Psychometric Analyses of the 2006 MCAS High School Chemistry Test[1,2]

Wendy Lam and Ronald K. Hambleton
University of Massachusetts Amherst

February 7, 2008

[1] Financial support for completing this research came from Measured Progress and the Massachusetts Department of Education. The authors are responsible for the contents of the report and for any errors or misinterpretations of the data that may remain. The authors want to acknowledge the assistance of Michael Nering with the IRT model fit analyses.

[2] **Center for Educational Assessment Research Report No. 646**. Amherst, MA: University of Massachusetts, Center for Educational Assessment.

1. Goal of the Psychometric Analyses

The primary goal of our work has been to provide readers with a number of worthwhile psychometric analyses of the 2006 MCAS high school Chemistry Test. These analyses provide more detail on the Chemistry Test than it was possible to provide in the summary report prepared by Hambleton, Zhao, Smith, Lam, and Deng (2008). These analyses include (1) an item analysis, (2) descriptive statistics on the test scores including break-outs for several subgroups of students, (3) classical reliability analyses for the test scores organized by item format, and for the total test, (4) two investigations of test dimensionality, (5) item response theory (IRT) item calibrations obtained from fitting the three-parameter logistic model to binary-scored items and the graded response model to polytomously-scored items, (6) various item and test level model fit findings, (7) test information and conditional standard errors, and (8) the identification of differentially functioning test items..

2. Description of the Chemistry Test

The MCAS 2006 Grade 9/10 Chemistry Test consists of 45 items assessing nine standards (sometimes called “curriculum strands”): Properties of Matter, Atomic Structure, Periodicity, Chemical Bonding, Chemical Reactions and Stoichiometry, Gases and Kinetic Molecular Theory, Solutions, Acids and Bases, and Equilibrium and Kinetics. The test was administered in a 2-day session in May of 2006, the first session consisted of the first 26 items on the test; and the second session consisted of the remaining 19 items. More information about the curriculum and the test items can be found at www.doe.mass.edu.

Table 2.1 presents the number of items, by item type, and the total number of items and score points for the MCAS 2006 Grade 9/10 Chemistry Test. There are 40 multiple choice items (each with four choices) and five polytomously-scored performance items (or sometimes called "constructed response items"). Multiple choice items were scored dichotomously; a score of 1 for a correct answer, 0 otherwise. Performance items were scored polytomously, with possible scores ranging from 0 to 4.

Table 2.1 Number of Items by Item Type on the Chemistry Test

Item Type	Points	Grade 9/10
Multiple Choice	1 or 0	40 items
Performance	0 to 4	5 items
Total Number of Test Items		45
Maximum Points on the Test		60

3. Classical Item Analyses

In total, 15,880 students were administered the Chemistry Test. However, several exclusion criteria were implemented so as to reduce the distortion of findings due to the use of student responses that would introduce systematic errors into the data analyses. First, students who had a total test score of 0 were excluded. Clearly, these students had not taken the test seriously, or perhaps were not even present for the test administration. Students who attempted less than four items from session 1 or if students did not attempt any of the items in session 2 were excluded from the psychometric analyses. After applying these exclusion rules, there were 14,997 students left in the dataset. Therefore, about 6% of the examinee data were excluded. All of these students would receive very low scores for their test performance, but they served no useful purpose for our psychometric analyses of the items and the test and so they were deleted. Their inclusion in the analyses would have inflated most of the important item and test statistics of interest such as item discrimination indices, reliability estimates, and IRT model fit.

Item Difficulty

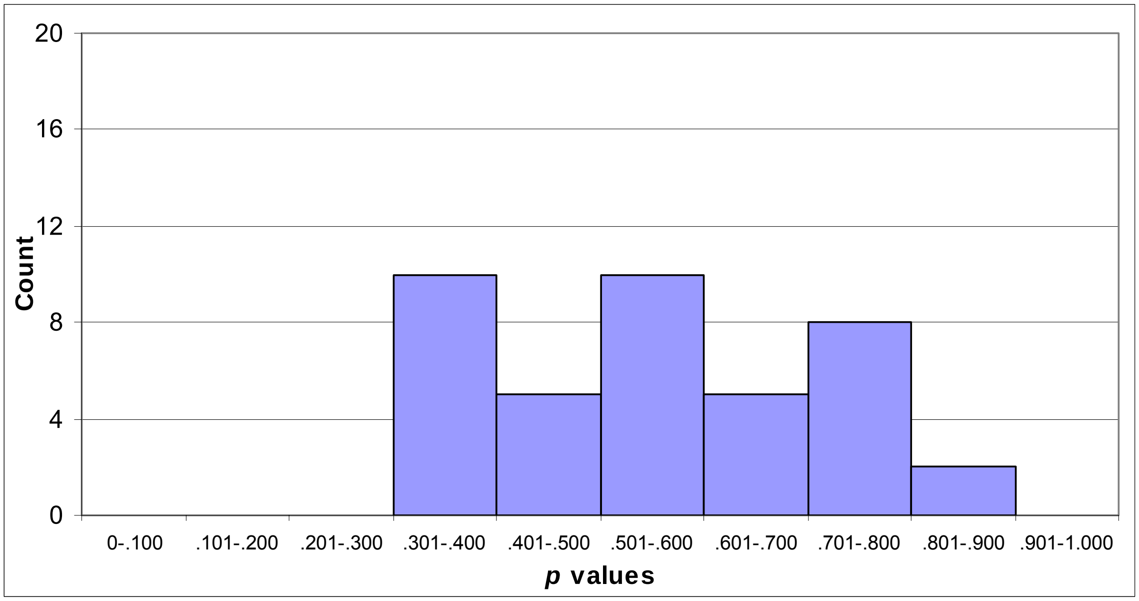
Item difficulty (*p*) is defined as the proportion of students answering an item correctly for dichotomous items (multiple choice items); or the average score for a polytomous item. It is also called the item mean score.

Table 3.1 presents the item difficulty for all multiple choice items based on valid student cases after the exclusion rules described above were applied. The item difficulty values range from .30 to .88 with an average of .56. This average is slightly higher than the overall test by .06 as only dichotomous items were included in the calculation. Students usually perform less well in polytomous items, therefore, making the overall performance lower when performance items were included. Averaged item difficulty for performance subtest items is .40, after rescaling to the same metric as those dichotomous items. Individual performance subtest item performance will be presented in Table 3.3. As presented in Figure 3.1, item difficulty for multiple choice items is uniformly distributed. There are 38% of the dichotomous items with a *p*-value less than or equal to .50, the other 62% are higher than .50. The range of difficulties seems appropriate to permit good measurement along the proficiency continuum.

Table 3.1 Distribution of Classical Item Difficulty Indices (N = 14,997)

Item[1]	*p*	Item[1]	*p*
1	.88	22	.52
2	.74	23	.60
3	.78	24	.62
4	.76	27	.80
5	.65	28	.74
6	.51	29	.56
7	.37	30	.64
8	.57	31	.73
9	.51	33	.73
10	.33	34	.38
12	.43	35	.37
13	.50	36	.72
14	.45	37	.40
15	.70	38	.59
16	.35	40	.48
17	.38	41	.30
18	.40	42	.62
19	.59	43	.37
20	.66	44	.48
21	.59	45	.42

Figure 3.1 Histogram Showing the Distribution of Classical Item Difficulty Indices



[1] This table only includes the multiple choice items.

Item Discrimination

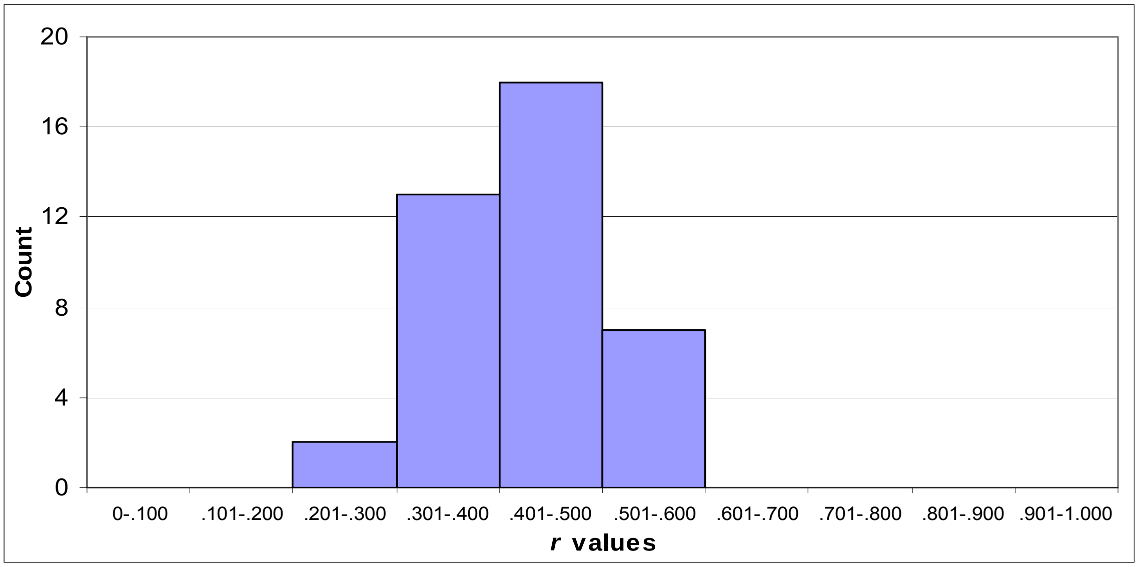
Item discrimination (*r*) is defined in this report as the correlation between item score and total test score. The correlation coefficient indicates the direction and strength of the relationship; it can range from -1.00 to 1.00.

Item discrimination for all multiple choice items are provided in Table 3.2, they are calculated after excluding all invalid students in the file. The distribution of the *r* values is presented graphically in Figure 3.2. All items are positively correlated to the total test score; the averaged item discrimination for all 40-multiple choice items is .42, with values ranging from .27 to .57. These statistical indicators suggest the items are excellent statistically. Individual item discrimination index for polytomous items are presented in Table 3.3. Averaged item discrimination index for this type of item is .76, higher than those multiple choice items.

Table 3.2 Distribution of Classical Item Discrimination Indices (N = 14,997)

Item[1]	*r*
1	.40
2	.44
3	.43
4	.53
5	.41
6	.47
7	.29
8	.34
9	.46
10	.37
12	.32
13	.47
14	.33
15	.56
16	.32
17	.27
18	.32
19	.45
20	.57
21	.41
22	.42
23	.45
24	.47
27	.34
28	.43
29	.41
30	.49
31	.52
33	.50
34	.41
35	.37
36	.57
37	.33
38	.46
40	.37
41	.36
42	.51
43	.48
44	.36
45	.52

Figure 3.2 Histogram Showing the Distribution of Classical Item Discrimination Indices



[1] This table only includes the multiple choice items.

Item Distractor Analyses

After excluding invalid cases based on the exclusion criteria discussed in the beginning of this section, 33% of the students (N = 4,980) were randomly chosen for an item distractor analysis.

The following item information and statistics are presented for each item in Table 3.3, and this time the polytomously scored items are included:

- Item – item number as it appeared on the test
- *p* – percent of students answering the dichotomous item correctly; or averaged points earned for the polytomous item
- *r* – correlation between score on an item with the total score
- Min – minimum score of the item
- Max – maximum score of the item
- Key – correct response for multiple choice items; key for performance
- Group – Total (all students from the random sample), High (top 25% of the total score based on the raw score distribution), Low (lowest 25% of the total score based on the raw score distribution)
- Percent of students endorsing each response option and omit rates for multiple choice items; or percent of students obtaining each score point and omit rates for the performance items.

Table 3.3 Classical Item Statistics for 2006 MCAS: Grade 9/10 Chemistry Test (N = 4,980)

Item	*p*	*r*	Min	Max	Key	Group	0	A/1	B/2	C/3	D/4	Omit
1	.87	.41	0	1	D	Total		5	3	5	87*	0
						High		1	0	0	99*	0
						Low		12	12	14	61*	1
2	.74	.45	0	1	A	Total		74*	2	15	8	0
						High		97*	0	2	1	0
						Low		44*	7	33	16	1
3	.79	.44	0	1	B	Total		9	79*	8	5	0
						High		2	97*	1	0	0
						Low		16	49*	21	13	1
4	.76	.53	0	1	A	Total		76*	9	8	6	0
						High		99*	0	0	0	0
						Low		38*	23	21	17	1
5	.65	.41	0	1	B	Total		12	65*	16	6	0
						High		5	92*	3	0	0
						Low		18	40*	28	13	1
6	.51	.48	0	1	B	Total		32	51*	5	12	0
						High		12	85*	0	3	0
						Low		41	23*	17	19	1
7	.37	.30	0	1	C	Total		29	20	37*	13	0
						High		14	19	61*	6	0
						Low		33	20	26*	21	1
8	.56	.32	0	1	B	Total		22	56*	8	13	0
						High		16	77*	2	5	0
						Low		27	36*	15	21	1
9	.51	.46	0	1	D	Total		9	26	13	51*	0
						High		3	11	5	81*	0
						Low		17	32	28	22*	1
10	.31	.35	0	1	D	Total		17	23	28	31*	1
						High		9	15	16	59*	0
						Low		23	27	30	19*	1
11	1.42	.72	0	4		Total	29	14	24	20	4	8
						High	4	7	30	45	13	1
						Low	62	12	4	0	0	21
12	.43	.32	0	1	A	Total		43*	30	7	19	0
						High		67*	15	3	15	0
						Low		26*	41	15	18	1
13	.50	.47	0	1	B	Total		10	50*	27	13	1
						High		5	85*	6	3	0
						Low		12	25*	39	22	1
14	.45	.32	0	1	A	Total		45*	11	27	17	1
						High		69*	3	21	7	0
						Low		28*	20	25	25	1
15	.70	.57	0	1	A	Total		70*	12	8	9	0
						High		98*	1	0	0	0
						Low		27*	27	21	23	1

Item	*p*	*r*	Min	Max	Key	Group	0	A/1	B/2	C/3	D/4	Omit
16	.34	.33	0	1	C	Total		37	24	34*	5	0
						High		29	9	59*	3	0
						Low		34	33	22*	11	1
17	.39	.26	0	1	B	Total		21	39*	22	18	1
						High		20	56*	12	11	0
						Low		22	25*	32	20	1
18	.40	.32	0	1	C	Total		26	5	40*	29	0
						High		9	1	64*	26	0
						Low		38	15	26*	21	0
19	.59	.46	0	1	B	Total		30	59*	6	5	0
						High		11	89*	0	0	0
						Low		40	29*	18	12	1
20	.66	.56	0	1	C	Total		13	13	66*	8	1
						High		1	1	96*	2	0
						Low		24	31	27*	17	1
21	.59	.41	0	1	D	Total		3	18	20	59*	0
						High		0	3	14	82*	0
						Low		11	32	27	29*	1
22	.51	.43	0	1	D	Total		8	27	13	51*	1
						High		1	19	3	77*	0
						Low		22	29	26	21*	1
23	.60	.47	0	1	A	Total		60*	13	15	11	1
						High		87*	3	3	7	0
						Low		27*	28	28	16	2
24	.62	.48	0	1	C	Total		10	17	62*	10	1
						High		6	2	90*	2	0
						Low		18	29	30*	21	2
25	1.49	.81	0	4		Total	18	26	19	18	8	11
						High	0	5	23	46	25	0
						Low	47	21	2	0	0	30
26	2.05	.80	0	4		Total	10	15	19	36	11	9
						High	0	1	6	57	36	0
						Low	31	26	10	2	0	32
27	.81	.33	0	1	A	Total		81*	9	5	5	0
						High		94*	5	1	0	0
						Low		60*	16	12	12	1
28	.74	.43	0	1	C	Total		3	8	74*	15	0
						High		0	1	95*	4	0
						Low		11	20	45*	24	1
29	.55	.43	0	1	D	Total		5	8	31	55*	0
						High		0	1	17	81*	0
						Low		15	22	37	26*	1
30	.65	.48	0	1	C	Total		6	17	65*	12	0
						High		2	4	92*	2	0
						Low		15	28	32*	25	1
31	.73	.50	0	1	D	Total		9	5	13	73*	0
						High		1	0	3	96*	0
						Low		23	15	27	35*	1

Item	*p*	*r*	Min	Max	Key	Group	0	A/1	B/2	C/3	D/4	Omit
32	1.35	.77	0	4		Total	22	23	25	9	9	12
						High	2	9	33	25	31	1
						Low	47	17	2	0	0	33
33	.73	.50	0	1	B	Total		9	73*	12	5	1
						High		1	97*	1	1	0
						Low		23	36*	26	11	3
34	.38	.41	0	1	D	Total		45	10	5	38*	1
						High		27	2	0	71*	0
						Low		43	19	15	19*	4
35	.37	.38	0	1	A	Total		37*	13	20	29	2
						High		64*	6	6	23	1
						Low		18*	23	30	26	4
36	.72	.57	0	1	A	Total		72*	9	8	10	1
						High		98*	0	1	1	0
						Low		28*	24	19	24	4
37	.39	.34	0	1	C	Total		17	32	39*	9	2
						High		8	20	66*	4	2
						Low		26	32	24*	14	4
38	.59	.46	0	1	B	Total		6	58*	21	12	2
						High		0	87*	10	3	0
						Low		18	27*	27	24	4
39	1.64	.74	0	4		Total	13	20	31	14	10	12
						High	1	8	28	29	33	0
						Low	35	21	9	0	0	35
40	.48	.36	0	1	A	Total		48*	20	15	10	6
						High		67*	21	7	3	2
						Low		21*	23	24	21	12
41	.31	.34	0	1	D	Total		12	28	24	31*	5
						High		8	20	14	57*	2
						Low		20	25	27	18*	11
42	.62	.51	0	1	D	Total		18	8	7	62*	5
						High		5	1	1	92*	1
						Low		27	17	19	26*	11
43	.38	.47	0	1	B	Total		35	38*	12	11	5
						High		18	75*	2	3	1
						Low		31	20*	22	16	11
44	.48	.35	0	1	C	Total		9	29	48*	8	5
						High		1	23	72*	2	2
						Low		20	25	28*	16	11
45	.43	.52	0	1	A	Total		43*	15	13	23	5
						High		81*	5	7	6	2
						Low		17*	23	20	30	11

Our impression from reviewing the distractor analysis is that the test items are of very good quality.

4. Basic Statistics and Reliability Analyses

Descriptive Statistics at Test Level

Table 4.1 presents the descriptive statistics at the test level for the overall group, then, it is broken down by gender, and also by ethnicity. There were around 1,100 more female students than male students who took the Chemistry Test. Test score performance was similar. Asian students performed better than all other ethnic groups, followed by White, Native American, Black, and Hispanic. One thing to note is that the *n*-counts from the gender or the ethnicity analyses do not add up to the overall as the demographic information was not complete in the dataset at the time we were analyzing the data.

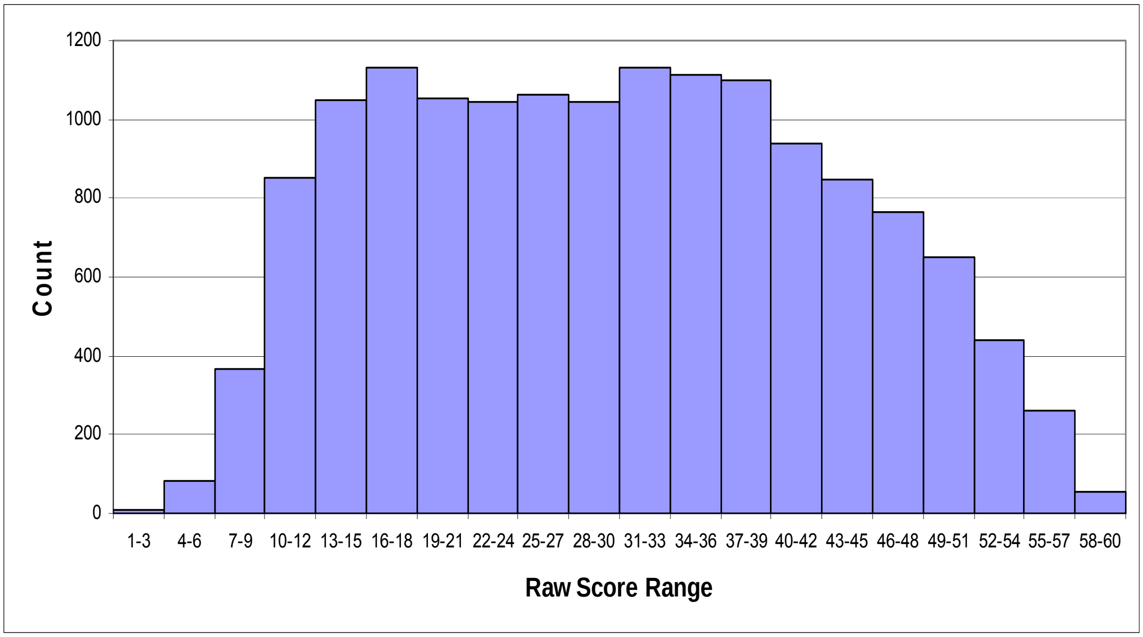
The raw score distribution for all students after applying the exclusion rules is presented in Figure 4.1. The distribution is relatively symmetric but platykurtic (Kurtosis = -.97), meaning that it is flat and less peaked about its mean than would be the case in a normal distribution.

Table 4.1 Descriptive Statistics for Overall 2006 MCAS: Grade 9/10 Chemistry Test

	N	$\overline{X}$	SD(*X*)	% of Points Earned
Overall	14,997	30.14	12.89	50%
Gender[1]				
Male	6,865	30.57	13.62	51%
Female	7,931	29.95	12.21	50%
Ethnicity[1]				
Asian	1,202	34.97	13.59	58%
Black	1,165	21.44	9.78	36%
Hispanic	1,376	19.53	9.53	33%
Native American	31	27.13	12.62	45%
White	11,015	32.09	12.38	53%

[1] N-count for gender and ethnicity groups does not add up to the overall.

