

Technical Report: Psychometric and DIF Analysis of
McDougal Littell Small Scale Study Math and
Algebra Assessment Items

Daniel Robertson, Ph.D.

EndVision Research
Logan, Utah

October 2005

This report documents psychometric analysis of assessment items used to formulate the dependent variables in the research studies reported in Callow-Heusser, Allred, Robertson & Borman (2005a) and Callow-Heusser, Allred, Robertson & Borman (2005b). The objective of these two studies was evaluation of the impact of McDougal Little Middle School Math and Algebra curricula on student achievement. For each study, dependent variables included student achievement in the relevant content area and student attitudes towards mathematics as an academic subject. Analyses were conducted using item response models, and were implemented in the BILOG-MG and MULTILOG software packages.

Psychometric analysis consisted of evaluation of individual items and of properties of the entire scales. Both Math and Algebra content test items were dichotomously scored, and hence two-parameter logistic models were used to evaluate items for both assessments. The attitudinal measure, identical for both studies, was scored on a five-point Likert-type scale, and items from this measure were evaluated using the Graded-response model.

1. Content Knowledge Assessments

1.1. Middle School Math

1.1a: Description of Instrument

The math content knowledge instrument was assembled for this study using publicly released items from past administrations of the National Assessment of Educational Progress (NAEP) by the U.S. Department of Education. Items were selected from the NAEP 8th Grade Mathematics item pool. Items classified at difficulty levels of high, medium, and low were selected in approximately equal proportions. The test contained 42 items, and so had a possible maximum score of 42.

BILOG-MG was used to run a 2-Parameter Logistic model on the item responses. Convergence was not reached using a 3PL model.

1.1b: Item Analysis

Table 1 displays descriptive statistics for each of the 42 items obtained from the first phase of the BILOG run. These data are used by BILOG in subsequent phases and are only displayed as initial estimates of item characteristics. ~~Data in the "Logit/1.7" column may be interpreted as initial estimates of the item discrimination or slope.~~ BILOG also produces estimates of both the Pearson and biserial item-test correlations (the biserial correlation displays less bias when the percent-correct is extreme).

Based on statistics in the percentage correct column, the most difficult items were 32, 33, 37, and 42.

Table 1. Item Statistics for Math Assessment

Item	No. Correct	Pct. Correct	Item-Scale Correlations	
			Pearson	Biserial
1	236	52.9	0.322	0.404
2	254	57.0	0.189	0.238
3	360	80.7	0.237	0.341
4	408	91.5	0.199	0.356
5	312	70.0	0.288	0.380
6	307	68.8	0.243	0.319
7	199	44.6	0.194	0.244
8	191	42.8	0.151	0.190
9	306	68.6	0.298	0.390
10	289	64.8	0.275	0.353
11	256	57.4	0.408	0.514
12	413	92.6	0.194	0.363
13	436	97.8	0.169	0.463
14	246	55.2	0.311	0.391
15	127	28.5	0.200	0.266
16	194	43.5	0.256	0.322
17	154	34.5	0.263	0.339
18	254	57.0	0.403	0.507
19	381	85.4	0.399	0.616
20	169	37.9	0.124	0.158
21	71	15.9	0.025	0.038
22	341	76.5	0.346	0.478
23	348	78.0	0.303	0.424
24	294	65.9	0.351	0.454
25	135	30.3	0.220	0.289
26	87	19.5	0.097	0.139
27	376	84.3	0.148	0.224
28	271	60.8	0.170	0.216
29	226	50.7	0.384	0.481
30	249	55.8	0.446	0.562
31	214	48.0	0.416	0.522
32	29	6.5	0.286	0.555
33	2	0.4	-0.037	-0.175
34	239	53.6	0.128	0.160
35	260	58.3	0.274	0.347
36	281	63.0	0.405	0.518
37	16	3.6	0.001	0.003
38	187	41.9	0.393	0.496
39	115	25.8	0.276	0.374
40	162	36.3	0.425	0.545
41	63	14.1	0.286	0.446
42	16	3.6	0.090	0.210

