Differential Item Functioning and Differential Test Functioning

In the

Idaho Standards Achievement Tests for Spring 2003

Carl Hauser and Gage Kingsbury

January, 2004

(Revised April 8, 2004)

Differential Item Functioning and Differential Test Functioning in the Idaho Standards Achievement Tests for Spring 2003

When a test question is presented to two students who have exactly the same achievement level, we would expect the two students to have an equal probability of answering the question correctly. If this is not the case (within sampling variability) for students who differ on ethnicity or gender, we might suspect the question is functioning differently for different groups of students. When this situation occurs at a group level and students across subgroups are matched for ability, we may have evidence of *differential item functioning* (DIF). However, it would be rare for there to be *identical* item functioning across groups. Thus, questions arise around how different item functioning must be across groups for the item to be considered differentially functioning -- or how differential item functioning is manifest in the resulting estimates of achievement.

The method used to detect DIF in this study is an item response theory (IRT) approach based on the work of Wright & Douglas (as cited in Draba, 1977), Linacre & Wright (1989), and implemented by Linacre (2003). When executed as part of a Winsteps (Linacre, 2003) analysis, this method is comparable to: a) carrying out a joint Rasch analysis that anchors all item difficulties to their NWEA item bank calibrations and, using all data from both Focal and Reference groups, producing anchor values for student abilities; b) carrying out a calibration analysis for the Reference group keeping the student ability estimates and scale structure anchored to produce Reference group item difficulty estimates; c) carrying out a calibration analysis for the Focal group keeping the student ability estimates and scale structure anchored to produce Focal group item difficulty estimates; and d) computing pairwise item difficulty differences (Focal group difficulty minus Reference group difficulty). The calibration analyses in steps b and c are computed for each item, as though all items, except the item currently targeted, are anchored at the bank calibrations as in the main analysis (step a). Mathematically, this approach will yield results comparable to the Mantel-Haenszel (1959) (MH) approach, when MH works as intended (Luppescu, 1993; Schulz, Perlman, Rice, & Wright, 1996). Its advantage is that it does not require the imposition of an arbitrary segmentation and matching scheme that affects the magnitude of the MH statistic (Linacre & Wright, 1989). In addition, this IRT approach offers more interpretable DIF estimates in that they are maintained in the original metric of the test rather in a rescaling of the log-odds ratio (Schultz, et. al., 1996).

Purpose. This study was designed to answer two key questions about the fixed portions of the fourth, eighth, and tenth grade spring 2003 *Idaho Standards Achievement Tests* (ISAT) in Reading, Language Usage and Mathematics:

1. To what extent do test items exhibit substantive differences in functioning across student groups formed based on ethnicity and based on gender?
2. What is the collective impact on test characteristics of any substantive differentially functioning items?

It is worth mentioning that DIF is not quite the same as what has commonly been called "bias". Differential item functioning is an item characteristic that occurs whenever groups differ noticeably in their correct answer frequency, when achievement level is matched. It may be used as part of the process to identify biased items, but is almost always viewed as secondary to a sensitivity review. Because of its statistical nature, many items that are identified as having DIF are not identified as having any difficulty in the sensitivity review. This type of item is rarely removed from use.

Method

Data sets.

For each grade level, 10,000 test records were randomly selected from the pool of complete ISAT records. This represented approximately 57, 55, and 61 percent of the available complete test records in grades 4, 8 and 10 respectively. A 'complete' test record was defined as one in which the student had taken all three content area tests and contained recognized ethnicity and gender codes. Two procedures were applied to this initial record set to insure the integrity of the test scores. First, all records were scanned to determine if the test had been terminated and then resumed at any point during the administration of the fixed portion of the test. If this was found to be the case, the record was eliminated. This was done to minimize the effects of test items or answers being shared with students from the time their test was terminated (temporarily paused) to the time the test was resumed.

Second, content area test records were also eliminated when their scores would not be considered valid if NWEA's standard scoring rules had been applied. Under these rules, a content area test record was eliminated when the proportion correct was less than chance (.25 for reading/language usage and .20 for mathematics) plus .05 or was equal to or greater than .92. This procedure eliminated between 3.6% (Grade 8, Language) and 10.9% (Grade 10, Mathematics) of the remaining test records in each content area set.

Ethnic group membership was used to define the groups for one set of DIF analyses. The minimum number of sub-group members for an analysis was set at 300. Under most circumstances, this is an adequate number of students in each grade level to allow stable estimates of item difficulties to be calculated. Only two ethnic groups (Caucasian and Hispanic) had an adequate number of members at each grade level when considering the entire state population. Thus, for the ethnic group analyses only Hispanic and non-Hispanic groups were formed. Records with ethnic group showing as African American, Asian/Pacific Islander, Native American/Alaskan Native, or Caucasian were selected for the non-Hispanic group. Records with ethnic group indicated as 'unknown' in the database were not included. In all analyses involving ethnicity, Hispanic students were considered the Focal group and non-Hispanic students were considered the Reference group.

The second set of DIF analyses used gender to define the groups. In all analyses involving gender, female students were considered the Focal group and male students were considered the Reference group.

The resulting numbers of content area test records by student grade, ethnicity, and gender appear in Table 1. As the table shows, Language Usage tests had the greatest percentage of tests with valid scores after data integrity procedures were applied. In mathematics across all three grades, approximately 15 percent of the test records were excluded. Test records from Hispanic students closely followed their pattern of representation in the general population with the highest percentage being in grade 4. Language usage was the only area where the percentage of Hispanic students with valid test records present slightly exceeded the percentage of Hispanic students statewide. For Reading and Mathematics, Hispanic students were slightly underrepresented.

Table 1. Numbers of student test records included and excluded in each analysis by ethnicity and gender.

		Reading										
		Ethnicity										
		non-Hispanic			Hispanic			Totals			Percentage Hispanic Included	State Percentage Hispanic
Grade	Gender	Incl'd	Exc-Lo	Exc-Hi	Incl'd	Exc-Lo	Exc-Hi	Incl'd	Exc-Lo	Exc-Hi		
4	Female	3880	97	106	490	60	2	4370	157	108	11.2	
	Male	4012	165	129	535	64	1	4547	229	130	11.8	
	Total	7892	262	235	1025	124	3	8917	386	238	11.5	15.2
8	Female	4082	95	122	361	48	0	4443	143	122	8.1	
	Male	4141	171	86	397	84	1	4538	255	87	8.7	
	Total	8223	266	208	758	132	1	8981	398	209	8.4	11.2
10	Female	3747	151	21	306	64	0	4053	215	21	7.5	
	Male	3795	195	30	294	108	1	4089	303	31	7.2	
	Total	7542	346	51	600	172	1	8142	518	52	7.4	9.9
	Grand Total	23657	874	494	2383	428	5	26040	1302	499		

		Language Usage										
4	Female	3955	61	97	515	37	2	4470	98	99	11.5	
	Male	4213	91	68	559	55	1	4772	146	69	11.7	
	Total	8168	152	165	1074	92	3	9242	244	168	11.6	15.2
8	Female	4207	36	76	403	17	1	4610	53	77	8.7	
	Male	4266	126	40	427	59	0	4693	185	40	9.1	
	Total	8473	162	116	830	76	1	9303	238	117	8.9	11.2
10	Female	3932	49	39	336	52	0	4268	101	39	7.9	
	Male	3954	131	27	520	62	1	4474	193	28	11.6	
	Total	7886	180	66	856	114	1	8742	294	67	9.8	9.9
	Grand Total	24527	494	347	2760	282	5	27287	776	352		

		Mathematics										
4	Female	3948	59	89	524	21	4	4472	80	93	11.7	
	Male	4162	44	157	588	21	5	4750	65	162	12.4	
	Total	8110	103	246	1112	42	9	9222	145	255	12.1	15.2
8	Female	3909	187	156	340	71	2	4249	258	158	8.0	
	Male	3925	200	225	387	80	5	4312	280	230	9.0	
	Total	7834	387	381	727	151	7	8561	538	388	8.5	11.2
10	Female	3532	309	33	289	87	0	3821	396	33	7.6	
	Male	3609	322	79	288	111	0	3897	433	79	7.4	
	Total	7141	631	112	577	198	0	7718	829	112	7.5	9.9
	Grand Total	23085	1121	739	2416	391	16	25501	1512	755		

Analysis.

    Analysis was carried out for each test in two phases, one for the detection of differential item functioning (DIF) and one for the detection of differential test functioning (DTF). Prior to these analyses, test items exhibiting unexpected functioning were identified and excluded from further consideration. This was accomplished by subjecting all items to a Rasch calibration procedure, which anchored all items to their original NWEA calibrations.

    The ISAT tests were designed as wide-range tests. This may cause some items to be too difficult or too easy for the whole group of students taking the test. These items were eliminated from consideration for the DIF study, because it is likely that their item statistics will vary due to difficulty, rather than DIF. An item was eliminated if the percentage of students responding correctly to it was less than chance plus 5 percent or greater than or equal to 92 percent. The result of this procedure was an item set for each test that was slightly reduced in size but more likely to reveal true instances of DIF, vis-à-vis, 'differential functioning' due to the characteristics of the item itself. A summary of the items removed under these criteria is shown in Table 2. Note: Goal (or strand) information about each item referenced in Table 2 as well as in Tables 3, 7, and 11 in the Results section can be obtained from tables in the appendix.

Table 2. Test items excluded from analyses due to unexpected performance.

Content Area	Grade	Item seq. number	RIT	Reason for exclusion
Reading	4	9	183	Percent correct > 92
		14	181	Percent correct > 92
		42	217	Percent correct < 30
	8	4	202	Percent correct > 92
		12	243	Percent correct < 30
	10	33	239	Percent correct < 30
		36	242	Percent correct < 30
		49	232	Percent correct < 30
Language Usage	4	23	204	Percent correct < 30
		33	211	Percent correct < 30
	8	41	238	Percent correct < 30
		42	239	Percent correct < 30
	10	1	192	Percent correct > 92
		49	231	Percent correct < 30
		55	237	Percent correct < 30
Mathematics	4	2	183	Percent correct > 92
		5	188	Percent correct > 92
		21	203	Percent correct > 92
		38	222	Percent correct < 25
		39	224	Percent correct < 25
	8	21	239	Percent correct < 25
	10	8	232	Percent correct > 92
		47	265	Percent correct < 25
		49	268	Percent correct < 25
		50	270	Percent correct < 25

DIF criterion level. For the purposes of this study, the magnitude of the difference between Reference group and Focal group item difficulty estimates was considered to be the primary indicator of an item evidencing DIF; that is, DIF large enough that the item may warrant further attention. Unfortunately there are few guidelines for considering how large a difference in difficulty estimates should be before it is, by itself, of practical importance. Wright and Douglas (1975) suggested a "rule of thumb" of .5 logits, which translates into 5 RIT. A 5 RIT difference is also treated by NWEA as the maximum allowable difference between two difficulty estimates of the same item within its calibration process. The criterion for DIF was set at a calibration difference of >= | 5 |.

Estimating the impact of DIF items on tests – Differential test functioning (DFT). Any comparison with two or more items identified as evidencing DIF in the previous phase was incorporated into the estimation of differential test functioning. For a particular comparison, the calibration difference (Focal group minus Reference group) for a DIF item was added to its NWEA item bank calibration. This produced an adjusted calibration' that was consistent with the Focal group performance on the item. The 'adjusted calibrations' were used to anchor the difficulties for their corresponding items in constructing a 'virtual test' comprised of these items along with the remaining non-DIF items that retained their original bank calibrations. A test characteristic curve (TCC) for each test (original and virtuals) was plotted from the raw score to RIT score relationship for the test. To the extent that the TCCs deviate from one another, differential test functioning would be present.

Results

Results are organized first by content area. Within each content area, the results of the DIF analyses are presented together as a set of six analyses; two comparison types (ethnic and gender) X three grade levels (4, 8 and 10). Within each grade level, scatter plots are used to illustrate how the identified DIF items (calibration difference >= | 5 |) relate to other items in the test and which group is advantaged by the difference in calibrations. Finally, for each grade level within a content area, test overall item summary information is provided for each test type (actual and virtual) along with test characteristic curves are presented.

Reading.
Reading items evidencing DIF under the adopted criterion are shown in Table 3. Across the grades, only two items evidenced gender-related DIF. As grade levels increase, the magnitudes of calibration differences increase slightly for ethnic-related DIF items.   Across all grades, 58% of the DIF items associated with ethnicity favored the reference groups (non-Hispanic and male).

Table 3. Reading test items evidencing noteworthy DIF by grade, ethnicity, and gender.

Grade 4

Seq Number	Hispanic Difficulty Est	Std Err	non-Hispanic Difficulty Est	Std Err	Difference	Std Err Difference	Bank Calibration	Adjusted Calibration
8	208.69	.721	213.71	.259	-5.02	.766	216	211
12	211.69	.750	216.78	.267	-5.08	.796	224	219
16	192.44	.742	186.73	.357	5.71	.824	188	194
17	195.59	.714	188.45	.340	7.14	.790	188	195
23	197.89	.701	192.21	.308	5.67	.765	196	202

Grade 8

	Hispanic		non-Hispanic					
28	214.62	.782	221.20	.244	-6.58	.819	223	216
36	207.10	.868	196.86	.445	10.25	.976	200	210

Grade 10

	Hispanic		non-Hispanic					
2	219.45	.891	229.41	.252	-9.96	.926	213	203
8	219.68	.890	212.04	.332	7.64	.950	215	223
19	214.85	.944	220.71	.273	-5.86	.983	221	215
24	228.42	.890	220.01	.276	8.41	.932	219	227
44	217.94	.904	211.14	.341	6.81	.967	224	231
	Female		**Male**					
12	228.47	.345	222.39	.358	6.07	.497	217	223
26	222.16	.364	227.93	.343	-5.77	.500	230	224

Grade 4. Reference group and Focal group calibrations for all grade 4 Reading items are shown in Figure 1 by comparison type (ethnicity and gender). The identity line running through each panel allows a quick visual assessment of the magnitude of calibration differences between the two groups; the further from the line the greater the difference. Open triangles mark the items with noteworthy DIF. The top panel of Figure 1 shows the slightly greater variation in the calibrations by ethnic group. Five of the eight items with DIF calibrated at or below 200 for both groups; 200 is the proficiency level for grade 4 Reading. Table 4 shows the relationship of the test administered to students in the spring of 2003 to the slightly shorter tests used for this study; the Studied (referred to here as the "original") and the ethnic adjusted virtual test. This table shows the items in the virtual test designed to fully favor the ethnic Focal group, had slightly more misfit than the administered test and was slightly less internally consistent from a raw score perspective (KR-20 = .78 vs. .7986). Figure 2 shows that the virtual test has the same raw score to RIT score relationship as the original test.