Figure 4.1 Test Score Distribution for the 2006 MCAS Chemistry Test (N = 14,997)



Descriptive Statistics at the Content Standards Level

The 2006 MCAS Chemistry Test for Grade 9/10 follows the 2001 curriculum, which has 10 different Chemistry standards; however, only 9 were being tested. Table 4.2 presents the descriptive statistics for each of these standards.

The total number of items and total number of points varies between standards. Standard 8 (Acids and Bases) and Standard 9 (Equilibrium and Kinetics) have the least items, they only consist of 3 and 2 multiple choice items, respectively. Standard 1 (Properties of Matters) has the most number of items, however, both Standard 2 (Atomic Structure) and Standard 4 (Chemical Bonding) weight more in the test. By comparing the percentage of points earned (Mean/Total possible points) between standards, students performed best in Standard 3 (Periodicity), and poorest in Standard 8 (Acids and Bases).

Table 4. 2 Descriptive Statistics by Content Standard for 2006 MCAS: Grade 9/10 Chemistry Test

Content Standard	Number of Items	Number of Points	$\overline{X}$	SD(X)	% of Points Earned
1. Properties of Matters	8	8	4.75	1.97	59%
2. Atomic Structure	7	10	4.62	2.58	46%
3. Periodicity	5	5	3.59	1.37	72%
4. Chemical Bonding	7	10	4.47	2.66	45%
5. Chemical Reactions and Stoichiometry	4	7	3.80	1.96	54%
6. Gases and Kinetic Molecular Theory	5	8	3.68	2.03	46%
7. Solutions	4	7	3.00	1.92	43%
8. Acids and Bases	3	3	1.13	.93	38%
9. Equilibrium and Kinetics	2	2	1.11	.66	56%
10. Thermochemistry (Enthalpy)	--	--	--	--	--

Correlation between Pairs of Content Standard Scores

All nine content standards are positively correlated to the total score; however, correlations between Standard 8 (Acids and Bases) and Standard 9 (Equilibrium and Kinetics) with the total score are lower as they only contribute 3 and 2 points, respectively, to the test. The relationships among content standards are all positive, ranging from .27 (Standard 9: Equilibrium and Kinetics and Standard 8: Acids and Bases) to .74 (Standard 4: Chemical Bonding and Standard 2: Atomic Structure). Results are presented in Table 4.3.

Table 4.3 Intercorrelations between Content Standards and Total Score for 2006 MCAS: Grade 9/10 Chemistry Test

Content Standard[1]	1.	2.	3.	4.	5.	6.	7.	8.	9.	Total Score
1. Properties of Matters	1.00									
2. Atomic Structure	.66	1.00								
3. Periodicity	.59	.61	1.00							
4. Chemical Bonding	.64	.74	.62	1.00						
5. Chemical Reactions and Stoichiometry	.65	.69	.62	.70	1.00					
6. Gases and Kinetic Molecular Theory	.63	.67	.57	.66	.66	1.00				
7. Solutions	.60	.64	.53	.64	.64	.62	1.00			
8. Acids and Bases	.35	.40	.31	.42	.36	.36	.36	1.00		
9. Equilibrium and Kinetics	.43	.43	.36	.42	.41	.40	.42	.27	1.00	
Total Score	.81	.88	.75	.88	.85	.82	.80	.50	.54	1.00

[1] Standard 10 (Thermochemistry (Enthalpy)) was not tested in the 2006 test.

Reliability

Reliability is a characteristic of test scores that refers to the degree of consistency in students' assessment results over time, in parallel forms, and items within the same test and raters. Cronbach's Coefficient Alpha (α) statistics can be used as an estimate for internal consistency for both multiple choice items and polytomous items. These statistics are calculated based on multiple choice items only (40 items), performance items only (5 items), and at the overall test level (45 items), as presented in Table 4.4. Reliability indices at the standards level are also presented in Table 4.5.

Note that Standard 8 (Acids and Bases) and Standard 9 (Equilibrium and Kinetics) have the least number of items, therefore, the reliability index for these two content standards, as expected, are much lower than the others. None of these reliabilities at the standard level are important because no scores are reported for students at the content standard level. In fact, the reliability levels indicate why this would not be a good idea in the future either unless the numbers of items per content standard were increased substantially.

Table 4.4 Reliability Indices for the Total Test and by Item Types for 2006 MCAS: Grade 9/10 Chemistry Test

	Coefficient α
Multiple Choice	.89
Performance	.88
Total Test	.92

Table 4.5 Reliability Index at the Content Standard Level for 2006 MCAS: Grade 9/10 Chemistry Test

Content Standard[1]	Coefficient α
1. Properties of Matters	.61
2. Atomic Structure	.63
3. Periodicity	.60
4. Chemical Bonding	.67
5. Chemical Reaction and Stoichiometry	.51
6. Gases and Kinetic Molecular	.50
7. Solutions	.48
8. Acids and Bases	.28
9. Equilibrium and Kinetics	.21

5. Test Dimensionality

Eigenvalue Plots

If the correlation between scores on the multiple choice items and performance items is high, then it lends credibility to treating the construct as unidimensional and moving forward with a unidimensional IRT model. It is an initial and easy calculation to the question of whether or not the test is unidimensional. Correlation between item types and the total raw score is presented in Table 5.1. Random samples of students are used for the analyses that follow (N = 4,980). The correlations suggest that the test item format is not increasing the dimensionality of the test as the correlation between multiple-

[1] Standard 10 (Thermochemistry (Enthalpy)) was not being tested.

choice test scores and performance test scores is very high (r=.83) before any correction for score unreliability is even applied.

Table 5.1 Correlation Between Item Types and Total Raw Score for 2006 MCAS: Grade 9/10 Chemistry Test

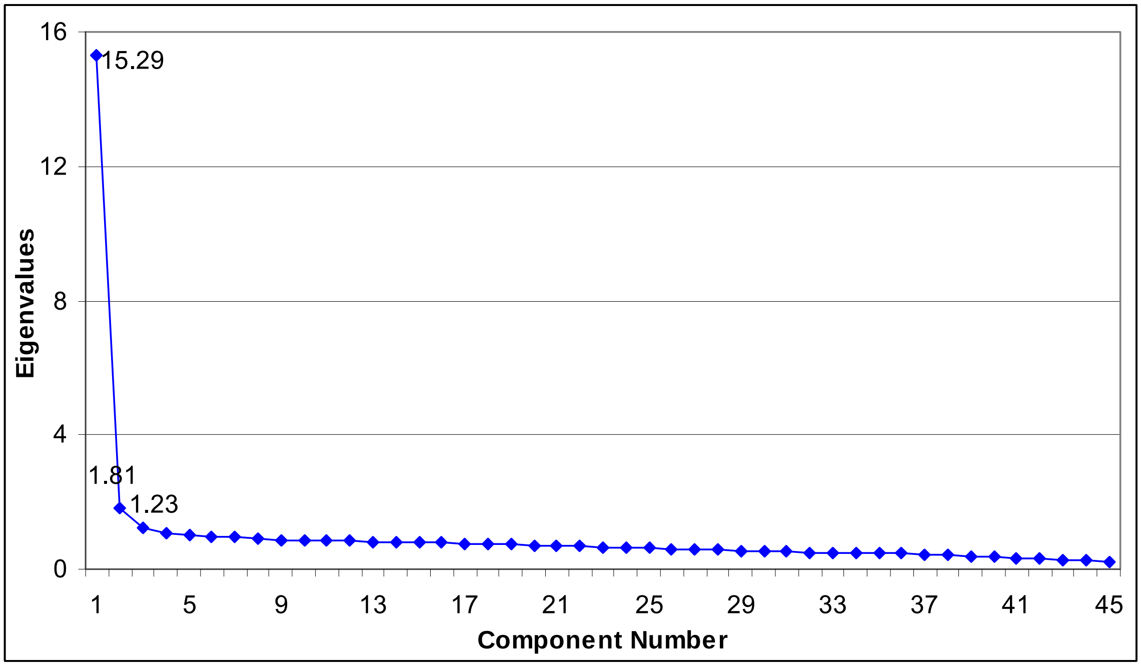
Item Type and Total Score	Multiple Choice	Performance	Total Raw Score
Multiple Choice	1.00		
Performance	.83	1.00	
Total Score	.97	.94	1.00

The 10 largest eigenvalues are listed in Table 5.2, and the 45 factors are plotted in Figure 5.1. The analysis shows that the Chemistry Test is dominated by a major first component, with a minor second factor. There is a significant drop in percent total test variance from the first to second eigenvalues, and slower decreasing rate for the remaining eigenvalues (only eight are shown here). This is more than sufficient to demonstrate the trend in the values. (The display shows the complete set.)

Table 5.2 Largest 10 Eigenvalues for the 2006 MCAS: Grade 9/10 Chemistry Test

Rank	Eigenvalue	Proportion of Variance Account for
1	15.29	34%
2	1.81	4%
3	1.23	3%
4	1.09	2%
5	1.03	2%
6	.98	2%
7	.97	2%
8	.91	2%
9	.88	2%
10	.86	2%

Figure 5.1 Eigenvalues Plot for the 2006 MCAS: Grade 9/10 Chemistry Test



With a dominant first factor (34% of the variability is explained by the first factor), and the first eigenvalue exceeding the second by a factor of more than 8 to 1, the evidence strongly supports the presence of a single factor. This evidence supports a decision to use a unidimensional IRT model in equating forms.

Parallel Analysis and SEM

Parallel analysis is conducted by generating 5,000 students' responses based on normal deviates with 10 replications to test if the second factor is due to random error. The result from the parallel analysis confirms that the Chemistry Test does have a minor second factor. Parallel analysis is presented in Figure 5.2, and factor loadings for a one-factor model are presented in Table 5.3. Again, the factor loadings highlight that a one-factor model provides an excellent accounting of the data.

Figure 5.2 Parallel Analysis for 2006 MCAS: Grade 9/10 Chemistry Using Random Normal Deviates

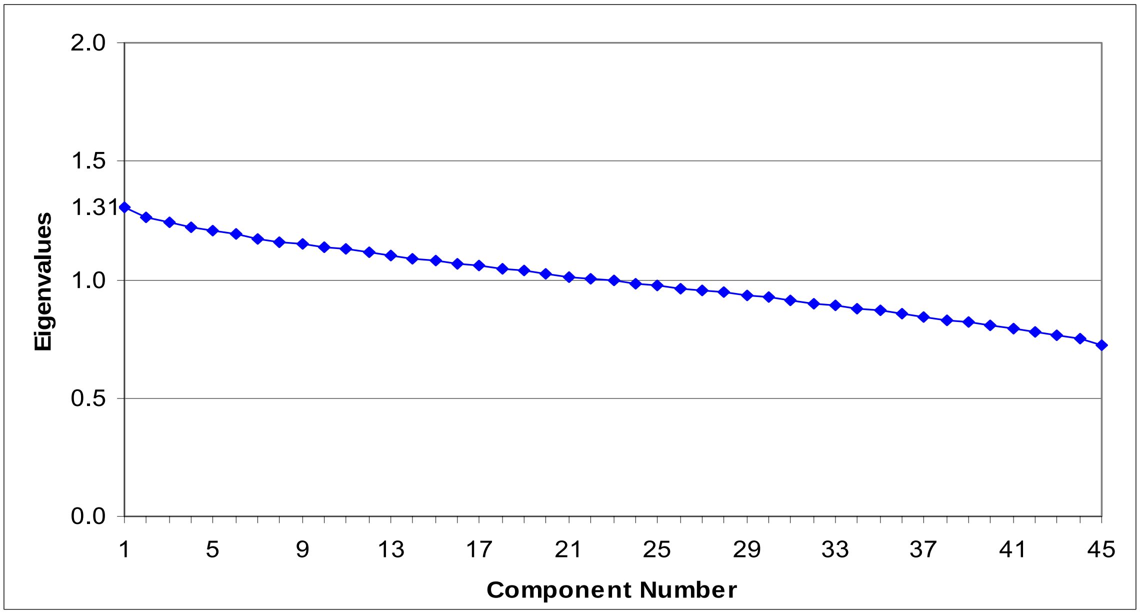


Table 5.3 Factor Loadings for a One Factor Model

Item	Factor Loading	Item	Factor Loading
1	.66	24	.61
2	.63	25	.88
3	.65	26	.88
4	.74	27	.55
5	.57	28	.60
6	.65	29	.55
7	.40	30	.65
8	.46	31	.74
9	.62	32	.85
10	.50	33	.70
11	.80	34	.53
12	.43	35	.49
13	.62	36	.79
14	.43	37	.44
15	.78	38	.61
16	.42	39	.84
17	.31	40	.49
18	.39	41	.52
19	.60	42	.70
20	.77	43	.72
21	.56	44	.51
22	.55	45	.73
23	.61		

6. Item Calibrations and Model Fit

Item Calibrations

The Parscale software program was used with a random sample of 4,980 examinees to fit an IRT model to the data. Three-parameter logistic model was used to calibrate multiple choice items, and the graded response model (GRM) was used to calibrate the polytomous items. The discriminant parameter (*a*), difficulty parameter (*b*), and the pseudo-guessing parameter (*c*) for multiple choice items are presented in Table 6.1; *a*-parameter, *b*-parameter, distance for each score point (d_1 to d_4) and thresholds for each score point (b_1 to b_4) for polytomous items are also presented in the table. Standard error (SE) of the *a*-, *b*-, *c*-parameters and the distance parameters (d_1 to d_4) are presented under their respective columns.

Table 6.1 *a*-, *b*-, *c*-Parameters, Distances and Threshold Estimates for Grade 9/10 Chemistry

Item	*a*	*b*	*c*	d_1	d_2	d_3	d_4	B_1	b_2	b_3	b_4
1	1.05	-1.57	.16								
	.05	.08	.05								
2	.96	-.47	.31								
	.06	.08	.03								
3	.95	-.73	.30								
	.06	.09	.04								
4	1.75	-.43	.34								
	.10	.04	.02								
5	1.31	.26	.40								
	.09	.05	.02								
6	1.30	.46	.22								
	.07	.03	.01								
7	1.30	1.25	.24								
	.11	.04	.01								
8	.66	.37	.27								
	.06	.10	.03								
9	.93	.31	.17								
	.06	.05	.02								
10	1.83	1.11	.20								
	.13	.03	.01								

Item	a	b	c	d_1	d_2	d_3	d_4	B_1	b_2	b_3	b_4
11	1.14	.67	.00	1.16	.67	-.27	-1.57	-.49	.00	.94	2.24
	.02	.01	.00	.02	.02	.02	.04				
12	.98	1.02	.27								
	.08	.05	.02								
13	1.41	.54	.26								
	.08	.03	.01								
14	1.30	1.06	.32								
	.11	.04	.01								
15	1.40	-.43	.20								
	.07	.04	.02								
16	1.43	1.26	.22								
	.11	.04	.01								
17	.89	1.43	.26								
	.09	.06	.01								
18	1.27	1.19	.28								
	.11	.04	.01								
19	1.01	.21	.27								
	.06	.05	.02								
20	1.59	-.15	.25								
	.08	.03	.02								
21	.65	-.15	.13								
	.04	.09	.03								
22	.65	.08	.09								
	.04	.07	.03								
23	.84	-.05	.17								
	.05	.07	.03								
24	.88	-.10	.20								
	.05	.07	.03								
25	1.55	.47	.00	1.16	.30	-.29	-1.17	-.69	.17	.76	1.63
	.02	.01	.00	.02	.01	.02	.02				
26	1.53	-.05	.00	1.08	.48	-.16	-1.40	-1.13	-.53	.11	1.35
	.02	.01	.00	.02	.02	.01	.02				
27	.62	-1.30	.22								
	.04	.18	.07								
28	1.04	-.29	.38								
	.07	.08	.03								
29	.62	-.05	.12								
	.04	.09	.03								
30	1.11	-.05	.26								
	.06	.05	.02								
31	1.20	-.52	.21								
	.06	.05	.03								
32	1.36	.63	.00	1.15	.41	-.53	-1.02	-.52	.22	1.16	1.65
	.02	.01	.00	.02	.02	.02	.03				
33	1.15	-.51	.24								
	.06	.06	.03								
34	1.27	.93	.19								
	.08	.03	.01								

Item	*a*	*b*	*c*	d_1	d_2	d_3	d_4	B_1	b_2	b_3	b_4
35	1.14	1.07	.22								
	.09	.04	.01								
36	1.38	-.55	.16								
	.06	.04	.02								
37	1.28	1.13	.27								
	.10	.04	.01								
38	.84	.01	.20								
	.05	.07	.03								
39	1.18	.34	.00	1.25	.50	-.53	-1.23	-.91	-.16	.87	1.57
	.02	.01	.00	.02	.02	.02	.03				
40	.41	-.14	.00								
	.04	.25	.08								
41	1.55	1.21	.18								
	.11	.03	.01								
42	.94	-.21	.15								
	.05	.06	.02								
43	1.91	.80	.17								
	.11	.02	.01								
44	.91	.78	.27								
	.07	.05	.02								
45	1.54	.57	.16								
	.08	.03	.01								

Model Fit

Chi-square statistics were used to, in a preliminary way, identify non-fitting items, and results are provided in Table 6.2. As seen in the table, at .05 alpha-level, there are 12 items (item 1, 10, 11, 21, 22, 25, 26, 27, 32, 36, 39, 40) that appear not to be fitting either the 3-PL or the GRM model. None of the polytomous items (Item 11, 25, 26, 32 and 39) are fitted by GRM model based on chi-square statistics but this is almost certainly because of the additional SRs used in the calculations. But it is well-known that these chi-square statistics tend to be inflated with large sample sizes, and so they provide far from conclusive information about model fit (Hambleton, Swaminathan & Rogers, 1991).

Table 6. 2 Chi-Square Item Fit Statistics for Grade 9/10 Chemistry

Item	Chi-sq	*df*	Prob	Not Fit
1	90.08	20	.00	*
2	32.70	26	.17	
3	34.00	26	.14	
4	32.18	22	.07	
5	37.81	28	.10	
6	37.23	30	.17	
7	21.49	30	.87	
8	32.29	30	.35	
9	37.46	30	.16	
10	46.84	30	.03	*
11	183.28	96	.00	*
12	27.02	30	.62	
13	40.25	30	.10	
14	35.54	30	.22	
15	30.93	24	.16	
16	33.78	30	.29	
17	22.07	30	.85	
18	34.55	30	.26	
19	29.81	30	.48	
20	29.45	25	.25	
21	47.10	30	.02	*
22	61.96	30	.00	*
23	24.21	30	.76	
24	29.56	30	.49	
25	186.59	84	.00	*
26	230.76	84	.00	*
27	75.78	29	.00	*
28	29.90	26	.27	
29	39.60	30	.11	
30	28.53	28	.44	
31	37.58	25	.05	
32	160.56	85	.00	*
33	24.90	26	.53	
34	19.60	30	.93	
35	34.56	30	.26	
36	39.03	24	.03	*
37	32.99	30	.32	
38	28.13	30	.56	
39	272.09	94	.00	*
40	104.12	30	.00	*
41	24.25	30	.76	
42	28.67	29	.48	
43	31.86	29	.33	
44	40.18	30	.10	
45	33.90	30	.29	

Since chi-square statistics are sensitive to sample size, graphical methods to determine model fit are preferred, and they are presented in Appendix A, Figure A.1. A few multiple choice items that appear on the Chemistry Test were not fit very well at the lower end of the proficiency continuum, for example, Item 1 to Item 6, Item 8, 13, 36, and 40. On the other hand, the polytomous items (Item 11, 25, 26, 32 and 39) were fit well by the GRM except for a small number of discrepancies. Small sample sizes are sometimes a problem in these graphical displays of model fit. Number of examinees, item *p*-value (for multiple choice questions) or item mean (for performance items) are presented in the displays for every item. In addition, the discrimination parameter (slope), difficulty parameter (location), and the pseudo-guessing parameter (the lower asymptote), along with the respective standard error of estimate are presented for the multiple choice items Only the slope and location parameters with their respective standard errors are presented for each polytomous item.

There were a total of 1,950 standardized residuals generated for the fit analysis, (using 30 quadrature points), and only about 2% of them were larger than two standard deviations. This result provides additional evidence that the unidimensional 3-PL model (for MCQ) and the GRM (for performance items) fit the MCAS Grade 9/10 Chemistry data very well.

Graphical comparisons of the relative frequency distribution and the cumulative frequency distribution for expected scores (assuming model fit) and the observed scores are presented in Figure 6.1 and Figure 6.2, and they highlight excellent model fit—the predictions could not be much better than they are.

