%	0.8	0.8	0.39	0.39	0.81	0.81
724_725	35	1	176	96.0%	4.0%	100.0%	0.9	0.8	0.34	0.36	0.84	0.84
724_725	36	1	176	85.2%	14.8%	100.0%	0.6	0.6	0.49	0.48	0.69	0.69
724_725	37	1	182	96.7%	3.3%	100.0%	0.8	0.8	0.38	0.37	0.88	0.88
724_725	38	1	179	86.0%	14.0%	100.0%	0.7	0.6	0.46	0.48	0.69	0.69
724_725	39	1	181	95.0%	5.0%	100.0%	0.6	0.6	0.49	0.48	0.89	0.89
724_725	40	1	178	94.9%	5.1%	100.0%	0.7	0.7	0.45	0.45	0.87	0.87
724_725	41	1	179	98.3%	1.7%	100.0%	0.7	0.7	0.44	0.44	0.96	0.96
724_725	42	1	181	100.0%	0.0%	100.0%	0.9	0.9	0.36	0.36	1.00	1.00
724_725	43	1	177	97.7%	2.3%	100.0%	0.8	0.8	0.41	0.40	0.93	0.93
724_725	44	1	179	94.4%	5.6%	100.0%	0.7	0.7	0.46	0.46	0.87	0.86
724_725	45	5	182	70.3%	29.1%	99.5%	1.3	1.3	0.99	1.03	0.85	0.73
726_727	31	2	174	91.4%	8.6%	100.0%	1.3	1.3	0.73	0.72	0.92	0.89
726_727	32	1	185	90.3%	9.7%	100.0%	0.6	0.6	0.49	0.50	0.80	0.80
726_727	33	1	181	96.7%	3.3%	100.0%	0.8	0.8	0.42	0.43	0.91	0.91
726_727	34	1	184	97.3%	2.7%	100.0%	0.8	0.8	0.43	0.42	0.92	0.92
726_727	35	1	173	91.9%	8.1%	100.0%	0.7	0.7	0.47	0.46	0.81	0.81
726_727	36	1	184	93.5%	6.5%	100.0%	0.7	0.7	0.44	0.45	0.84	0.84
726_727	37	1	174	94.8%	5.2%	100.0%	0.8	0.8	0.42	0.41	0.85	0.85
726_727	38	1	187	90.4%	9.6%	100.0%	0.6	0.6	0.49	0.49	0.80	0.80
726_727	39	1	187	93.0%	7.0%	100.0%	0.6	0.6	0.48	0.48	0.85	0.85
726_727	40	2	186	80.1%	19.4%	99.5%	1.3	1.3	0.83	0.83	0.84	0.76
726_727	41	5	198	82.3%	17.7%	100.0%	1.0	1.0	0.93	0.93	0.90	0.83
728	31	1	88	97.7%	2.3%	100.0%	0.9	0.9	0.29	0.25	0.85	0.85

Form	Item	Score Points	Total N-Count	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intraclass Corr.	Weighted Kappa
				Exact	Adj.	Total	First Read	Second Read	First Read	Second Read		
728	32	1	91	87.9%	12.1%	100.0%	0.6	0.6	0.49	0.48	0.74	0.74
728	33	2	90	62.2%	32.2%	94.4%	1.4	1.4	0.68	0.67	0.41	0.37
728	34	1	91	90.1%	9.9%	100.0%	0.7	0.7	0.45	0.44	0.75	0.75
728	35	1	94	96.8%	3.2%	100.0%	0.8	0.8	0.38	0.39	0.89	0.89
728	36	1	90	100.0%	0.0%	100.0%	0.7	0.7	0.46	0.46	1.00	1.00
728	37	1	89	97.8%	2.2%	100.0%	0.8	0.8	0.43	0.43	0.94	0.94
728	38	1	92	81.5%	18.5%	100.0%	0.4	0.5	0.49	0.50	0.63	0.63
728	39	1	91	93.4%	6.6%	100.0%	0.3	0.3	0.46	0.48	0.85	0.85
728	40	1	92	84.8%	15.2%	100.0%	0.5	0.5	0.50	0.50	0.70	0.69
728	41	1	93	89.2%	10.8%	100.0%	0.6	0.6	0.50	0.49	0.78	0.78
728	42	1	93	88.2%	11.8%	100.0%	0.6	0.6	0.49	0.50	0.76	0.76
728	43	5	89	79.8%	20.2%	100.0%	1.0	1.0	0.90	0.90	0.88	0.79

Appendix D: Partial-Credit Model Item Analysis

The first five columns from the left contain the test name, form name, item type, item number on the form, and maximum points possible for the item. The sixth column contains the number of students that the item was administered to. The remaining six columns contain the Rasch Item Difficulty, step difficulties (for multi–point items only), and the INFIT Rasch model fit statistic. Items without statistics are DNS (Do Not Score) status items.

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	701	MC	01	1	836	−1.1500						0.85
2014_GLHG	701	MC	02	1	836	−0.6300						0.99
2014_GLHG	701	MC	03	1	836	−0.0400						1.00
2014_GLHG	701	MC	04	1	836	0.6000						1.00
2014_GLHG	701	MC	05	1	836	0.3500						1.10
2014_GLHG	701	MC	06	1	836	−0.3400						1.10
2014_GLHG	701	MC	07	1	836	−0.0800						1.02
2014_GLHG	701	MC	08	1	836	0.1300						1.19
2014_GLHG	701	MC	09	1	836	−1.7000						1.04
2014_GLHG	701	MC	10	1	836	−0.5400						0.94
2014_GLHG	701	MC	11	1	836	0.2000						0.99
2014_GLHG	701	MC	12	1	836	−1.2500						0.87
2014_GLHG	701	MC	13	1	836	0.2600						0.94
2014_GLHG	701	MC	14	1	836	0.5600						1.14
2014_GLHG	701	MC	15	1	836	−0.5300						0.92
2014_GLHG	701	MC	16	1	836	0.0200						1.11
2014_GLHG	701	MC	17	1	836	0.2200						1.02
2014_GLHG	701	MC	18	1	836	−0.2300						0.89
2014_GLHG	701	MC	19	1	836	0.0100						0.89
2014_GLHG	701	MC	20	1	836	0.7500						1.25
2014_GLHG	701	MC	21	1	836	0.0300						0.96

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	701	MC	22	1	836	−0.7100						0.98
2014_GLHG	701	MC	23	1	836	−0.3400						0.93
2014_GLHG	701	MC	24	1	836	−0.0200						1.14
2014_GLHG	701	MC	25	1	836	0.1100						1.05
2014_GLHG	701	MC	26	1	836	0.0900						1.11
2014_GLHG	701	MC	27	1	836	1.0100						1.02
2014_GLHG	701	MC	28	1	836	−0.3700						0.89
2014_GLHG	701	MC	29	1	836	0.8400						1.09
2014_GLHG	701	MC	30	1	836	0.0200						0.87
2014_GLHG	702	MC	01	1	848	0.2525						1.03
2014_GLHG	702	MC	02	1	848	−1.5956						0.93
2014_GLHG	702	MC	03	1	848	0.9770						0.99
2014_GLHG	702	MC	04	1	848	0.0726						1.19
2014_GLHG	702	MC	05	1	848	−1.1888						0.89
2014_GLHG	702	MC	06	1	848	−1.5561						0.86
2014_GLHG	702	MC	07	1	848	0.2883						1.01
2014_GLHG	702	MC	08	1	848	0.4372						1.18
2014_GLHG	702	MC	09	1	848	−0.3158						1.08
2014_GLHG	702	MC	10	1	848	0.2107						1.17
2014_GLHG	702	MC	11	1	848	−0.5166						0.89
2014_GLHG	702	MC	12	1	848	−0.6996						0.87
2014_GLHG	702	MC	13	1	848	0.6767						1.01
2014_GLHG	702	MC	14	1	848	0.3955						1.11
2014_GLHG	702	MC	15	1	848	−0.8343						0.96
2014_GLHG	702	MC	16	1	848	−0.1283						0.89
2014_GLHG	702	MC	17	1	848	−0.0608						1.13
2014_GLHG	702	MC	18	1	848	0.4968						1.06
2014_GLHG	702	MC	19	1	848	−0.5099						0.86

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	702	MC	20	1	848	0.7858						1.02
2014_GLHG	702	MC	21	1	848	0.3240						1.01
2014_GLHG	702	MC	22	1	848	0.7858						1.04
2014_GLHG	702	MC	23	1	848	0.1148						0.90
2014_GLHG	702	MC	24	1	848	0.0061						0.98
2014_GLHG	702	MC	25	1	848	−0.4574						0.99
2014_GLHG	702	MC	26	1	848	−0.3861						0.92
2014_GLHG	702	MC	27	1	848	−0.6719						0.84
2014_GLHG	702	MC	28	1	848	0.3240						0.93
2014_GLHG	702	MC	29	1	848	0.2286						0.93
2014_GLHG	702	MC	30	1	848	0.1208						1.02
2014_GLHG	702	MC	31	1	848	0.9085						1.11
2014_GLHG	702	MC	32	1	848	0.4670						1.08
2014_GLHG	703	MC	01	1	855	−1.6193						0.97
2014_GLHG	703	MC	02	1	855	−0.5198						1.04
2014_GLHG	703	MC	03	1	855	−0.0946						1.32
2014_GLHG	703	MC	04	1	855	−0.5657						1.12
2014_GLHG	703	MC	05	1	855	−0.4808						0.90
2014_GLHG	703	MC	06	1	855	0.3215						1.00
2014_GLHG	703	MC	07	1	855	−0.7691						0.92
2014_GLHG	703	MC	08	1	855	−0.2907						0.98
2014_GLHG	703	MC	09	1	855	0.3156						1.14
2014_GLHG	703	MC	10	1	855	1.1969						1.17
2014_GLHG	703	MC	11	1	855	0.2390						0.99
2014_GLHG	703	MC	12	1	855	−0.0343						0.88
2014_GLHG	703	MC	13	1	855	−0.7275						0.86
2014_GLHG	703	MC	14	1	855	0.2449						1.14
2014_GLHG	703	MC	15	1	855	−0.4103						0.95