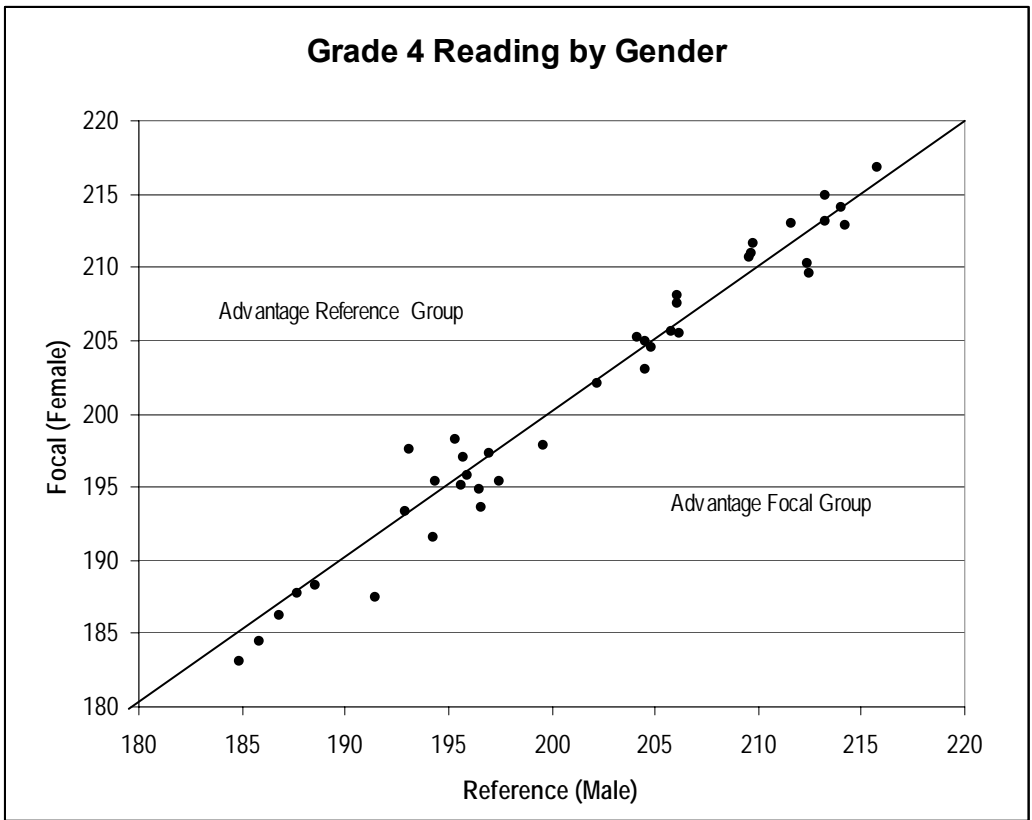**Figure 1. Item calibrations for grade 4 Reading by comparison type.**

Grade 4 Reading by Ethnicity

Focal (Hispanic)

Advantage Reference Group

Advantage Focal Group

Note: Open markers are most discrepant items between groups

Reference (non-Hispanic)

Grade 4 Reading by Gender

Focal (Female)

Advantage Reference Group

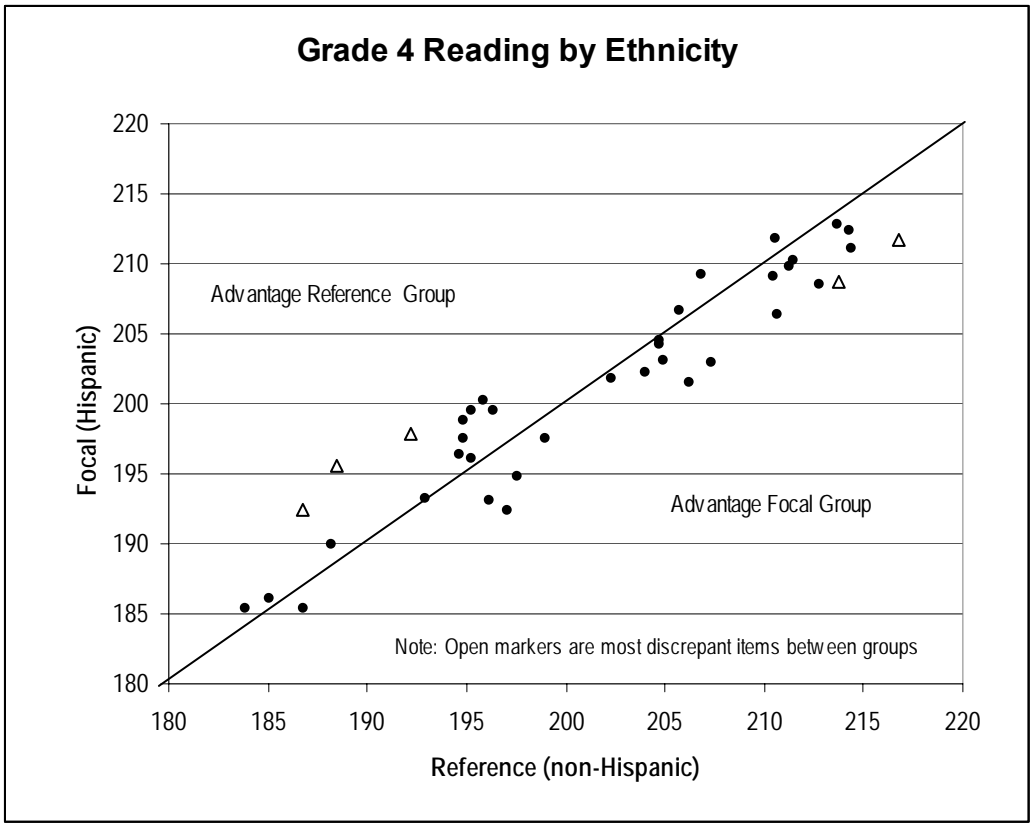Advantage Focal Group

Reference (Male)

Table 4. Grade 4 Reading tests - Actual and virtual

Test	Item difficulties			Standardized item fit		Internal consistency (KR-20)	Proportion correct for 'proficient'
	Mean	sd	n	Mean	sd		
Administered	200.3	11.15	42	0.1	1.0	.79	.50
Studied (original)	201.2	10.42	39	0.3	0.9	.78	.49
Ethnic adjusted	201.4	9.49	39	0.2	1.0	.78	.49

Figure 2. Test characteristic curves for original and ethnic adjusted test for grade 4 Reading.
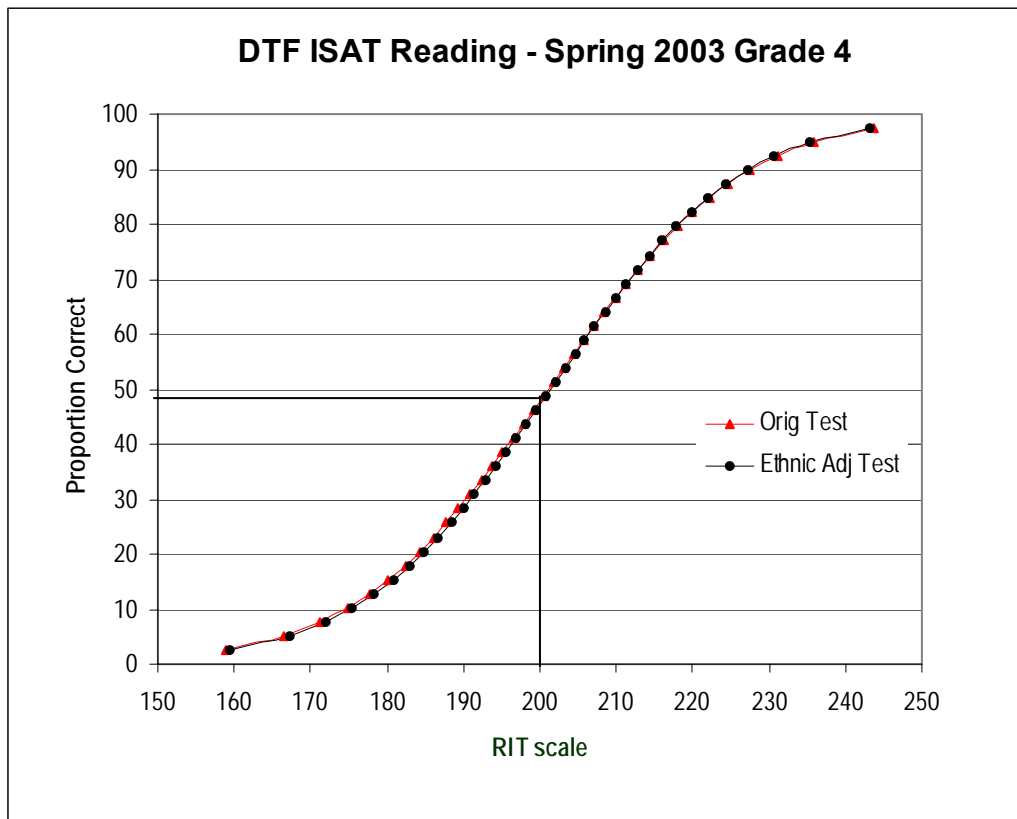
Grade 8. Figure 3 shows the grade 8 Reading calibrations, by group, for both comparison types. Only two ethnic-related DIF items were observed. No gender-related DIF items were observed. The top panel of the figure shows the two items for all Reading comparisons; one with a 10.25 RIT difference favoring non-Hispanic students and a one with a 6.58 RIT difference favoring Hispanic students. Table 5 illustrates, again, that the administered test had slightly better fitting items than the Studied test, overall. The lone virtual test did not deviate from the administered test or the Studied test. The virtual test designed to fully favor the ethnic Focal group has the same raw score to RIT score relationship as the original test.   Figure 4 shows the test characteristic curves for the Studied and the ethnic adjusted tests.

Figure 3. Item calibrations for grade 8 Reading by comparison type.

Grade 8 Reading by Ethnicity

Advantage Reference Group

Advantage Focal Group

Note: Open markers are most discrepant items between groups

Focal (Hispanic)

Reference (non-Hispanic)

Grade 8 Reading by Gender

Advantage Reference Group

Advantage Focal Group
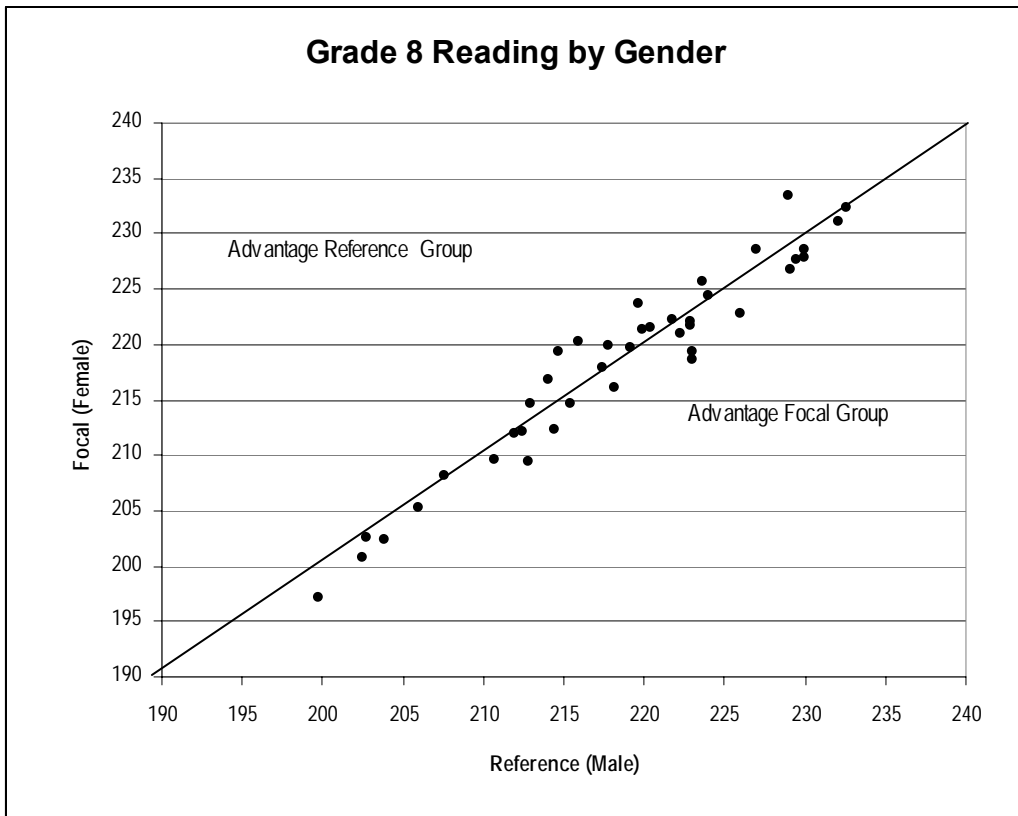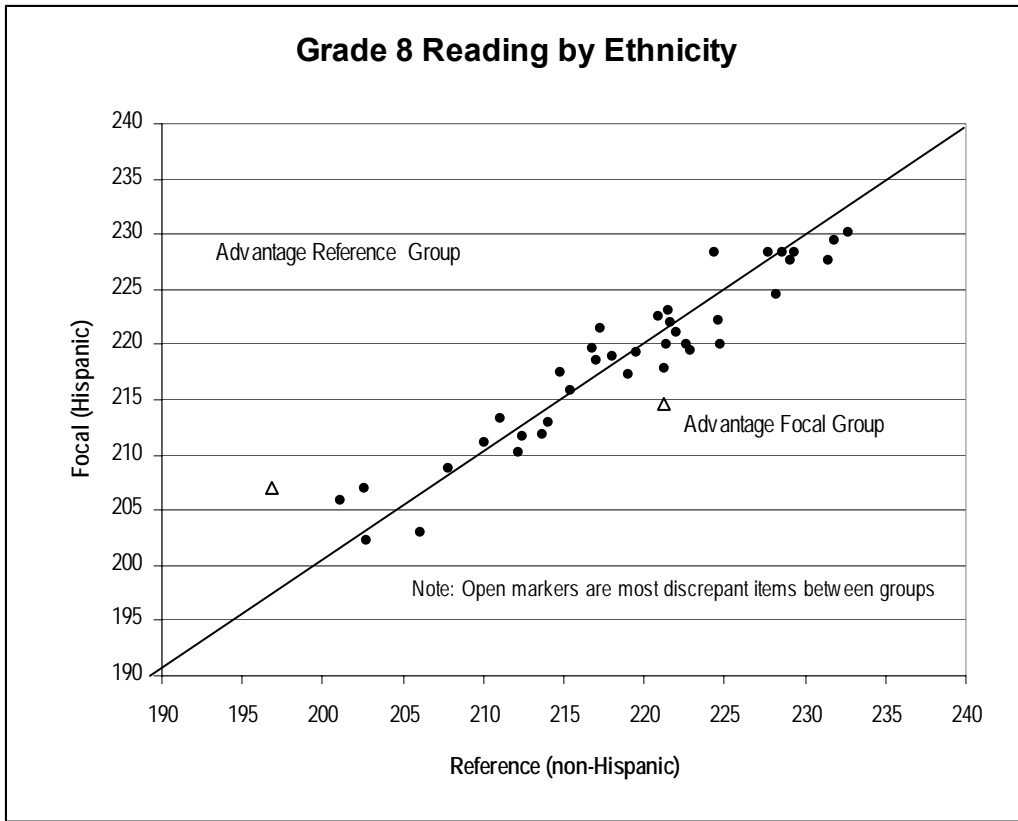
Focal (Female)