Figure 6.1 Relative Frequency Distribution for Grade 9/10 Chemistry

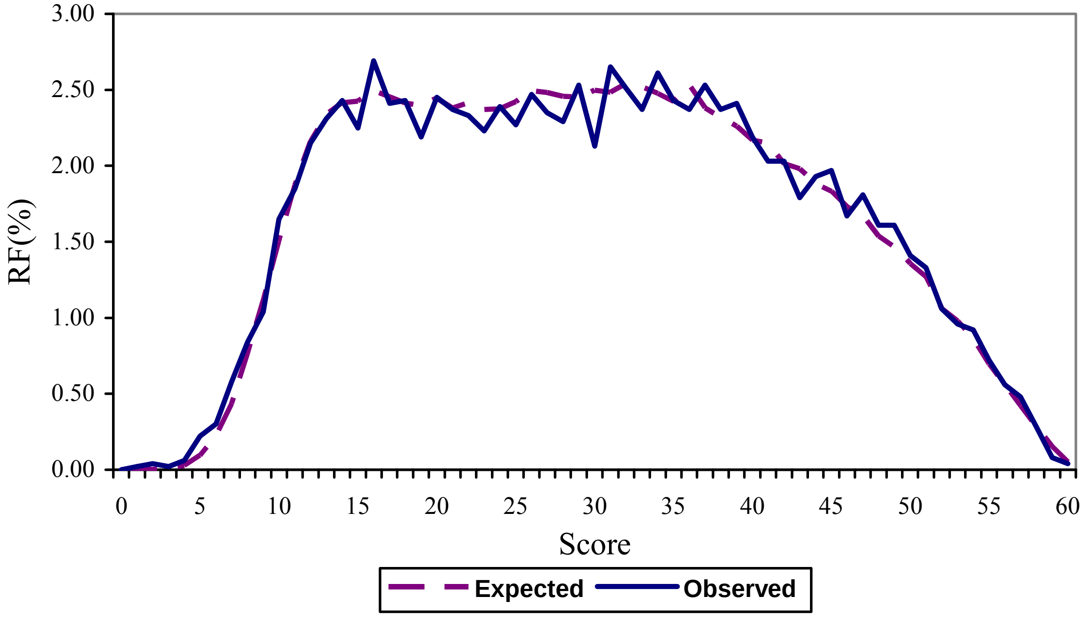
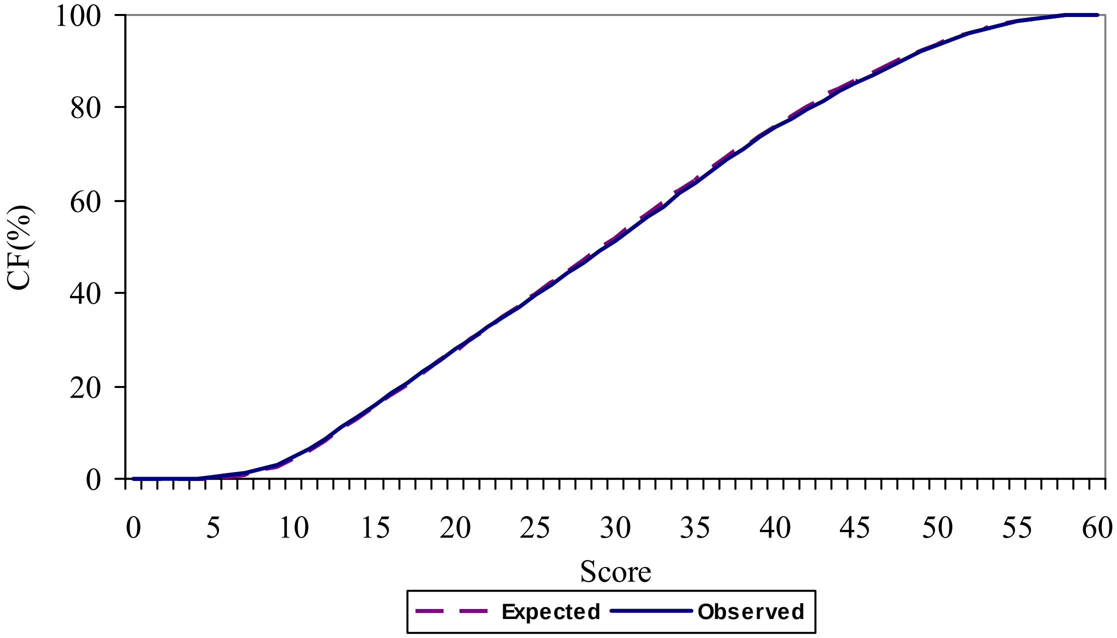


Figure 6.2 Cumulative Frequency Distribution for Grade 9/10 Chemistry



7. Test Information and Conditional Standard Errors

Test characteristics curve (TCC), test information function (TIF), and conditional standard error of measurement (CSEM) are presented in Figures 7.1 to 7.3, respectively. These analyses are important because they highlight the level of information achieved across the score reporting scale with the current Chemistry Test, and the associated conditional standard errors. The three figures reveal that test information is excellent, and correspondingly measurement errors are acceptably low for most regions along the proficiency scale where students are performing.

Figure 7. 1 Test Characteristics Curve (TCC) for 2006 MCAS Grade 9/10 Chemistry

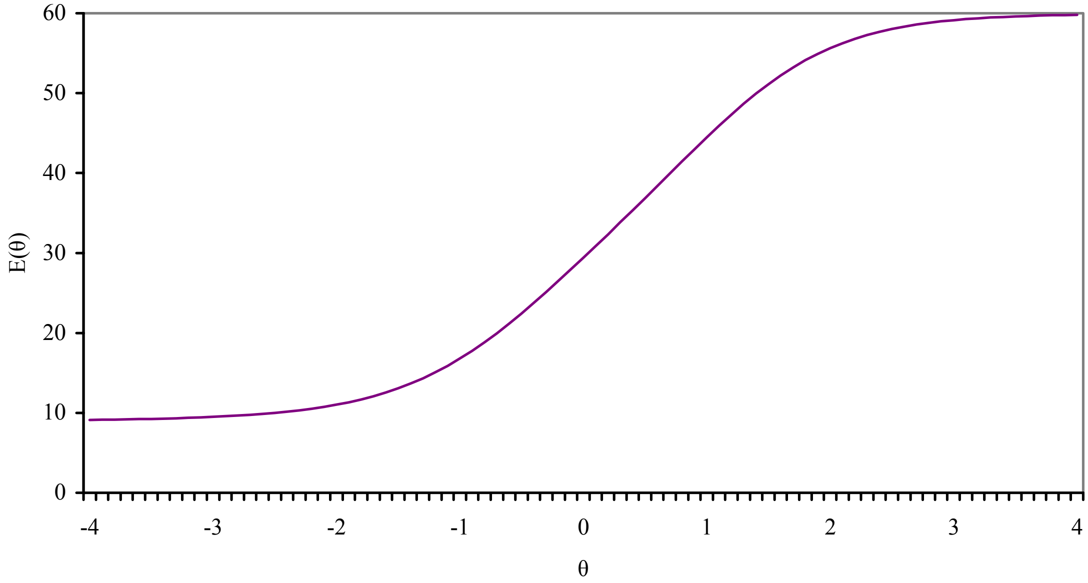


Figure 7.2 Test Information Function (TIF) for 2006 MCAS Grade 9/10 Chemistry

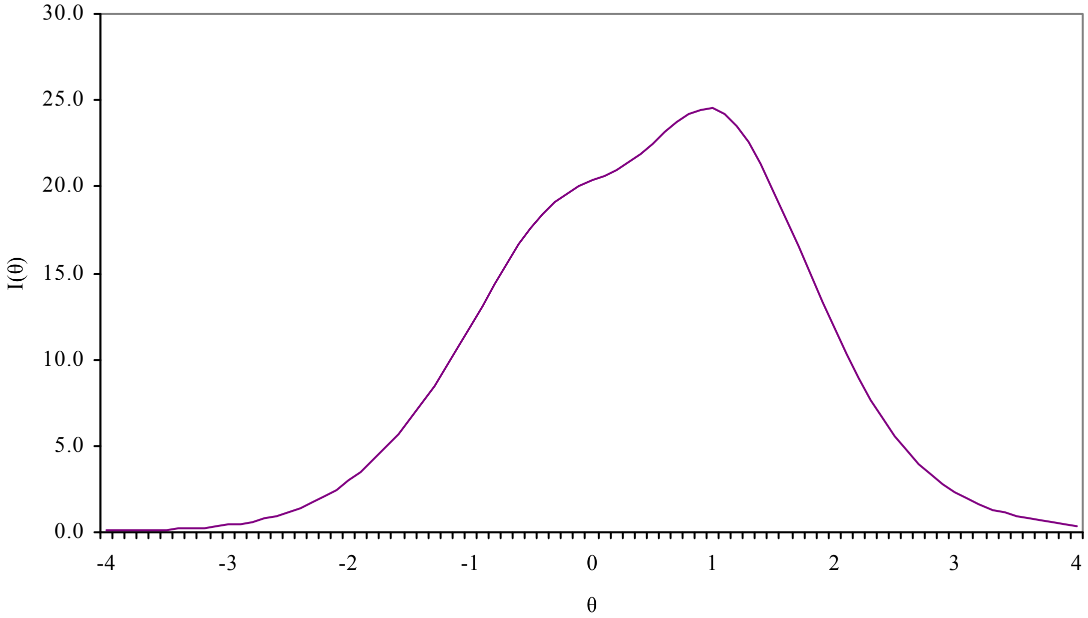
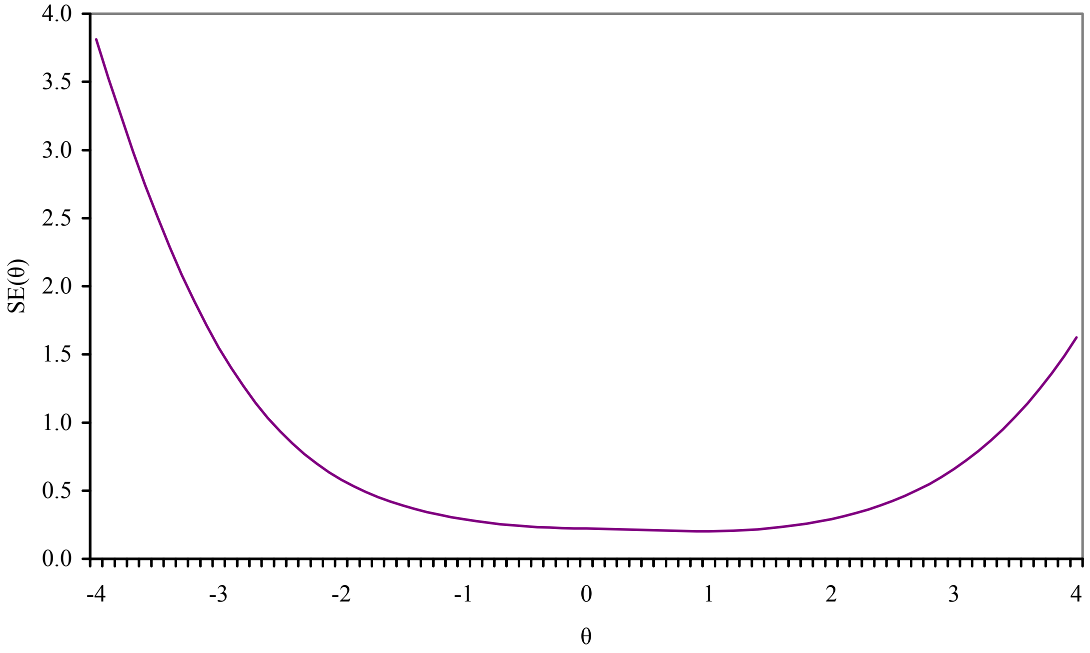


Figure 7.3 Conditional Standard Error of Measurement (CSEM) for 2006 MCAS Grade 9/10 Chemistry



8. Identification of Differentially Functioning Items

A weighted two-stage conditional *p*-value comparison procedure (see for example, Zenisky, Hambleton & Robin, 2003; Zenisky, Hambleton & Robin, 2004; Zenisky & Hambleton, 2007) was used to identify DIF items in the Chemistry Test between the reference group (male in gender DIF; and White in ethnic DIF) and the focal group (Black, Hispanic or Asian in ethnic analyses). Results from our analyses do not provide comparisons for all possible test scores, as the *n*-count for some of the score points are too small for any meaningful comparisons. Items were flagged as potentially DIF items at stage 1 if the unsigned DIF (UDIF) index was less than -.075 or greater than .075. Potential DIF items were flagged if the UDIF index was less than -.10 or greater than .10 in stage 2. This amounts to identifying items at stage 2 with sufficient conditional differences to account for about 1/10th of a point on the test score scale.

The number of DIF items flagged at stages 1 and 2 is summarized in Table 8.1. Only two items in total at Stage 2, and this is the important stage for identifying DIF, were identified from the four DIF analyses of 45 items each. This is a very small number and may be reflecting little more than chance.

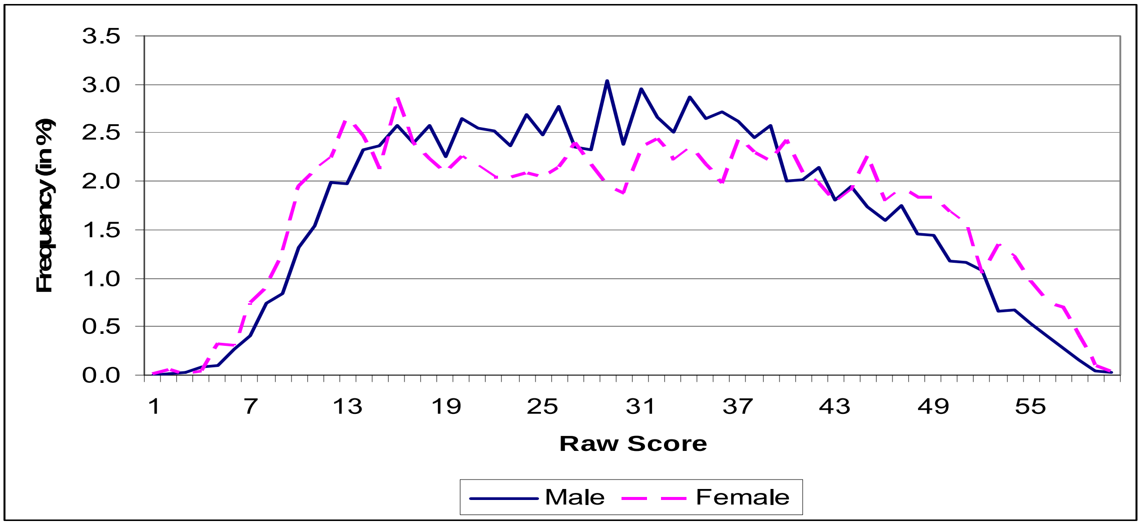
Table 8.1 Number of DIF Items Across Stage 1 and Stage 2, Reported by DIF Analysis

	Number of DIF items	
DIF Analysis	Stage 1	Stage 2
Male vs Female	2	0
White vs Black	18	1
White vs Hispanic	8	1
White vs Asian	16	0

Gender DIF

Descriptive statistics for the male and female groups of students are provided in Table 4.1 under the section of Basic Statistics and Reliability Analysis. As seen in the table, males and females perform similarly on the Chemistry Test. A plot of the total test score distribution for males and females is given in Figure 8.1 and show only small differences in shape. Also, females show a bit more score variability.

Figure 8.1 Total Test Score Distribution for Male and Female Students



A list of SDIF and UDIF indices across stage one and stage two is presented in Table 8.2. The complete set of DIF indices is presented in Figure 8.2. In addition, gender DIF is also presented in Figure 8.3. Neither Figure 8.2 nor Figure 8.3 reveal any patterns in the results due to item position or item content.

Table 8.2 Summary of DIF Indices: Males/Females

Item	Stage 1[1]		Stage 2[1]	
	SDIF	UDIF	SDIF	UDIF
1	-.029	-.035	-.027	-.033
2	.011	.034	.014	.036
3	.017	.035	.019	.036
4	.015	.033	.017	.030
5	.046	.062	.048	.063
6	-.017	-.044	-.013	-.044
7	.015	.048	.017	.043
8	.075	.084	.077	.084
9	-.014	-.040	-.011	-.046
10	.045	.054	.047	.055
11	-.123	-.038	-.028	-.034
12	.032	.056	.035	.048
13	.015	.038	.017	.042
14	.044	.062	.047	.063
15	-.037	-.047	-.034	-.044
16	.079	.081	.081	.088
17	-.010	-.045	-.008	-.034
18	-.007	-.033	-.005	-.037
19	.040	.054	.043	.055
20	.060	.068	.063	.070
21	-.048	-.059	-.046	-.059
22	.002	.047	.005	.043
23	-.004	-.041	.000	-.038
24	-.002	-.034	.000	-.048
25	-.041	-.021	-.007	-.017
26	-.111	-.029	-.025	-.027
27	.021	.032	.022	.036
28	.003	.030	.005	.032
29	.052	.061	.055	.065
30	.048	.056	.051	.064
31	.004	.034	.007	.031
32	-.132	-.036	-.030	-.032
33	-.011	-.034	-.009	-.029
34	.012	.042	.014	.045
35	.006	.047	.008	.048
36	-.031	-.041	-.027	-.035
37	.011	.038	.013	.043
38	-.040	-.055	-.038	-.055
39	-.110	-.031	-.025	-.032
40	.025	.052	.027	.057
41	.009	.035	.011	.034
42	-.005	-.045	-.003	-.043
43	.039	.050	.042	.055
44	.018	.061	.021	.056
45	.029	.043	.033	.044

[1] Items were flagged based on UDIF at the .075 level in stage 1 and .10 in stage 2.

Figure 8. 2 Gender DIF indices (MCQ: 1-10, 12-24, 27-31, 33-38, 40-45 and Constructed Response: 11, 25-26, 32, 39)

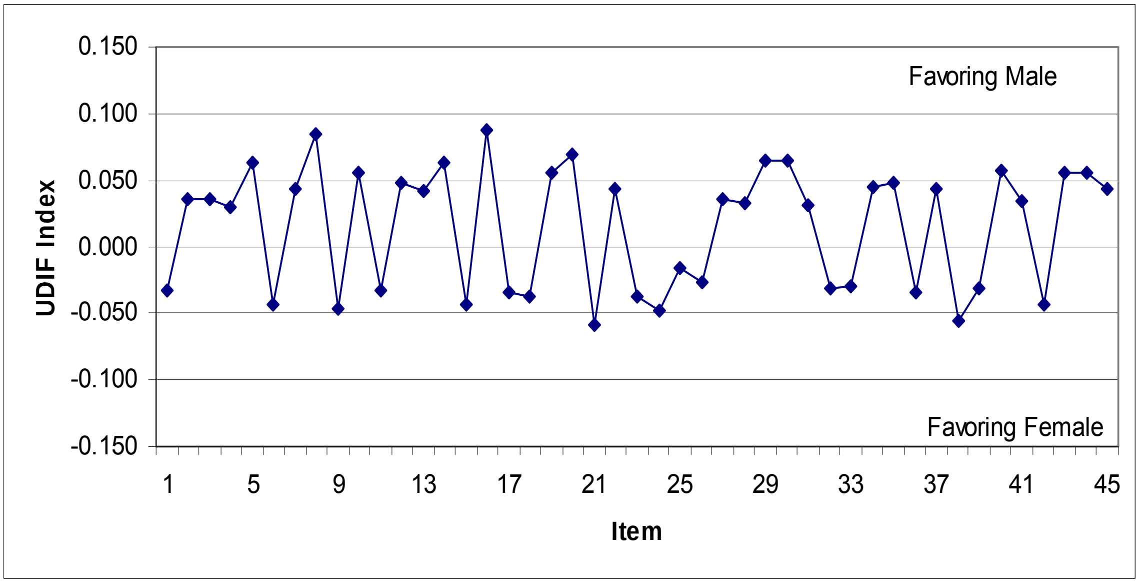
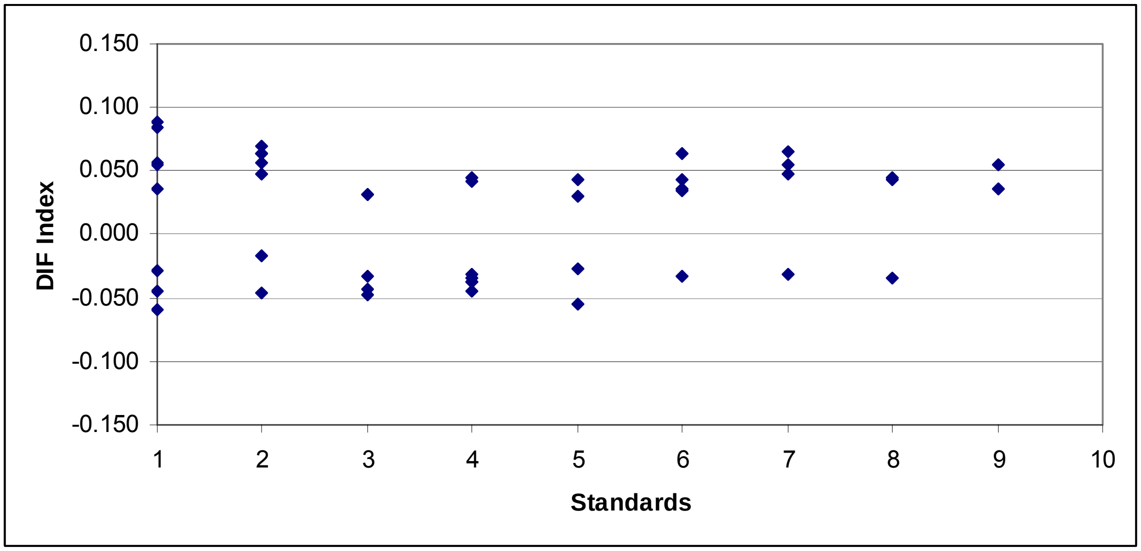


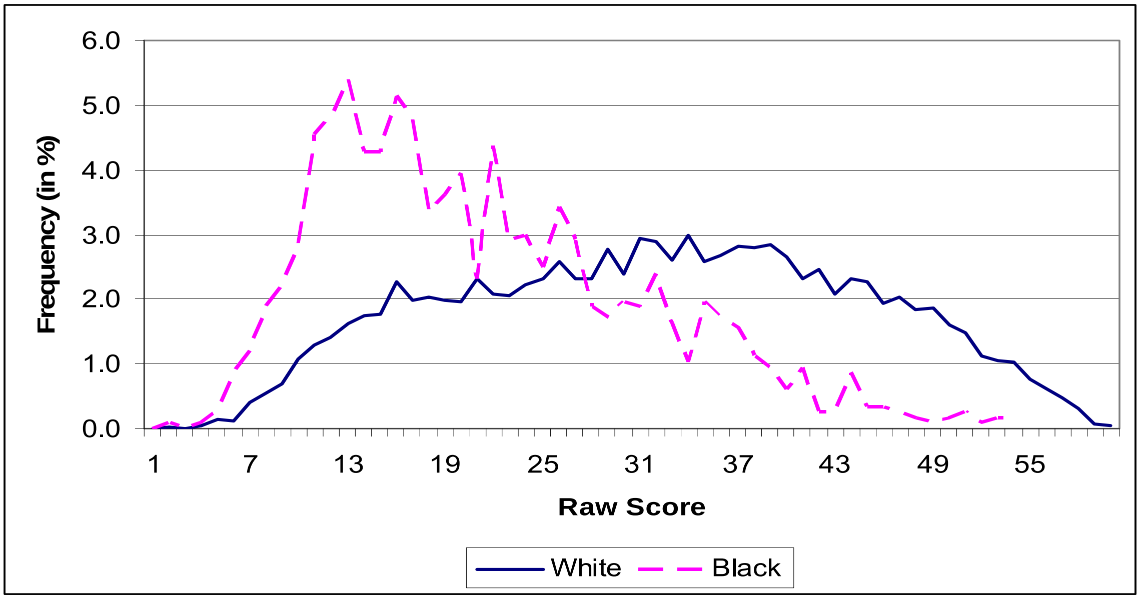
Figure 8.3 Gender DIF Organized by (Content) Standard



Ethnicity DIF – White vs. Black

Descriptive statistics for the White and Black groups of students are provided in Table 4.1 under the section of Basic Statistics and Reliability Analysis. As seen in the table, White students perform around 10 raw score points higher than Black students on the Chemistry Test. Plot of the total score distributions for the two groups is given in Figure 8.4.