Table 2 displays IRT item parameters, and Figure 1 displays item characteristic curves. The Location parameter in Table 2 represents the difficulty of the item, with higher location values representing more difficult items. The location parameter may be interpreted as the latent trait value associated with . Item 13, for example, was extremely easy, while Items 33 and 37 are extremely difficult. The average item location was 0.222 (SD = 2.274). The Slope parameter represents the discriminating power of the item, or how well the item differentiates between examinees of higher and lower ability on the basis of examinee responses to that item. Items with higher slope values discriminate better between high and low ability examinees. For example, Items 19, 32, and 40 were effective at discriminating between higher and lower ability examinees, while Items 20, 21, and 37 were not. The average item slope was 0.509 (SD = 0.183). The Loading parameter is a single-factor item factor loading, and may be interpreted analogously to an item-scale correlation – it is the correlation between item response and the latent ability. The fit of the 2PL IRT model to item responses may be assessed via the Chi Square values, for which non-significant p-values indicate better-fitting items. Four items had p-values smaller than 0.10.

In general, most items appear to operating efficiently. Notable exceptions are items 21, 33, 37, and 42, which may be difficult beyond the reasonable expected ability of students at this age/ability level.

Table 2: Item Parameters for Math Assessment

Item	Location	Slope	Loading	ChiSq	Sig
1	-.154	.503	.449	5.8	.669
2	-.556	.315	.301	5.1	.742
3	-2.211	.421	.388	6.8	.660
4	-3.084	.512	.456	6.0	.427
5	-1.165	.488	.439	12.3	.138
6	-1.277	.402	.373	5.4	.802
7	.414	.333	.316	2.9	.941
8	.656	.273	.264	6.7	.573
9	-1.053	.503	.449	14.1	.079
10	-.890	.455	.414	5.6	.593
11	-.316	.698	.572	4.0	.785
12	-3.092	.551	.483	10.7	.097
13	-3.620	.742	.596	0.4	.939
14	-.277	.502	.449	12.4	.134
15	1.575	.376	.352	4.9	.669
16	.426	.403	.374	2.3	.944
17	.970	.437	.400	5.1	.651
18	-.293	.707	.577	5.0	.665
19	-1.569	.930	.681	3.4	.764
20	1.206	.252	.244	14.3	.074
21	4.602	.219	.214	10.5	.162
22	-1.370	.611	.522	5.0	.754
23	-1.661	.519	.460	5.7	.768
24	-.823	.557	.487	3.8	.799
25	1.343	.404	.375	4.6	.594
26	3.145	.278	.268	8.7	.278
27	-3.017	.351	.331	3.5	.943
28	-.899	.304	.291	3.1	.960
29	-.025	.647	.543	8.4	.301

30	-.234	.757	.604	5.5	.593
31	.093	.701	.574	5.8	.560
32	2.253	.954	.690	2.7	.617
33	6.623	.513	.457	0.0	.000
34	-.363	.241	.234	6.8	.556
35	-.504	.436	.400	7.5	.481
36	-.575	.685	.565	8.6	.282
37	6.341	.318	.303	1.8	.767
38	.376	.645	.542	4.1	.774
39	1.434	.500	.447	5.1	.533
40	.559	.810	.629	6.0	.424
41	1.976	.656	.548	1.9	.863
42	4.356	.487	.438	1.5	.820