Reference (Male)

Table 5. Grade 8 Reading tests - Actual and virtual

Test	Item difficulties			Standardized item fit		Internal consistency (KR-20)	Proportion correct for 'proficient'
	Mean	sd	n	Mean	sd		
Administered	218.5	10.69	42	0.2	1.0	.80	.50
Studied (original)	218.3	9.77	39	0.3	0.9	.79	.50
Ethnic adjusted	218.5	8.89	39	0.3	1.0	.78	.50

Figure 4.   Test characteristic curves for original, ethnic adjusted and gender adjusted test for grade 8 Reading.
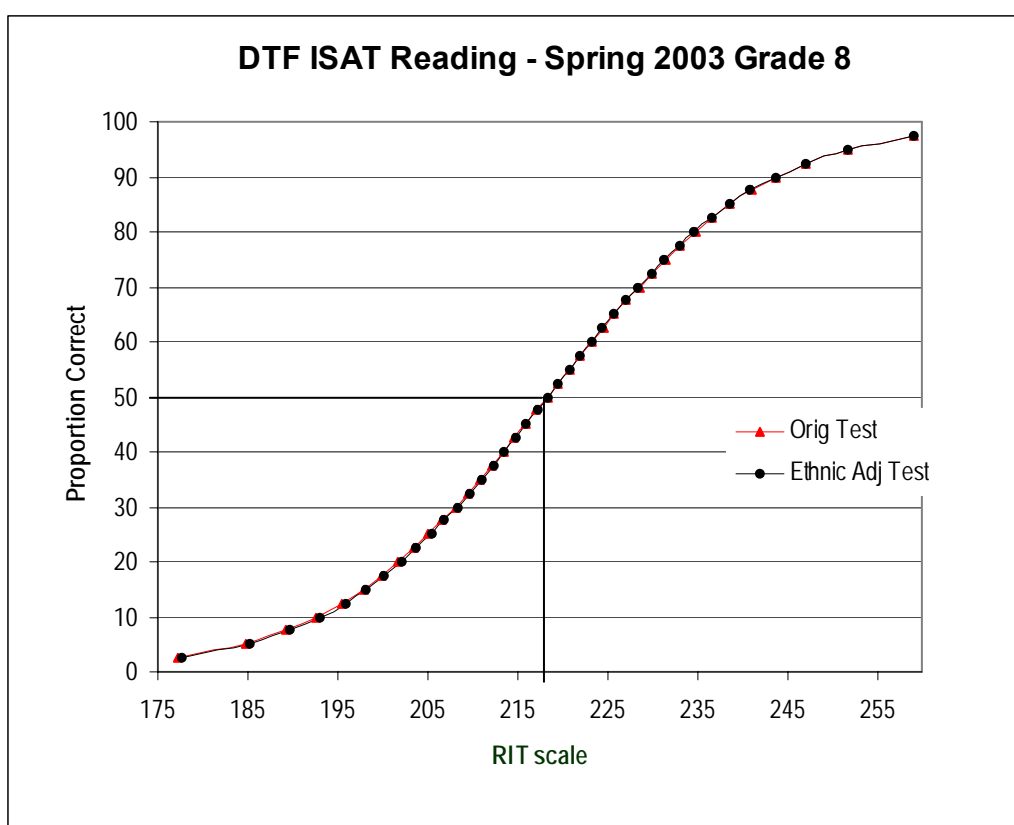
Grade 10. Figure 5 provides the grade 10 Reading calibrations for the ethnic and gender comparisons. For the ethnic comparison, the most discrepant items were toward the lower end of the grade 10 difficulty range. Four of five of these items were at or below the grade 10 proficiency cut-off (RIT 224). Only two DIF items were observed for the gender comparison. Table 6 shows that the there was a slight drop in item difficulty from the test actually administered to the Studied test and the virtual test adjusted for ethnicity. Overall item fit was somewhat better for the virtual tests relative to the Studied test. The administered test showed the best overall fit. Once again, as shown in Figure 6, there was no difference between the Studied test and the virtual tests in term of the proportion of items required to be answered correctly to be considered proficient.

Figure 5. Item calibrations for grade 10 Reading by comparison type.

Grade 10 Reading by Ethnicity

Advantage Reference Group

Advantage Focal Group

Note: Open markers are most discrepant items between groups

Focal (Hispanic)

Reference (non-Hispanic)

Grade 10 Reading by Gender

Advantage Reference Group

Advantage Focal Group

Note: Open markers are most discrepant items between groups
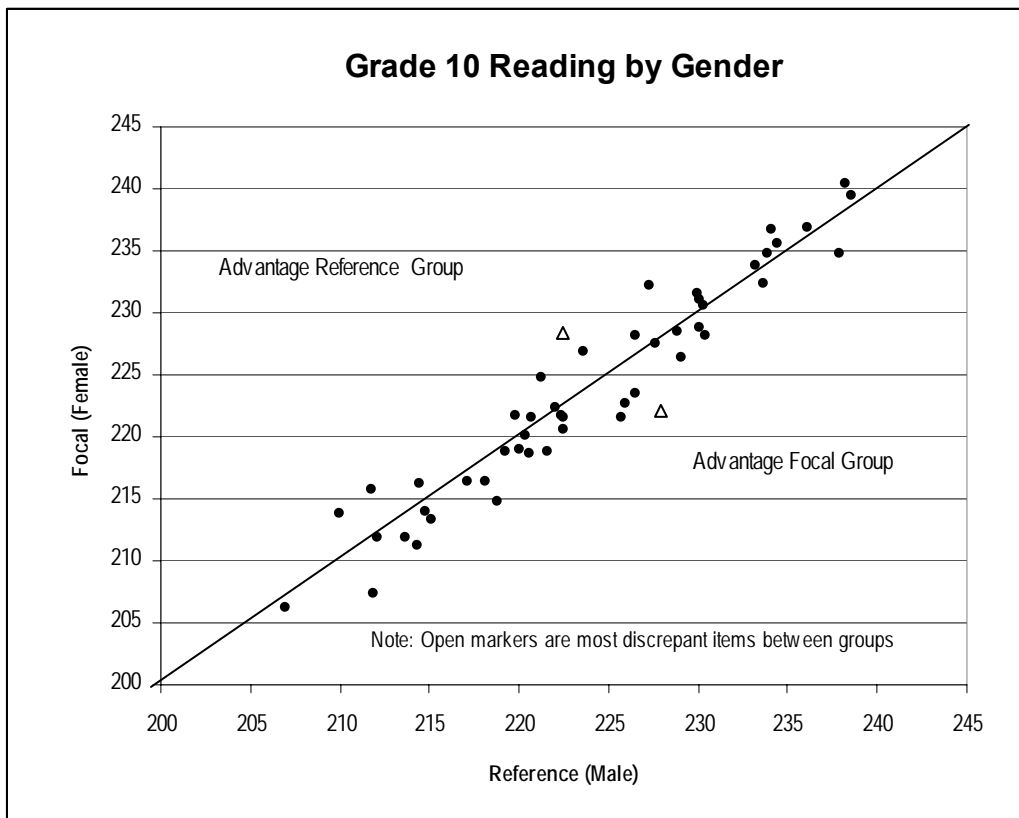
Focal (Female)

Reference (Male)

Table 6. Grade 10 Reading tests - Actual and virtual

Test	Item difficulties			Standardized item fit		Internal consistency (KR-20)	Proportion correct for 'proficient'
	Mean	sd	n	Mean	sd		
Administered	224.3	10.45	55	0.2	1.0	.82	.47
Studied (original)	223.6	10.08	52	0.5	0.9	.82	.50
Ethnic adjusted	223.6	10.40	52	0.3	1.0	.83	.50
Gender adjusted	223.6	10.29	52	0.3	1.0	.83	.50

Figure 6. Test characteristic curves for original, ethnic adjusted and gender adjusted tests for grade 10 Reading.
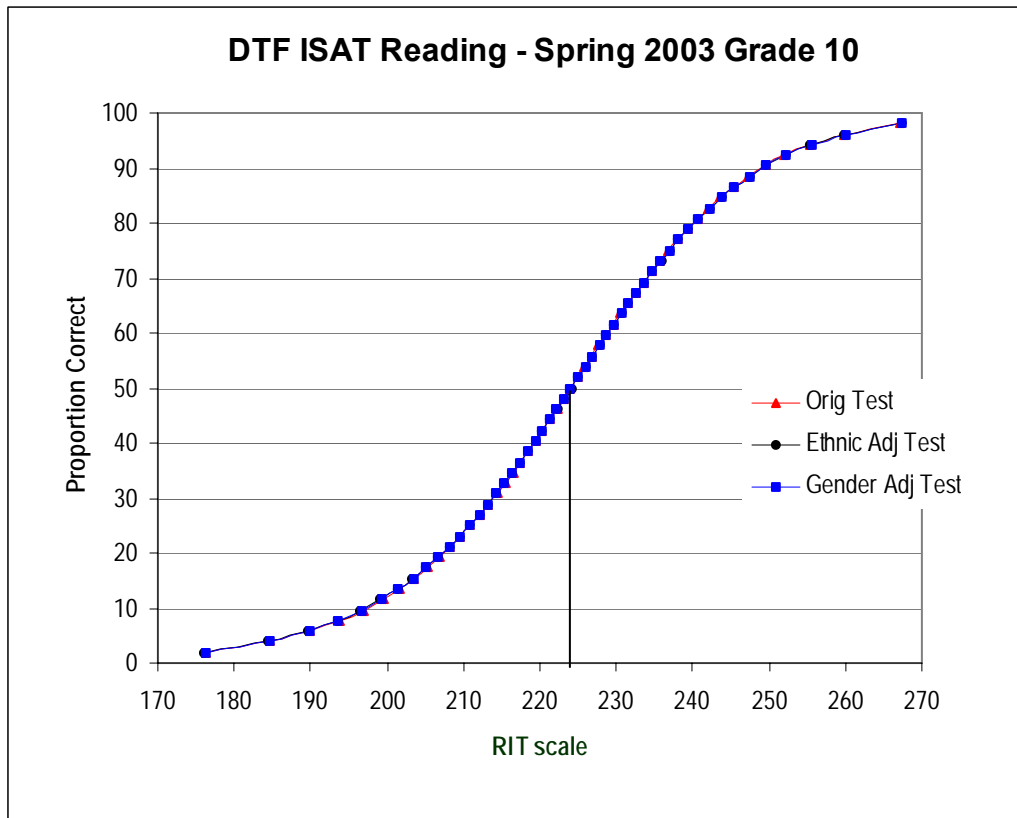
Language Usage.

Of the three content areas, only Language Usage resulted in a comparison set in which all items evidencing DIF were found easier by the focal ethnic group (grade 8). These results were not consistent with comparisons at grades 4 and 10 where differences tended to favor the Reference groups. Table 7 summarizes these items and all others evidencing DIF. In the only grades to evidence gender-related DIF items, all items favored females.

Table 7. Language Usage test items evidencing noteworthy DIF by grade, ethnicity, and gender.

Grade 4

Seq Number	Hispanic Difficulty Est	Std Err	non-Hispanic Difficulty Est	Std Err	Difference	Std Err Difference	Bank Calibration	Adjusted Calibration
5	193.93	.711	187.58	.350	6.36	.792	185	191
9	187.22	.823	181.93	.425	5.30	.927	187	192
16	195.30	.696	200.65	.257	-5.35	.742	193	188
38	200.49	.662	207.05	.243	-6.56	.705	215	208

Grade 8

	Hispanic Difficulty Est	Std Err	non-Hispanic Difficulty Est	Std Err	Difference	Std Err Difference	Bank Calibration	Adjusted Calibration
8	202.26	.342	210.47	.768	-8.21	.840	204	196
12	207.28	.292	214.20	.736	-6.92	.792	209	202
18	206.24	.301	211.48	.757	-5.24	.815	213	208

	Female		Male					
2	195.26	.632	200.80	.450	-5.54	.776	200	194
3	197.72	.571	203.72	.410	-6.00	.703	201	195

Grade 10

	Hispanic Difficulty Est	Std Err	non-Hispanic Difficulty Est	Std Err	Difference	Std Err Difference	Bank Calibration	Adjusted Calibration
17	209.18	.949	202.37	.399	6.81	1.029	214	221
20	221.74	.820	215.98	.272	5.76	.864	216	222
22	222.67	.821	215.87	.273	6.80	.865	218	225
25	223.14	.821	233.31	.251	-10.17	.859	220	210
32	221.08	.821	215.40	.275	5.67	.866	210	216
50	229.06	.858	234.16	.254	-5.10	.895	233	228
53	241.26	1.118	235.21	.257	6.06	1.147	235	241

	Female		Male					
3	199.49	.637	204.49	.476	-5.00	.795	205	200
18	197.68	.685	206.10	.452	-8.42	.821	216	208

Grade 4. All grade 4 Language Usage item calibrations are plotted by ethnic and gender comparison groups in Figure 7. With the exception of the ethnic-related DIF items, most items are found quite close to the identity lines. In view of these calibrations differences, little difference would be expected between the original test and the virtual test designed to fully favor the Focal group. This is exactly the case as shown in Figure 8.