Figure 8.4 Total Score Distribution for White and Black Students

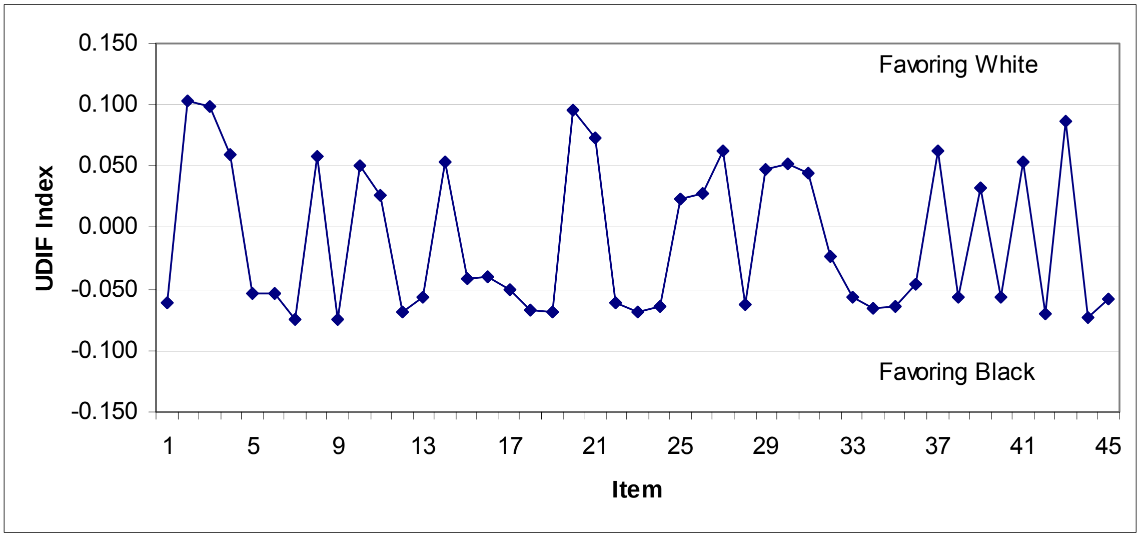


A list of SDIF and UDIF indices across stage one and stage two is presented in Table 8.3. And, the presentation of the complete set of DIF indices is presented in Figure 8.5. One multiple choice item (item 2) was flagged as DIF after stage two; the performance on this item for the two groups is shown in Figure 8.6. In the region on the test score scale where Black students are located, they tended to perform less well than the White students of similar overall performance on item 2. The plot is a bit erratic because of small numbers of Black students. It is very erratic at high scores and this pattern is definitely due to a small Black sample at the high end of the score scale. Ethnic DIF between White and Black students organized by content category reported in Figure 8.7 does not reveal a pattern in the data.

Table 8.3 Summary of DIF Indices: White/Black

	Stage 1[1]		Stage 2[1]	
Item	SDIF	UDIF	SDIF	UDIF
1	-.050	-.061	-.048	-.061
2	.057	.079	.080	.104
3	.077	.082	.093	.098
4	.049	.074	.047	.060
5	-.038	-.076	-.019	-.054
6	-.015	-.068	-.015	-.054
7	-.051	-.096	-.035	-.074
8	.006	.086	.033	.058
9	-.080	-.096	-.066	-.074
10	.030	.059	.029	.050
11	.049	.033	.022	.026
12	-.001	-.074	-.001	-.069
13	-.020	-.066	-.024	-.056
14	.005	.076	.025	.054
15	-.002	-.062	-.002	-.042
16	-.013	-.053	-.011	-.040
17	-.004	-.067	-.010	-.050
18	-.008	-.064	-.018	-.067
19	-.039	-.065	-.035	-.068
20	.049	.097	.067	.095
21	.012	.079	.027	.073
22	-.026	-.070	-.028	-.061
23	-.015	-.072	-.023	-.068
24	-.001	-.065	-.002	-.064
25	.075	.028	.065	.023
26	.096	.040	.063	.029
27	.025	.076	.042	.063
28	-.059	-.078	-.046	-.062
29	.004	.076	.027	.047
30	-.005	-.082	.016	.052
31	.006	.055	.005	.045
32	-.012	-.024	-.030	-.023
33	-.009	-.073	-.008	-.056
34	-.046	-.081	-.032	-.066
35	-.031	-.073	-.033	-.064
36	-.014	-.050	-.017	-.046
37	.012	.059	.007	.062
38	-.050	-.088	-.034	-.057
39	.107	.036	.089	.032
40	-.043	-.076	-.028	-.057
41	.007	.065	.006	.053
42	-.071	-.097	-.055	-.070
43	.065	.083	.081	.086
44	-.013	-.067	-.013	-.073
45	-.018	-.075	-.015	-.058

Figure 8. 5 White/Black DIF Indices (MCQ: 1-10, 12-24, 27-31, 33-38, 40-45 and Constructed Response: 11, 25-26, 32, 39)



[1] Items were flagged based on UDIF at the .075 level in stage 1 and .10 in stage 2.

Figure 8.6 Conditional *p*-Value Plot for Item 2

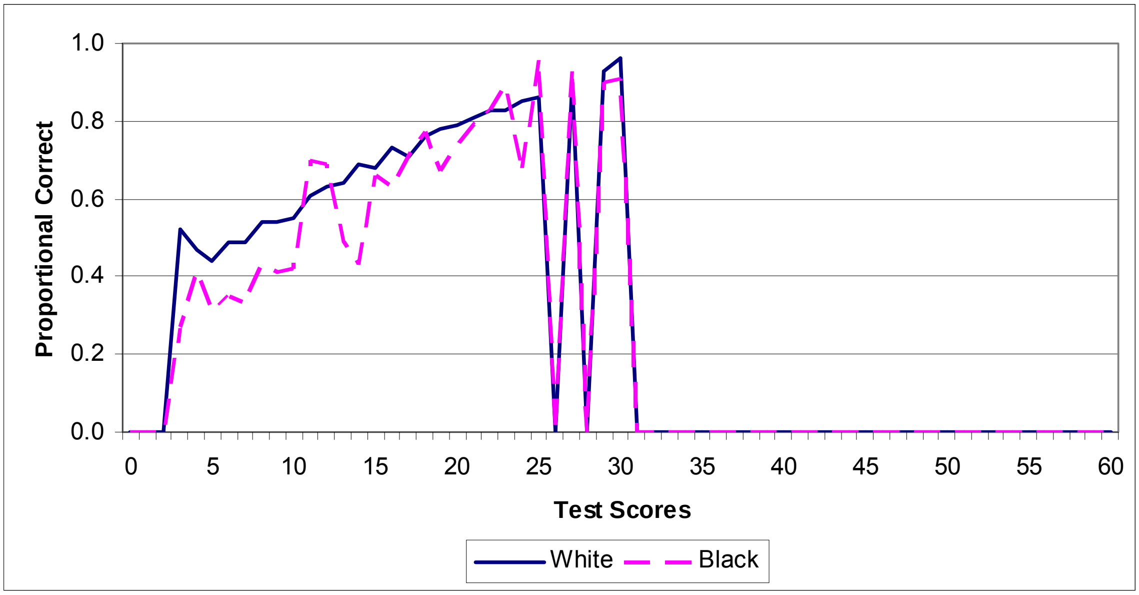
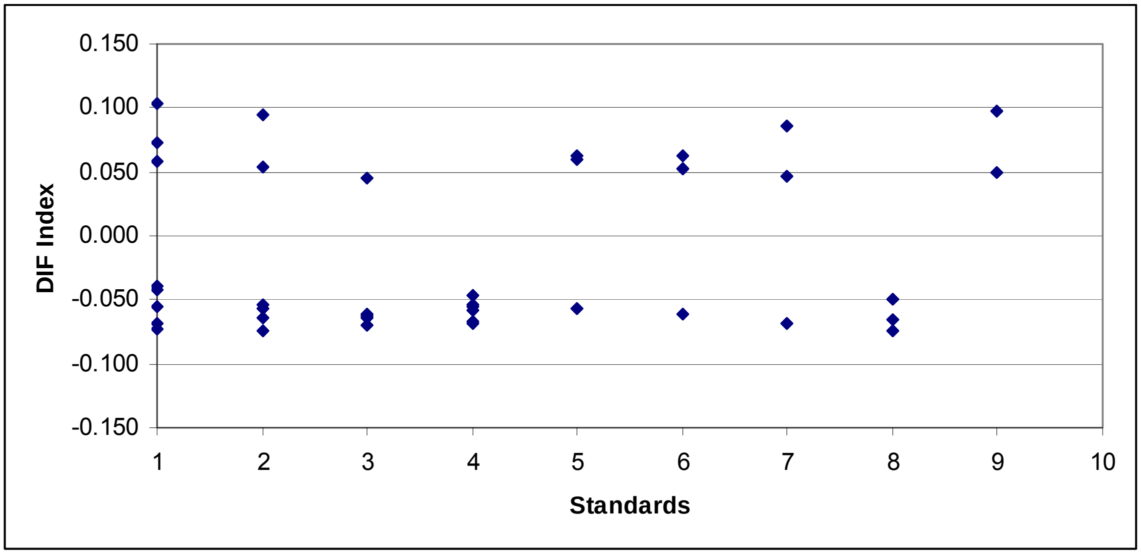


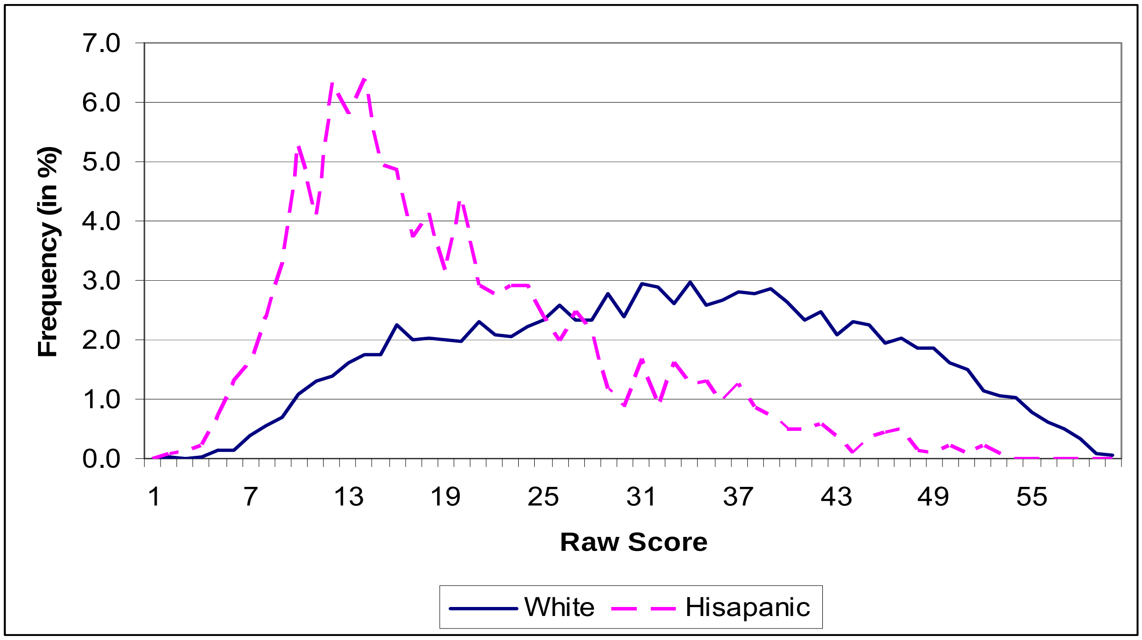
Figure 8.7 White/Black DIF Organized by (Content) Standard



Ethnicity DIF – White vs. Hispanic

Descriptive statistics for the White and Hispanic groups were presented in Table 4.1 under the section of Basic Statistics and Reliability Analysis. As seen in the table, White students performed around 13 raw score points higher than Hispanic students on the Chemistry Test. Plot of the total score distributions for the groups are given in Figure 8.8.

Figure 8.8 Total Test Score Distribution for White and Hispanic Students



A list of SDIF and UDIF indices across stage one and stage two is presented in Table 8.4. The complete set of DIF indices is presented in Figure 8.9. One multiple choice item (Item 3) was flagged as DIF at stage 2, and the conditional p value comparison is shown in Figure 8.10. The item favored White students over the range of the score scale where Hispanic students were located. In addition, ethnic DIF between White and Hispanic students is also organized by items measuring the content standards and shown in Figure 8.11. No pattern is evident.

Table 8.4 Summary of DIF Indices: White/Hispanic

	Stage 1[1]		Stage 2[1]	
Item	SDIF	UDIF	SDIF	UDIF
1	-.020	-.051	-.022	-.047
2	.043	.073	.041	.069
3	.076	.080	.092	.111
4	.038	.067	.036	.070
5	-.054	-.085	-.030	-.066
6	-.042	-.072	-.046	-.068
7	-.044	-.096	-.025	-.062
8	-.009	-.058	-.012	-.071
9	-.048	-.088	-.027	-.064
10	-.012	-.057	-.017	-.053
11	.089	.032	.079	.032
12	-.009	-.070	-.013	-.065
13	-.030	-.064	-.029	-.066
14	-.006	-.063	-.006	-.063
15	.013	.060	.009	.035
16	-.007	-.047	-.006	-.042
17	-.027	-.070	-.030	-.074
18	-.047	-.079	-.027	-.071
19	-.015	-.064	-.020	-.062
20	.045	.085	.067	.070
21	.000	-.069	-.008	-.073
22	.008	.060	.005	.074
23	-.047	-.073	-.047	-.069
24	-.007	-.074	-.009	-.067
25	.104	.029	.098	.030
26	.185	.047	.159	.042
27	-.015	-.061	-.021	-.057
28	-.027	-.068	-.029	-.069
29	-.033	-.062	-.036	-.054
30	.000	.057	.002	.057
31	-.038	-.075	-.041	-.077
32	.035	.022	.023	.019
33	-.024	-.075	-.029	-.065
34	-.017	-.057	-.026	-.053
35	-.007	-.056	-.010	-.052
36	-.009	-.058	-.012	-.050
37	.012	.057	.012	.059
38	-.047	-.082	-.024	-.063
39	.113	.031	.088	.031
40	-.034	-.074	-.035	-.069
41	-.028	-.073	-.030	-.060
42	-.050	-.082	-.033	-.059
43	.020	.073	.017	.058
44	-.009	-.058	-.015	-.043
45	-.019	-.061	-.021	-.048

[1] Items were flagged based on UDIF at the .075 level in stage 1 and .10 in stage 2.

Figure 8.9 White/Hispanic DIF Indices (MCQ: 1-10, 12-24, 27-31, 33-38, 40-45 and Constructed Response: 11, 25-26, 32, 39)

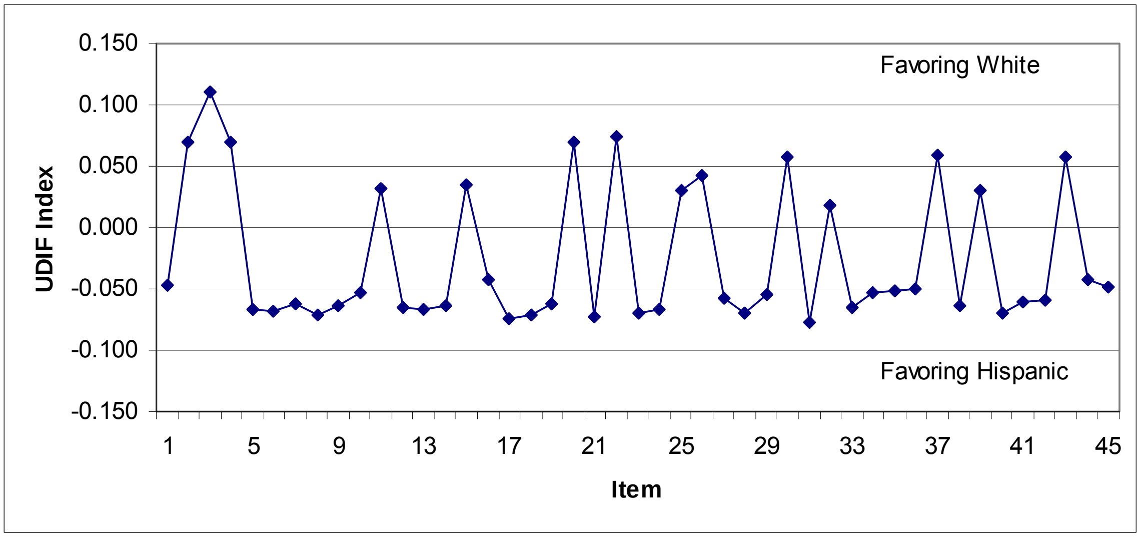


Figure 8.10 Conditional *p*-value Plot for Item 3

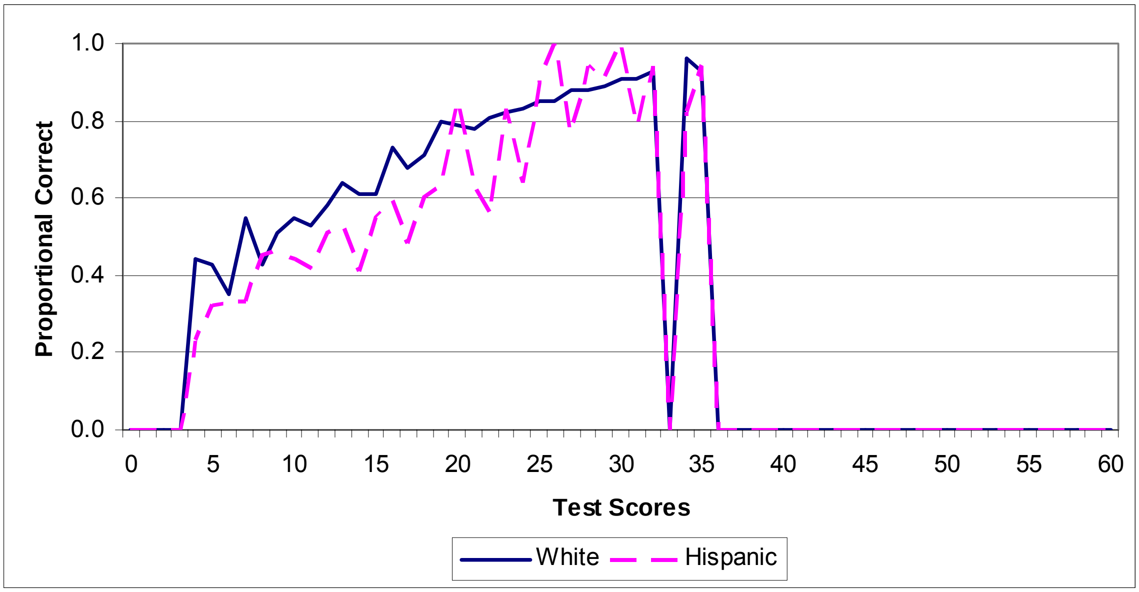
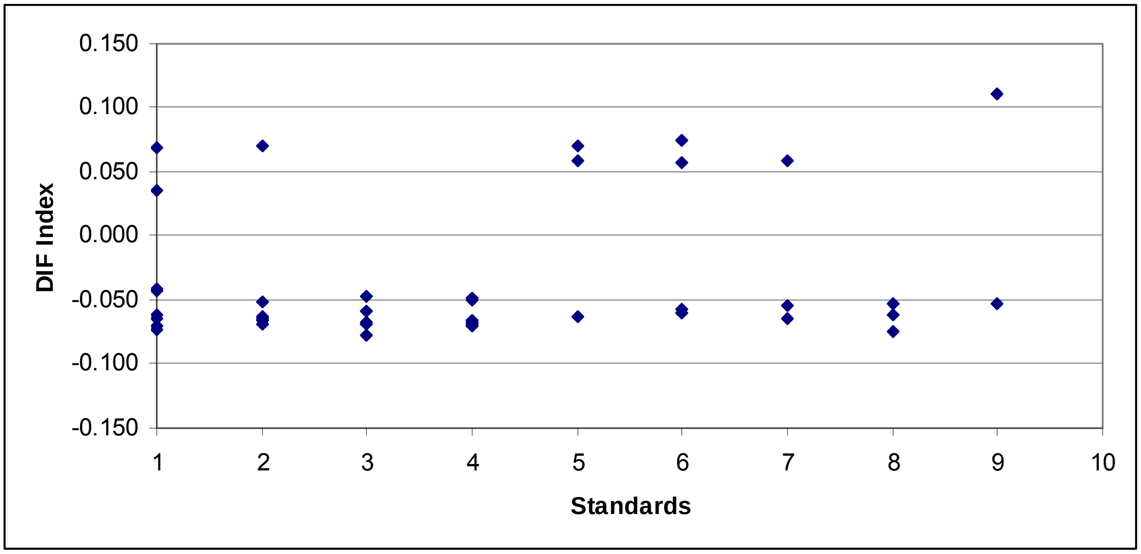


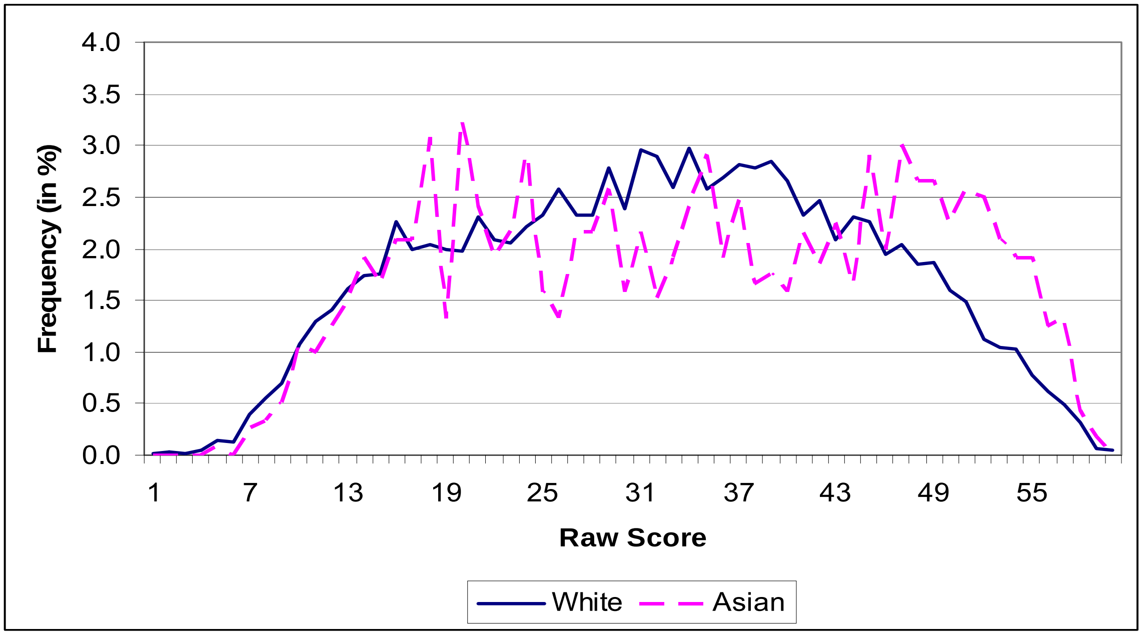
Figure 8.11 White/Hispanic DIF Organized by (Content) Standard



Ethnicity DIF – White vs. Asian

Descriptive statistics for the White and Asian groups were provided in Table 4.1 under the section, Basic Statistics and Reliability Analysis. As seen in the table, Asian students performed slightly better than the White students (around 2 test score points). The plot of the total score distributions for the groups is shown in Figure 8.12. The erratic distribution for the Asian students is due to the modest sample size.