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	703	MC	16	1	855	0.4099						0.95
2014_GLHG	703	MC	17	1	855	0.8883						1.09
2014_GLHG	703	MC	18	1	855	−0.4230						1.00
2014_GLHG	703	MC	19	1	855	−0.1007						1.06
2014_GLHG	703	MC	20	1	855	−0.8828						0.79
2014_GLHG	703	MC	21	1	855	0.0016						1.07
2014_GLHG	703	MC	22	1	855	0.6229						1.12
2014_GLHG	703	MC	23	1	855	−0.1553						0.98
2014_GLHG	703	MC	24	1	855	−0.4615						0.85
2014_GLHG	703	MC	25	1	855	−0.2721						0.87
2014_GLHG	703	MC	26	1	855	0.0851						0.92
2014_GLHG	703	MC	27	1	855	0.6706						1.15
2014_GLHG	703	MC	28	1	855	−0.5198						0.83
2014_GLHG	703	MC	29	1	855	0.1977						1.02
2014_GLHG	703	MC	30	1	855	−0.5329						0.97
2014_GLHG	703	MC	31	1	855	−0.1370						0.89
2014_GLHG	703	MC	32	1	855	−0.1249						0.99
2014_GLHG	704	MC	01	1	829	0.2834						1.15
2014_GLHG	704	MC	02	1	829	−0.5646						1.05
2014_GLHG	704	MC	03	1	829	0.1593						1.08
2014_GLHG	704	MC	04	1	829	−0.5098						1.17
2014_GLHG	704	MC	05	1	829	0.1966						0.94
2014_GLHG	704	MC	06	1	829	−0.0031						1.03
2014_GLHG	704	MC	07	1	829	−1.5772						0.89
2014_GLHG	704	MC	08	1	829	−0.6200						0.91
2014_GLHG	704	MC	09	1	829	0.1157						1.25
2014_GLHG	704	MC	10	1	829	0.3391						1.02
2014_GLHG	704	MC	11	1	829	−0.3297						0.89

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	704	MC	12	1	829	−1.4124						0.80
2014_GLHG	704	MC	13	1	829	−0.6340						0.96
2014_GLHG	704	MC	14	1	829	0.0408						1.00
2014_GLHG	704	MC	15	1	829	0.5252						1.22
2014_GLHG	704	MC	16	1	829	−0.7479						0.95
2014_GLHG	704	MC	17	1	829	0.0533						0.95
2014_GLHG	704	MC	18	1	829	−0.7407						0.87
2014_GLHG	704	MC	19	1	829	1.1418						1.10
2014_GLHG	704	MC	20	1	829	0.5501						1.04
2014_GLHG	704	MC	21	1	829	0.9369						1.09
2014_GLHG	704	MC	22	1	829	0.3453						0.89
2014_GLHG	704	MC	23	1	829	−0.2515						1.02
2014_GLHG	704	MC	24	1	829	−0.1678						0.95
2014_GLHG	704	MC	25	1	829	−0.5508						0.83
2014_GLHG	704	MC	26	1	829	−0.4156						1.08
2014_GLHG	704	MC	27	1	829	0.3763						0.96
2014_GLHG	704	MC	28	1	829	0.2834						0.88
2014_GLHG	704	MC	29	1	829	0.3391						0.96
2014_GLHG	704	MC	30	1	829	0.6001						0.93
2014_GLHG	704	MC	31	1	829	0.0157						0.90
2014_GLHG	704	MC	32	1	829	0.5128						1.14
2014_GLHG	705	MC	01	1	829	−0.5258						1.10
2014_GLHG	705	MC	02	1	829	−1.1031						1.03
2014_GLHG	705	MC	03	1	829	−1.5776						0.93
2014_GLHG	705	MC	04	1	829	1.2816						1.15
2014_GLHG	705	MC	05	1	829	−1.2390						1.00
2014_GLHG	705	MC	06	1	829	−0.2898						0.99
2014_GLHG	705	MC	07	1	829	−0.1852						1.01

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	705	MC	08	1	829	−0.6809						0.91
2014_GLHG	705	MC	09	1	829	−1.0297						0.97
2014_GLHG	705	MC	10	1	829	0.1078						1.04
2014_GLHG	705	MC	11	1	829	−0.1594						1.03
2014_GLHG	705	MC	12	1	829	−0.8885						0.94
2014_GLHG	705	MC	13	1	829	0.5448						1.05
2014_GLHG	705	MC	14	1	829	−1.8515						0.88
2014_GLHG	705	MC	15	1	829	−0.3963						0.88
2014_GLHG	705	MC	16	1	829	−1.2653						0.88
2014_GLHG	705	MC	17	1	829	−0.6167						1.02
2014_GLHG	705	MC	18	1	829	0.3328						1.20
2014_GLHG	705	MC	19	1	829	0.8938						1.27
2014_GLHG	705	MC	20	1	829	0.7020						1.21
2014_GLHG	705	MC	21	1	829	−0.2635						0.90
2014_GLHG	705	MC	22	1	829	0.3701						1.12
2014_GLHG	705	MC	23	1	829	0.5011						1.17
2014_GLHG	705	MC	24	1	829	−0.8885						0.88
2014_GLHG	705	MC	25	1	829	0.4761						0.91
2014_GLHG	705	MC	26	1	829	0.1392						0.84
2014_GLHG	705	MC	27	1	829	0.6894						0.99
2014_GLHG	705	MC	28	1	829	−0.6167						0.77
2014_GLHG	705	MC	29	1	829	−0.3162						0.95
2014_GLHG	705	MC	30	1	829	−0.1658						0.95
2014_GLHG	705	MC	31	1	829	−0.7757						0.94
2014_GLHG	705	MC	32	1	829	−0.1982						0.94
2014_GLHG	706	MC	01	1	821	−1.7977						1.08
2014_GLHG	706	MC	02	1	821	−0.2387						1.04
2014_GLHG	706	MC	03	1	821	0.8218						1.03

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	706	MC	04	1	821	−1.2767						0.94
2014_GLHG	706	MC	05	1	821	−1.0407						0.95
2014_GLHG	706	MC	06	1	821	−0.4287						0.93
2014_GLHG	706	MC	07	1	821	0.1289						1.06
2014_GLHG	706	MC	08	1	821	−0.0135						0.83
2014_GLHG	706	MC	09	1	821	0.1642						1.13
2014_GLHG	706	MC	10	1	821	0.0401						0.90
2014_GLHG	706	MC	11	1	821	1.1546						1.17
2014_GLHG	706	MC	12	1	821	0.2697						1.17
2014_GLHG	706	MC	13	1	821	0.2756						1.02
2014_GLHG	706	MC	14	1	821	−0.7615						0.90
2014_GLHG	706	MC	15	1	821	0.7740						1.08
2014_GLHG	706	MC	16	1	821	−0.5875						0.91
2014_GLHG	706	MC	17	1	821	0.0401						1.06
2014_GLHG	706	MC	18	1	821	0.5031						1.25
2014_GLHG	706	MC	19	1	821	−0.3581						0.98
2014_GLHG	706	MC	20	1	821	0.7323						1.06
2014_GLHG	706	MC	21	1	821	1.0165						1.21
2014_GLHG	706	MC	22	1	821	0.0044						0.92
2014_GLHG	706	MC	23	1	821	−0.4158						0.82
2014_GLHG	706	MC	24	1	821	−0.9933						0.85
2014_GLHG	706	MC	25	1	821	0.2814						1.09
2014_GLHG	706	MC	26	1	821	−0.3200						0.90
2014_GLHG	706	MC	27	1	821	0.0580						1.06
2014_GLHG	706	MC	28	1	821	−0.0978						0.90
2014_GLHG	706	MC	29	1	821	0.9489						0.98
2014_GLHG	706	MC	30	1	821	−0.3264						0.81
2014_GLHG	706	MC	31	1	821	1.3639						1.06

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	706	MC	32	1	821	−0.1525						0.86
2014_GLHG	707	MC	01	1	845	−0.0247						1.01
2014_GLHG	707	MC	02	1	845	0.8723						1.02
2014_GLHG	707	MC	03	1	845	0.7911						1.27
2014_GLHG	707	MC	04	1	845	0.1749						1.07
2014_GLHG	707	MC	05	1	845	−0.3749						0.99
2014_GLHG	707	MC	06	1	845	−0.0710						1.07
2014_GLHG	707	MC	07	1	845	−0.8769						1.13
2014_GLHG	707	MC	08	1	845	0.1523						1.14
2014_GLHG	707	MC	09	1	845	−1.4566						0.97
2014_GLHG	707	MC	10	1	845	0.0784						0.94
2014_GLHG	707	MC	11	1	845	−0.7027						0.87
2014_GLHG	707	MC	12	1	845	−0.2416						0.82
2014_GLHG	707	MC	13	1	845	−0.2776						1.01
2014_GLHG	707	MC	14	1	845	0.9133						0.97
2014_GLHG	707	MC	15	1	845	0.1183						1.05
2014_GLHG	707	MC	16	1	845	0.3495						0.89
2014_GLHG	707	MC	17	1	845	0.6141						1.03
2014_GLHG	707	MC	18	1	845	0.1749						0.83
2014_GLHG	707	MC	19	1	845	−0.2297						0.96
2014_GLHG	707	MC	20	1	845	−0.2596						0.91
2014_GLHG	707	MC	21	1	845	1.0560						1.10
2014_GLHG	707	MC	22	1	845	−0.5899						0.94
2014_GLHG	707	MC	23	1	845	−0.1293						0.89
2014_GLHG	707	MC	24	1	845	0.8085						1.18
2014_GLHG	707	MC	25	1	845	−0.0594						0.88
2014_GLHG	707	MC	26	1	845	−0.2356						0.88
2014_GLHG	707	MC	27	1	845	0.8665						0.98