Figure 7. Item calibrations for grade 4 Language Usage by comparison type.
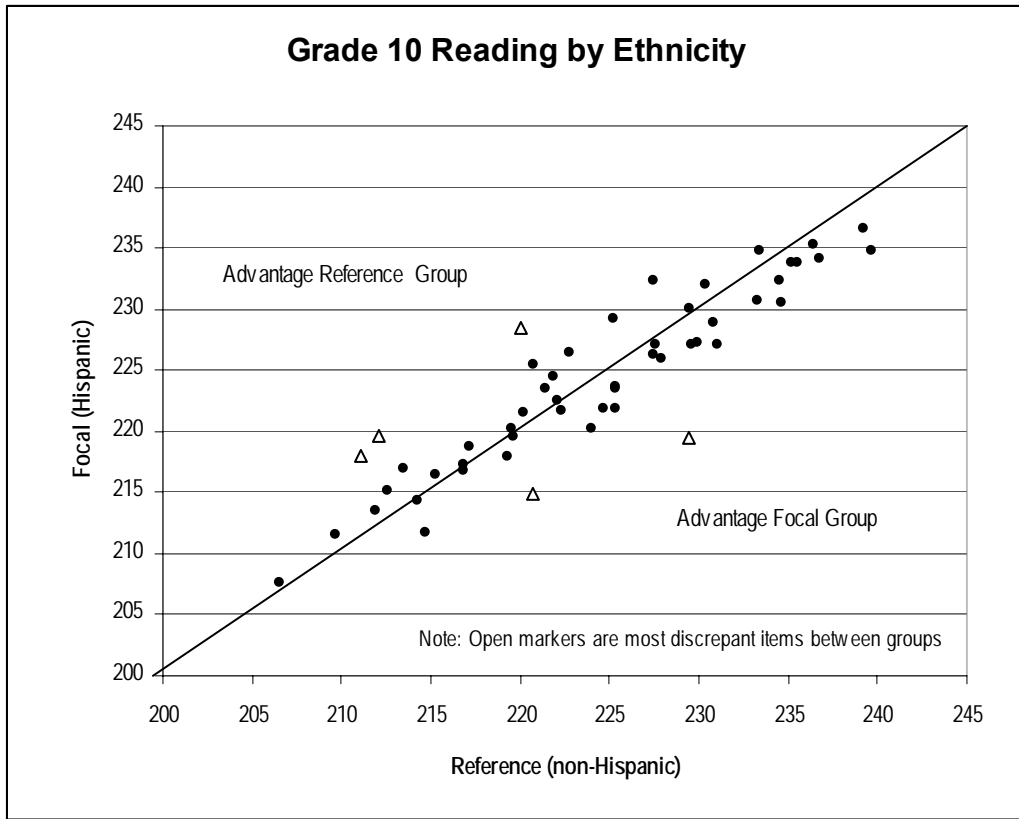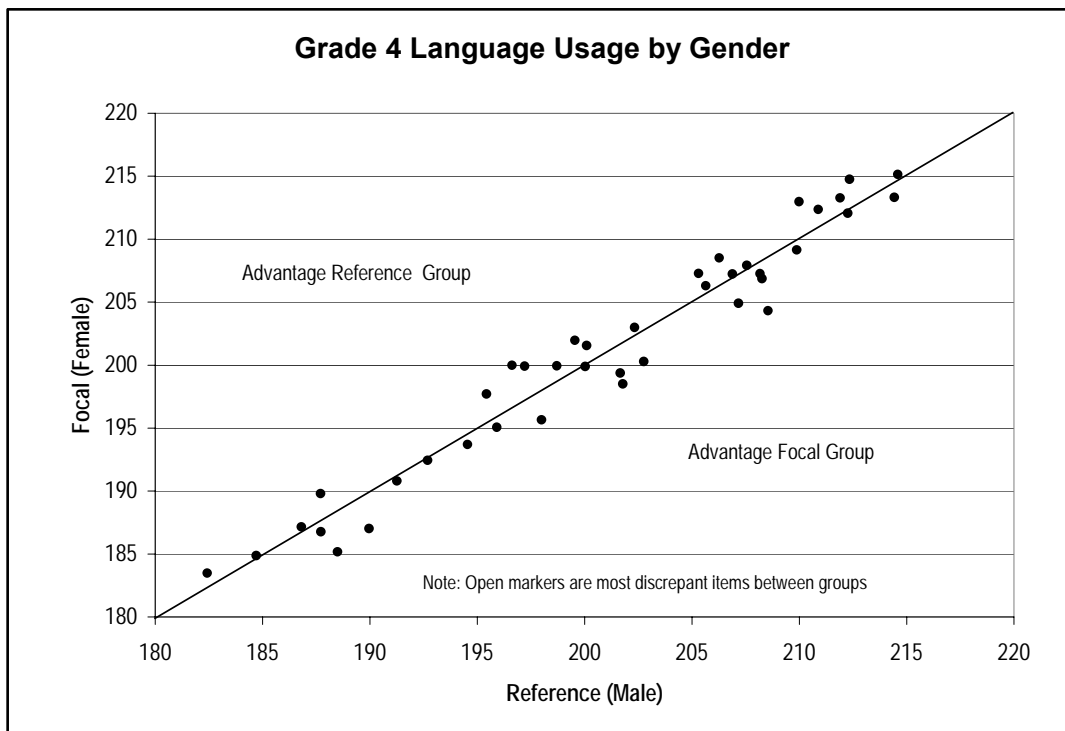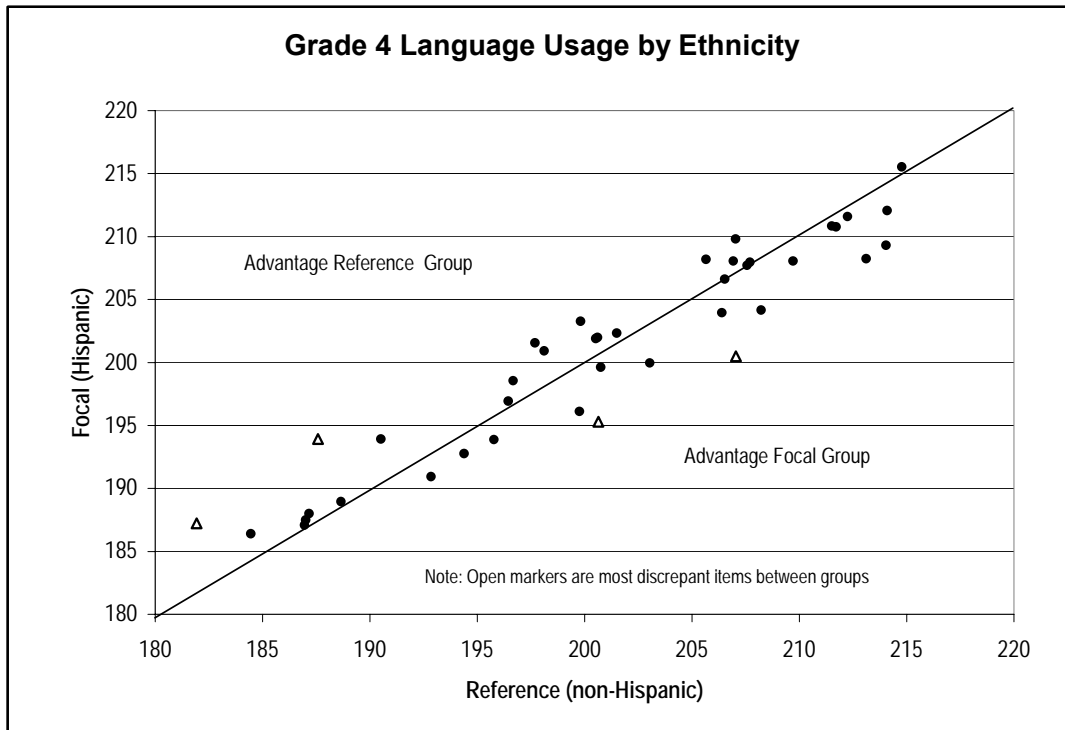
Table 8. Grade 4 Language Usage tests - Actual and virtual

Test	Item difficulties			Standardized item fit		Internal consistency (KR-20)	Proportion correct for 'proficient'
	Mean	sd	n	Mean	sd		
Administered	201.1	10.26	42	0.4	0.9	.79	.50
Studied (original)	200.8	10.13	40	0.4	0.9	.79	.50
Ethnic adjusted	200.7	9.87	40	0.4	0.9	.79	.50

Figure 8.   Test characteristic curves for original, ethnic adjusted and gender adjusted tests for grade 4 Language Usage.
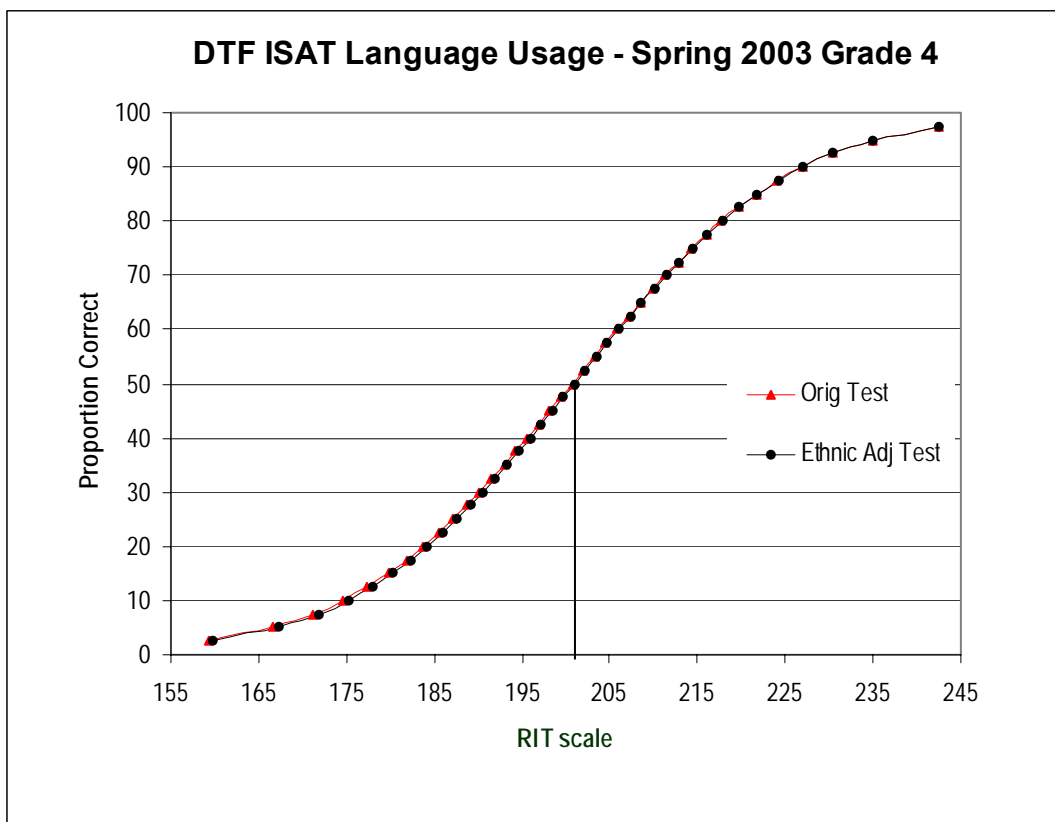
Grade 8. Language usage item calibrations for grade 8 by comparison type are plotted in Figure 9. Like the grade 4 items, most differences between groups are small and therefore appear close to the identity line. Like the grade 4 comparisons, all items showing ethnic-related DIF appear in the lower portion of the test's difficulty range and favor the Reference group. Gender-related DIF items were also in the lower part of the test range but favored the focal group. The effects that these differences would have on virtual tests designed to favor the Focal groups are shown in Figure 10. The Studied test and the two virtual tests would all require 23 items answered correctly to be considered proficient.

Figure 9. Item calibrations for grade 8 Language Usage by comparison type.

Grade 8 Language Usage by Ethnicity

Advantage Reference Group

Advantage Focal Group

Note: Open markers are most discrepant items between groups

Focal (Hispanic)

Reference (non-Hispanic)

Grade 8 Language Usage by Gender

Advantage Reference Group

Advantage Focal Group

Note: Open markers are most discrepant items between groups
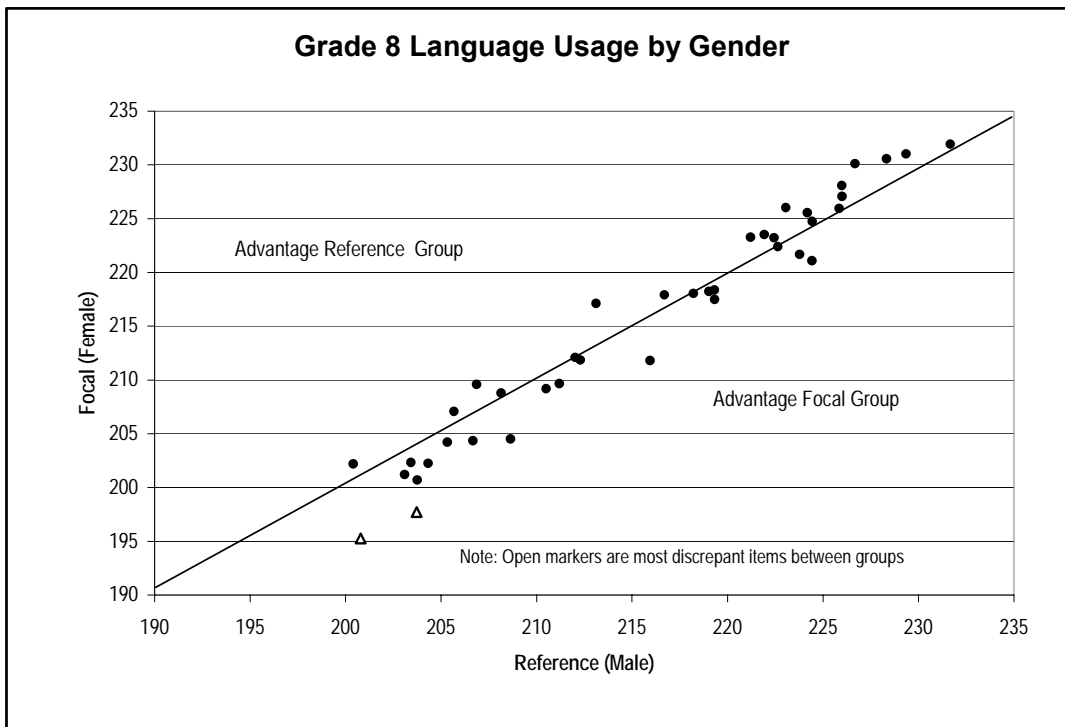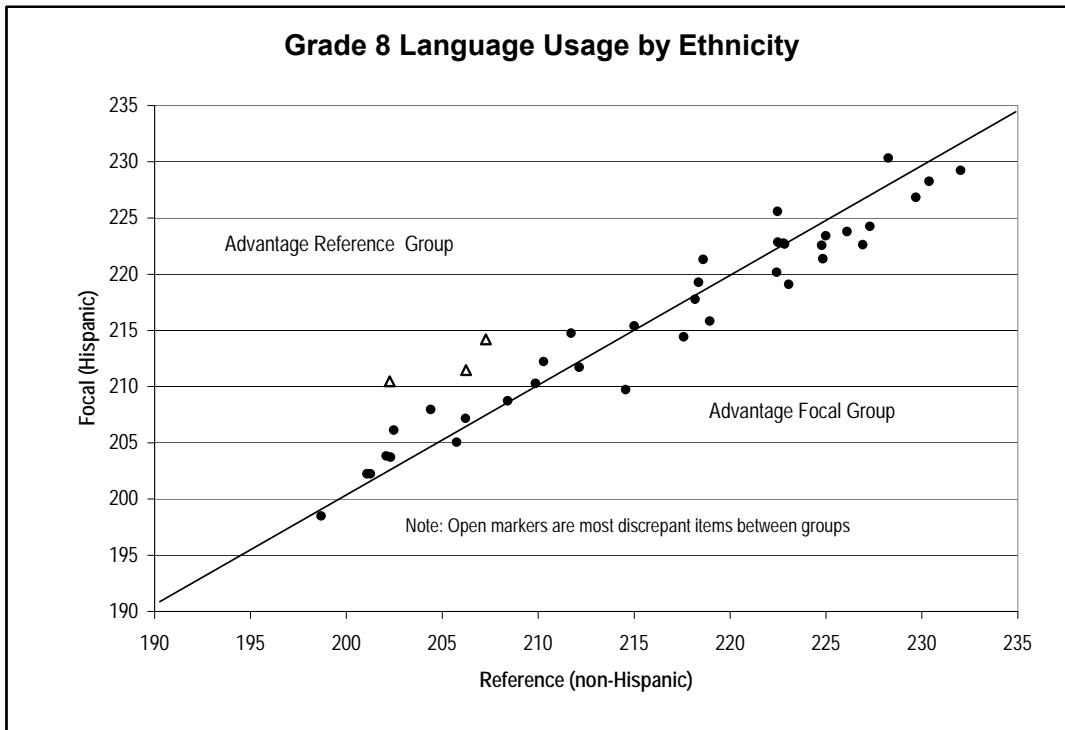
Focal (Female)