Figure 8.12 Total Test Score Distribution for White and Asian Students

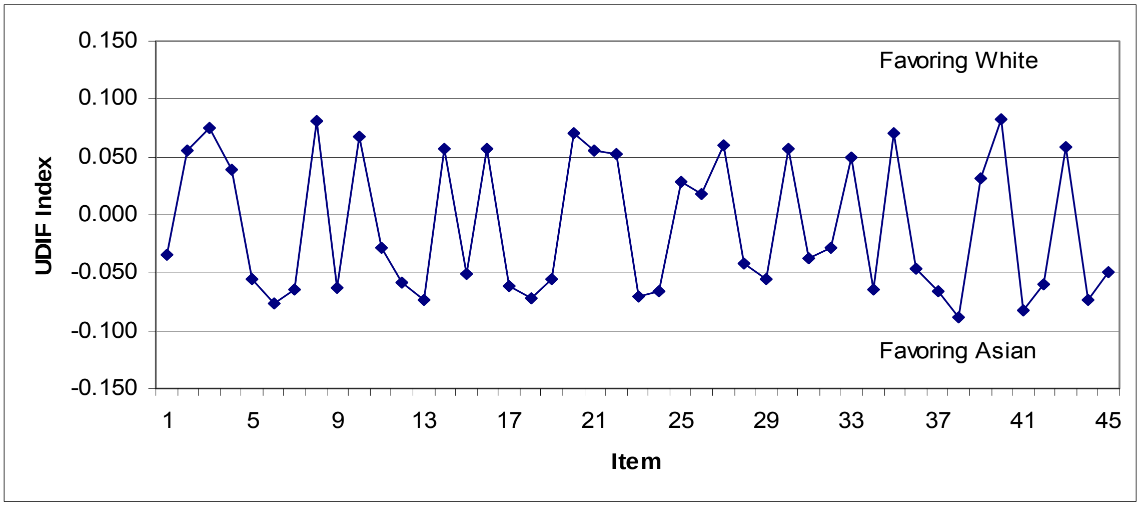


A list of SDIF and UDIF indices across stage one and stage two is presented in Table 8.5. The complete set of DIF indices at stage 2 is presented in Figure 8.13. None of the items on the test was flagged as DIF at stage 2. DIF between White and Asian students for items organized by the content standards is shown in Figure 8.14. As with all the other comparisons in our study, there were no patterns showing up in the data.

Table 8. 5 Summary of DIF Indices: White/Asian

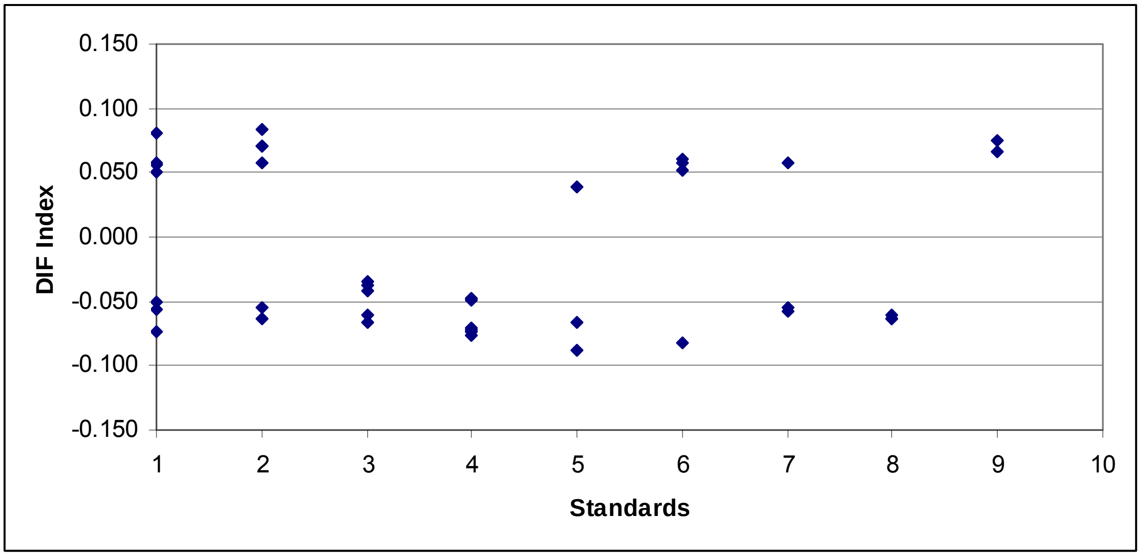
	Stage 1[1]		Stage 2[1]	
Item	SDIF	UDIF	SDIF	UDIF
1	-.015	-.031	-.017	-.035
2	.007	.062	.006	.056
3	.064	.083	.061	.075
4	.016	.052	.013	.039
5	-.017	-.063	-.016	-.055
6	-.046	-.069	-.046	-.076
7	-.026	-.072	-.029	-.064
8	.055	.080	.046	.081
9	-.010	-.081	-.014	-.063
10	.038	.072	.038	.067
11	-.026	-.038	-.025	-.028
12	-.029	-.091	-.034	-.058
13	-.033	-.078	-.034	-.074
14	.027	.066	.028	.057
15	-.034	-.061	-.030	-.051
16	.027	.093	.022	.057
17	-.009	-.096	-.012	-.061
18	-.014	-.085	-.023	-.072
19	-.001	-.077	-.003	-.056
20	.052	.073	.049	.070
21	.020	.072	.020	.056
22	.005	.082	.001	.052
23	-.021	-.073	-.020	-.070
24	-.027	-.069	-.030	-.066
25	.047	.037	.046	.029
26	.014	.028	.012	.018
27	.031	.063	.032	.060
28	-.012	-.061	-.013	-.042
29	.001	.076	-.004	-.055
30	.019	.074	.019	.057
31	-.022	-.051	-.021	-.037
32	-.054	-.033	-.048	-.029
33	.025	.052	.027	.050
34	-.036	-.071	-.035	-.064
35	.006	.067	.006	.070
36	-.021	-.047	-.022	-.047
37	-.011	-.085	-.017	-.066
38	-.054	-.091	-.055	-.088
39	.065	.040	.069	.031
40	.049	.095	.041	.083
41	-.038	-.086	-.047	-.082
42	-.009	-.067	-.013	-.060
43	.003	.068	.001	.058
44	-.002	-.077	-.007	-.073
45	-.001	-.075	.000	-.049

Figure 8.13 White/Asian DIF indices (MCQ: 1-10, 12-24, 27-31, 33-38, 40-45 and Constructed Response: 11, 25-26, 32, 39)



[1] Items were flagged based on UDIF at the .075 level in stage 1 and .10 in stage 2.

Figure 8.14 White/Asian DIF Organized by (Content) Standard



Summary

In summary, only two multiple choice items were flagged as functioning differently between the reference and the focal group. Item numbers for the DIF items, total number of flagged items, and the direction of DIF for each comparison are summarized in Table 8.6. No items were identified in the gender and White/Asian DIF analyses.

Table 8.6 Summary of DIF Items After Stage 2

	Favoring Majority		Favoring Minority	
	Item Number	Number of Items	Item Number	Number of Items
Male vs Female	--	0	--	0
White vs Black	2	1	--	0
White vs Hispanic	3	1	--	0
White vs Asian	--	0	--	0

9. Conclusions

As stated at the beginning of this report, the primary goal was to provide some useful psychometric analyses that might help in the evaluation and the on-going development of the MCAS 2006 Chemistry Test. The report began with a brief description of the Chemistry Test structure. Then, exclusion criteria were introduced to obtain a valid dataset for the psychometric analyses we carried out.

Classical approaches were used to analyze the 2006 MCAS Chemistry Test at the test-level, and also at the item-level. The test was definitely on the difficult side. Individual item difficulty and discrimination indices were compiled and a full distractor analysis was completed for the 45 items. The finding was that the test items are in excellent statistical shape. Analyses of the overall Chemistry Test results were also reported for gender groups and ethnic groups. These latter results were helpful to us in our study to detect potentially biased test items. We also carried out some analyses for items organized by the content strands. These may be of value to the DOE. The analyses revealed high overlap in student performance across the content strands, thus supporting the unidimensionality assumption of the Chemistry Test. Reliability results were reported too and these showed high values, high enough to support reporting the total test scores.

Modern test theoretic approaches (that is, item response theory based approaches) were also used to evaluate the psychometric quality of the Chemistry Test. Eigenvalue plot and structural equation modeling approaches were used to check the dimensionality of the item response data. These analyses are important because one of the important assumptions for IRT modeling is the assumption of test unidimensionality. Chemistry Test data showed a strong first factor and a minor second factor, and the fits of the IRT models were excellent with only a few misfitting items when a 3-PL/GRM was fitted to the data. Reviewing the test information function and the associated level of measurement error along the Chemistry proficiency continuum confirmed that the level of information being provided by the current test is high, and correspondingly measurement errors were acceptably low for most regions along the reporting score continuum.

DIF results also indicated that the Chemistry Test is of high psychometric quality. Only two items exhibiting DIF could be found. One multiple-choice item appeared in the White/Black comparison and another one appeared in the White/Hispanic comparison. Both of these items were slightly favoring the White group. Two items in total appearing in four different comparisons involving 45 items is a very small level of DIF, and could be due to chance factors only.

Our summary is that the Chemistry Test appears to be in excellent shape psychometrically with very few problems. Even the limited level of DIF we discovered may be little more than sampling error. Follow-up study of the two items would be desirable.

References

Hambleton, R. K., Zhao, Y., Smith, Z., Lam, W., & Deng, N. (2008). *Psychometric analyses of the 2006 MCAS high school science tests* (Center for Educational Assessment Research Report No. 649). Amherst, MA: University of Massachusetts, Center for Educational Assessment.

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

Zenisky, A. L., & Hambleton, R. K. (2007). *Differential item functioning analyses with STDIF: User's guide* (Unpublished report). Amherst, MA: University of Massachusetts, Center for Educational Assessment.

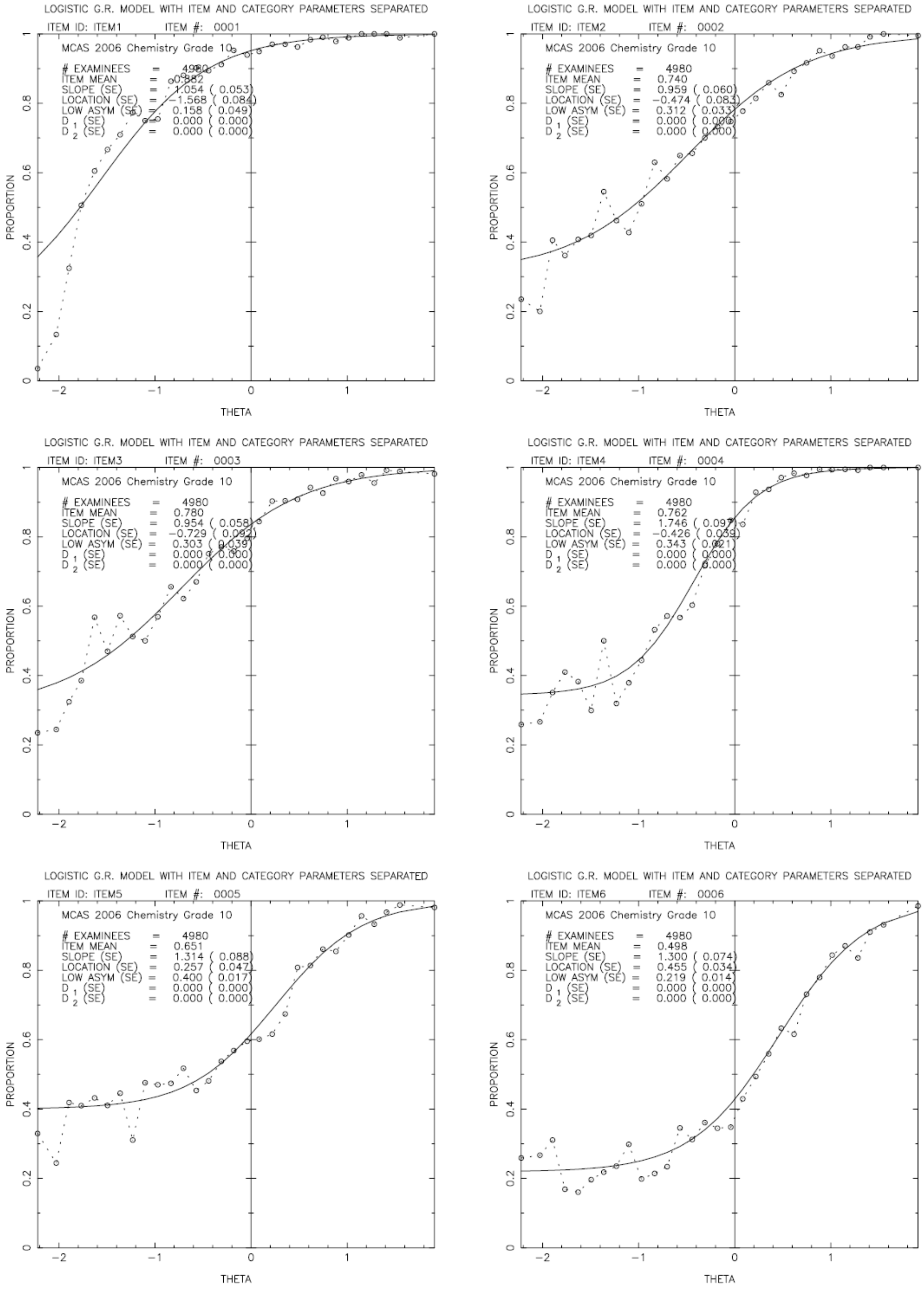
Zenisky, A., Hambleton, R. K., & Robin, F. (2003). Detection of differential item functioning in large-scale state assessments: A study evaluating a two-stage approach. *Educational and Psychological Measurement, 63*(1), 51-64.

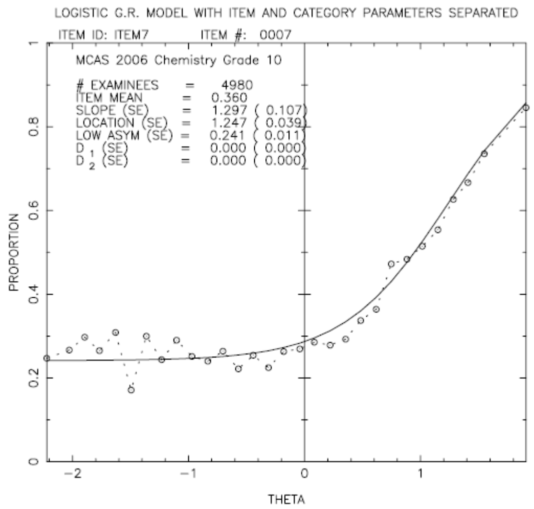
Zenisky, A., R. K. Hambleton, R. K., & Robin, F. (2004). DIF detection and interpretation in large-scale science assessments: Informing item writing practices. *Educational Assessment, 9*(1&2), 61-78.

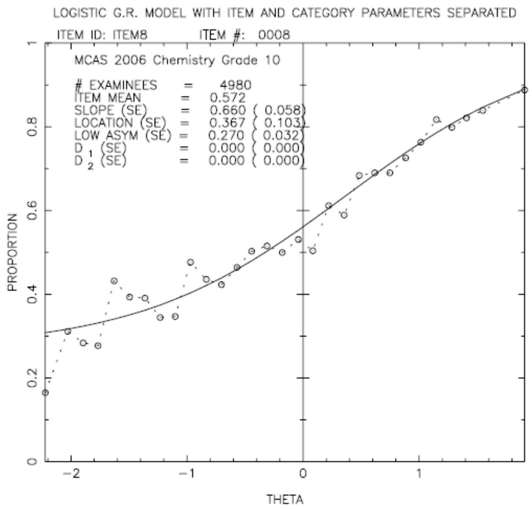
Appendix A

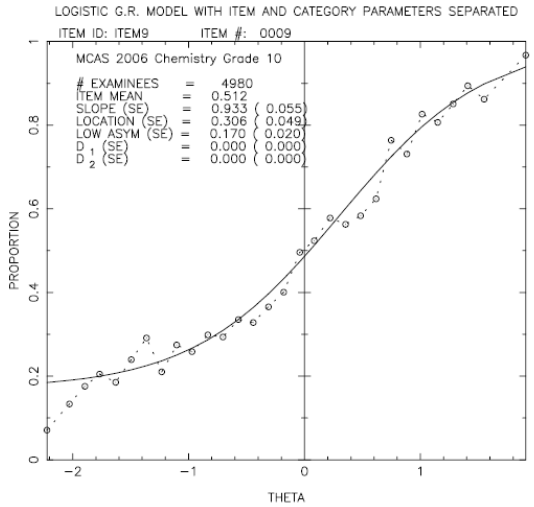
IRT Model Fit at the Item Level

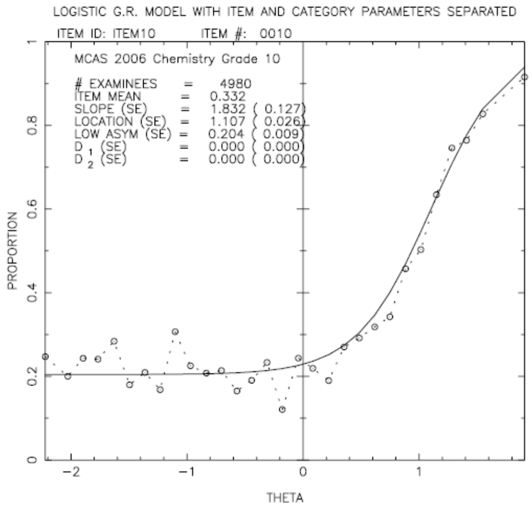
Figure A.1 Item Overall Model Fit Plot for the Grade 9/10 Chemistry Test

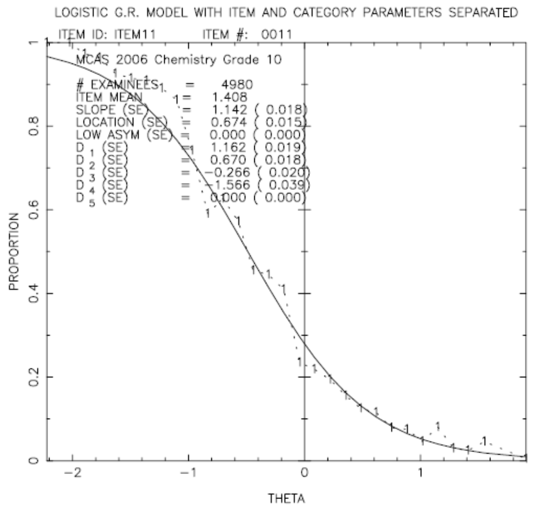


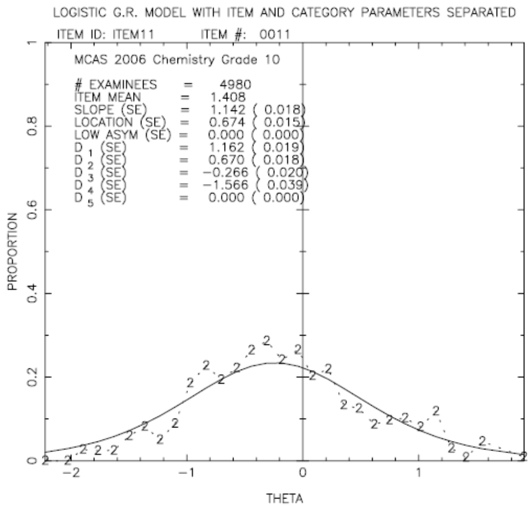

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM7 ITEM #: 0007
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.360
SLOPE (SE) = 1.297 (0.107)
LOCATION (SE) = 1.247 (0.039)
LOW ASYM (SE) = 0.241 (0.011)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