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	707	MC	28	1	845	0.5689						1.28
2014_GLHG	707	MC	29	1	845	0.5068						1.02
2014_GLHG	707	MC	30	1	845	0.1409						1.00
2014_GLHG	707	MC	31	1	845	1.0380						1.03
2014_GLHG	707	MC	32	1	845	−0.1176						0.85
2014_GLHG	708	MC	01	1	833	−0.6222						0.98
2014_GLHG	708	MC	02	1	833	−1.4688						0.95
2014_GLHG	708	MC	03	1	833	−1.3125						1.05
2014_GLHG	708	MC	04	1	833	−0.4648						1.02
2014_GLHG	708	MC	05	1	833	0.0480						0.99
2014_GLHG	708	MC	06	1	833	0.9196						1.04
2014_GLHG	708	MC	07	1	833	0.1885						1.05
2014_GLHG	708	MC	08	1	833	0.1243						0.87
2014_GLHG	708	MC	09	1	833	1.9692						1.09
2014_GLHG	708	MC	10	1	833	0.4038						0.97
2014_GLHG	708	MC	11	1	833	−0.2456						1.00
2014_GLHG	708	MC	12	1	833	2.9899						1.07
2014_GLHG	708	MC	13	1	833	0.2118						0.94
2014_GLHG	708	MC	14	1	833	−0.2702						0.92
2014_GLHG	708	MC	15	1	833	0.6672						0.96
2014_GLHG	708	MC	16	1	833	0.3631						1.14
2014_GLHG	708	MC	17	1	833	0.5437						1.14
2014_GLHG	708	MC	18	1	833	0.5496						1.02
2014_GLHG	708	MC	19	1	833	0.6082						0.92
2014_GLHG	708	MC	20	1	833	1.2890						1.29
2014_GLHG	708	MC	21	1	833	0.9936						0.97
2014_GLHG	708	MC	22	1	833	0.4736						0.95
2014_GLHG	708	MC	23	1	833	−0.1063						0.87

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	708	MC	24	1	833	0.7265						1.15
2014_GLHG	708	MC	25	1	833	0.3863						1.03
2014_GLHG	708	MC	26	1	833	0.2060						1.00
2014_GLHG	708	MC	27	1	833	0.2875						0.97
2014_GLHG	708	MC	28	1	833	−0.1183						0.93
2014_GLHG	708	MC	29	1	833	−0.7732						0.83
2014_GLHG	708	MC	30	1	833	−0.2150						0.94
2014_GLHG	708	MC	31	1	833	0.7087						0.96
2014_GLHG	708	MC	32	1	833	0.4911						0.97
2014_GLHG	709	MC	01	1	833	0.6229						1.10
2014_GLHG	709	MC	02	1	833	2.3180						1.27
2014_GLHG	709	MC	03	1	833	−0.1990						1.03
2014_GLHG	709	MC	04	1	833	−1.5141						0.87
2014_GLHG	709	MC	05	1	833	−0.9200						0.93
2014_GLHG	709	MC	06	1	833	1.3438						1.07
2014_GLHG	709	MC	07	1	833	0.3709						1.04
2014_GLHG	709	MC	08	1	833	−0.1198						0.93
2014_GLHG	709	MC	09	1	833	−0.7391						0.97
2014_GLHG	709	MC	10	1	833	1.4323						1.00
2014_GLHG	709	MC	11	1	833	−0.3042						0.99
2014_GLHG	709	MC	12	1	833	−0.6018						0.90
2014_GLHG	709	MC	13	1	833	−0.1806						1.01
2014_GLHG	709	MC	14	1	833	−0.4245						0.89
2014_GLHG	709	MC	15	1	833	0.1071						1.16
2014_GLHG	709	MC	16	1	833	−0.7815						0.90
2014_GLHG	709	MC	17	1	833	0.8315						1.00
2014_GLHG	709	MC	18	1	833	−0.5484						0.84
2014_GLHG	709	MC	19	1	833	0.3066						1.00

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	709	MC	20	1	833	−0.5750						0.82
2014_GLHG	709	MC	21	1	833	1.0463						1.01
2014_GLHG	709	MC	22	1	833	0.0420						1.03
2014_GLHG	709	MC	23	1	833	0.4937						1.07
2014_GLHG	709	MC	24	1	833	0.0716						0.95
2014_GLHG	709	MC	25	1	833	−0.0836						0.84
2014_GLHG	709	MC	26	1	833	1.0526						1.10
2014_GLHG	709	MC	27	1	833	−0.4181						0.95
2014_GLHG	709	MC	28	1	833	0.4586						1.19
2014_GLHG	709	MC	29	1	833	−0.2297						0.85
2014_GLHG	709	MC	30	1	833	0.1660						1.09
2014_GLHG	709	MC	31	1	833	1.2641						1.25
2014_GLHG	709	MC	32	1	833	0.7834						0.82
2014_GLHG	710_711	ESS	THM	5	732	2.2259	−1.6674	−0.0514	1.7188			1.02
2014_GLHG	712_713	ESS	THM	5	708	2.6686	−2.5724	−1.3399	0.2649	3.6475		1.14
2014_GLHG	714_715	ESS	THM	5	712	1.5681	−1.5959	0.1053	1.4905			1.10
2014_GLHG	716_717	ESS	THM	5	678	1.9592	−1.6048	0.0887	1.5161			1.02
2014_GLHG	718_719	ESS	THM	5	674	2.2265	−1.7061	−0.2406	1.9467			0.95
2014_GLHG	720	ESS	THM	5	360	2.6928	−2.4250	−0.8468	0.4129	2.8589		1.06
2014_GLHG	721	ESS	THM	5	349	2.7445	−1.6161	−0.4228	2.0389			0.87
2014_GLHG	722_723	SCF	01	1	991	−1.4381						1.03
2014_GLHG	722_723	SCF	02	1	991	0.6298						0.95
2014_GLHG	722_723	SCF	03	2	991	−0.5013	0.3048	−0.3048				1.12
2014_GLHG	722_723	SCF	04	2	991	−0.3926	0.2780	−0.2780				1.12
2014_GLHG	722_723	SCF	5a	1	991	−0.6011						0.97
2014_GLHG	722_723	SCF	5b	1	991	−0.6264						0.92
2014_GLHG	722_723	SCF	06	1	991	−1.0853						0.92
2014_GLHG	722_723	SCF	07	1	991	0.6048						0.94

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	722_723	SCF	08	1	991	−0.3189						0.84
2014_GLHG	722_723	SCF	09	1	991	0.2925						1.01
2014_GLHG	722_723	ESS	DBQ	5	991	2.9122	−2.4548	−1.9115	0.1448	4.2216		1.15
2014_GLHG	724_725	SCF	1a	1	1017	−0.7270						1.00
2014_GLHG	724_725	SCF	1b	1	1017	0.0263						0.90
2014_GLHG	724_725	SCF	02	1	1017	0.3388						0.97
2014_GLHG	724_725	SCF	3a	1	1017	−1.0072						0.86
2014_GLHG	724_725	SCF	3b	1	1017	−1.0809						0.81
2014_GLHG	724_725	SCF	4a	1	1017	0.0577						1.05
2014_GLHG	724_725	SCF	4b	1	1017	−1.2849						0.81
2014_GLHG	724_725	SCF	5a	1	1017	−0.0371						0.89
2014_GLHG	724_725	SCF	5b	1	1017	0.2072						0.91
2014_GLHG	724_725	SCF	06	1	1017	−0.4852						0.92
2014_GLHG	724_725	SCF	7a	1	1017	−0.7531						0.89
2014_GLHG	724_725	SCF	7b	1	1017	−1.0659						0.80
2014_GLHG	724_725	SCF	8a	1	1017	−0.7077						0.84
2014_GLHG	724_725	SCF	8b	1	1017	−0.2603						0.84
2014_GLHG	724_725	ESS	DBQ	5	1017	1.4065	−0.9558	−0.4739	1.4297			1.23
2014_GLHG	726_727	SCF	01	2	995	−0.0756	−0.5029	0.5029				1.05
2014_GLHG	726_727	SCF	02	1	995	0.3618						0.95
2014_GLHG	726_727	SCF	3a	1	995	−0.6024						0.98
2014_GLHG	726_727	SCF	3b	1	995	−0.7788						0.94
2014_GLHG	726_727	SCF	04	1	995	−0.3304						0.89
2014_GLHG	726_727	SCF	05	1	995	−0.3188						0.87
2014_GLHG	726_727	SCF	06	1	995	−0.7788						0.85
2014_GLHG	726_727	SCF	07	1	995	−0.0725						0.82
2014_GLHG	726_727	SCF	08	1	995	0.0827						0.93
2014_GLHG	726_727	SCF	09	2	995	0.1832	0.3109	−0.3109				1.10

Test	Form	Type	Item	Max	N-Count	RID	S1	S2	S3	S4	S5	INFIT
2014_GLHG	726_727	ESS	DBQ	5	995	1.9506	−1.2923	−0.7227	2.0150			1.03
2014_GLHG	728	SCF	01	1	491	−1.4674						0.93
2014_GLHG	728	SCF	02	1	491	−0.3001						1.06
2014_GLHG	728	SCF	03	2	491	−0.2768	−0.4929	0.4929				1.11
2014_GLHG	728	SCF	04	1	491	−0.5897						0.93
2014_GLHG	728	SCF	5a	1	491	−0.9656						0.88
2014_GLHG	728	SCF	5b	1	491	−0.3001						0.87
2014_GLHG	728	SCF	06	1	491	−0.5179						0.94
2014_GLHG	728	SCF	7a	1	491	1.0426						1.03
2014_GLHG	728	SCF	7b	1	491	1.3288						0.92
2014_GLHG	728	SCF	08	1	491	0.4083						1.03
2014_GLHG	728	SCF	9a	1	491	0.3983						1.04
2014_GLHG	728	SCF	9b	1	491	0.1967						0.99
2014_GLHG	728	ESS	DBQ	5	491	1.6173	−0.9939	−0.4930	1.4870			1.00

Appendix E: DIF Statistics

The first four columns from the left contain the test name, form ID, item type, and item sequence number within the form. The next three columns contain the Mantel-Haenszel DIF statistical values (note that the *MH* Delta statistic cannot be calculated for scaffold and essay items). The final two columns will only have values if the item displays possible moderate or severe DIF; if so, the degree of DIF (B/BB = moderate; C/CC = severe) and the favored group will be shown. Items without statistics are DNS (Do Not Score) status items.