Reference (Male)

Table 9. Grade 8 Language Usage tests - Actual and virtual

Test	Item difficulties			Standardized item fit		Internal consistency (KR-20)	Proportion correct for 'proficient'
	Mean	sd	n	Mean	sd		
Administered	216.8	10.73	42	0.1	1.0	.81	.55
Studied (original)	215.7	9.69	40	0.3	1.0	.79	.58
Ethnic adjusted	215.2	10.25	40	0.3	0.9	.79	.58
Gender adjusted	215.4	10.21	40	0.3	1.0	.79	.58

Figure 10.   Test characteristic curves for original, ethnic adjusted and gender adjusted tests for grade 8 Language Usage.
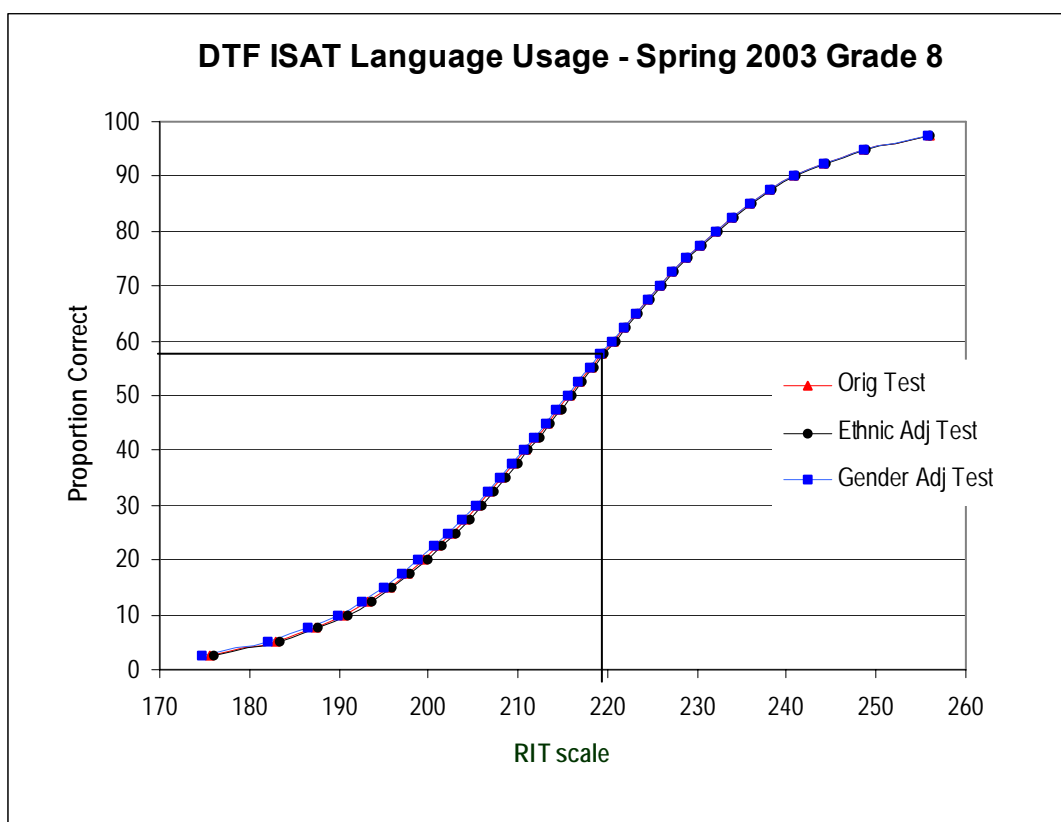
Grade 10. Figure 11 provides the plots of grade 10 Language Usage item calibrations by comparison type. These plots are similar to those for grade 8 for gender related DIF but not for ethnic-related DIF. Ethnic-related DIF items were spread across the difficulty range while gender-related DIF items were in the lower end of the RIT scale. In the ethnic comparisons, DIF item were generally easier for the Reference group while the Focal group (females) tended to find gender-related DIF items easier. In both virtual tests designed to fully favor the Focal groups, a student would need to answer 27 items correctly to be considered proficient. On the Studied test, 27 items would also have to be answered correctly to attain the same performance category. Figure 12 presents the test characteristic curves for all three tests.

Figure 11. Item calibrations for grade 10 Language Usage by comparison type.

Grade 10 Language Usage by Ethnicity

Advantage Reference Group

Advantage Focal Group

Note: Open markers are most discrepant items between groups

Focal (Hispanic)

Reference (non-Hispanic)

Grade 10 Language Usage by Gender

Advantage Reference Group

Advantage Focal Group

Note: Open markers are most discrepant items between groups
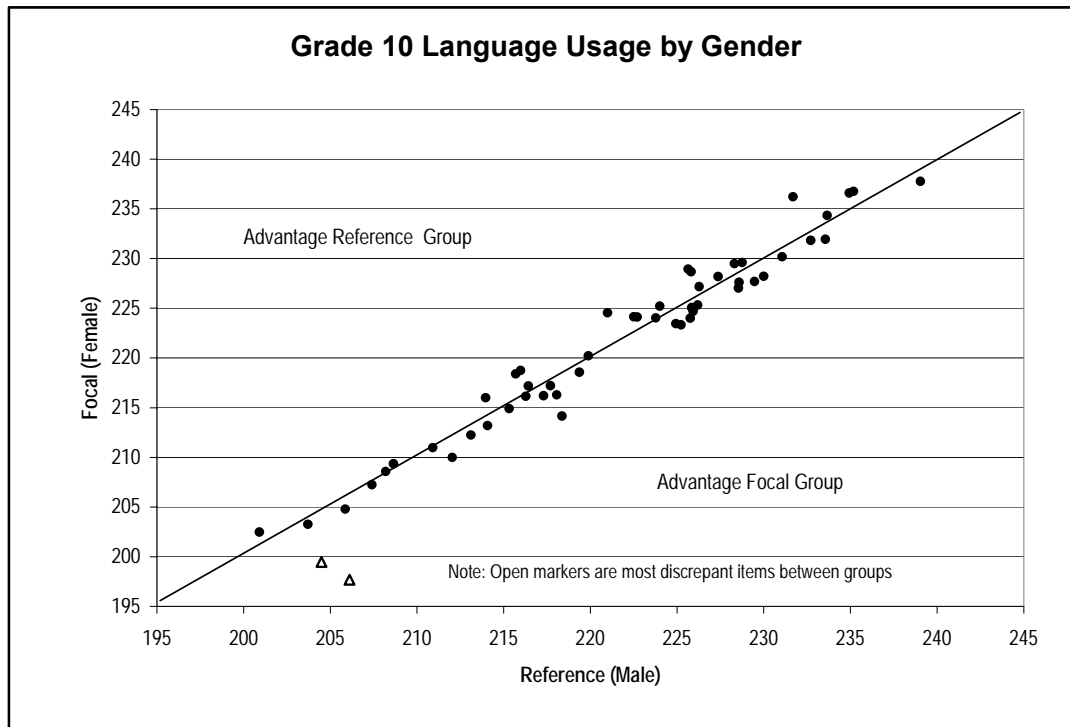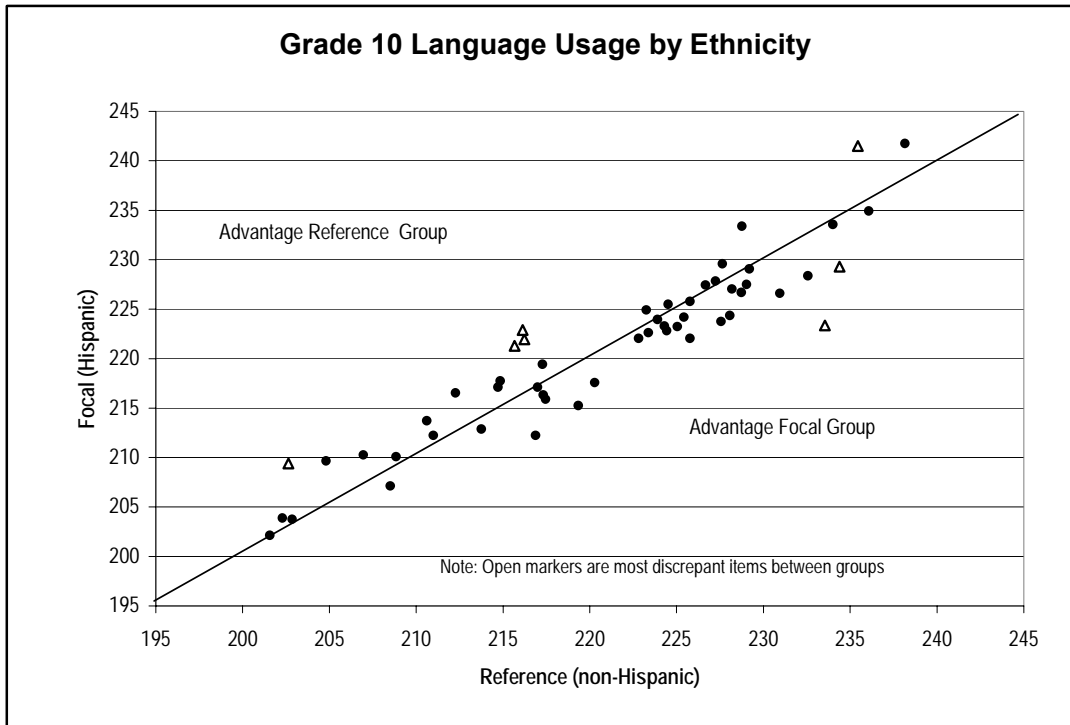
Focal (Female)

Reference (Male)

18

Table 10. Grade 10 Language Usage tests - Actual and virtual

Test	Item difficulties			Standardized item fit		Internal consistency (KR-20)	Proportion correct for 'proficient'
	Mean	sd	n	Mean	sd		
Administered	222.0	9.33	56	0.1	1.0	.80	.50
Studied (original)	220.9	10.03	53	0.2	1.0	.82	.51
Ethnic adjusted	220.6	9.94	53	0.2	1.0	.82	.51
Gender adjusted	220.7	10.30	53	0.2	1.0	.82	.51

Figure 12.   Test characteristic curves for original, ethnic adjusted and gender adjusted tests for grade 10 Language Usage.
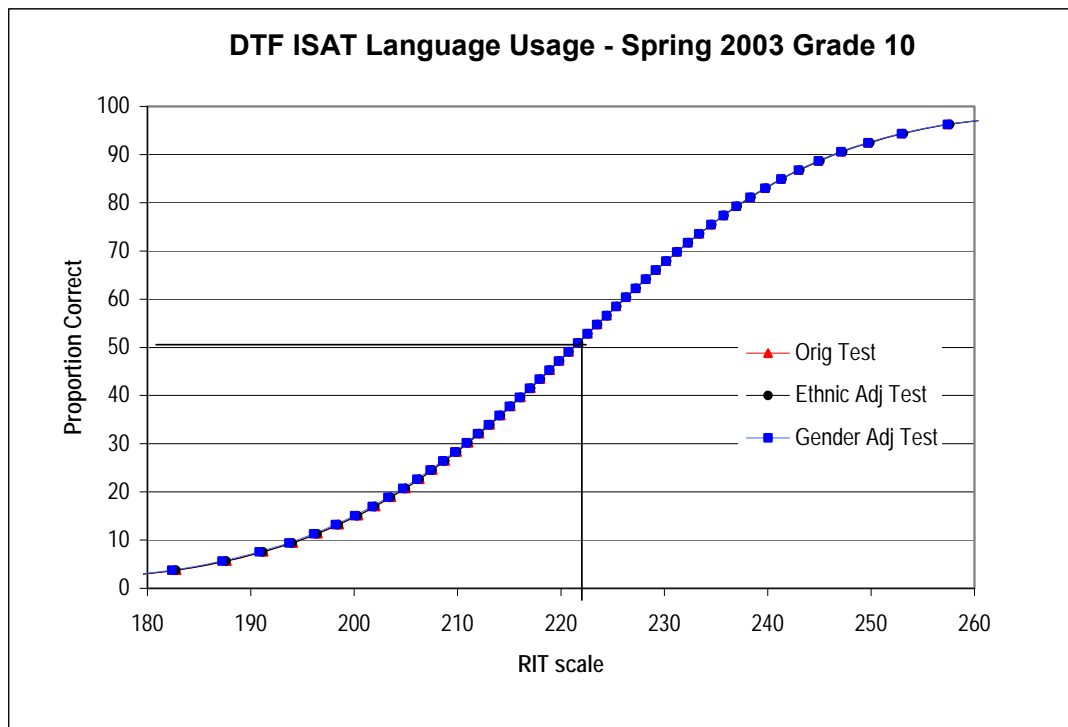
Mathematics.

In contrast to Language Usage, Mathematics was the content area with the greatest number of items evidencing DIF. There were twice as many items observed with DIF in the ethnicity comparisons for grade 10 (n=6) as there were for either grade 4 or grade 8. For the gender comparisons, a larger difference was observed (grade 8, n=2; grade 10, n=14). The magnitude of differences for gender-related DIF in grade 10 tended to be larger than for ethnic-related DIF in grade 10. Eight of the 14 gender-related DIF item in grade 10 had calibration differences greater than | 6 |. For the ethnic-related DIF items in grade 10, only two of six had differences this large. These differences are presented in Table 11.