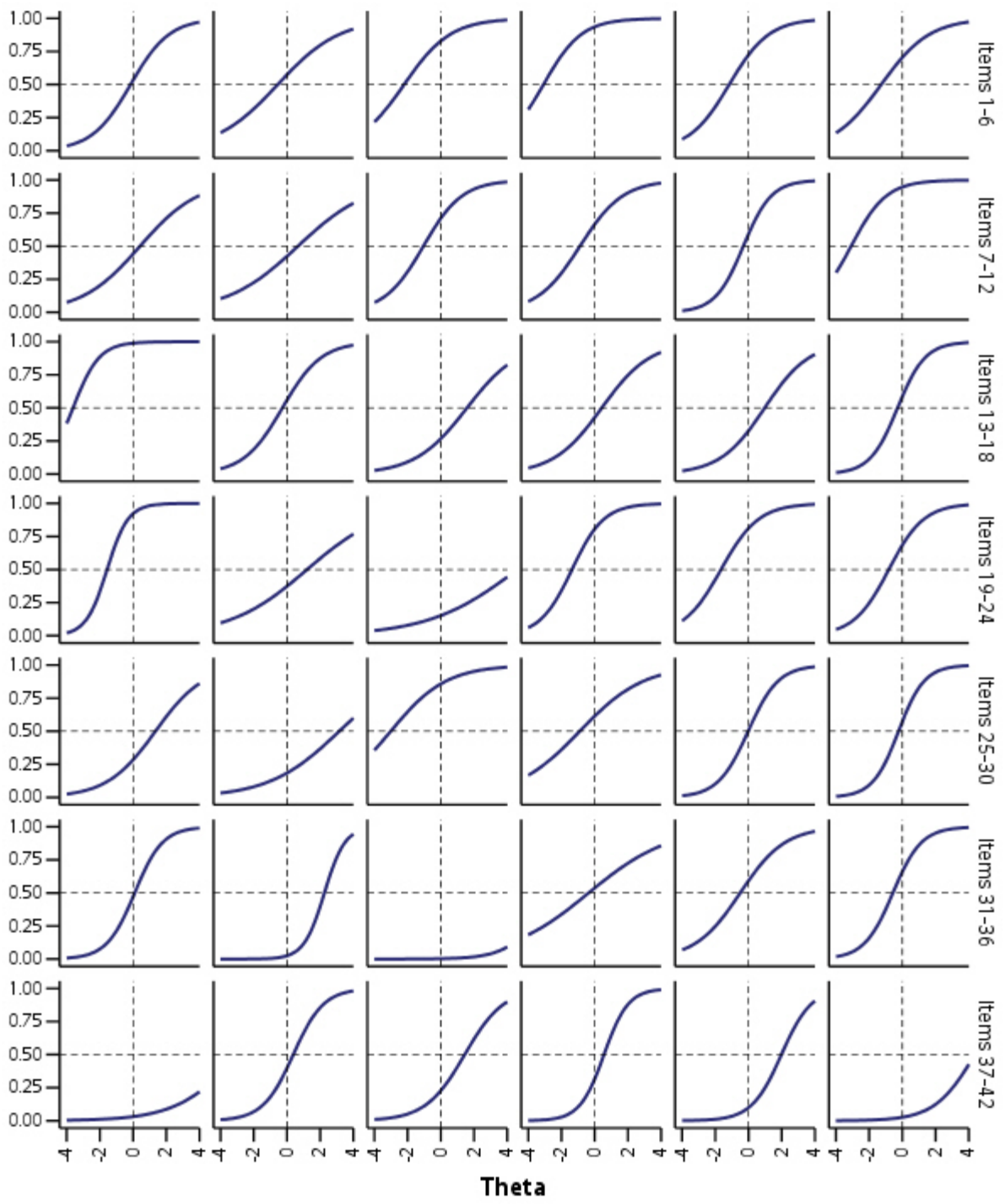


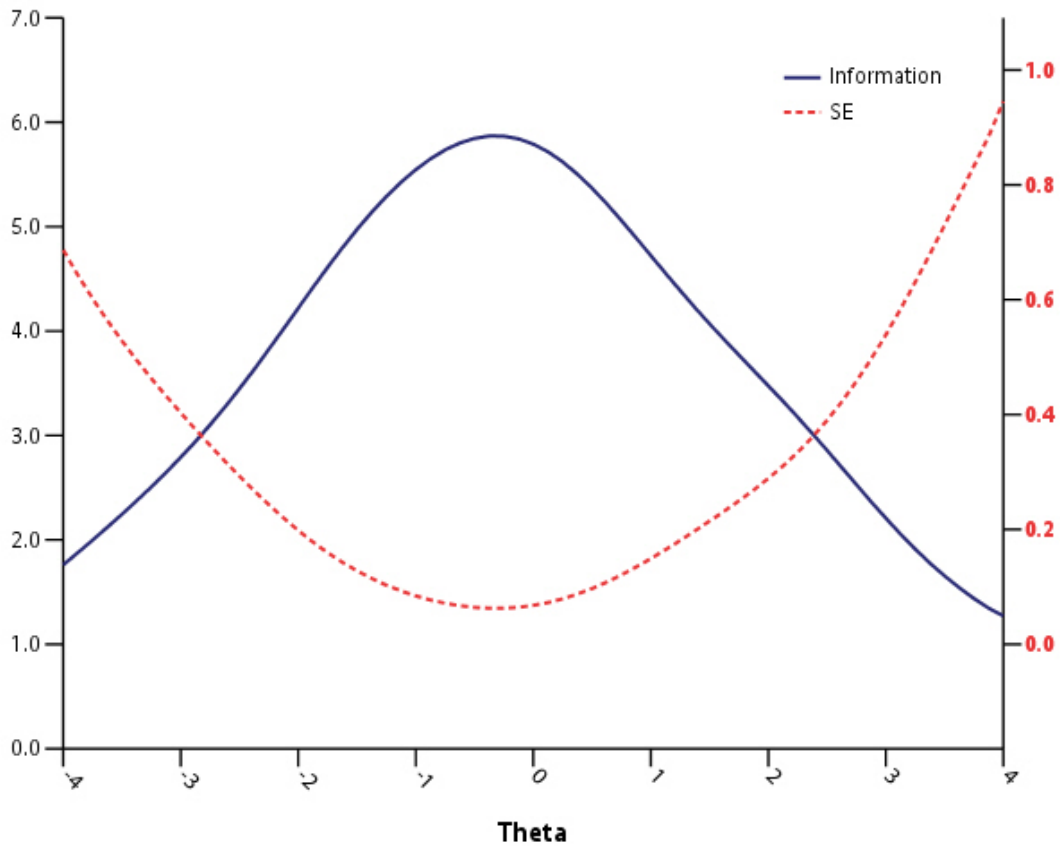
Figure 1: Math Assessment Item Characteristics Curves