Test	Form	Type	Item	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
2014_GLHG	701	MC	01	0.83	2.87	0.10		
2014_GLHG	701	MC	02	0.12	0.09	0.02		
2014_GLHG	701	MC	03	0.22	0.32	0.04		
2014_GLHG	701	MC	04	0.45	1.37	0.07		
2014_GLHG	701	MC	05	−0.01	0.00	−0.02		
2014_GLHG	701	MC	06	0.62	2.57	0.12		
2014_GLHG	701	MC	07	−0.25	0.42	−0.03		
2014_GLHG	701	MC	08	−0.10	0.08	0.00		
2014_GLHG	701	MC	09	0.34	0.44	0.04		
2014_GLHG	701	MC	10	−0.16	0.15	−0.04		
2014_GLHG	701	MC	11	1.15	8.98	0.20	B	F
2014_GLHG	701	MC	12	1.25	6.25	0.14	B	F
2014_GLHG	701	MC	13	−0.17	0.18	−0.02		
2014_GLHG	701	MC	14	0.34	0.92	0.05		
2014_GLHG	701	MC	15	−0.12	0.08	−0.01		
2014_GLHG	701	MC	16	−0.14	0.15	−0.02		
2014_GLHG	701	MC	17	−0.36	0.91	−0.07		
2014_GLHG	701	MC	18	−0.23	0.30	−0.03		
2014_GLHG	701	MC	19	−0.35	0.72	−0.05		
2014_GLHG	701	MC	20	−0.02	0.00	−0.03		
2014_GLHG	701	MC	21	−0.09	0.06	−0.02		
2014_GLHG	701	MC	22	0.07	0.03	0.00		
2014_GLHG	701	MC	23	0.16	0.16	0.02		
2014_GLHG	701	MC	24	0.09	0.06	0.04		
2014_GLHG	701	MC	25	−0.50	1.86	−0.10		
2014_GLHG	701	MC	26	0.14	0.14	0.01		
2014_GLHG	701	MC	27	−1.18	8.75	−0.19	B	M
2014_GLHG	701	MC	28	−0.24	0.33	−0.02		

Test	Form	Type	Item	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
2014_GLHG	701	MC	29	−1.51	15.80	−0.24	C	M
2014_GLHG	701	MC	30	0.62	2.13	0.09		
2014_GLHG	702	MC	01	−0.19	0.26	−0.06		
2014_GLHG	702	MC	02	−0.36	0.47	−0.08		
2014_GLHG	702	MC	03	−0.34	0.73	−0.04		
2014_GLHG	702	MC	04	−0.97	7.24	−0.20		
2014_GLHG	702	MC	05	−0.22	0.20	−0.03		
2014_GLHG	702	MC	06	0.52	0.88	0.04		
2014_GLHG	702	MC	07	−0.29	0.58	−0.01		
2014_GLHG	702	MC	08	0.06	0.03	0.02		
2014_GLHG	702	MC	09	−0.04	0.01	−0.03		
2014_GLHG	702	MC	10	0.57	2.57	0.09		
2014_GLHG	702	MC	11	−0.66	2.25	−0.10		
2014_GLHG	702	MC	12	−1.37	9.01	−0.20	B	M
2014_GLHG	702	MC	13	−0.06	0.02	0.02		
2014_GLHG	702	MC	14	0.07	0.04	0.02		
2014_GLHG	702	MC	15	0.45	1.07	0.10		
2014_GLHG	702	MC	16	−0.30	0.51	−0.04		
2014_GLHG	702	MC	17	0.21	0.35	0.06		
2014_GLHG	702	MC	18	−0.37	0.99	−0.06		
2014_GLHG	702	MC	19	0.11	0.06	0.01		
2014_GLHG	702	MC	20	−0.15	0.14	0.00		
2014_GLHG	702	MC	21	0.39	1.02	0.05		
2014_GLHG	702	MC	22	0.48	1.57	0.07		
2014_GLHG	702	MC	23	−0.29	0.51	−0.04		
2014_GLHG	702	MC	24	−0.34	0.71	−0.06		
2014_GLHG	702	MC	25	0.17	0.17	0.01		
2014_GLHG	702	MC	26	0.85	3.87	0.12		
2014_GLHG	702	MC	27	0.49	1.08	0.05		
2014_GLHG	702	MC	28	0.50	1.52	0.09		
2014_GLHG	702	MC	29	−0.32	0.66	−0.04		
2014_GLHG	702	MC	30	0.54	1.98	0.09		
2014_GLHG	702	MC	31	0.94	6.55	0.16		
2014_GLHG	702	MC	32	−0.24	0.43	−0.03		
2014_GLHG	703	MC	01	−0.12	0.06	−0.01		
2014_GLHG	703	MC	02	0.32	0.66	0.04		
2014_GLHG	703	MC	03	−0.35	0.97	−0.07		
2014_GLHG	703	MC	04	−0.48	1.60	−0.07		
2014_GLHG	703	MC	05	0.27	0.41	0.03		

Test	Form	Type	Item	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
2014_GLHG	703	MC	06	0.08	0.04	0.02		
2014_GLHG	703	MC	07	−0.07	0.03	−0.02		
2014_GLHG	703	MC	08	−1.26	10.59	−0.21	B	M
2014_GLHG	703	MC	09	0.09	0.07	0.01		
2014_GLHG	703	MC	10	−0.17	0.20	−0.03		
2014_GLHG	703	MC	11	0.10	0.07	0.02		
2014_GLHG	703	MC	12	−0.02	0.00	0.01		
2014_GLHG	703	MC	13	−0.43	0.90	−0.06		
2014_GLHG	703	MC	14	0.23	0.41	0.05		
2014_GLHG	703	MC	15	0.07	0.03	0.02		
2014_GLHG	703	MC	16	−0.41	1.07	−0.04		
2014_GLHG	703	MC	17	−0.65	3.07	−0.13		
2014_GLHG	703	MC	18	0.58	2.09	0.09		
2014_GLHG	703	MC	19	0.42	1.34	0.06		
2014_GLHG	703	MC	20	0.74	2.31	0.08		
2014_GLHG	703	MC	21	−0.42	1.25	−0.06		
2014_GLHG	703	MC	22	−0.27	0.57	−0.05		
2014_GLHG	703	MC	23	−0.84	4.50	−0.13		
2014_GLHG	703	MC	24	0.29	0.43	0.03		
2014_GLHG	703	MC	25	0.45	1.14	0.06		
2014_GLHG	703	MC	26	−0.33	0.68	−0.05		
2014_GLHG	703	MC	27	1.10	9.35	0.21	B	F
2014_GLHG	703	MC	28	0.12	0.07	0.01		
2014_GLHG	703	MC	29	−0.39	1.09	−0.08		
2014_GLHG	703	MC	30	1.13	7.81	0.19	B	F
2014_GLHG	703	MC	31	0.28	0.47	0.05		
2014_GLHG	703	MC	32	0.32	0.70	0.05		
2014_GLHG	704	MC	01	−0.91	6.34	−0.16		
2014_GLHG	704	MC	02	−0.12	0.09	0.00		
2014_GLHG	704	MC	03	−0.45	1.44	−0.07		
2014_GLHG	704	MC	04	0.20	0.27	0.04		
2014_GLHG	704	MC	05	−0.27	0.44	−0.03		
2014_GLHG	704	MC	06	0.59	2.34	0.12		
2014_GLHG	704	MC	07	−1.93	12.73	−0.22	C	M
2014_GLHG	704	MC	08	−1.56	12.96	−0.21	C	M
2014_GLHG	704	MC	09	0.50	1.96	0.09		
2014_GLHG	704	MC	10	0.12	0.09	0.03		
2014_GLHG	704	MC	11	−0.99	5.23	−0.15		
2014_GLHG	704	MC	12	0.93	2.81	0.09		