Table 11. Mathematics test items evidencing noteworthy DIF by grade, ethnicity, and gender.

Grade 4

Seq Number	Hispanic Difficulty Est	Std Err	non-Hispanic Difficulty Est	Std Err	Difference	Std Err Difference	Bank Calibration	Adjusted Calibration
6	197.82	.775	190.78	.373	7.03	.860	189	196
34	217.16	.791	211.00	.277	6.16	.839	212	218
35	216.45	.784	223.28	.292	-6.82	.837	219	212

Grade 8

	Hispanic		non-Hispanic					
4	215.31	1.037	209.60	.421	5.71	1.120	215	221
9	231.37	.964	225.90	.312	5.47	1.013	225	230
19	233.82	.979	227.42	.308	6.40	1.026	238	244
	Female		**Male**					
4	215.73	.489	203.15	.665	12.58	.826	215	228
12	225.43	.422	230.60	.405	-5.17	.586	230	225

Grade 10

	Hispanic		non-Hispanic					
12	231.00	1.334	225.28	.469	5.72	1.414	236	242
16	250.24	1.288	242.20	.372	8.04	1.341	244	252
27	231.39	1.328	236.73	.391	-5.34	1.385	249	244
33	246.55	1.262	251.56	.361	-5.02	1.313	253	248
42	258.23	1.399	265.45	.391	-7.23	1.453	260	253
55	248.67	1.275	242.90	.370	5.78	1.328	239	245
	Female		**Male**					
6	245.41	.497	237.96	.522	7.45	.720	230	237
9	232.80	.554	238.06	.521	-5.26	.760	234	229
13	254.44	.497	244.51	.498	9.94	.703	238	248
20	266.12	.544	260.52	.510	5.60	.746	247	253
21	234.88	.539	242.91	.502	-8.03	.736	247	239
27	239.02	.516	233.36	.551	5.66	.755	249	255
29	262.31	.523	252.67	.492	9.64	.718	250	260
30	236.25	.531	243.06	.501	-6.82	.730	250	243
35	216.69	.767	221.92	.678	-5.23	1.023	255	250
39	257.86	.505	263.37	.522	-5.51	.726	256	250
44	267.09	.550	272.29	.581	-5.21	.80	262	257
51	237.07	.526	229.80	.581	7.26	.784	229	236
54	273.95	.609	264.14	.526	9.81	.804	236	246
55	247.34	.494	239.20	.516	8.13	.714	239	247

Grade 4. Item calibrations for grade 4 Mathematics are plotted in Figure 13 for each comparison type. The close proximity of items to their respective identity lines illustrates the small differences between the calibrations within each comparison type. As shown in Table 12 and Figure 14, these differences have no impact on the original test: the original test and the virtual test that fully favors the Focus have the same test characteristic curves.

Figure 13. Item calibrations for grade 4 Mathematics by comparison type.

Grade 4 Mathematics by Ethnicity

Grade 4 Mathematics by Gender

Table 12. Grade 4 Mathematics tests - Actual and virtual

Test	Item difficulties			Standardized item fit		Internal consistency (KR-20)	Proportion correct for 'proficient'
	Mean	sd	n	Mean	sd		
Administered	204.0	12.22	42	0.0	1.0	.85	.52
Studied (original)	204.5	11.28	37	0.2	1.0	.82	.51
Ethnic adjusted	204.1	10.79	34	0.1	1.0	.81	.51

Figure 14. Test characteristic curves for original, ethnic adjusted and gender adjusted tests for grade 4 Mathematics.
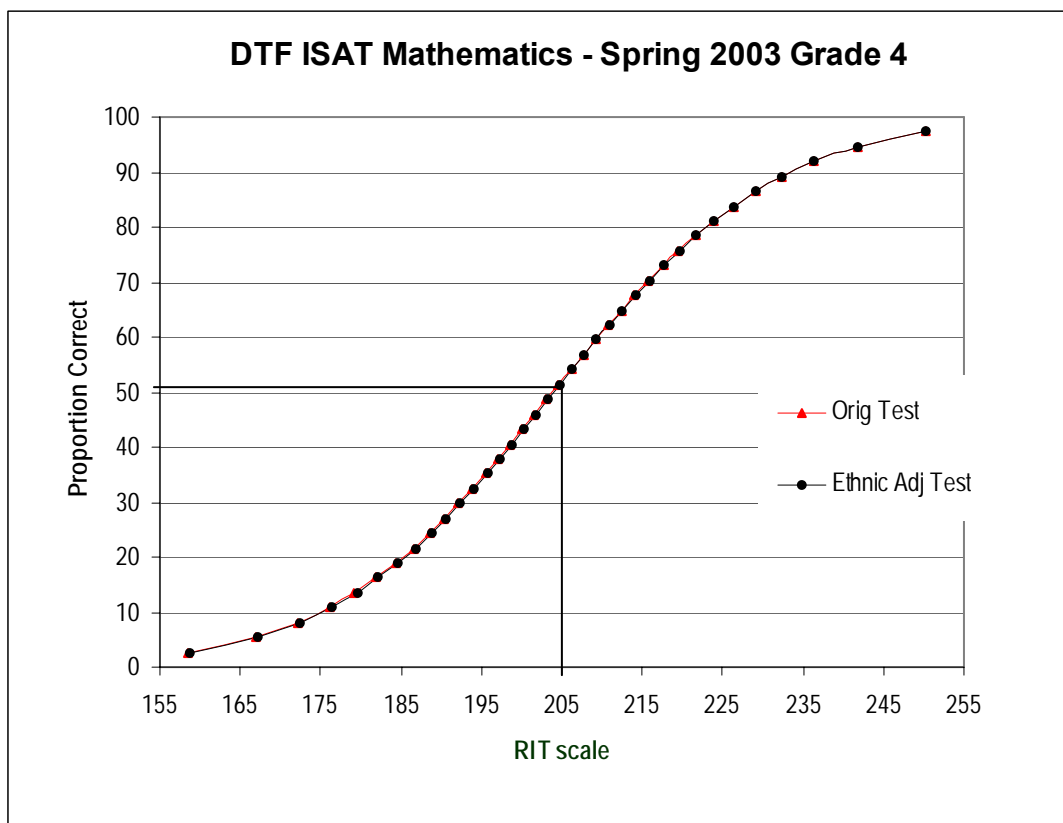
Grade 8. Plots of all grade 8 Mathematics calibrations by comparison type appear in Figure 15. All DIF items but one favor the Reference group. Table 13 shows that the Studied test had overall item fit which matched the Administered test and was only slightly better than the two virtual tests. Both virtual tests and the Studied test would require 23 items to be answered correctly to reach the defined proficiency level of 233. The original test would require 24 items to be answered correctly.

Figure 15. Item calibrations for grade 8 Mathematics by comparison type.
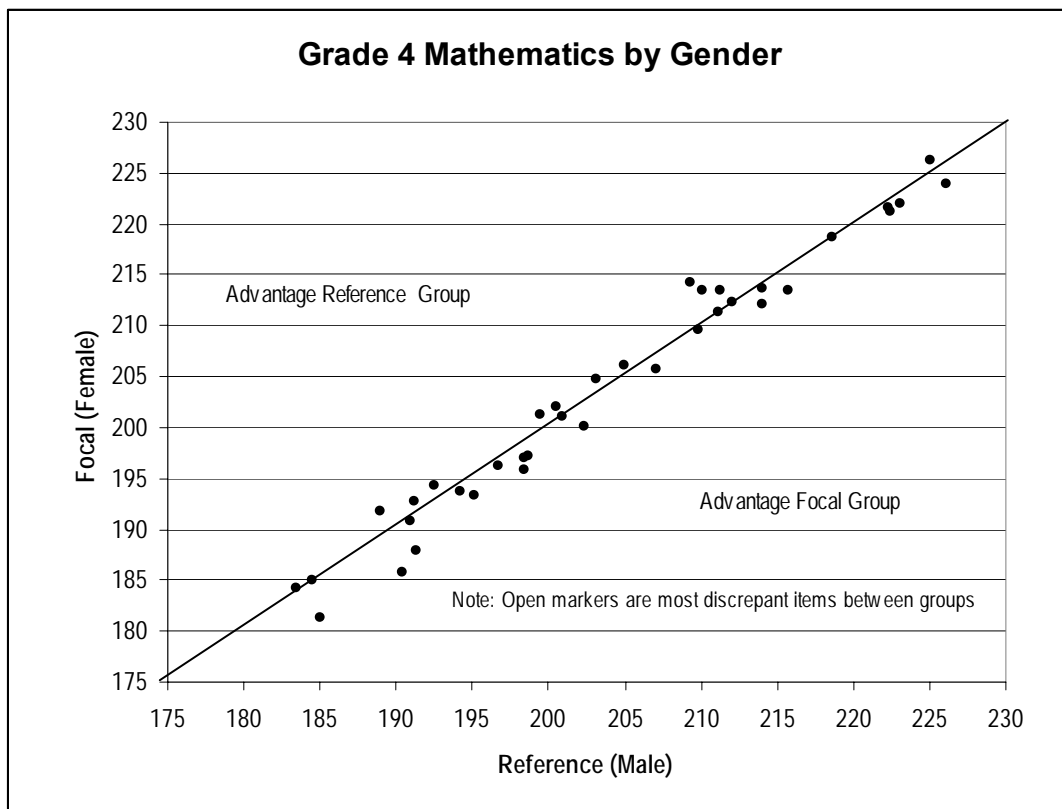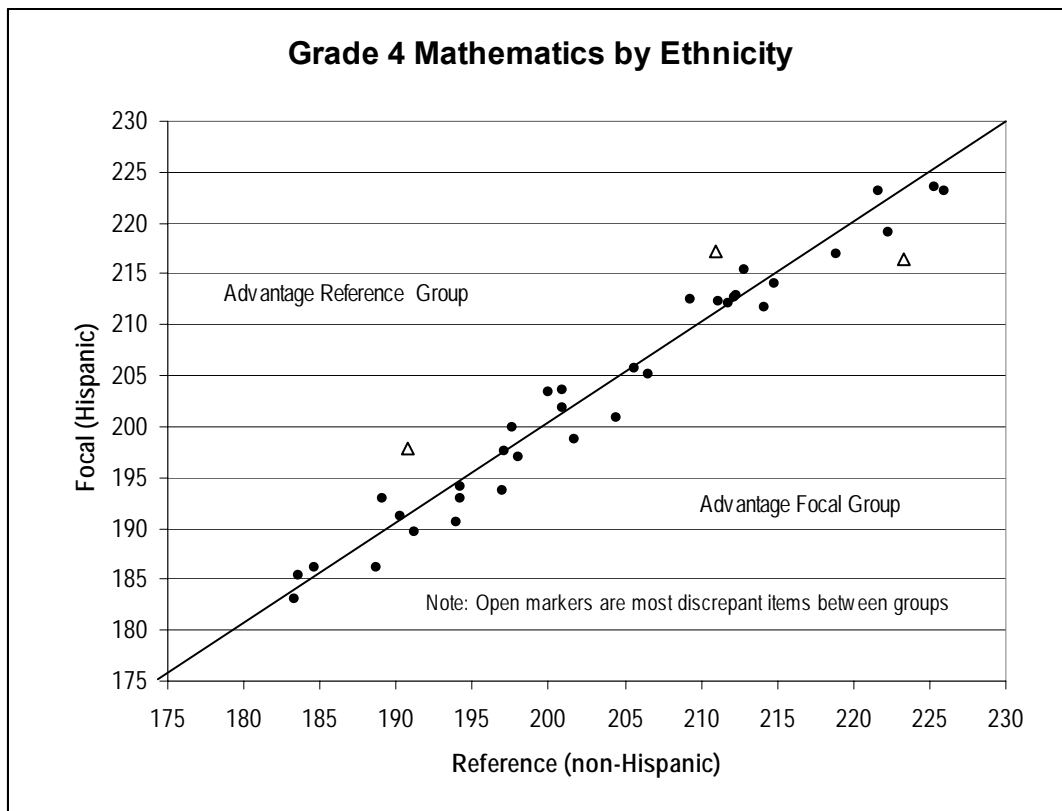
Table 13. Grade 8 Mathematics tests - Actual and virtual

Test	Item difficulties			Standardized item fit		Internal consistency (KR-20)	Proportion correct for 'proficient'
	Mean	sd	n	Mean	sd		
Administered	229.3	11.81	42	0.3	1.0	.89	.57
Studied (original)	229.4	11.78	40	0.3	1.0	.86	.56
Ethnic adjusted	229.4	12.27	40	0.4	0.9	.86	.56
Gender adjusted	229.2	12.10	40	0.4	0.9	.86	.56

Figure 16. Test characteristic curves for original, ethnic adjusted and gender adjusted tests for grade 8 Mathematics.
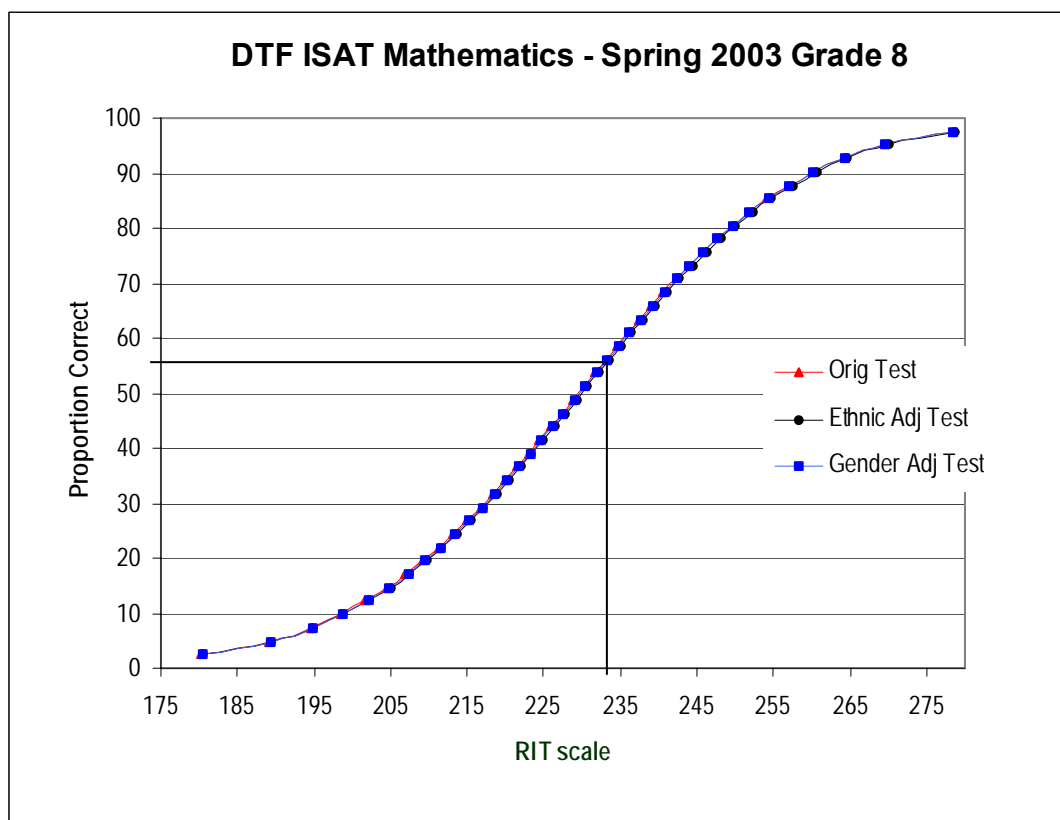
Grade 10. Figure 17 contains the plots of all grade 10 Mathematics item calibration differences by comparison type. The items showing DIF have a somewhat similar pattern in each comparison set. For the ethnic-related DIF items, half the items favored the Focal group and were fairly evenly spread across a 27 RIT range from 258 down, while items favoring the Reference group covered a 17 point range from 242 down. For the gender-related DIF items, those favoring the Focal group difficulties were spread across nearly the entire 50 RIT range of the test item difficulties. Gender-related DIF items favoring the Reference group had difficulties within about a 35 RIT range. The magnitude of calibration differences for gender-related DIF items favoring males tended to be slightly larger than those favoring females. In both comparison types (ethnic and gender), the mean calibration

differences across all items were less than .1 RIT between the Reference and the Focal groups. However, the variance of these differences was close to twice that for the other content areas and grade levels. These differences were large enough for the gender-related virtual test to function slightly differently from the Studied test as well as the ethnic-adjusted virtual test. The required number of items to be answered correctly to reach the proficient level of 242 was 24 for the Studied test and the ethnic-adjusted virtual test but 23 for the gender-adjusted virtual test.