1.1c: Scale Analysis

PARAMETER	MEAN	STN DEV
SLOPE	0.509	0.183
LOG(SLOPE)	-0.739	0.371
THRESHOLD	0.222	2.274

MAXIMUM INFORMATION APPROXIMATELY 0.5868D+01 AT -0.3750

FOR A NORMAL POPULATION WITH MEAN 0.000 AND S.D. 1.000
 AVERAGE INFORMATION= 0.5233D+01 AND RELIABILITY INDEX= 0.840



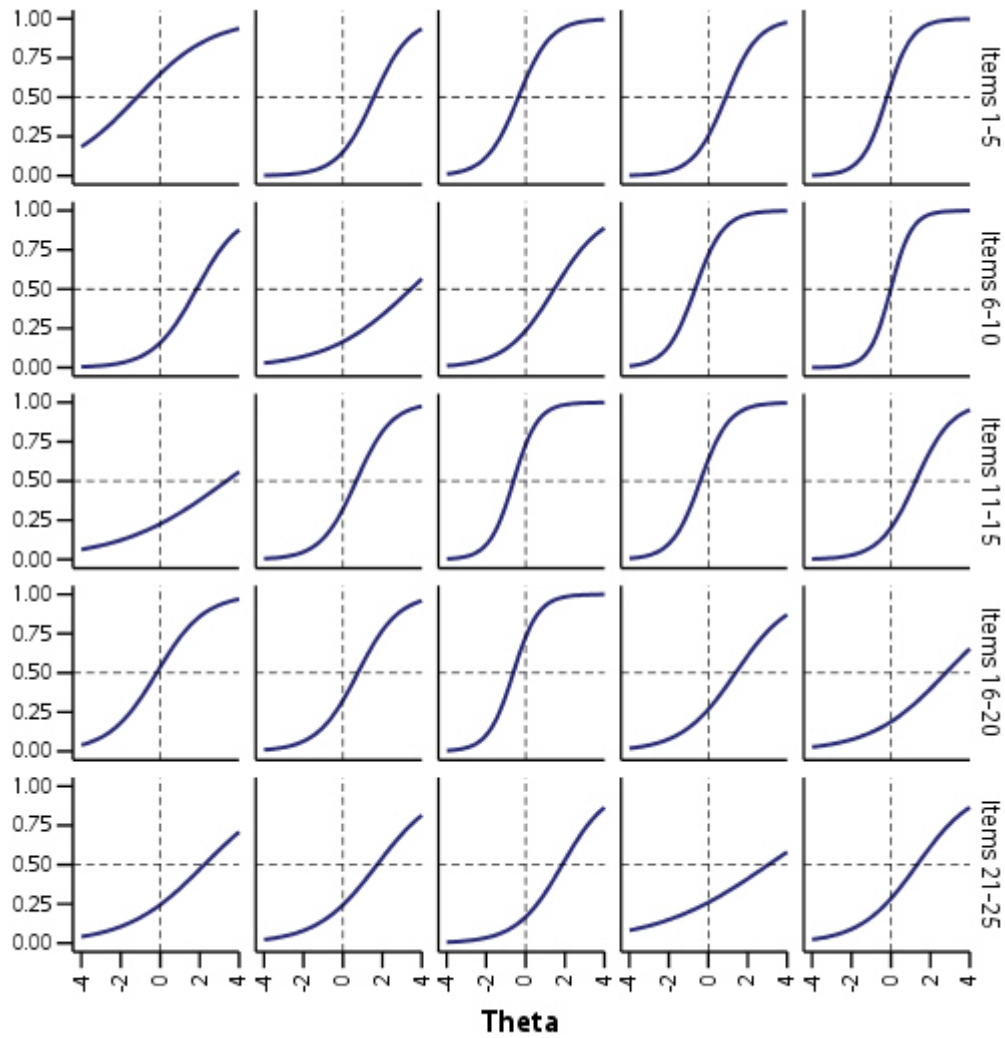
1.1d: DIF Analysis

1.2. Middle School Algebra

1.1a: Description of Instrument

The algebra content knowledge assessment was a standardized Algebra achievement test published by Educational Testing Service. The test contained 25 items, and so had a possible maximum score of 25.