Test	Form	Type	Item	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
2014_GLHG	704	MC	13	0.40	0.88	0.06		
2014_GLHG	704	MC	14	−0.05	0.02	−0.01		
2014_GLHG	704	MC	15	0.45	1.58	0.09		
2014_GLHG	704	MC	16	−0.03	0.01	−0.03		
2014_GLHG	704	MC	17	−0.01	0.00	0.00		
2014_GLHG	704	MC	18	0.22	0.25	0.02		
2014_GLHG	704	MC	19	−0.31	0.65	−0.05		
2014_GLHG	704	MC	20	−0.54	2.09	−0.12		
2014_GLHG	704	MC	21	0.71	3.47	0.12		
2014_GLHG	704	MC	22	−0.47	1.29	−0.08		
2014_GLHG	704	MC	23	0.59	2.33	0.10		
2014_GLHG	704	MC	24	0.95	5.62	0.16		
2014_GLHG	704	MC	25	0.12	0.07	0.02		
2014_GLHG	704	MC	26	0.75	3.62	0.14		
2014_GLHG	704	MC	27	−0.31	0.61	−0.04		
2014_GLHG	704	MC	28	−0.40	0.95	−0.06		
2014_GLHG	704	MC	29	0.40	1.05	0.05		
2014_GLHG	704	MC	30	0.15	0.14	0.02		
2014_GLHG	704	MC	31	0.24	0.35	0.03		
2014_GLHG	704	MC	32	0.17	0.23	0.04		
2014_GLHG	705	MC	01	−0.24	0.35	−0.07		
2014_GLHG	705	MC	02	0.05	0.01	0.04		
2014_GLHG	705	MC	03	0.72	1.72	0.08		
2014_GLHG	705	MC	04	−0.29	0.55	−0.02		
2014_GLHG	705	MC	05	−1.04	4.69	−0.15	B	M
2014_GLHG	705	MC	06	−0.42	1.13	−0.08		
2014_GLHG	705	MC	07	0.32	0.62	0.05		
2014_GLHG	705	MC	08	0.83	3.32	0.11		
2014_GLHG	705	MC	09	−0.46	1.07	−0.08		
2014_GLHG	705	MC	10	−0.53	1.85	−0.07		
2014_GLHG	705	MC	11	−1.07	7.59	−0.20	B	M
2014_GLHG	705	MC	12	−1.11	6.09	−0.18	B	M
2014_GLHG	705	MC	13	0.16	0.17	0.05		
2014_GLHG	705	MC	14	0.53	0.74	0.07		
2014_GLHG	705	MC	15	−0.75	2.79	−0.09		
2014_GLHG	705	MC	16	0.23	0.21	0.04		
2014_GLHG	705	MC	17	0.70	3.00	0.12		
2014_GLHG	705	MC	18	−0.34	0.92	−0.07		
2014_GLHG	705	MC	19	0.12	0.11	0.02		

Test	Form	Type	Item	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
2014_GLHG	705	MC	20	−1.02	7.56	−0.18	B	M
2014_GLHG	705	MC	21	0.84	4.22	0.13		
2014_GLHG	705	MC	22	−0.45	1.46	−0.11		
2014_GLHG	705	MC	23	0.00	0.00	−0.02		
2014_GLHG	705	MC	24	−0.30	0.42	−0.05		
2014_GLHG	705	MC	25	−0.01	0.00	0.00		
2014_GLHG	705	MC	26	0.28	0.41	0.03		
2014_GLHG	705	MC	27	0.50	1.56	0.06		
2014_GLHG	705	MC	28	0.85	2.82	0.09		
2014_GLHG	705	MC	29	0.32	0.63	0.07		
2014_GLHG	705	MC	30	1.06	7.10	0.17	B	F
2014_GLHG	705	MC	31	1.06	5.52	0.15	B	F
2014_GLHG	705	MC	32	0.62	2.23	0.11		
2014_GLHG	706	MC	01	0.31	0.38	0.04		
2014_GLHG	706	MC	02	0.08	0.05	0.01		
2014_GLHG	706	MC	03	0.05	0.02	0.01		
2014_GLHG	706	MC	04	−0.49	1.06	−0.07		
2014_GLHG	706	MC	05	−0.31	0.45	−0.04		
2014_GLHG	706	MC	06	−0.16	0.16	−0.03		
2014_GLHG	706	MC	07	−0.06	0.02	−0.03		
2014_GLHG	706	MC	08	0.63	2.23	0.09		
2014_GLHG	706	MC	09	−0.69	3.90	−0.15		
2014_GLHG	706	MC	10	−0.32	0.64	−0.05		
2014_GLHG	706	MC	11	0.59	2.57	0.10		
2014_GLHG	706	MC	12	0.04	0.02	0.00		
2014_GLHG	706	MC	13	0.13	0.12	0.03		
2014_GLHG	706	MC	14	−0.32	0.54	−0.05		
2014_GLHG	706	MC	15	0.25	0.48	0.04		
2014_GLHG	706	MC	16	0.54	1.59	0.09		
2014_GLHG	706	MC	17	0.12	0.11	0.04		
2014_GLHG	706	MC	18	0.34	0.93	0.07		
2014_GLHG	706	MC	19	−0.06	0.02	−0.01		
2014_GLHG	706	MC	20	−1.27	11.80	−0.23	B	M
2014_GLHG	706	MC	21	−0.32	0.84	−0.04		
2014_GLHG	706	MC	22	−0.60	2.38	−0.10		
2014_GLHG	706	MC	23	1.08	5.72	0.14	B	F
2014_GLHG	706	MC	24	−0.18	0.14	−0.02		
2014_GLHG	706	MC	25	0.00	0.00	0.00		
2014_GLHG	706	MC	26	−0.58	1.93	−0.09		

Test	Form	Type	Item	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
2014_GLHG	706	MC	27	0.48	1.60	0.08		
2014_GLHG	706	MC	28	−1.11	7.36	−0.18	B	M
2014_GLHG	706	MC	29	0.21	0.30	0.04		
2014_GLHG	706	MC	30	1.13	6.31	0.16	B	F
2014_GLHG	706	MC	31	0.56	1.97	0.11		
2014_GLHG	706	MC	32	0.39	0.86	0.05		
2014_GLHG	707	MC	01	−0.57	2.37	−0.12		
2014_GLHG	707	MC	02	0.87	5.28	0.18		
2014_GLHG	707	MC	03	−0.45	1.73	−0.10		
2014_GLHG	707	MC	04	0.79	4.72	0.13		
2014_GLHG	707	MC	05	0.16	0.18	0.03		
2014_GLHG	707	MC	06	0.44	1.48	0.09		
2014_GLHG	707	MC	07	0.13	0.11	0.03		
2014_GLHG	707	MC	08	−0.72	4.24	−0.14		
2014_GLHG	707	MC	09	0.41	0.70	0.06		
2014_GLHG	707	MC	10	1.24	10.35	0.22	B	F
2014_GLHG	707	MC	11	−0.60	1.85	−0.07		
2014_GLHG	707	MC	12	−1.42	10.29	−0.18	B	M
2014_GLHG	707	MC	13	0.73	3.75	0.13		
2014_GLHG	707	MC	14	−0.24	0.40	−0.05		
2014_GLHG	707	MC	15	0.28	0.60	0.08		
2014_GLHG	707	MC	16	−0.49	1.58	−0.08		
2014_GLHG	707	MC	17	0.33	0.84	0.05		
2014_GLHG	707	MC	18	−0.86	4.43	−0.11		
2014_GLHG	707	MC	19	−0.24	0.41	−0.04		
2014_GLHG	707	MC	20	−0.73	3.21	−0.09		
2014_GLHG	707	MC	21	−0.41	1.30	−0.08		
2014_GLHG	707	MC	22	−0.02	0.00	0.01		
2014_GLHG	707	MC	23	−0.63	2.47	−0.10		
2014_GLHG	707	MC	24	0.30	0.74	0.08		
2014_GLHG	707	MC	25	0.10	0.06	0.02		
2014_GLHG	707	MC	26	−0.15	0.13	−0.04		
2014_GLHG	707	MC	27	−0.08	0.04	−0.02		
2014_GLHG	707	MC	28	0.67	3.86	0.15		
2014_GLHG	707	MC	29	−0.15	0.16	−0.01		
2014_GLHG	707	MC	30	0.16	0.19	0.05		
2014_GLHG	707	MC	31	−0.41	1.22	−0.06		
2014_GLHG	707	MC	32	1.02	5.85	0.15	B	F
2014_GLHG	708	MC	01	−0.15	0.13	−0.03		

Test	Form	Type	Item	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
2014_GLHG	708	MC	02	−0.95	3.67	−0.10		
2014_GLHG	708	MC	03	0.00	0.00	0.00		
2014_GLHG	708	MC	04	−0.06	0.02	0.00		
2014_GLHG	708	MC	05	−1.38	12.79	−0.23	B	M
2014_GLHG	708	MC	06	0.44	1.28	0.08		
2014_GLHG	708	MC	07	−0.99	6.93	−0.16		
2014_GLHG	708	MC	08	−0.38	0.87	−0.05		
2014_GLHG	708	MC	09	−0.58	1.74	−0.10		
2014_GLHG	708	MC	10	−0.52	1.86	−0.09		
2014_GLHG	708	MC	11	−0.48	1.55	−0.10		
2014_GLHG	708	MC	12	−0.08	0.02	−0.01		
2014_GLHG	708	MC	13	1.00	6.76	0.18		
2014_GLHG	708	MC	14	−0.05	0.02	−0.01		
2014_GLHG	708	MC	15	0.14	0.12	0.03		
2014_GLHG	708	MC	16	0.50	1.96	0.09		
2014_GLHG	708	MC	17	0.58	2.64	0.12		
2014_GLHG	708	MC	18	0.42	1.23	0.08		
2014_GLHG	708	MC	19	−0.87	4.69	−0.14		
2014_GLHG	708	MC	20	1.09	8.72	0.20	B	F
2014_GLHG	708	MC	21	−0.68	2.92	−0.10		
2014_GLHG	708	MC	22	−0.08	0.05	−0.04		
2014_GLHG	708	MC	23	−0.19	0.21	−0.02		
2014_GLHG	708	MC	24	−1.47	16.21	−0.30	B	M
2014_GLHG	708	MC	25	0.15	0.16	0.03		
2014_GLHG	708	MC	26	1.26	11.38	0.23	B	F
2014_GLHG	708	MC	27	0.64	2.85	0.10		
2014_GLHG	708	MC	28	0.47	1.46	0.06		
2014_GLHG	708	MC	29	1.42	8.95	0.18	B	F
2014_GLHG	708	MC	30	0.71	3.11	0.11		
2014_GLHG	708	MC	31	−0.10	0.06	−0.02		
2014_GLHG	708	MC	32	−0.12	0.10	−0.03		
2014_GLHG	709	MC	01	−0.50	1.84	−0.09		
2014_GLHG	709	MC	02	1.00	5.03	0.14	B	F
2014_GLHG	709	MC	03	0.03	0.00	0.01		
2014_GLHG	709	MC	04	1.18	4.91	0.13	B	F
2014_GLHG	709	MC	05	0.33	0.56	0.05		
2014_GLHG	709	MC	06	−0.88	4.91	−0.15		
2014_GLHG	709	MC	07	−0.07	0.04	−0.01		
2014_GLHG	709	MC	08	−0.80	3.98	−0.12		