Figure 17. Item calibrations for grade 10 Mathematics by comparison type.

Grade 10 Mathematics by Ethnicity

Advantage Reference Group

Advantage Focal Group

Note: Open markers are most discrepant items between groups

Focal (Hispanic)

Reference (non-Hispanic)

Grade 10 Mathematics by Gender

Advantage Reference Group

Advantage Focal Group

Note: Open markers are most discrepant items between groups
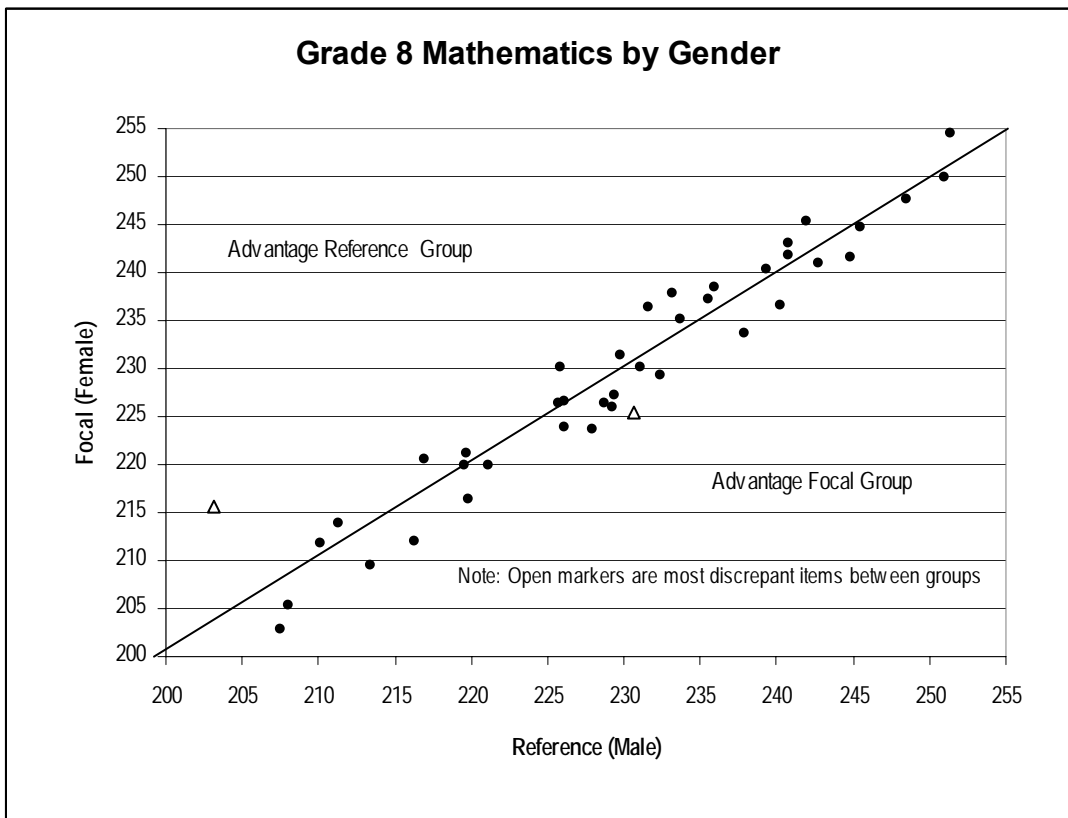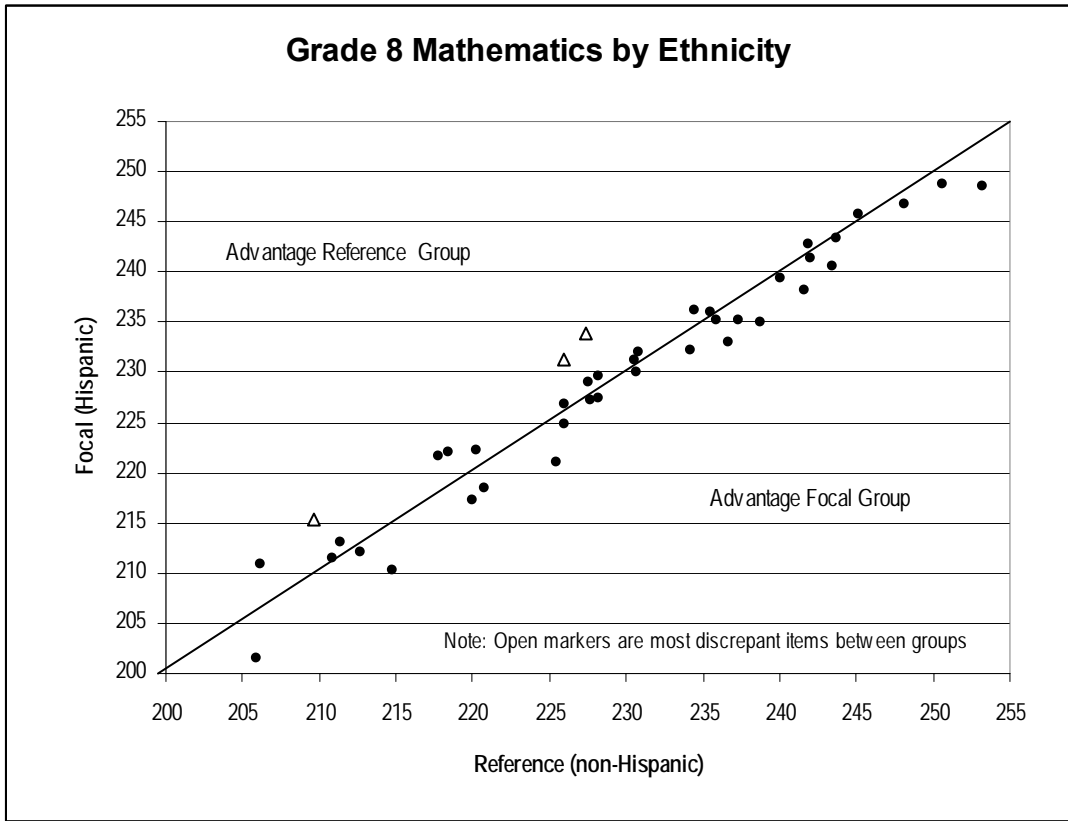
Focal (Female)

Reference (Male)
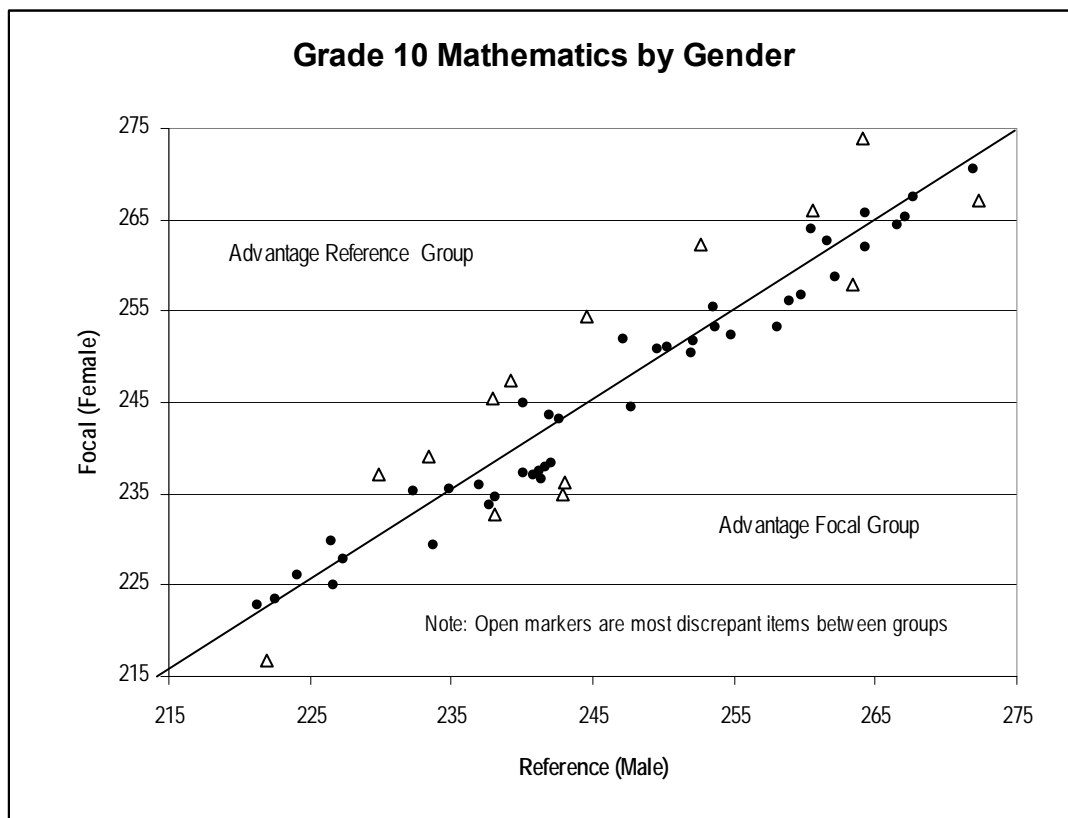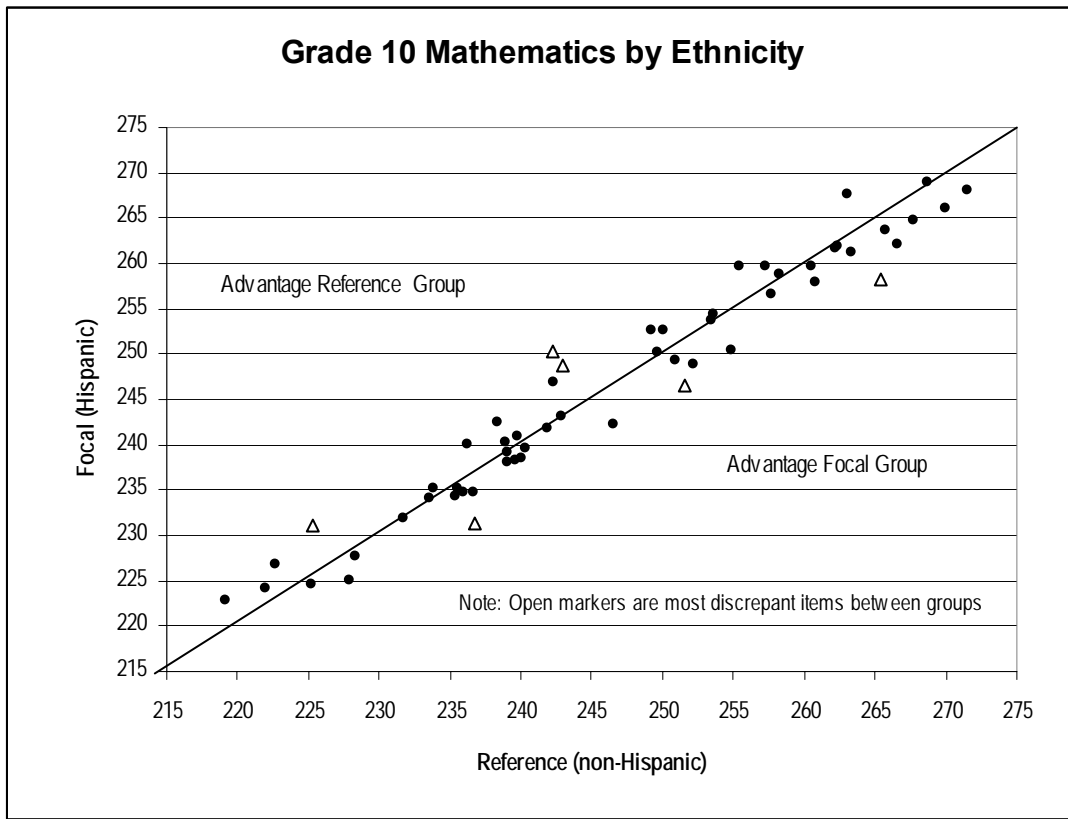
Table 14. Grade 10 Mathematics tests - Actual and virtual

Test	Item difficulties			Standardized item fit		Internal consistency (KR-20)	Proportion correct for 'proficient'
	Mean	sd	n	Mean	sd		
Administered	247.8	12.40	60	0.0	1.0	.90	.40
Studied (original)	245.6	15.64	56	0.3	1.0	.90	.43
Ethnic adjusted	246.8	13.30	52	0.2	1.0	.90	.43
Gender adjusted	245.9	14.90	52	0.5	0.9	.90	.41