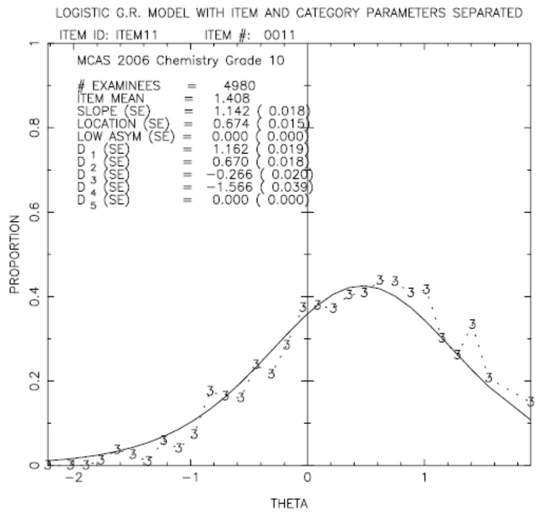

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM8 ITEM #: 0008
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.572
SLOPE (SE) = 0.660 (0.058)
LOCATION (SE) = 0.367 (0.103)
LOW ASYM (SE) = 0.270 (0.032)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

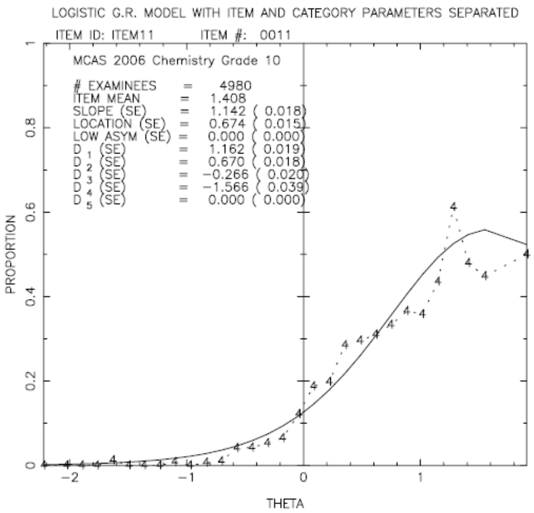

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM9 ITEM #: 0009
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.512
SLOPE (SE) = 0.933 (0.055)
LOCATION (SE) = 0.306 (0.049)
LOW ASYM (SE) = 0.170 (0.020)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

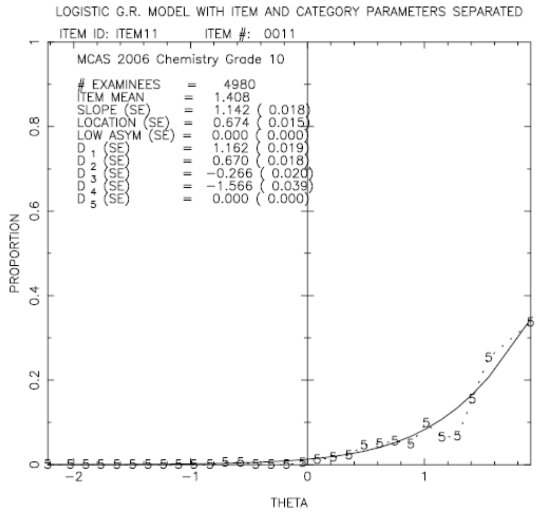

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM10 ITEM #: 0010
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.332
SLOPE (SE) = 1.832 (0.127)
LOCATION (SE) = 1.107 (0.026)
LOW ASYM (SE) = 0.204 (0.009)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

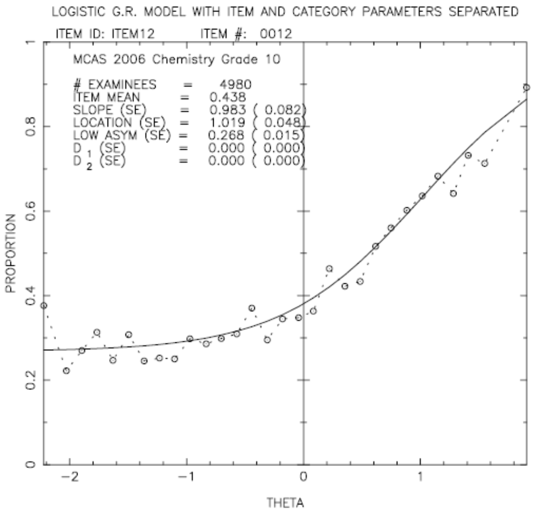

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM11 ITEM #: 0011
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.408
SLOPE (SE) = 1.142 (0.018)
LOCATION (SE) = 0.674 (0.015)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.162 (0.019)
D 2 (SE) = 0.670 (0.018)
D 3 (SE) = −0.266 (0.020)
D 4 (SE) = −1.566 (0.039)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

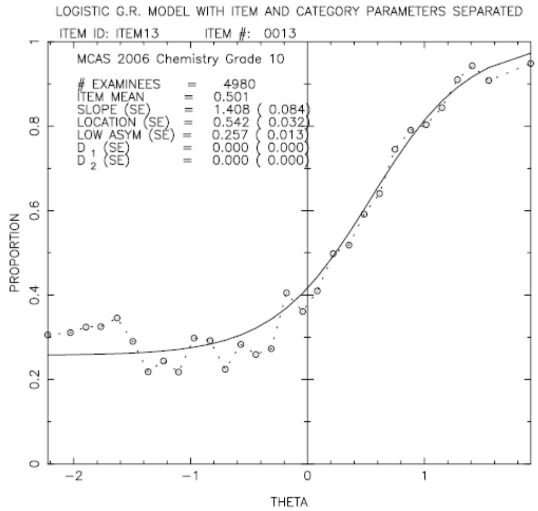

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM11 ITEM #: 0011
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.408
SLOPE (SE) = 1.142 (0.018)
LOCATION (SE) = 0.674 (0.015)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.162 (0.019)
D 2 (SE) = 0.670 (0.018)
D 3 (SE) = −0.266 (0.020)
D 4 (SE) = −1.566 (0.039)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

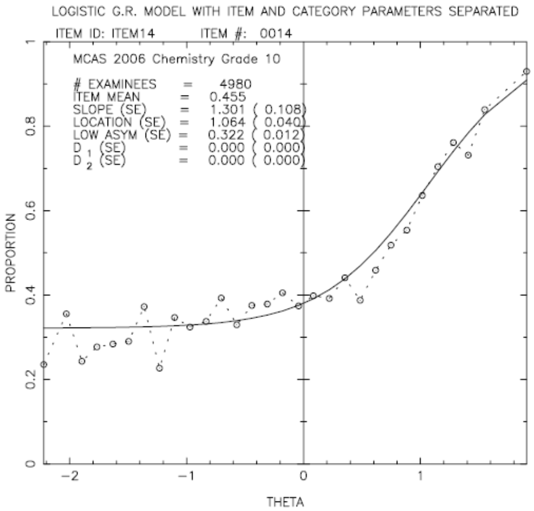

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM11 ITEM #: 0011
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.408
SLOPE (SE) = 1.142 (0.018)
LOCATION (SE) = 0.674 (0.015)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.162 (0.019)
D 2 (SE) = 0.670 (0.018)
D 3 (SE) = −0.266 (0.020)
D 4 (SE) = −1.566 (0.039)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

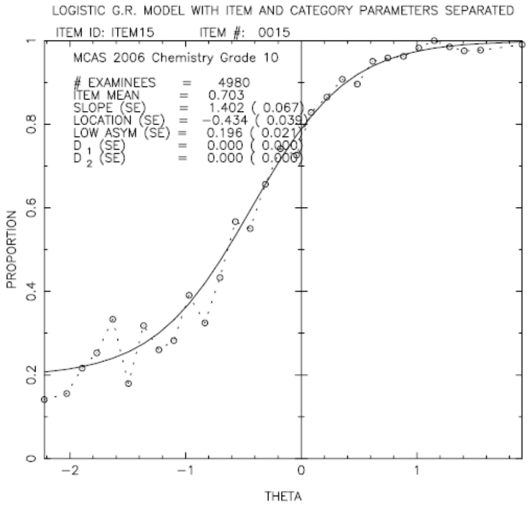

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM11 ITEM #: 0011
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.408
SLOPE (SE) = 1.142 (0.018)
LOCATION (SE) = 0.674 (0.015)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.162 (0.019)
D 2 (SE) = 0.670 (0.018)
D 3 (SE) = −0.266 (0.020)
D 4 (SE) = −1.566 (0.039)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

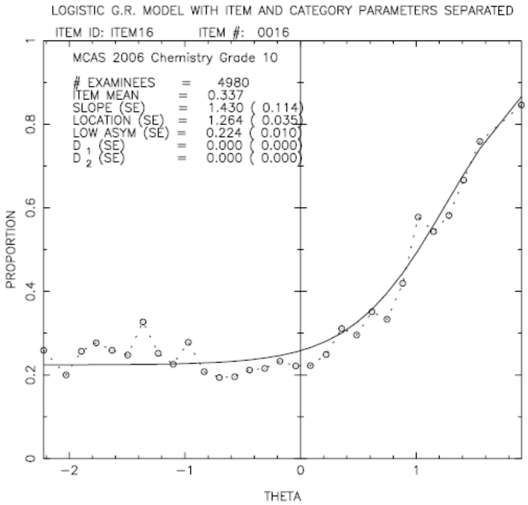

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM11 ITEM #: 0011
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.408
SLOPE (SE) = 1.142 (0.018)
LOCATION (SE) = 0.674 (0.015)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.162 (0.019)
D 2 (SE) = 0.670 (0.018)
D 3 (SE) = −0.266 (0.020)
D 4 (SE) = −1.566 (0.039)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

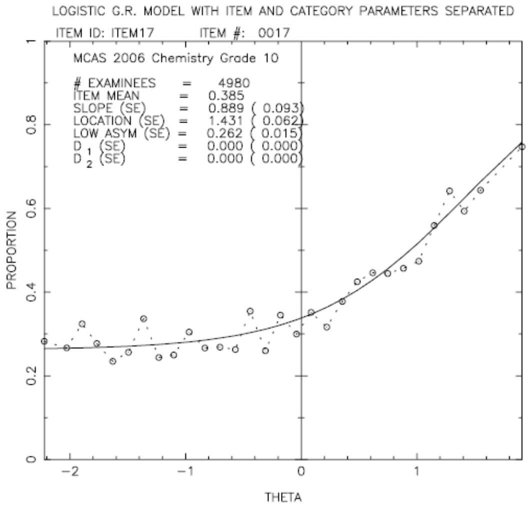

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM12 ITEM #: 0012
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.438
SLOPE (SE) = 0.983 (0.082)
LOCATION (SE) = 1.019 (0.048)
LOW ASYM (SE) = 0.268 (0.015)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

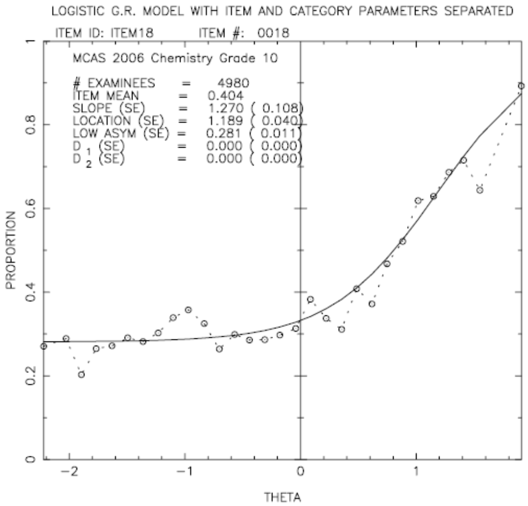

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM13 ITEM #: 0013
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.501
SLOPE (SE) = 1.408 (0.084)
LOCATION (SE) = 0.542 (0.032)
LOW ASYM (SE) = 0.257 (0.013)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

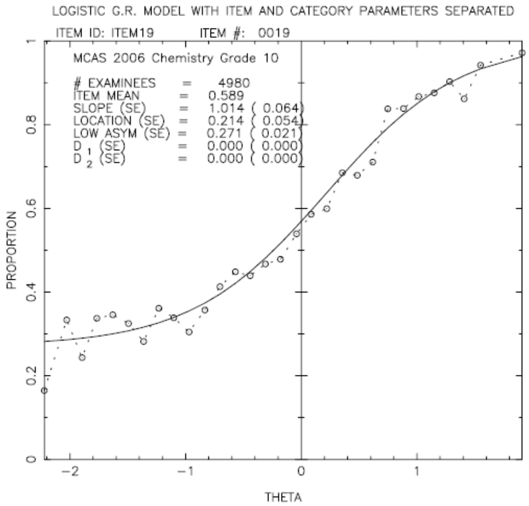

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM14 ITEM #: 0014
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.455
SLOPE (SE) = 1.301 (0.108)
LOCATION (SE) = 1.064 (0.040)
LOW ASYM (SE) = 0.322 (0.012)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

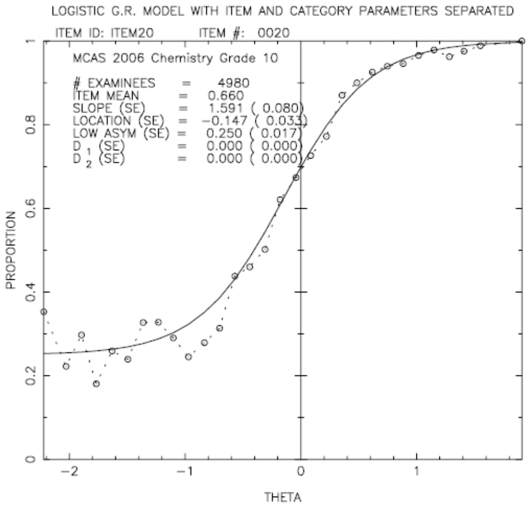

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM15 ITEM #: 0015
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.703
SLOPE (SE) = 1.402 (0.067)
LOCATION (SE) = −0.434 (0.039)
LOW ASYM (SE) = 0.196 (0.021)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

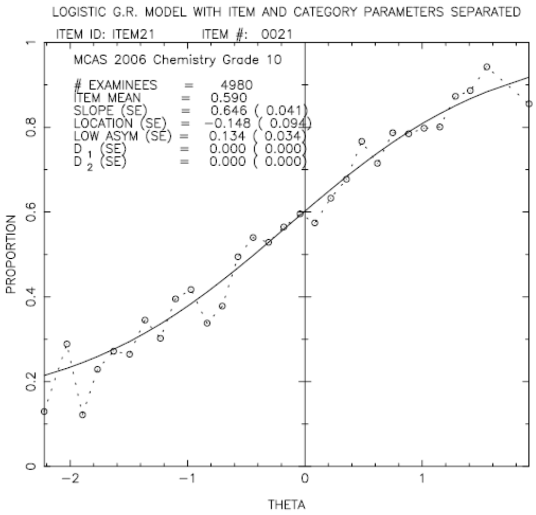

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM16 ITEM #: 0016
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.337
SLOPE (SE) = 1.430 (0.114)
LOCATION (SE) = 1.264 (0.035)
LOW ASYM (SE) = 0.224 (0.010)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

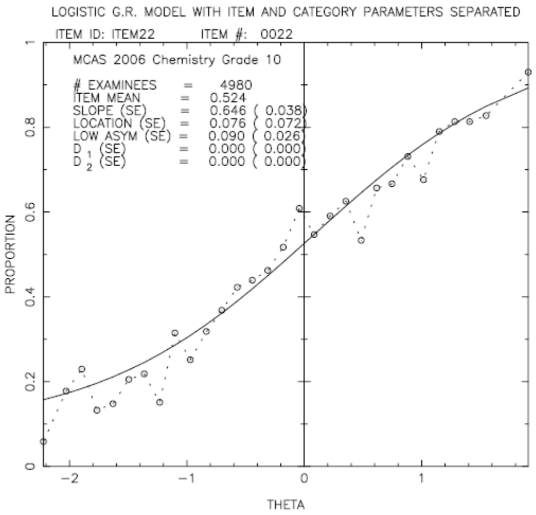

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM17 ITEM #: 0017
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.385
SLOPE (SE) = 0.889 (0.093)
LOCATION (SE) = 1.431 (0.062)
LOW ASYM (SE) = 0.262 (0.015)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

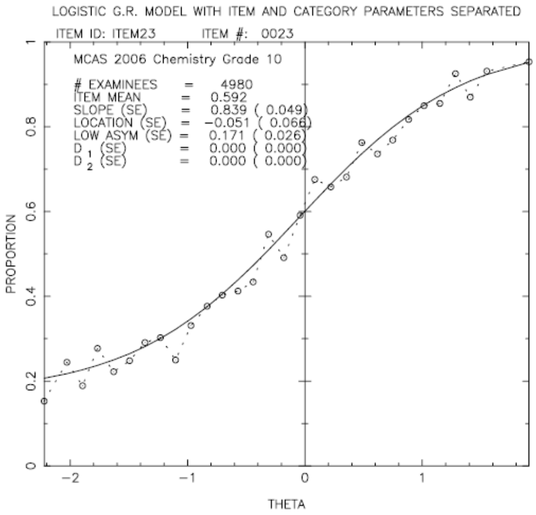

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM18 ITEM #: 0018
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.404
SLOPE (SE) = 1.270 (0.108)
LOCATION (SE) = 1.189 (0.040)
LOW ASYM (SE) = 0.281 (0.011)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

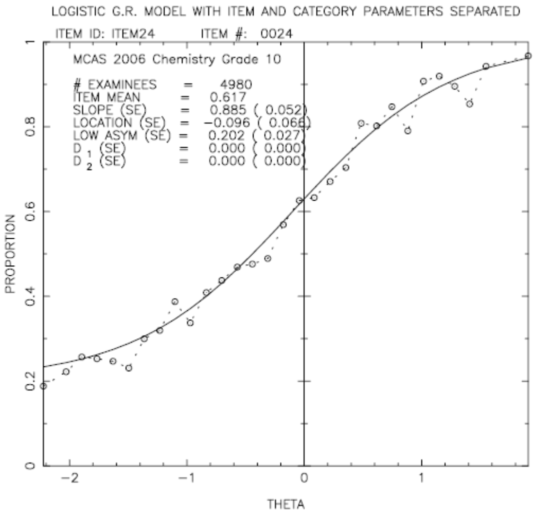

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM19 ITEM #: 0019
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.589
SLOPE (SE) = 1.014 (0.064)
LOCATION (SE) = 0.214 (0.054)
LOW ASYM (SE) = 0.271 (0.021)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

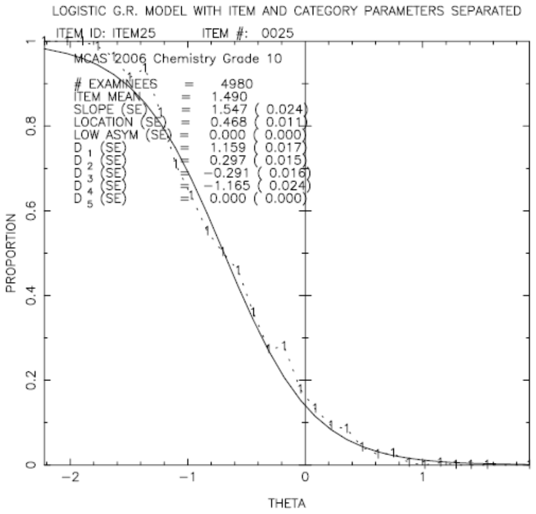

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM20 ITEM #: 0020
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.660
SLOPE (SE) = 1.591 (0.080)
LOCATION (SE) = −0.147 (0.033)
LOW ASYM (SE) = 0.250 (0.017)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

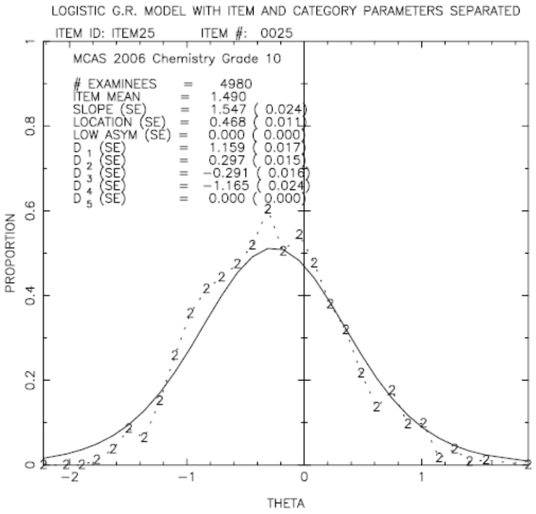

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM21 ITEM #: 0021
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.590
SLOPE (SE) = 0.646 (0.041)
LOCATION (SE) = −0.148 (0.094)
LOW ASYM (SE) = 0.134 (0.034)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