Test	Form	Type	Item	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
2014_GLHG	709	MC	09	0.18	0.17	0.01		
2014_GLHG	709	MC	10	−0.15	0.13	−0.04		
2014_GLHG	709	MC	11	−0.62	2.45	−0.10		
2014_GLHG	709	MC	12	0.57	1.71	0.08		
2014_GLHG	709	MC	13	−0.66	2.94	−0.12		
2014_GLHG	709	MC	14	0.63	2.18	0.09		
2014_GLHG	709	MC	15	0.05	0.02	0.00		
2014_GLHG	709	MC	16	0.82	3.58	0.11		
2014_GLHG	709	MC	17	−0.55	1.98	−0.10		
2014_GLHG	709	MC	18	0.05	0.01	0.00		
2014_GLHG	709	MC	19	−0.36	0.92	−0.07		
2014_GLHG	709	MC	20	0.36	0.66	0.05		
2014_GLHG	709	MC	21	−0.04	0.01	0.01		
2014_GLHG	709	MC	22	−0.50	1.91	−0.08		
2014_GLHG	709	MC	23	−0.46	1.63	−0.07		
2014_GLHG	709	MC	24	−0.40	1.02	−0.07		
2014_GLHG	709	MC	25	1.21	7.74	0.18	B	F
2014_GLHG	709	MC	26	0.29	0.61	0.07		
2014_GLHG	709	MC	27	0.39	0.96	0.08		
2014_GLHG	709	MC	28	−0.70	4.06	−0.14		
2014_GLHG	709	MC	29	−0.21	0.23	−0.03		
2014_GLHG	709	MC	30	0.41	1.26	0.09		
2014_GLHG	709	MC	31	1.30	12.17	0.22	B	F
2014_GLHG	709	MC	32	−0.26	0.35	−0.04		
2014_GLHG	710_711	ESS	THM		37.89	0.12		
2014_GLHG	712_713	ESS	THM		11.03	−0.07		
2014_GLHG	714_715	ESS	THM		2.85	−0.04		
2014_GLHG	716_717	ESS	THM		8.80	−0.06		
2014_GLHG	718_719	ESS	THM		2.86	−0.04		
2014_GLHG	720	ESS	THM		0.18	0.01		
2014_GLHG	721	ESS	THM		6.81	0.06		
2014_GLHG	722_723	SCF	01		0.88	−0.03		
2014_GLHG	722_723	SCF	02		11.94	0.07		
2014_GLHG	722_723	SCF	03		7.51	−0.06		
2014_GLHG	722_723	SCF	04		6.81	0.06		
2014_GLHG	722_723	SCF	5a		14.81	0.08		
2014_GLHG	722_723	SCF	5b		0.35	−0.02		
2014_GLHG	722_723	SCF	06		15.68	0.09		
2014_GLHG	722_723	SCF	07		0.63	0.02		

Test	Form	Type	Item	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
2014_GLHG	722_723	SCF	08		1.17	−0.03		
2014_GLHG	722_723	SCF	09		20.63	−0.10		
2014_GLHG	722_723	ESS	DBQ		12.75	0.08		
2014_GLHG	724_725	SCF	1a		9.28	−0.06		
2014_GLHG	724_725	SCF	1b		2.35	0.04		
2014_GLHG	724_725	SCF	02		2.35	−0.03		
2014_GLHG	724_725	SCF	3a		4.14	0.04		
2014_GLHG	724_725	SCF	3b		5.08	−0.05		
2014_GLHG	724_725	SCF	4a		0.35	−0.01		
2014_GLHG	724_725	SCF	4b		4.50	0.06		
2014_GLHG	724_725	SCF	5a		6.76	0.06		
2014_GLHG	724_725	SCF	5b		23.45	−0.10		
2014_GLHG	724_725	SCF	06		12.78	0.07		
2014_GLHG	724_725	SCF	7a		38.94	−0.14		
2014_GLHG	724_725	SCF	7b		6.41	0.05		
2014_GLHG	724_725	SCF	8a		4.55	0.21	BB	F
2014_GLHG	724_725	SCF	8b		12.17	0.31	CC	F
2014_GLHG	724_725	ESS	DBQ		20.87	0.45	CC	F
2014_GLHG	726_727	SCF	01		0.02	0.02		
2014_GLHG	726_727	SCF	02		3.02	0.15		
2014_GLHG	726_727	SCF	3a		0.56	0.08		
2014_GLHG	726_727	SCF	3b		1.93	0.09		
2014_GLHG	726_727	SCF	04		5.99	0.26	CC	F
2014_GLHG	726_727	SCF	05		0.47	0.10		
2014_GLHG	726_727	SCF	06		7.60	0.31	CC	F
2014_GLHG	726_727	SCF	07		3.26	0.09		
2014_GLHG	726_727	SCF	08		0.85	−0.09		
2014_GLHG	726_727	SCF	09		13.99	0.37	CC	F
2014_GLHG	726_727	ESS	DBQ		1.96	0.19		
2014_GLHG	728	SCF	01		13.37	0.36	CC	F
2014_GLHG	728	SCF	02		1.49	0.09		
2014_GLHG	728	SCF	03		7.89	0.24	BB	F
2014_GLHG	728	SCF	04		8.91	0.24	BB	F
2014_GLHG	728	SCF	5a		6.11	0.24	BB	F
2014_GLHG	728	SCF	5b		1.36	0.10		
2014_GLHG	728	SCF	06		0.18	0.04		
2014_GLHG	728	SCF	7a		7.30	0.26	CC	F
2014_GLHG	728	SCF	7b		8.55	0.30	CC	F
2014_GLHG	728	SCF	08		7.34	0.25	BB	F

Test	Form	Type	Item	MH Delta	MH Chi-Sq	Effect Size	DIF Category	Favored Group
2014_GLHG	728	SCF	9a		2.76	0.15		
2014_GLHG	728	SCF	9b		14.64	0.33	CC	F
2014_GLHG	728	ESS	DBQ		14.85	0.38	CC	F

DIF category meanings: A/AA = negligible, B/BB = moderate, C/CC = severe.

Favored group meanings: F = Female, M = Male.

Appendix F: Operational Test Maps

January 2014

Position	Item Type	Max Points	Weight	Std	Key Idea	PI	Mean	Point-Biserial	RID	INFIT
1	MC	1	1	4			0.65	0.48	-0.1568	0.95
2	MC	1	1	3			0.42	0.32	1.0225	1.15
3	MC	1	1	3			0.78	0.41	-0.9652	0.98
4	MC	1	1	2			0.68	0.54	-0.3350	0.90
5	MC	1	1	2			0.52	0.45	0.5149	1.02
6	MC	1	1	3			0.77	0.46	-0.8237	0.95
7	MC	1	1	2			0.64	0.36	-0.0681	1.09
8	MC	1	1	2			0.49	0.38	0.6490	1.13
9	MC	1	1	3			0.57	0.43	0.2841	1.04
10	MC	1	1	4			0.60	0.38	0.1209	1.10
11	MC	1	1	3			0.63	0.45	-0.0239	0.99
12	MC	1	1	2			0.59	0.46	0.1428	1.01
13	MC	1	1	2			0.51	0.36	0.5683	1.14
14	MC	1	1	3			0.72	0.51	-0.5442	0.92
15	MC	1	1	2			0.61	0.55	0.0703	0.89
16	MC	1	1	3			0.60	0.47	0.1319	0.99
17	MC	1	1	4			0.49	0.45	0.6486	1.02
18	MC	1	1	2			0.60	0.45	0.1099	1.01
19	MC	1	1	5			0.61	0.52	0.0437	0.93
20	MC	1	1	2			0.68	0.47	-0.3350	0.97
21	MC	1	1	5			0.56	0.44	0.3326	1.03
22	MC	1	1	2			0.71	0.52	-0.4938	0.90
23	MC	1	1	3			0.76	0.57	-0.7751	0.83
24	MC	1	1	3			0.70	0.43	-0.4442	1.01
25	MC	1	1	4			0.66	0.59	-0.1765	0.84
26	MC	1	1	2			0.71	0.45	-0.4876	0.98
27	MC	1	1	2			0.75	0.49	-0.6818	0.92
28	MC	1	1	2			0.61	0.54	-0.0099	0.95
29	MC	1	1	2			0.60	0.45	0.0932	1.00
30	MC	1	1	3			0.52	0.31	0.5064	1.15
31	MC	1	1	5			0.74	0.55	-0.6426	0.86
32	MC	1	1	3			0.72	0.47	-0.5032	0.96
33	MC	1	1	3			0.68	0.45	-0.3282	0.98
34	MC	1	1	4			0.60	0.42	0.1363	1.03
35	MC	1	1	2			0.72	0.47	-0.5281	0.94