1.1b: Item Analysis

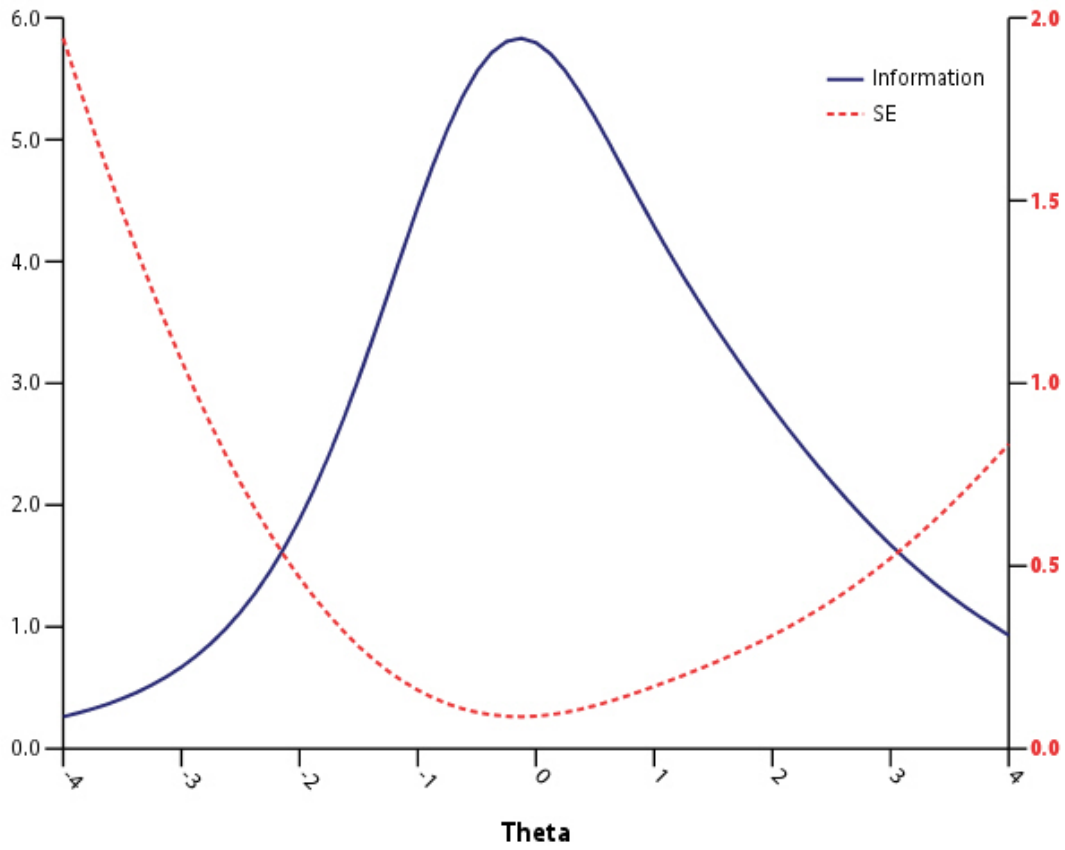


1.1c: Scale Analysis

PARAMETER	MEAN	STN DEV
SLOPE	0.573	0.252
LOG(SLOPE)	-0.660	0.479
THRESHOLD	1.014	1.353

MAXIMUM INFORMATION APPROXIMATELY 0.5832D+01 AT -0.1250

FOR A NORMAL POPULATION WITH MEAN 0.000 AND S.D. 1.000
 AVERAGE INFORMATION= 0.4688D+01 AND RELIABILITY INDEX= 0.824

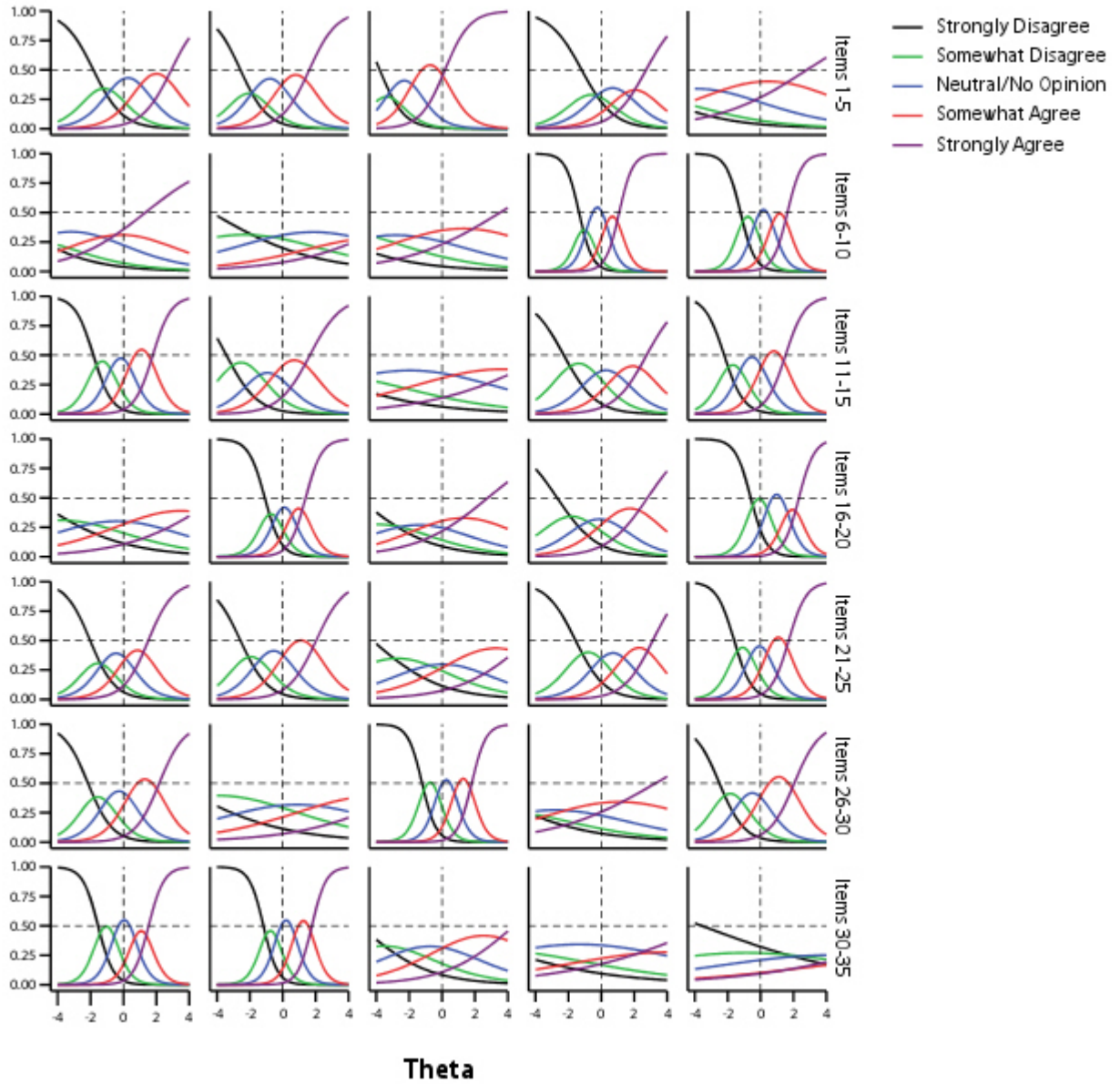