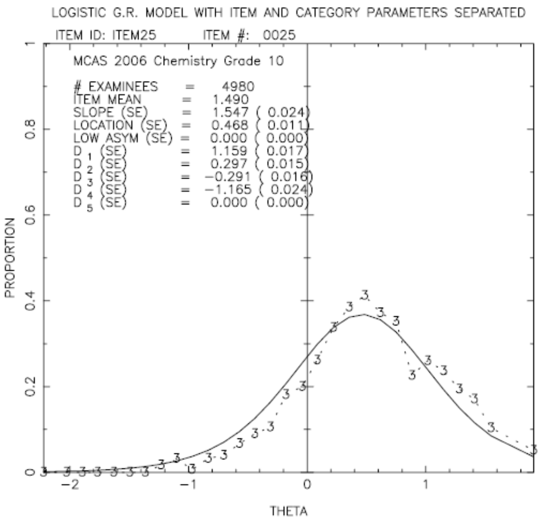

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM22 ITEM #: 0022
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.524
SLOPE (SE) = 0.646 (0.038)
LOCATION (SE) = 0.076 (0.072)
LOW ASYM (SE) = 0.090 (0.026)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

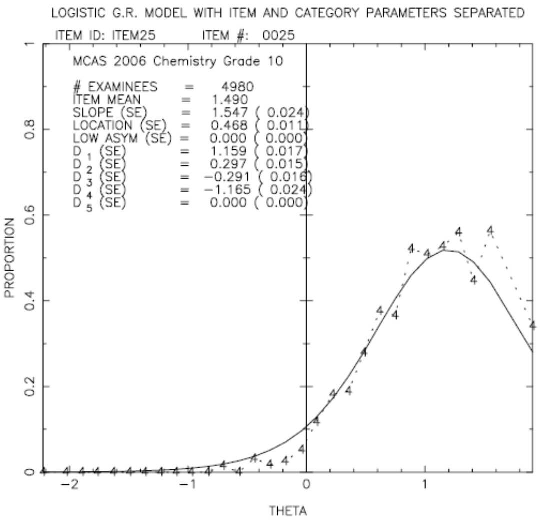

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM23 ITEM #: 0023
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.592
SLOPE (SE) = 0.839 (0.049)
LOCATION (SE) = −0.051 (0.066)
LOW ASYM (SE) = 0.171 (0.026)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

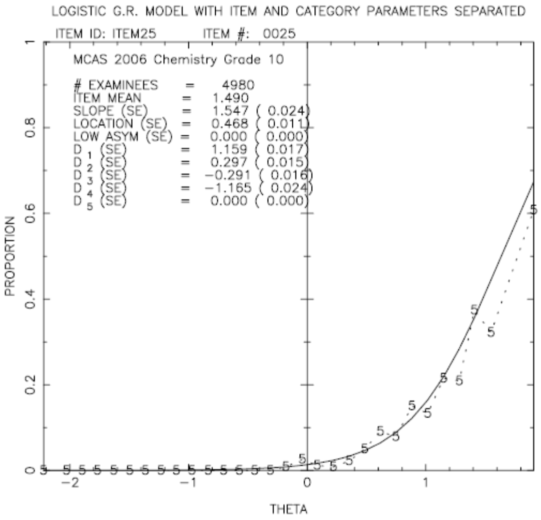

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM24 ITEM #: 0024
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.617
SLOPE (SE) = 0.885 (0.052)
LOCATION (SE) = −0.096 (0.066)
LOW ASYM (SE) = 0.202 (0.027)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

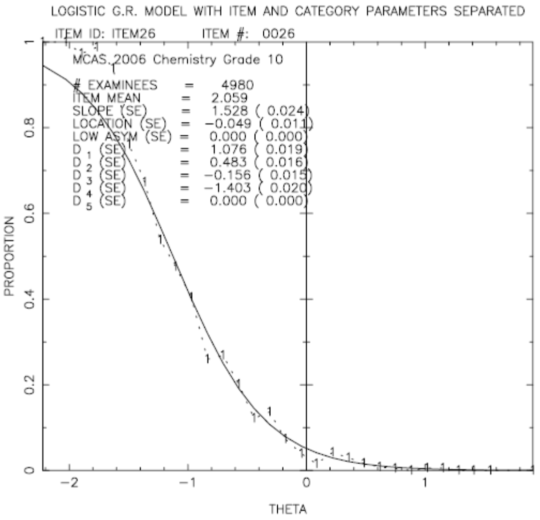

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM25 ITEM #: 0025
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.490
SLOPE (SE) = 1.547 (0.024)
LOCATION (SE) = 0.468 (0.011)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.159 (0.017)
D 2 (SE) = 0.297 (0.015)
D 3 (SE) = −0.291 (0.016)
D 4 (SE) = −1.165 (0.024)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

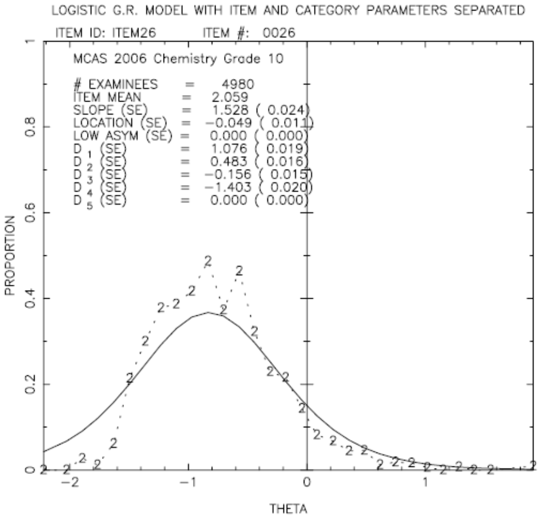

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM25 ITEM #: 0025
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.490
SLOPE (SE) = 1.547 (0.024)
LOCATION (SE) = 0.468 (0.011)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.159 (0.017)
D 2 (SE) = 0.297 (0.015)
D 3 (SE) = −0.291 (0.016)
D 4 (SE) = −1.165 (0.024)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

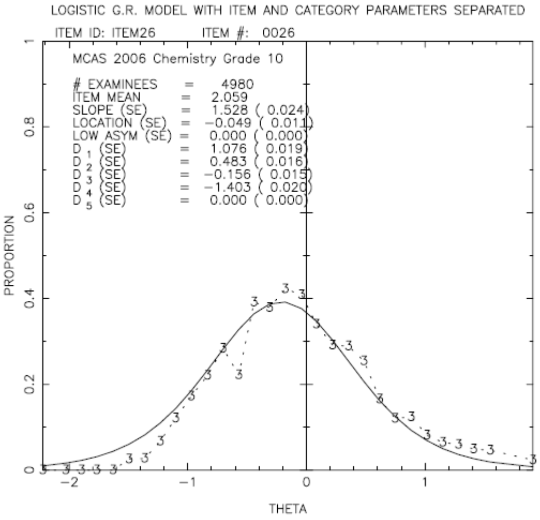

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM25 ITEM #: 0025
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.490
SLOPE (SE) = 1.547 (0.024)
LOCATION (SE) = 0.468 (0.011)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.159 (0.017)
D 2 (SE) = 0.297 (0.015)
D 3 (SE) = −0.291 (0.016)
D 4 (SE) = −1.165 (0.024)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

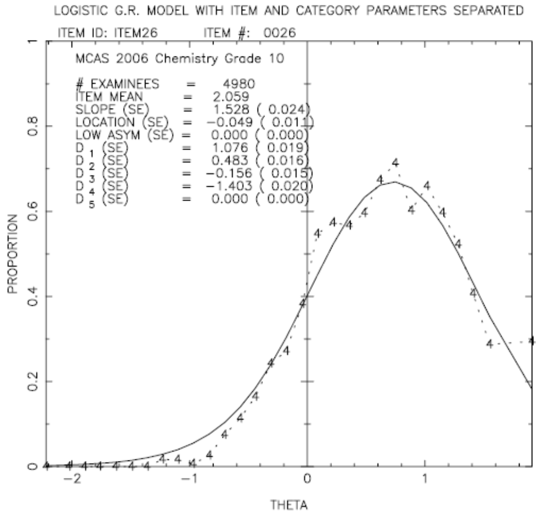

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM25 ITEM #: 0025
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.490
SLOPE (SE) = 1.547 (0.024)
LOCATION (SE) = 0.468 (0.011)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.159 (0.017)
D 2 (SE) = 0.297 (0.015)
D 3 (SE) = −0.291 (0.016)
D 4 (SE) = −1.165 (0.024)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

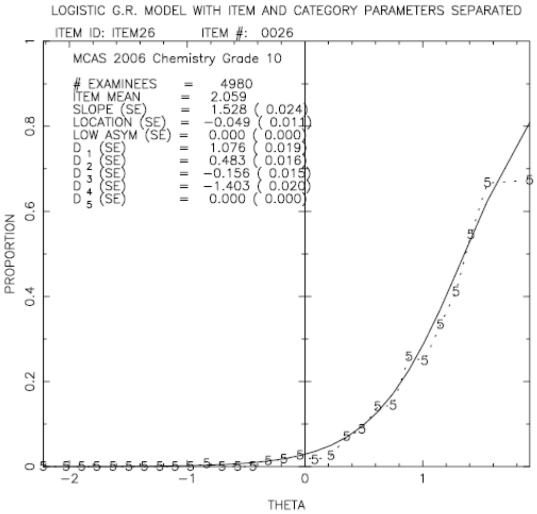

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM25 ITEM #: 0025
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.490
SLOPE (SE) = 1.547 (0.024)
LOCATION (SE) = 0.468 (0.011)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.159 (0.017)
D 2 (SE) = 0.297 (0.015)
D 3 (SE) = −0.291 (0.016)
D 4 (SE) = −1.165 (0.024)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

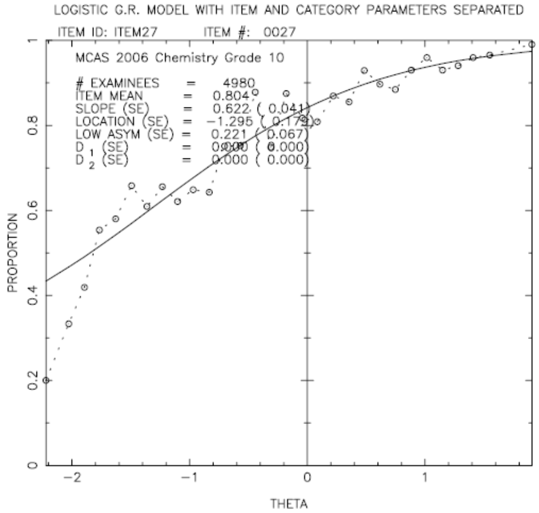

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM26 ITEM #: 0026
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 2.059
SLOPE (SE) = 1.528 (0.024)
LOCATION (SE) = −0.049 (0.011)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.076 (0.019)
D 2 (SE) = 0.483 (0.016)
D 3 (SE) = −0.156 (0.015)
D 4 (SE) = −1.403 (0.020)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

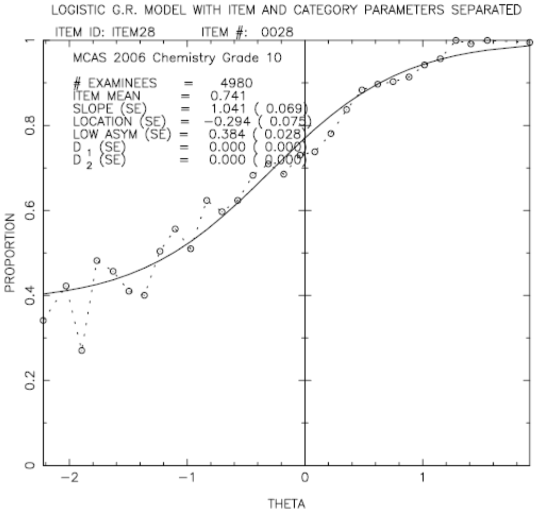

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM26 ITEM #: 0026
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 2.059
SLOPE (SE) = 1.528 (0.024)
LOCATION (SE) = −0.049 (0.011)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.076 (0.019)
D 2 (SE) = 0.483 (0.016)
D 3 (SE) = −0.156 (0.015)
D 4 (SE) = −1.403 (0.020)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

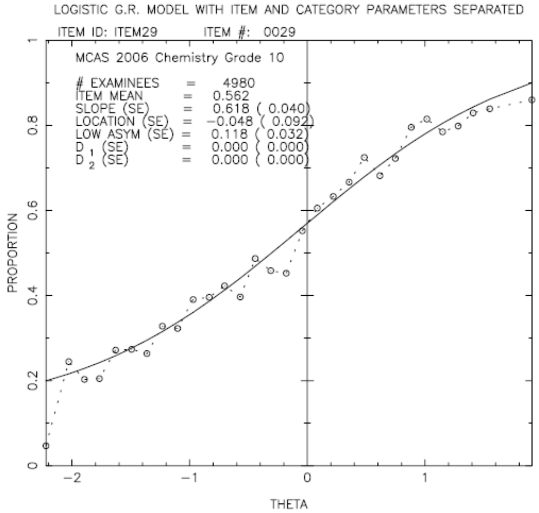

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM26 ITEM #: 0026
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 2.059
SLOPE (SE) = 1.528 (0.024)
LOCATION (SE) = −0.049 (0.011)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.076 (0.019)
D 2 (SE) = 0.483 (0.016)
D 3 (SE) = −0.156 (0.015)
D 4 (SE) = −1.403 (0.020)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

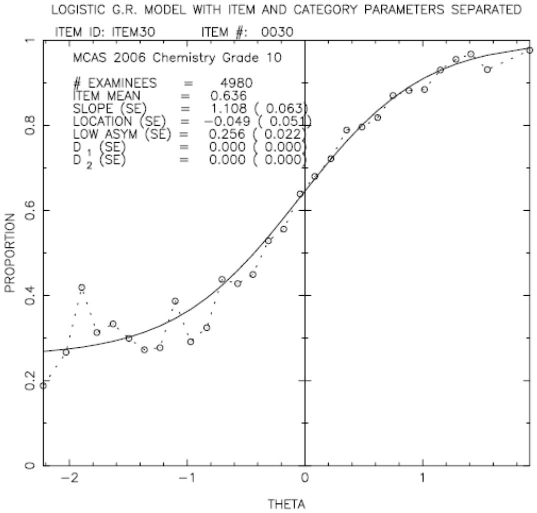

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM26 ITEM #: 0026
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 2.059
SLOPE (SE) = 1.528 (0.024)
LOCATION (SE) = −0.049 (0.011)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.076 (0.019)
D 2 (SE) = 0.483 (0.016)
D 3 (SE) = −0.156 (0.015)
D 4 (SE) = −1.403 (0.020)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

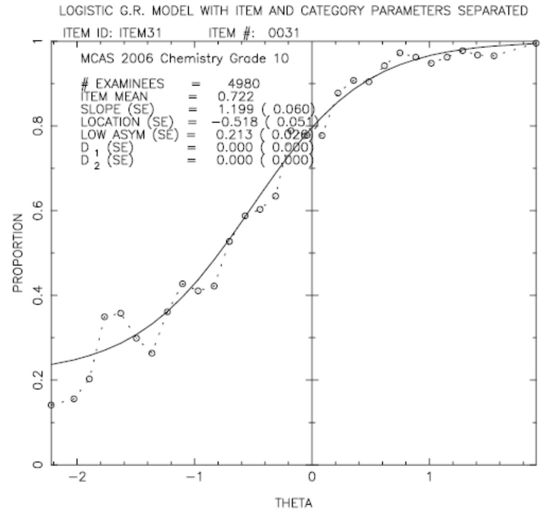

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM26 ITEM #: 0026
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 2.059
SLOPE (SE) = 1.528 (0.024)
LOCATION (SE) = −0.049 (0.011)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.076 (0.019)
D 2 (SE) = 0.483 (0.016)
D 3 (SE) = −0.156 (0.015)
D 4 (SE) = −1.403 (0.020)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

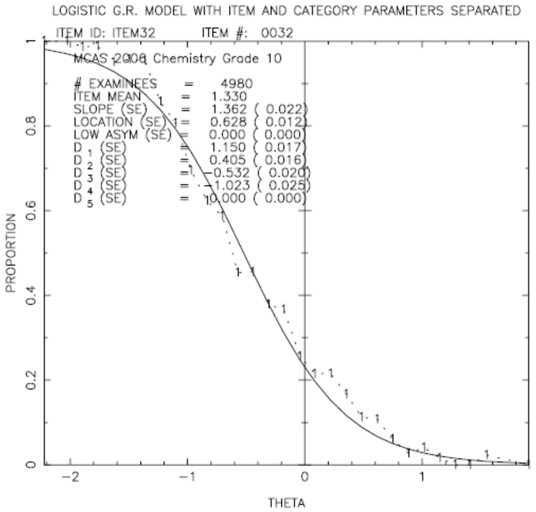

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM27 ITEM #: 0027
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.804
SLOPE (SE) = 0.622 (0.041)
LOCATION (SE) = −1.295 (0.179)
LOW ASYM (SE) = 0.221 (0.067)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

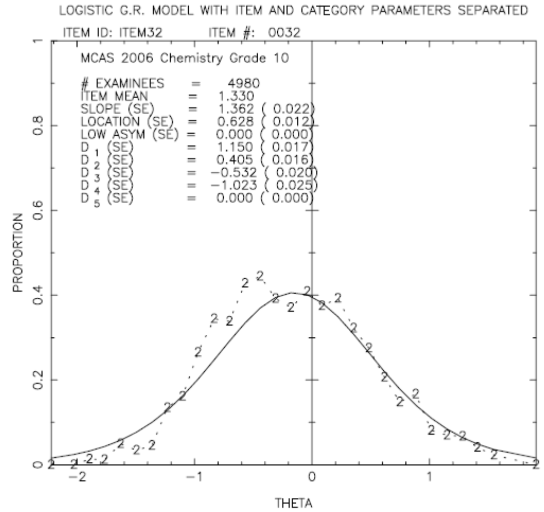

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM28 ITEM #: 0028
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.741
SLOPE (SE) = 1.041 (0.069)
LOCATION (SE) = −0.294 (0.075)
LOW ASYM (SE) = 0.384 (0.028)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

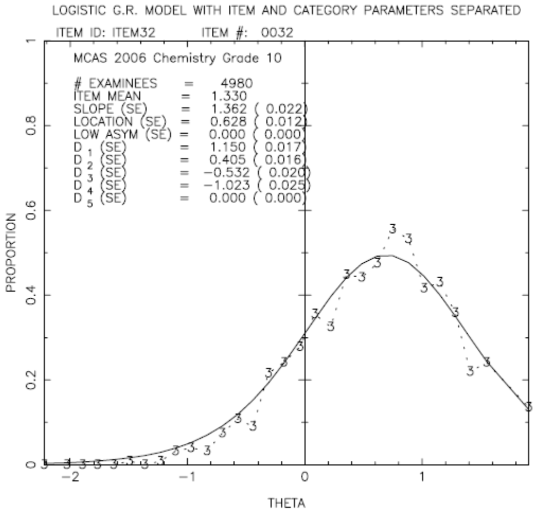

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM29 ITEM #: 0029
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.562
SLOPE (SE) = 0.618 (0.040)
LOCATION (SE) = −0.048 (0.092)
LOW ASYM (SE) = 0.118 (0.032)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

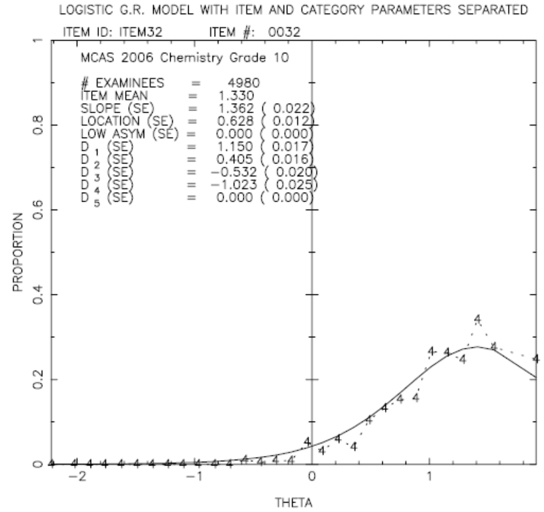

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM30 ITEM #: 0030
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.636
SLOPE (SE) = 1.108 (0.063)
LOCATION (SE) = −0.049 (0.051)
LOW ASYM (SE) = 0.256 (0.022)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA

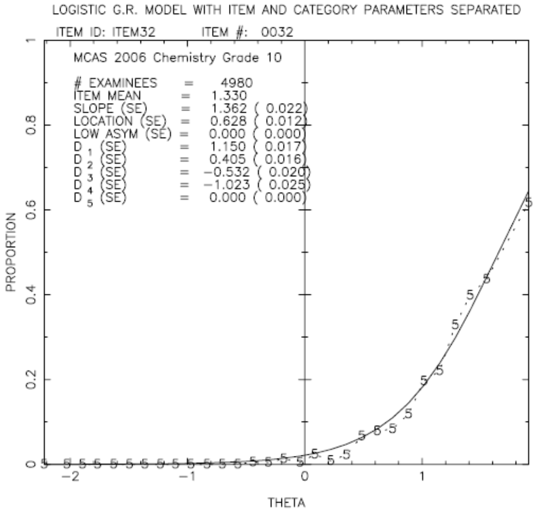

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM31 ITEM #: 0031
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.722
SLOPE (SE) = 1.199 (0.060)
LOCATION (SE) = −0.518 (0.051)
LOW ASYM (SE) = 0.213 (0.028)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