Position	Item Type	Max Points	Weight	Std	Key Idea	PI	Mean	Point-Biserial	RID	INFIT
36	MC	1	1	3			0.63	0.51	-0.0545	0.93
37	MC	1	1	2			0.78	0.55	-0.8872	0.84
38	MC	1	1	4			0.61	0.48	0.0878	0.96
39	MC	1	1	3			0.68	0.48	-0.3101	0.95
40	MC	1	1	2			0.55	0.39	0.3753	1.07
41	MC	1	1	2			0.72	0.45	-0.5156	0.97
42	MC	1	1	3			0.67	0.49	-0.2697	0.93
43	MC	1	1	2			0.62	0.42	0.0061	1.03
44	MC	1	1	2			0.58	0.50	0.2166	0.94
45	MC	1	1	5			0.46	0.36	0.8261	1.10
46	MC	1	1	2			0.55	0.35	0.3858	1.12
47	MC	1	1	2			0.61	0.53	0.0661	0.90
48	MC	1	1	2			0.50	0.54	0.6162	0.89
49	MC	1	1	4			0.46	0.39	0.7945	1.07
50	MC	1	1	3			0.54	0.49	0.4121	0.95
THM	ESS	5	3	2, 5			1.01	0.55	1.9428	0.99
DBQ	SCF	2	1	3			1.80	0.41	-1.3990	0.99
DBQ	SCF	1	1	2,3			0.84	0.42	-1.1855	0.93
DBQ	SCF	1	1	3,5			0.83	0.35	-1.1528	0.99
DBQ	SCF	1	1	3			0.76	0.41	-0.6273	0.99
DBQ	SCF	1	1	3,4			0.89	0.46	-1.6917	0.86
DBQ	SCF	1	1	2,3			0.91	0.44	-1.9622	0.87
DBQ	SCF	1	1	3,5			0.80	0.32	-0.8879	1.04
DBQ	SCF	1	1	2,3			0.84	0.41	-1.2188	0.93
DBQ	SCF	1	1	3			0.79	0.44	-0.8250	0.94
DBQ	SCF	1	1	3,4			0.73	0.45	-0.4680	0.96
DBQ	SCF	2	1	2			1.59	0.56	-0.5569	0.96
DBQ	ESS	5	3	2,3,4,5			1.19	0.55	2.7832	1.06

June 2014

Position	Item Type	Max Points	Weight	Std	Key Idea	PI	Mean	Point-Biserial	RID	INFIT
1	MC	1	1	3			0.65	0.55	-0.1446	0.89
2	MC	1	1	2			0.43	0.40	0.9573	1.10
3	MC	1	1	4			0.79	0.45	-0.9756	0.94
4	MC	1	1	3			0.74	0.31	-0.6761	1.13
5	MC	1	1	2			0.71	0.39	-0.5133	1.06
6	MC	1	1	3			0.68	0.32	-0.3446	1.15
7	MC	1	1	5			0.58	0.44	0.1762	1.04
8	MC	1	1	2			0.72	0.49	-0.5690	0.93
9	MC	1	1	2			0.68	0.45	-0.3183	0.99
10	MC	1	1	3			0.60	0.46	0.0728	1.01
11	MC	1	1	3			0.52	0.51	0.5236	0.97
12	MC	1	1	4			0.74	0.53	-0.6980	0.88
13	MC	1	1	2			0.56	0.51	0.3086	0.96
14	MC	1	1	2			0.55	0.54	0.3685	0.92
15	MC	1	1	3			0.70	0.56	-0.4381	0.85
16	MC	1	1	5			0.40	0.39	1.0998	1.11
17	MC	1	1	3			0.48	0.38	0.7209	1.13
18	MC	1	1	3			0.58	0.58	0.2064	0.87
19	MC	1	1	2			0.58	0.52	0.2064	0.95
20	MC	1	1	4			0.54	0.53	0.2977	0.96
21	MC	1	1	2			0.63	0.52	-0.0381	0.93
22	MC	1	1	5			0.49	0.43	0.6609	1.07
23	MC	1	1	2			0.64	0.50	-0.1319	0.94
24	MC	1	1	3			0.60	0.48	0.0789	0.99
25	MC	1	1	5			0.51	0.56	0.5838	0.91
26	MC	1	1	3			0.58	0.34	0.2445	1.19
27	MC	1	1	4			0.75	0.50	-0.7304	0.93
28	MC	1	1	4			0.56	0.39	0.3522	1.13
29	MC	1	1	2			0.54	0.52	0.4356	0.96
30	MC	1	1	2			0.68	0.32	-0.2919	1.18
31	MC	1	1	2			0.67	0.47	-0.2268	1.00
32	MC	1	1	2			0.63	0.45	-0.0179	1.04
33	MC	1	1	3			0.61	0.45	0.0991	1.05
34	MC	1	1	4			0.72	0.46	-0.5495	1.00
35	MC	1	1	2			0.60	0.49	0.1357	1.00
36	MC	1	1	3			0.79	0.52	-1.0220	0.88
37	MC	1	1	5			0.78	0.57	-0.8998	0.83

Position	Item Type	Max Points	Weight	Std	Key Idea	PI	Mean	Point-Biserial	RID	INFIT
38	MC	1	1	3			0.67	0.49	-0.2332	0.97
39	MC	1	1	4			0.50	0.43	0.6313	1.09
40	MC	1	1	5			0.53	0.56	0.5186	0.91
41	MC	1	1	2			0.63	0.48	-0.0491	1.00
42	MC	1	1	2			0.72	0.46	-0.5425	0.99
43	MC	1	1	2			0.60	0.63	0.1478	0.82
44	MC	1	1	3			0.51	0.47	0.6194	1.04
45	MC	1	1	2			0.47	0.34	0.7858	1.20
46	MC	1	1	3			0.54	0.54	0.4296	0.93
47	MC	1	1	2			0.65	0.54	-0.1509	0.90
48	MC	1	1	2			0.53	0.56	0.4831	0.90
49	MC	1	1	2			0.74	0.51	-0.6495	0.93
50	MC	1	1	3			0.55	0.50	0.3820	0.98
THM	ESS	5	3	2, 4, 5			1.15	0.58	1.376	1.06
1	SCF	1	1	2, 3, 4			0.57	0.40	0.4340	1.05
2	SCF	1	1	2, 4			0.60	0.40	0.258	1.05
3	SCF	1	1	2, 4			0.88	0.35	-1.6466	0.96
4a	SCF	1	1	2, 4			0.78	0.56	-0.7444	0.84
4b	SCF	1	1	2, 3, 4			0.75	0.53	-0.5883	0.87
5	SCF	1	1	2, 4			0.82	0.52	-1.0217	0.87
6	SCF	1	1	2, 3, 4			0.57	0.47	0.4284	0.98
7a	SCF	1	1	2, 3			0.80	0.46	-0.9364	0.93
7b	SCF	1	1	2, 5			0.75	0.51	-0.5811	0.90
8	SCF	1	1	2			0.71	0.56	-0.3534	0.86
9	SCF	1	1	2, 3			0.78	0.50	-0.7909	0.90
DBQ	ESS	5	3	2,3,4,5			1.35	0.63	1.0995	1.03

August 2014

Position	Item Type	Max Points	Weight	Std	Key Idea	PI	Mean	Point-Biserial	RID	INFIT
1	MC	1	1	3			0.75	0.49	-0.6966	0.92
2	MC	1	1	2			0.74	0.51	-0.6397	0.90
3	MC	1	1	3			0.72	0.40	-0.5224	1.01
4	MC	1	1	2			0.41	0.44	1.0786	0.99
5	MC	1	1	2			0.66	0.50	-0.2915	0.93
6	MC	1	1	2			0.69	0.51	-0.3319	0.91
7	MC	1	1	2			0.61	0.51	0.0767	0.93
8	MC	1	1	3			0.65	0.54	-0.1141	0.88
9	MC	1	1	3			0.50	0.35	0.6313	1.12
10	MC	1	1	3			0.44	0.45	0.9217	0.99
11	MC	1	1	3			0.49	0.50	0.6712	0.92
12	MC	1	1	2			0.66	0.52	-0.2061	0.91
13	MC	1	1	3			0.44	0.38	0.8930	1.07
14	MC	1	1	4			0.41	0.36	0.9957	1.10
15	MC	1	1	2			0.62	0.34	0.0355	1.12
16	MC	1	1	2			0.60	0.45	0.1118	1.00
17	MC	1	1	4			0.62	0.38	0.0296	1.06
18	MC	1	1	5			0.56	0.39	0.3365	1.07
19	MC	1	1	3			0.53	0.53	0.4785	0.91
20	MC	1	1	2			0.64	0.59	-0.0598	0.83
21	MC	1	1	5			0.64	0.54	-0.0839	0.89
22	MC	1	1	4			0.66	0.26	-0.2271	1.19
23	MC	1	1	4			0.78	0.52	-0.8676	0.86
24	MC	1	1	3			0.44	0.47	0.9045	0.97
25	MC	1	1	2			0.67	0.37	-0.2234	1.08
26	MC	1	1	3			0.70	0.39	-0.3684	1.05
27	MC	1	1	2			0.62	0.33	0.0469	1.13
28	MC	1	1	5			0.59	0.34	0.1803	1.12
29	MC	1	1	2			0.62	0.49	0.0223	0.96
30	MC	1	1	2			0.55	0.46	0.3355	1.01
31	MC	1	1	4			0.65	0.37	-0.1274	1.08
32	MC	1	1	5			0.52	0.27	0.5469	1.21
33	MC	1	1	2			0.60	0.41	0.1260	1.04
34	MC	1	1	3			0.41	0.46	1.0790	0.96
35	MC	1	1	2			0.68	0.49	-0.2755	0.94
36	MC	1	1	2			0.49	0.34	0.6581	1.14
37	MC	1	1	2			0.76	0.43	-0.7851	0.96
38	MC	1	1	3			0.66	0.55	-0.1533	0.86