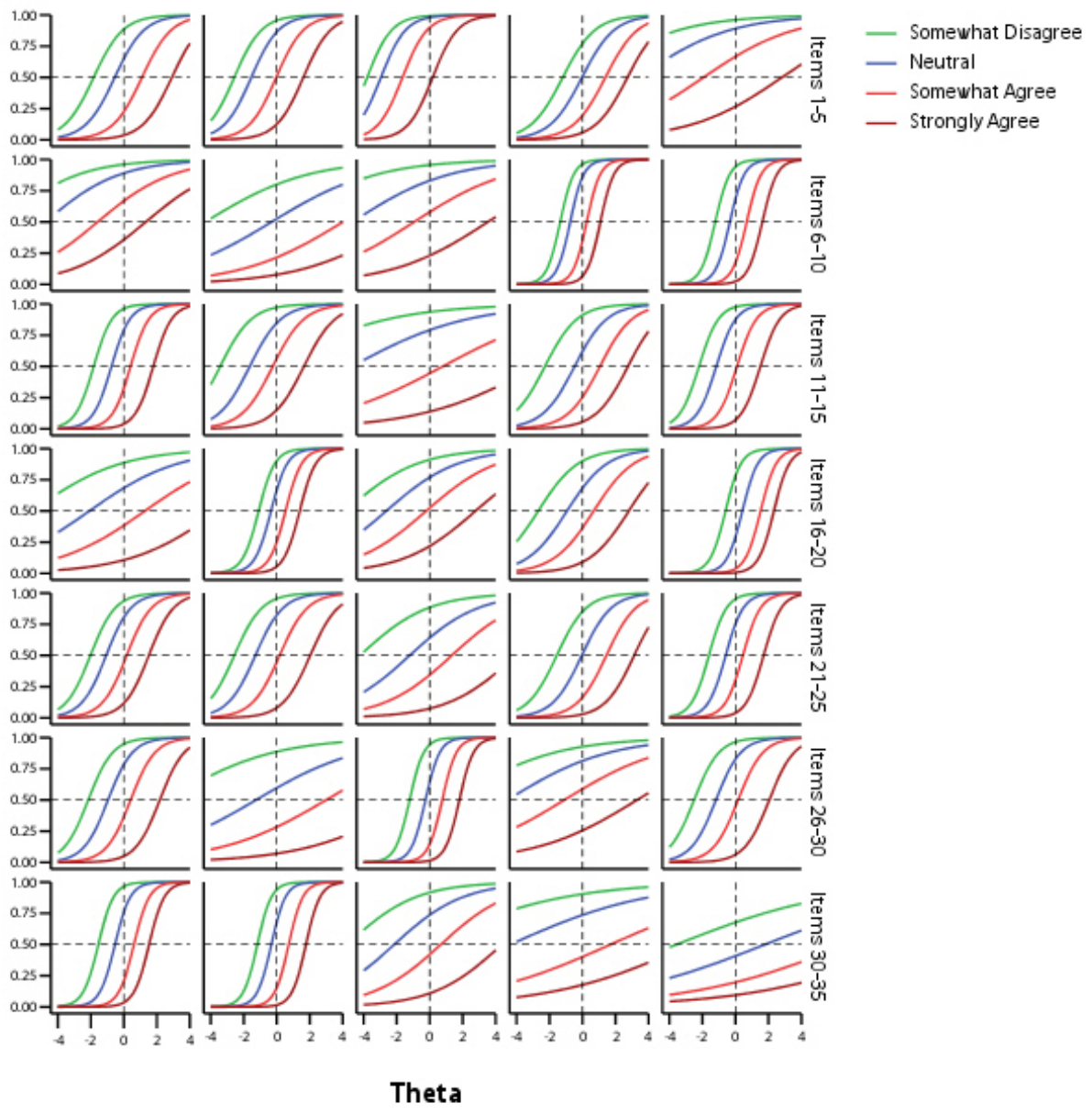
1.1d: DIF Analysis

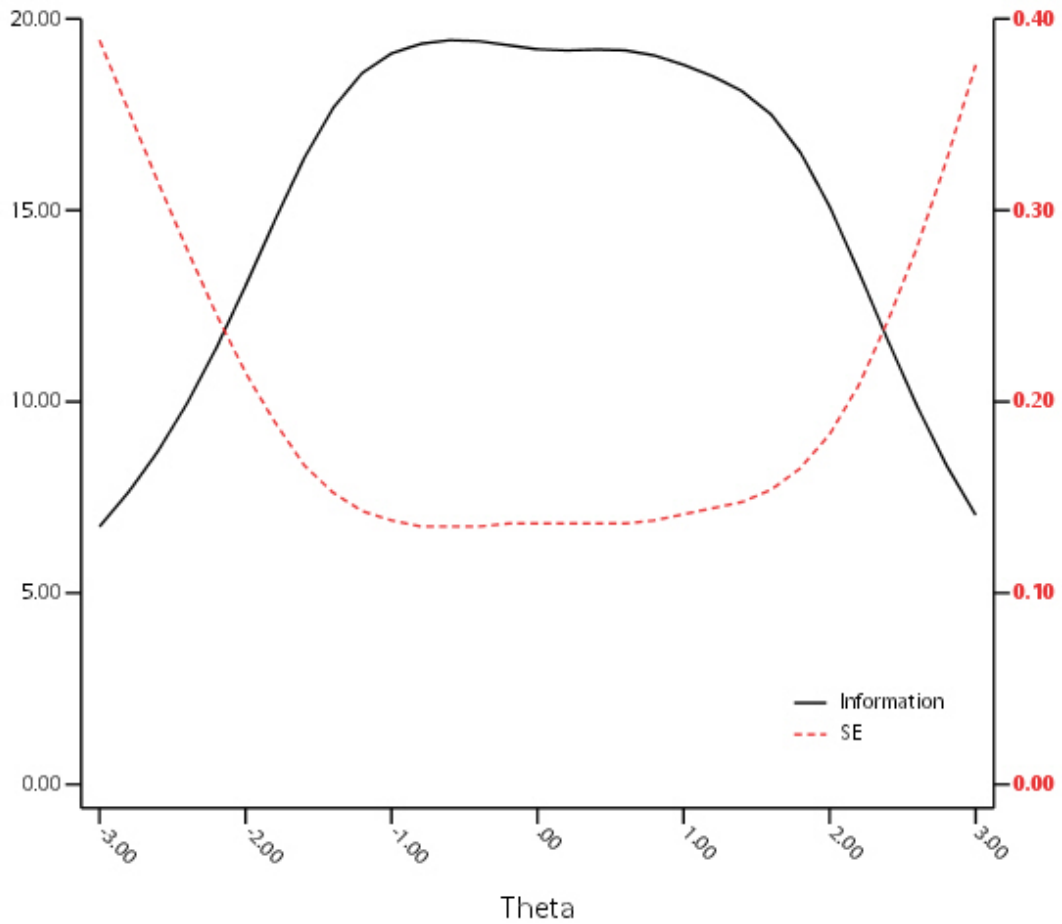
2. Attitudinal Assessments

Item	Slope (discrimination)	Approximate max for Neutral (location)
1	1.120	.250
2	1.204	-.750
3	1.330	-2.250
4	1.019	.625
5	.359	-3.875
6	.441	-3.125
7	.321	1.875
8	.341	-2.750
9	2.447	-.250
10	2.287	.125
11	1.790	-.125
12	1.050	-1.000
13	.283	-2.000
14	1.021	.250

15	1.689	-.500
16	.371	-.375
17	2.036	.125
18	.458	-1.375
19	.819	-.125
20	2.124	1.000
21	1.330	-.500
22	1.191	-.625
23	.482	.000
24	1.119	.750
25	1.869	-.125
26	1.350	-.250
27	.310	.875
28	2.324	.250
29	.323	-2.875
30	1.323	-.500
31	2.160	.000
32	2.341	.250
33	.480	-.750
34	.235	-1.375
35	.208	4.000







References

- Callow-Heusser, C., Allred, D., Robertson, D. J., & Borman, G., (2005a, September). *McDougal Littell Evidence-Based Small Scale Study Final Report: 7th Grade Math*. Logan, Utah: EndVision Research.
- Callow-Heusser, C., Allred, D., Robertson, D. J., & Borman, G., (2005b, September). *McDougal Littell Evidence-Based Small Scale Study Final Report: 7th Grade Algebra*. Logan, Utah: EndVision Research.