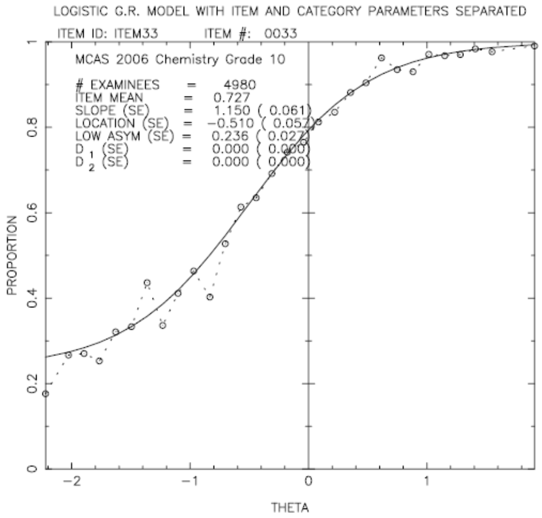

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM32 ITEM #: 0032
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.330
SLOPE (SE) = 1.362 (0.022)
LOCATION (SE) = 0.628 (0.012)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.150 (0.017)
D 2 (SE) = 0.405 (0.016)
D 3 (SE) = −0.532 (0.020)
D 4 (SE) = −1.023 (0.025)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

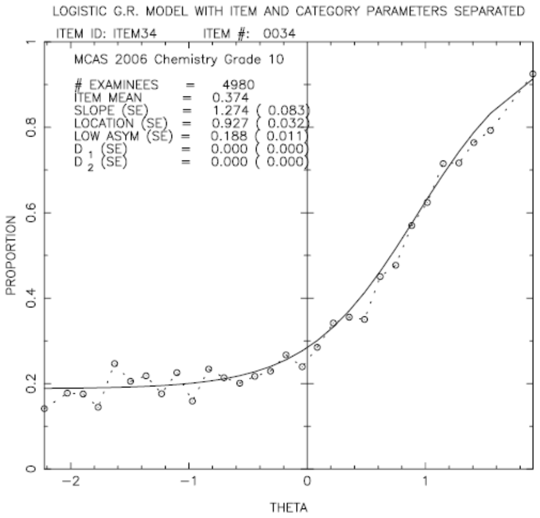

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM32 ITEM #: 0032
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.330
SLOPE (SE) = 1.362 (0.022)
LOCATION (SE) = 0.628 (0.012)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.150 (0.017)
D 2 (SE) = 0.405 (0.016)
D 3 (SE) = −0.532 (0.020)
D 4 (SE) = −1.023 (0.025)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1
2

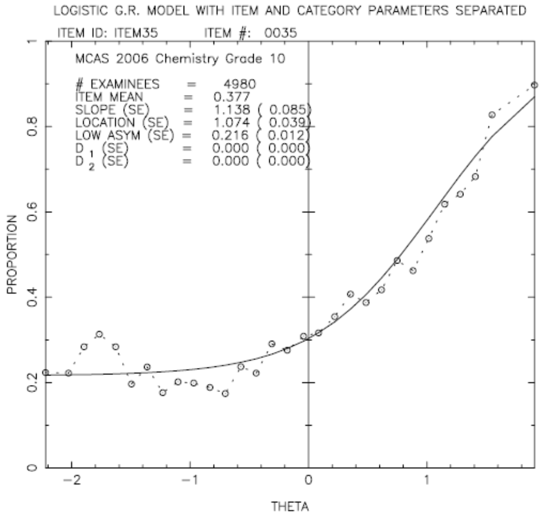

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM32 ITEM #: 0032
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.330
SLOPE (SE) = 1.362 (0.022)
LOCATION (SE) = 0.628 (0.012)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.150 (0.017)
D 2 (SE) = 0.405 (0.016)
D 3 (SE) = −0.532 (0.020)
D 4 (SE) = −1.023 (0.025)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

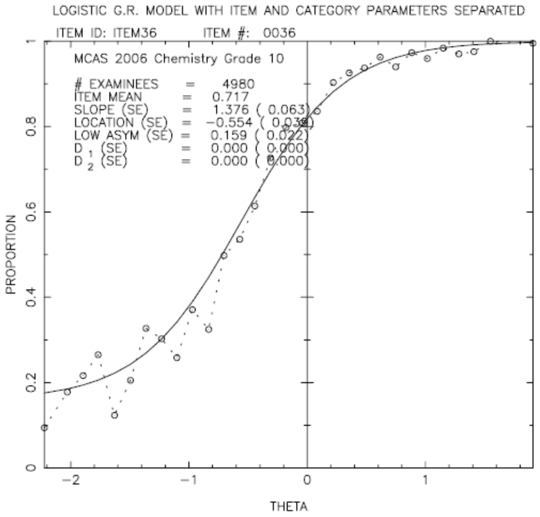

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM32 ITEM #: 0032
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.330
SLOPE (SE) = 1.362 (0.022)
LOCATION (SE) = 0.628 (0.012)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.150 (0.017)
D 2 (SE) = 0.405 (0.016)
D 3 (SE) = −0.532 (0.020)
D 4 (SE) = −1.023 (0.025)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

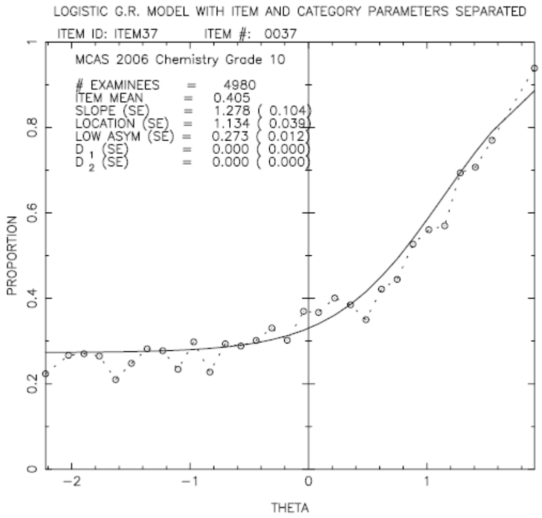

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM32 ITEM #: 0032
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.330
SLOPE (SE) = 1.362 (0.022)
LOCATION (SE) = 0.628 (0.012)
LOW ASYM (SE) = 0.000 (0.000)
D 1 (SE) = 1.150 (0.017)
D 2 (SE) = 0.405 (0.016)
D 3 (SE) = −0.532 (0.020)
D 4 (SE) = −1.023 (0.025)
D 5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

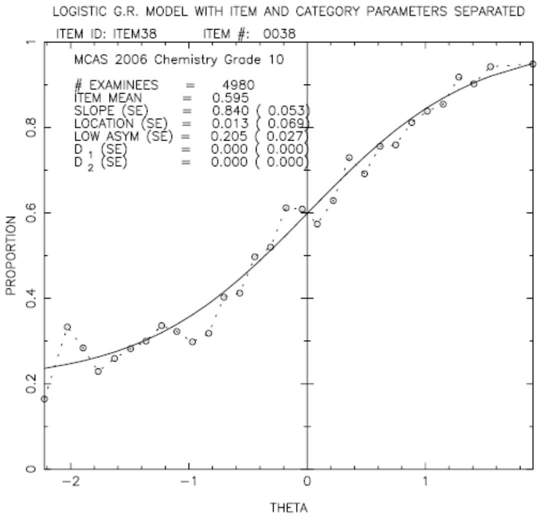

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM33 ITEM #: 0033
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.727
SLOPE (SE) = 1.150 (0.061)
LOCATION (SE) = −0.510 (0.052)
LOW ASYM (SE) = 0.236 (0.027)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

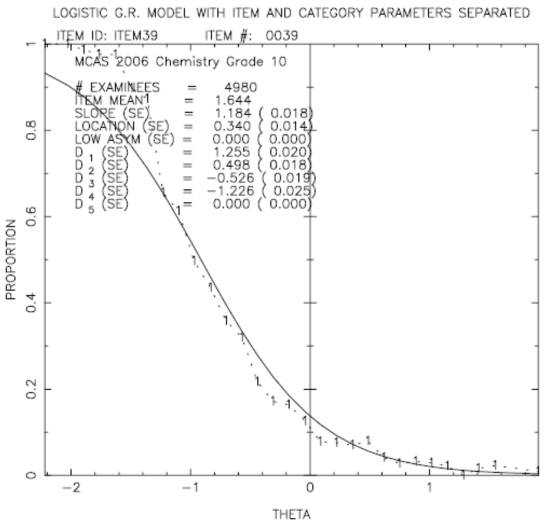

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM34 ITEM #: 0034
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.374
SLOPE (SE) = 1.274 (0.083)
LOCATION (SE) = 0.927 (0.032)
LOW ASYM (SE) = 0.188 (0.011)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

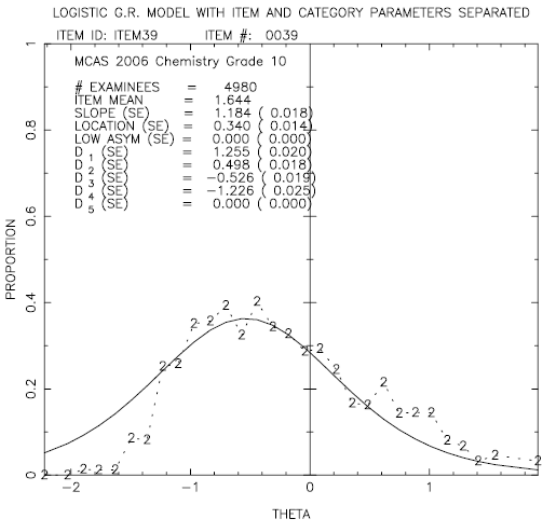

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM35 ITEM #: 0035
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.377
SLOPE (SE) = 1.138 (0.085)
LOCATION (SE) = 1.074 (0.039)
LOW ASYM (SE) = 0.216 (0.012)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

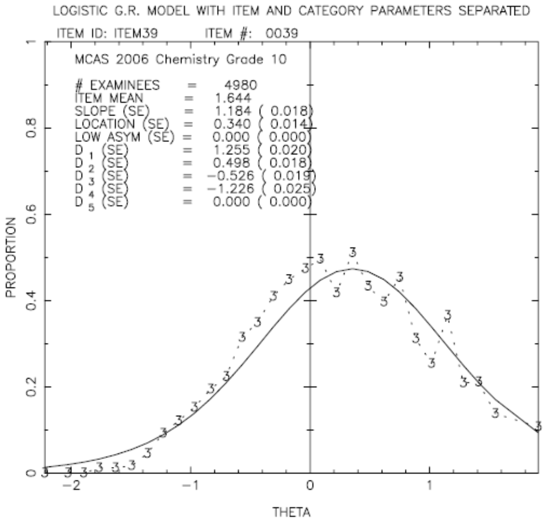

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM36 ITEM #: 0036
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.717
SLOPE (SE) = 1.376 (0.063)
LOCATION (SE) = −0.554 (0.039)
LOW ASYM (SE) = 0.159 (0.022)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

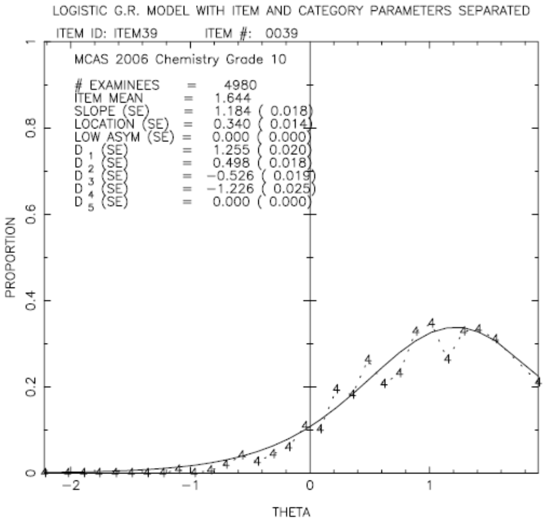

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM37 ITEM #: 0037
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.405
SLOPE (SE) = 1.278 (0.104)
LOCATION (SE) = 1.134 (0.039)
LOW ASYM (SE) = 0.273 (0.012)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

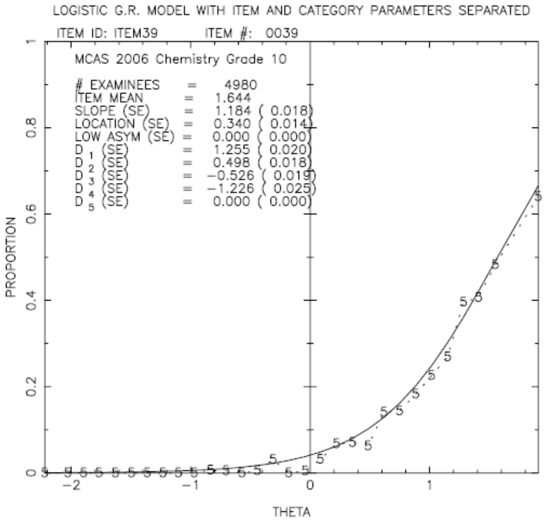

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM38 ITEM #: 0038
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.595
SLOPE (SE) = 0.840 (0.053)
LOCATION (SE) = 0.013 (0.069)
LOW ASYM (SE) = 0.205 (0.027)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

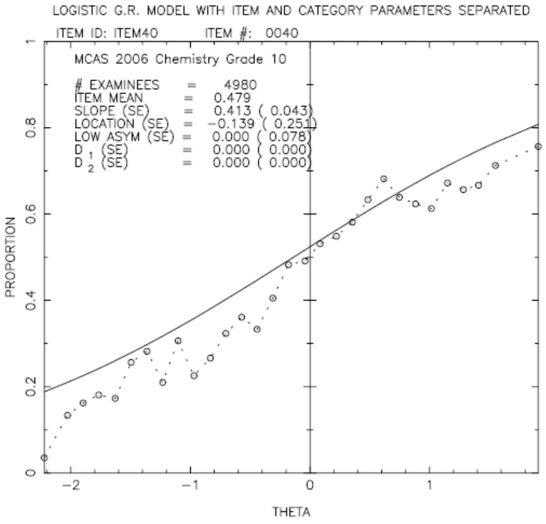

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM39 ITEM #: 0039
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.644
SLOPE (SE) = 1.184 (0.018)
LOCATION (SE) = 0.340 (0.014)
LOW ASYM (SE) = 0.000 (0.000)
D1 (SE) = 1.255 (0.020)
D2 (SE) = 0.498 (0.018)
D3 (SE) = −0.526 (0.019)
D4 (SE) = −1.226 (0.025)
D5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

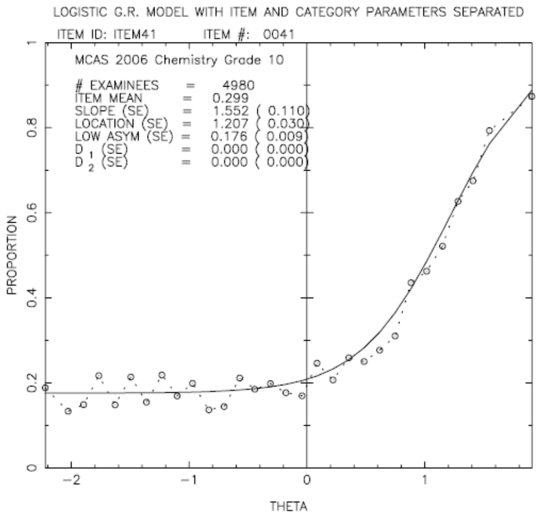

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM39 ITEM #: 0039
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.644
SLOPE (SE) = 1.184 (0.018)
LOCATION (SE) = 0.340 (0.014)
LOW ASYM (SE) = 0.000 (0.000)
D1 (SE) = 1.255 (0.020)
D2 (SE) = 0.498 (0.018)
D3 (SE) = −0.526 (0.019)
D4 (SE) = −1.226 (0.025)
D5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

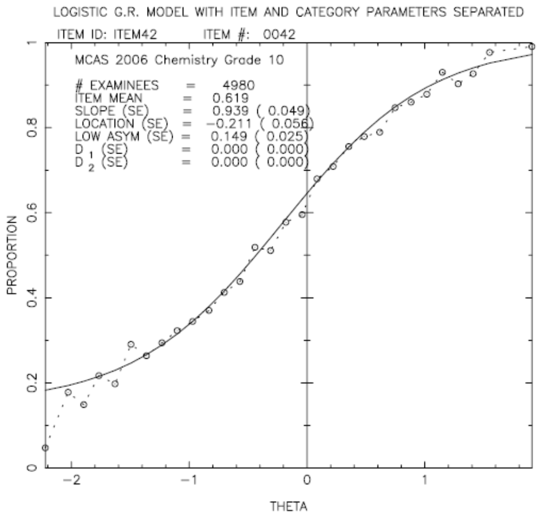

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM39 ITEM #: 0039
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.644
SLOPE (SE) = 1.184 (0.018)
LOCATION (SE) = 0.340 (0.014)
LOW ASYM (SE) = 0.000 (0.000)
D1 (SE) = 1.255 (0.020)
D2 (SE) = 0.498 (0.018)
D3 (SE) = −0.526 (0.019)
D4 (SE) = −1.226 (0.025)
D5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

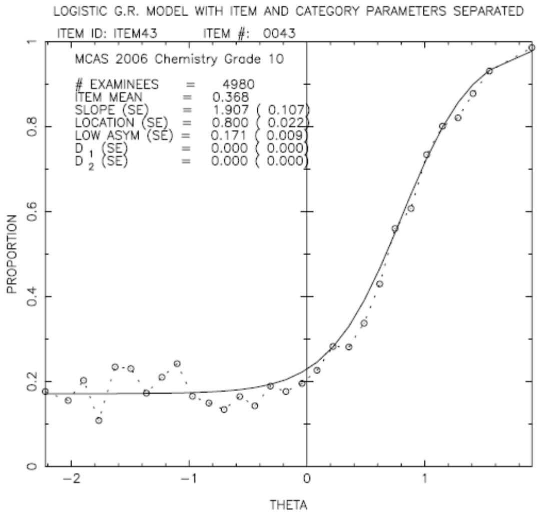

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM39 ITEM #: 0039
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.644
SLOPE (SE) = 1.184 (0.018)
LOCATION (SE) = 0.340 (0.014)
LOW ASYM (SE) = 0.000 (0.000)
D1 (SE) = 1.255 (0.020)
D2 (SE) = 0.498 (0.018)
D3 (SE) = −0.526 (0.019)
D4 (SE) = −1.226 (0.025)
D5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

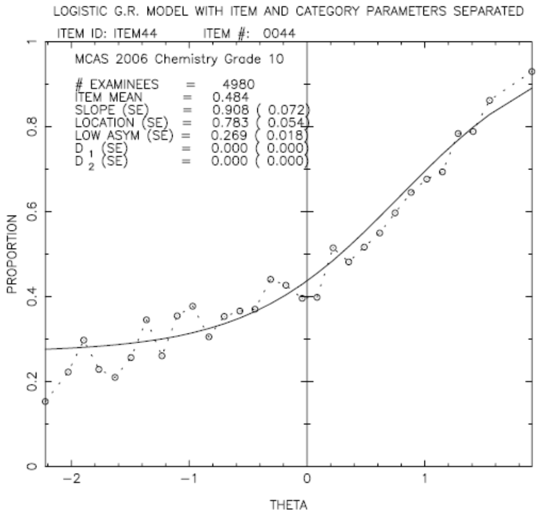

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM39 ITEM #: 0039
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 1.644
SLOPE (SE) = 1.184 (0.018)
LOCATION (SE) = 0.340 (0.014)
LOW ASYM (SE) = 0.000 (0.000)
D1 (SE) = 1.255 (0.020)
D2 (SE) = 0.498 (0.018)
D3 (SE) = −0.526 (0.019)
D4 (SE) = −1.226 (0.025)
D5 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1

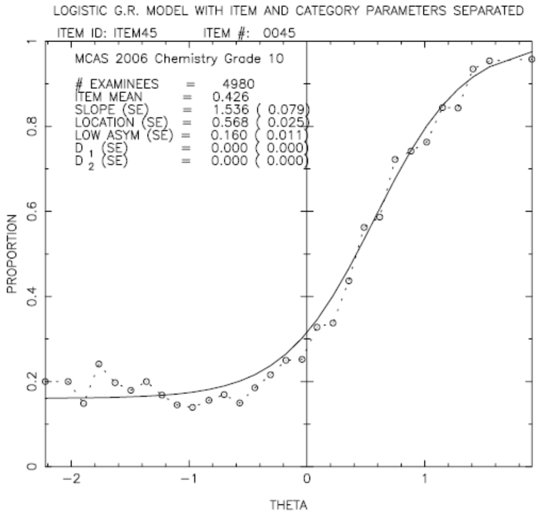

LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM40 ITEM #: 0040
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.479
SLOPE (SE) = 0.413 (0.043)
LOCATION (SE) = −0.139 (0.251)
LOW ASYM (SE) = 0.000 (0.078)
D1 (SE) = 0.000 (0.000)
D2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
THETA
−2
−1
0
1


LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM41 ITEM #: 0041
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.299
SLOPE (SE) = 1.552 (0.110)
LOCATION (SE) = 1.207 (0.030)
LOW ASYM (SE) = 0.176 (0.009)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA


LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM42 ITEM #: 0042
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.619
SLOPE (SE) = 0.939 (0.049)
LOCATION (SE) = −0.211 (0.056)
LOW ASYM (SE) = 0.149 (0.025)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA


LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM43 ITEM #: 0043
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.368
SLOPE (SE) = 1.907 (0.107)
LOCATION (SE) = 0.800 (0.022)
LOW ASYM (SE) = 0.171 (0.009)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA


LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM44 ITEM #: 0044
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.484
SLOPE (SE) = 0.908 (0.072)
LOCATION (SE) = 0.783 (0.054)
LOW ASYM (SE) = 0.269 (0.018)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA


LOGISTIC G.R. MODEL WITH ITEM AND CATEGORY PARAMETERS SEPARATED
ITEM ID: ITEM45 ITEM #: 0045
MCAS 2006 Chemistry Grade 10
# EXAMINEES = 4980
ITEM MEAN = 0.426
SLOPE (SE) = 1.536 (0.079)
LOCATION (SE) = 0.568 (0.025)
LOW ASYM (SE) = 0.160 (0.011)
D 1 (SE) = 0.000 (0.000)
D 2 (SE) = 0.000 (0.000)
PROPORTION
0
0.2
0.4
0.6
0.8
1
−2
−1
0
1
THETA