Position	Item Type	Max Points	Weight	Std	Key Idea	PI	Mean	Point-Biserial	RID	INFIT
39	MC	1	1	2			0.68	0.44	-0.2624	1.00
40	MC	1	1	3			0.66	0.55	-0.1719	0.88
41	MC	1	1	3			0.63	0.52	-0.0147	0.92
42	MC	1	1	2			0.71	0.39	-0.4777	1.04
43	MC	1	1	2			0.73	0.63	-0.5694	0.77
44	MC	1	1	3			0.67	0.51	-0.2105	0.92
45	MC	1	1	2			0.68	0.51	-0.2821	0.91
46	MC	1	1	2			0.57	0.46	0.3116	1.00
47	MC	1	1	2			0.69	0.50	-0.3617	0.92
48	MC	1	1	2			0.52	0.43	0.5176	1.03
49	MC	1	1	2			0.44	0.46	0.9469	0.98
50	MC	1	1	3			0.47	0.44	0.7872	1.00
THM	ESS	5	3	2, 3, 4, 5			1.36	0.58	0.866	1.10
1	SCF	1	1	2, 3			0.84	0.42	-1.2322	0.98
2a	SCF	1	1	2, 3			0.89	0.47	-1.781	0.87
2b	SCF	1	1	2, 5			0.81	0.54	-1.0267	0.85
3a	SCF	1	1	2			0.90	0.46	-1.9230	0.88
3b	SCF	1	1	2, 5			0.89	0.51	-1.7949	0.82
4	SCF	1	1	2, 3			0.92	0.41	-2.2175	0.87
5a	SCF	1	1	3, 4			0.75	0.53	-0.5701	0.89
5b	SCF	1	1	3, 4			0.87	0.48	-1.5309	0.88
6	SCF	1	1	2, 4			0.85	0.57	-1.3229	0.80
7	SCF	1	1	2, 3, 4			0.86	0.52	-1.4625	0.83
8a	SCF	1	1	4			0.83	0.50	-1.1645	0.89
8b	SCF	1	1	4			0.81	0.52	-1.0267	0.87
9	SCF	2	1	2, 5			1.30	0.53	0.0729	1.15
DBQ	ESS	5	3	2, 3, 4, 5			1.45	0.62	0.9655	1.05

Appendix G: Scoring Tables

January 2014

Raw Score	Ability	Scale Score
0	−5.777	0.000
1	−4.560	0.660
2	−3.848	1.359
3	−3.424	2.032
4	−3.118	2.706
5	−2.877	3.463
6	−2.676	4.317
7	−2.503	5.114
8	−2.351	5.967
9	−2.215	6.850
10	−2.090	7.734
11	−1.976	8.690
12	−1.870	9.684
13	−1.770	10.679
14	−1.677	11.685
15	−1.588	12.760
16	−1.503	13.873
17	−1.423	14.986
18	−1.345	16.107
19	−1.270	17.226
20	−1.198	18.425
21	−1.128	19.651
22	−1.060	20.878
23	−0.994	22.110
24	−0.930	23.351
25	−0.867	24.677
26	−0.805	25.916
27	−0.745	27.238
28	−0.686	28.567
29	−0.627	29.895
30	−0.570	31.229
31	−0.513	32.544
32	−0.457	33.872
33	−0.402	35.296
34	−0.347	36.601
35	−0.293	38.023
36	−0.239	39.331
37	−0.186	40.734
38	−0.132	42.051
39	−0.079	43.451
40	−0.026	44.849
41	0.026	46.232
42	0.079	47.536
43	0.132	48.931
44	0.185	50.320
45	0.238	51.628
46	0.291	53.011
47	0.345	54.391
48	0.399	55.672
49	0.453	57.057
50	0.508	58.451
51	0.564	59.739
52	0.620	61.047
53	0.677	62.443
54	0.735	63.739
55	0.794	65.035
56	0.854	66.353
57	0.915	67.675
58	0.977	68.992
59	1.041	70.215
60	1.106	71.546
61	1.173	72.800
62	1.242	74.062
63	1.313	75.326
64	1.386	76.609
65	1.461	77.863
66	1.539	79.115
67	1.620	80.388
68	1.703	81.627
69	1.790	82.799
70	1.881	84.059
71	1.975	85.242
72	2.072	86.442
73	2.174	87.605
74	2.279	88.729
75	2.389	89.869
76	2.502	90.966
77	2.620	91.983
78	2.741	92.989
79	2.866	93.840
80	2.994	94.655
81	3.127	95.378
82	3.264	95.992
83	3.407	96.556
84	3.556	97.046
85	3.713	97.449
86	3.883	97.817
87	4.068	98.155
88	4.277	98.472
89	4.521	98.736
90	4.822	98.951
91	5.233	99.207
92	5.920	99.570
93	7.109	100.000

June 2014

Raw Score	Ability	Scale Score
0	−5.610	0.000
1	−4.392	0.825
2	−3.679	1.612
3	−3.253	2.397
4	−2.945	3.236
5	−2.702	4.199
6	−2.499	5.131
7	−2.325	6.123
8	−2.171	7.143
9	−2.033	8.185
10	−1.907	9.324
11	−1.792	10.465
12	−1.684	11.605
13	−1.583	12.819
14	−1.488	14.074
15	−1.399	15.320
16	−1.313	16.579
17	−1.232	17.852
18	−1.153	19.209
19	−1.078	20.558
20	−1.005	21.905
21	−0.935	23.254
22	−0.866	24.684
23	−0.800	26.020
24	−0.735	27.447

Raw Score	Ability	Scale Score
25	−0.672	28.867
26	−0.611	30.282
27	−0.550	31.686
28	−0.491	33.075
29	−0.432	34.518
30	−0.375	35.937
31	−0.318	37.366
32	−0.262	38.760
33	−0.207	40.179
34	−0.152	41.553
35	−0.098	42.962
36	−0.044	44.387
37	0.010	45.803
38	0.063	47.133
39	0.116	48.520
40	0.170	49.923
41	0.223	51.245
42	0.276	52.608
43	0.329	53.989
44	0.382	55.279
45	0.436	56.622
46	0.490	57.997
47	0.545	59.295
48	0.600	60.570
49	0.655	61.907

Raw Score	Ability	Scale Score
50	0.711	63.208
51	0.768	64.468
52	0.826	65.739
53	0.884	67.016
54	0.944	68.295
55	1.004	69.515
56	1.066	70.744
57	1.130	71.991
58	1.194	73.195
59	1.261	74.403
60	1.329	75.613
61	1.399	76.841
62	1.471	78.020
63	1.545	79.222
64	1.622	80.425
65	1.702	81.600
66	1.784	82.713
67	1.869	83.901
68	1.957	85.029
69	2.049	86.159
70	2.144	87.284
71	2.243	88.344
72	2.346	89.422
73	2.452	90.505
74	2.562	91.490

Raw Score	Ability	Scale Score
75	2.675	92.449
76	2.793	93.347
77	2.914	94.151
78	3.039	94.922
79	3.168	95.570
80	3.301	96.149
81	3.441	96.682
82	3.587	97.128
83	3.742	97.511
84	3.910	97.875
85	4.093	98.198
86	4.300	98.498
87	4.543	98.760
88	4.843	98.964
89	5.254	99.220
90	5.940	99.577
91	7.130	100.000

August 2014

Raw Score	Ability	Scale Score
0	−5.918	0.000
1	−4.695	0.527
2	−3.974	1.236
3	−3.541	1.840
4	−3.226	2.460
5	−2.975	3.136
6	−2.766	3.910
7	−2.585	4.731
8	−2.426	5.523
9	−2.282	6.403
10	−2.150	7.295
11	−2.029	8.219
12	−1.917	9.235
13	−1.811	10.264
14	−1.712	11.301
15	−1.617	12.387
16	−1.528	13.546
17	−1.442	14.714
18	−1.360	15.887
19	−1.281	17.063
20	−1.205	18.309
21	−1.131	19.596
22	−1.060	20.884
23	−0.991	22.176
24	−0.923	23.477
25	−0.858	24.870
26	−0.793	26.168
27	−0.731	27.556
28	−0.669	28.939
29	−0.609	30.321
30	−0.550	31.697
31	−0.491	33.060
32	−0.434	34.477
33	−0.377	35.880
34	−0.321	37.286
35	−0.266	38.673
36	−0.211	40.069
37	−0.157	41.441
38	−0.103	42.831
39	−0.049	44.248
40	0.004	45.659
41	0.058	46.996
42	0.111	48.366
43	0.164	49.767
44	0.217	51.099
45	0.270	52.453
46	0.323	53.835
47	0.377	55.140
48	0.430	56.474
49	0.484	57.851
50	0.539	59.167
51	0.594	60.444
52	0.650	61.779
53	0.706	63.095
54	0.763	64.362
55	0.821	65.639
56	0.880	66.925
57	0.940	68.211
58	1.001	69.447
59	1.063	70.677
60	1.127	71.936
61	1.192	73.149
62	1.259	74.364
63	1.327	75.584
64	1.398	76.818
65	1.470	78.006
66	1.545	79.216
67	1.622	80.427
68	1.702	81.607
69	1.785	82.726
70	1.870	83.920
71	1.959	85.051
72	2.051	86.185
73	2.147	87.311
74	2.246	88.375
75	2.349	89.455
76	2.455	90.539
77	2.565	91.519
78	2.679	92.478
79	2.796	93.371
80	2.917	94.174
81	3.042	94.943
82	3.171	95.586
83	3.305	96.163
84	3.444	96.693
85	3.590	97.136
86	3.745	97.518
87	3.912	97.881
88	4.096	98.203
89	4.303	98.500
90	4.545	98.763
91	4.846	98.965
92	5.256	99.222
93	5.943	99.578
94	7.133	100.000