Figure 18. Test characteristic curves for original, ethnic adjusted and gender adjusted tests for grade 10 Mathematics.
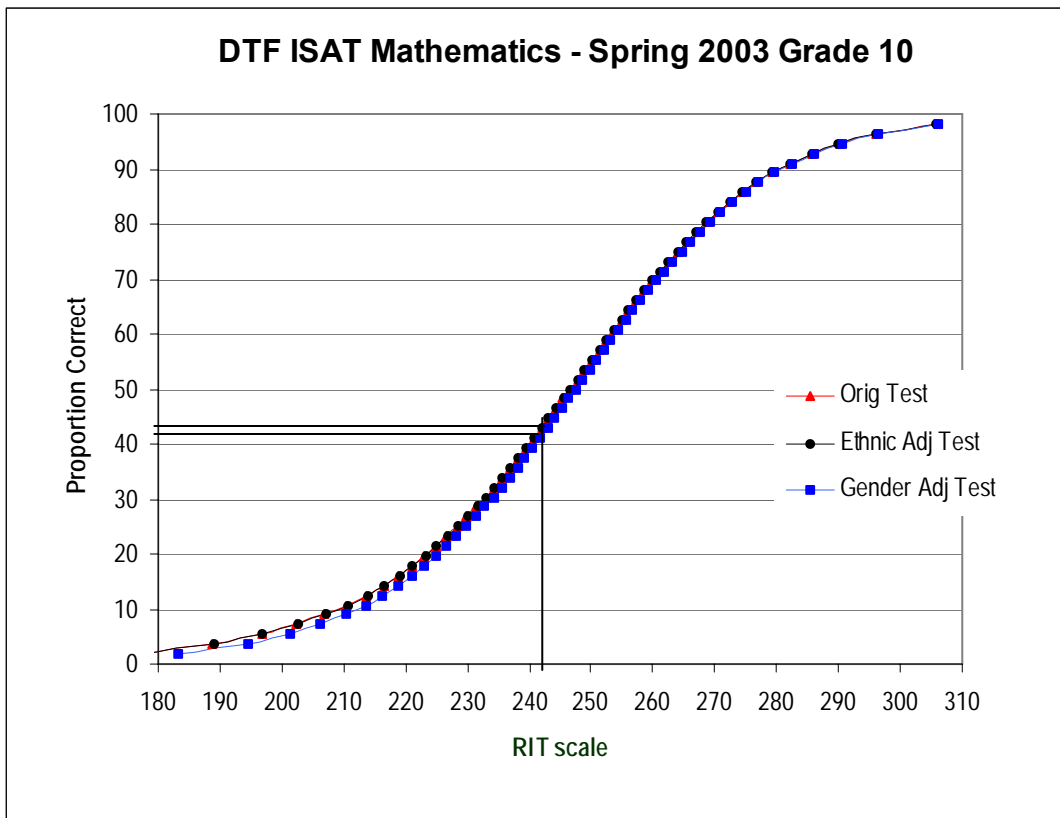
Conclusions and Discussion

The first of two questions guiding this study was, to what extent do test items exhibit substantive differences in functioning across student groups formed based on ethnicity and based on gender? Using the criterion rule of calibration differences $>= | 5 |$ (.5 logits) to define DIF, between zero percent (grades 4-10, gender, Reading, Language and Mathematics) and 25 percent (grade 10, gender, Mathematics) of the items in any one test were observed as exhibiting DIF. Just exactly what these levels of DIF mean or what subsequent action they suggest is open to question. As an initial step, examination of the DIF identified items by content experts to identify possible sources of DIF will be taken. The results of this examination may point to other steps.

Perhaps the most interesting finding in this study was the number of DIF items found in the grade 10 Mathematics test. The number was greater than expected and merits discussion. As Boldt (2000) has pointed out, DIF studies are primarily observational while item writers and test developers are often looking for help in eliminating or at least reducing DIF items in tests. This study is also primarily observational. It did not pursue the root causes of DIF noted in grade 10 Mathematics. We can, however, point to several areas to explore in this regard. One obvious area to explore is *item content* or *characteristics*.

A cursory examination of the grade 10 Mathematics ISAT does not provide many clues. When looking for item content that might introduce bias, only one of the DIF items stood out in any way. The item named the baseball star Willie Stargel with a number of "at bats" and a "number of hits". The question asked what his batting average was. Males found this item much easier. Beyond this item, there were only two other items that males found substantially easier that had discernible common characteristics. Both items involved measurement and both pertained to scale conversion (e.g. 1" equals 20 miles).

Another area that may deserve some attention in terms of explaining the observed levels of DIF is that the test's content was not well matched to the full range of student skills being tested. A logical and plausible argument for this possibility is not difficult to make. In grade 10 students can span a wide range of mathematics ability, from levels consistent with basic mathematics to ability levels consistent with Algebra 2 or even pre-calculus. Even though the test spans the equivalent of four Achievement Level Tests (50 RIT), the test predominantly focused on upper level basic mathematics and pre-Algebra with some Algebra 1 and some basic Geometry. This suggests that the test for high and particularly for low achieving students may not be well matched to the students' most recent instructional content. Recall that nearly 10% of the grade 10 Mathematics test records had scores that were at or below the chance level. This percentage is more than six times the percentage excluded at grade 4 and about 1.7 times the percentage excluded for grade 8. Related to the test content-instructional content match issue, are the potential effects of differences in student course taking.

Some suggestion that differences in course taking may hold some value for explaining DIF in grade 10 Mathematics is revealed by looking at the item content strands where DIF occurred. All three of the Algebra strand items that exhibited DIF, were found to be easier for females. However, all the items that exhibited DIF in measurement (n = 3), data analysis and probability (n = 2) strands were easier for males.

Finally, it may that the criterion established for DIF was not as useful for grade 10 Mathematics as it was for grade 10 Reading and Language Usage given the differences in the natural variation in performance in these domains. Average Reading and Language Usage scores were at the 57th and the 60th percentiles, respectively, whereas the average Mathematics score was at the 45th percentile with about 33% more variability. This suggests that more variability in the estimates of Mathematics item difficulties might be expected. It also suggests that using the same 'substantive' criterion as for Reading and Language Usage to evaluate the magnitude of differences in difficulty estimates might be too severe for practical purposes.

The second question guiding this study was, what is the collective impact on test characteristics of any substantive differentially functioning items? This question has a very clear and direct answer; *viz.*, very little. As shown in Tables 4-6, 8-10, and 12-14, the proportion correct needed to be

considered to be "proficient" on the 'studied (original)' test and the 'virtual' tests were never more than two percentage points apart and were typically equal. For any particular comparison this translates, at most, to a difference of a single item being answered correctly. Thus, even though the virtual tests were designed to fully favor the Focal groups, the balance in the direction of differences in calibrations counteracted this intended bias. This supports the contention that the overall level of differential item functioning within any particular test was virtually nil with respect to its impact on the larger issue of differential test functioning.

References

Boldt, D. M. (2000). A SIBTEST approach to testing DIF hypotheses using experimentally designed test items. *Journal of Educational Measurement, 37,* (4), 307-327.

Draba, R. E. (1977). *The identification and interpretation of item bias.* (Research Memorandum No. 25). Chicago: University of Chicago, Statistical Laboratory, Department of Education.

Linacre, J. M. (2003). Winsteps-Ministep: Rasch Model Computer Programs, Version 3.47. www.winsteps.com.

Linacre, J. M. & Wright, B. D. (1989). Mantel-Haenszel DIF and PROX are Equivalent. *Rasch Measurement Transactions, 1989, 3,* (2), 52-53.

Luppescu, S. (1993). DIF detection examined. *Rasch Measurement Transactions, 1993, 7,* (2), 285-286.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719-748.

Schulz, E. M., Perlman, C., Rice, W. K., Wright, B. D. (1996). An empirical comparison Rasch and Mantel-Haenszel procedures for assessing differential item functioning. In G. Englehard & M. Wilson (Eds.), *Objective measurement: Theory into practice, Volume 3* (pp. 65-81). Norwood, NJ: Ablex.

Appendix

ISAT Core Items for the Spring 2003 Tests

The tables included in this appendix provide references to the test items making up the spring 2003 ISAT. Entries in the columns titled "Seq" in these tables correspond entries in columns titled "Seq Number" in Tables 2, 3, 7, and 11 in the same content area-grade level combination.

ISAT Grade 4 Core Reading Items - Spring 2003

Seq	Item ID	RIT	Goal ID	Goal Area
1	842290	183	4	Interpretive Comprehension
2	53601	195	3	Literal Comprehension
3	281050	199	2	Vocabulary
4	59400	202	3	Literal Comprehension
5	293250	209	4	Interpretive Comprehension
6	800150	211	3	Literal Comprehension
7	281441	215	2	Vocabulary
8	10020330	216	3	Literal Comprehension
9	10024250	183	2	Vocabulary
10	868990	219	6	Literacy Analysis
11	77601	182	5	Evaluative Comprehension
12	870370	224	5	Evaluative Comprehension
13	10028860	185	1	Word Analysis
14	10029360	181	4	Interpretive Comprehension
15	10029110	188	3	Literal Comprehension
16	10031470	188	6	Literacy Analysis
17	292870	188	5	Evaluative Comprehension
18	10024270	189	1	Word Analysis
19	10028990	191	2	Vocabulary
20	10030400	195	1	Word Analysis
21	282830	193	4	Interpretive Comprehension
22	10030750	195	3	Literal Comprehension
23	10031450	196	6	Literacy Analysis
24	10031070	197	6	Literacy Analysis
25	10031060	197	4	Interpretive Comprehension
26	10028880	198	1	Word Analysis
27	44000	199	5	Evaluative Comprehension
28	24600	201	2	Vocabulary
29	282770	200	5	Evaluative Comprehension
30	281270	202	5	Evaluative Comprehension
31	10028900	202	5	Evaluative Comprehension
32	811160	202	4	Interpretive Comprehension
33	10024740	204	6	Literacy Analysis
34	10020240	206	1	Word Analysis
35	815980	213	1	Word Analysis
36	801490	207	1	Word Analysis
37	10027050	207	2	Vocabulary
38	404470	208	6	Literacy Analysis
39	10031030	210	3	Literal Comprehension
40	10029390	214	4	Interpretive Comprehension
41	818390	215	6	Literacy Analysis
42	298930	217	2	Vocabulary

ISAT Grade 8 Core Reading Items - Spring 2003

Seq	Item ID	RIT	Goal ID	Goal Area
1	22401	217	1	Word Analysis
2	26600	207	1	Word Analysis
3	79801	201	3	Literal Comprehension
4	85000	202	5	Evaluative Comprehension
5	94400	203	5	Evaluative Comprehension
6	282210	220	4	Interpretive Comprehension
7	282790	226	6	Literacy Analysis
8	282940	221	3	Literal Comprehension
9	292190	220	5	Evaluative Comprehension
10	292231	227	4	Interpretive Comprehension
11	293260	217	4	Interpretive Comprehension
12	293410	243	2	Vocabulary
13	294710	223	3	Literal Comprehension
14	410400	226	2	Vocabulary
15	808560	215	5	Evaluative Comprehension
16	811290	222	4	Interpretive Comprehension
17	816080	222	2	Vocabulary
18	816490	229	3	Literal Comprehension
19	820740	225	1	Word Analysis
20	828010	216	1	Word Analysis
21	831950	224	5	Evaluative Comprehension
22	836980	235	5	Evaluative Comprehension
23	841670	225	5	Evaluative Comprehension
24	10018970	213	6	Literacy Analysis
25	10020250	212	4	Interpretive Comprehension
26	10020360	210	4	Interpretive Comprehension
27	10020390	229	3	Literal Comprehension
28	10020400	223	3	Literal Comprehension
29	10020520	206	4	Interpretive Comprehension
30	10020630	219	3	Literal Comprehension
31	10024290	211	2	Vocabulary
32	10025370	235	1	Word Analysis
33	10025380	236	6	Literacy Analysis
34	10026550	206	6	Literacy Analysis
35	10026730	200	1	Word Analysis
36	10027210	200	2	Vocabulary
37	10027490	217	6	Literacy Analysis
38	10028440	208	6	Literacy Analysis
39	10028530	212	2	Vocabulary
40	10030320	221	1	Word Analysis
41	10031140	236	6	Literacy Analysis
42	10032420	217	2	Vocabulary

ISAT Grade 10 Core Reading Items - Spring 2003

Seq	Item ID	RIT	Goal ID	Goal Area
1	243640	192	4	Interpretive Comprehension
2	10027140	213	1	Word Analysis
3	10020210	213	3	Literal Comprehension
4	37800	214	1	Word Analysis
5	10020490	224	4	Interpretive Comprehension
6	10021680	214	3	Literal Comprehension
7	10021670	225	3	Literal Comprehension
8	293360	215	5	Evaluative Comprehension
9	630190	185	2	Vocabulary
10	10030660	217	6	Literacy Analysis
11	292170	217	4	Interpretive Comprehension
12	10026750	217	2	Vocabulary
13	10011450	218	6	Literacy Analysis
14	62400	216	3	Literal Comprehension
15	852550	220	4	Interpretive Comprehension
16	852540	229	4	Interpretive Comprehension
17	10018940	218	5	Evaluative Comprehension
18	850740	220	6	Literacy Analysis
19	815830	221	1	Word Analysis
20	282090	225	4	Interpretive Comprehension
21	282100	227	4	Interpretive Comprehension
22	294600	218	5	Evaluative Comprehension
23	10020460	227	5	Evaluative Comprehension
24	823790	219	2	Vocabulary
25	41000	221	1	Word Analysis
26	411180	230	3	Literal Comprehension
27	10020510	230	3	Literal Comprehension
28	300760	222	2	Vocabulary
29	290470	231	2	Vocabulary
30	291040	220	5	Evaluative Comprehension
31	10032460	232	6	Literacy Analysis
32	852660	237	4	Interpretive Comprehension
33	852650	239	4	Interpretive Comprehension
34	10020480	233	6	Literacy Analysis
35	801140	222	1	Word Analysis
36	10021550	242	1	Word Analysis
37	300730	222	1	Word Analysis
38	10011090	234	3	Literal Comprehension
39	10011120	234	3	Literal Comprehension
40	806121	223	2	Vocabulary
41	804880	221	3	Literal Comprehension
42	291180	224	1	Word Analysis
43	817470	224	4	Interpretive Comprehension
44	290280	224	2	Vocabulary
45	816160	228	1	Word Analysis
46	10031310	229	6	Literacy Analysis
47	10028650	230	5	Evaluative Comprehension
48	10025420	238	6	Literacy Analysis
49	814600	232	5	Evaluative Comprehension
50	804610	232	5	Evaluative Comprehension
51	804600	232	5	Evaluative Comprehension
52	298370	235	2	Vocabulary
53	10024830	237	6	Literacy Analysis
54	10031150	237	6	Literacy Analysis
55	10025400	240	2	Vocabulary

ISAT Grade 4 Core Language Usage Items - Spring 2003

Seq	Item ID	RIT	Goal ID	Goal Area
1	10029620	191	2	Composition and Structure
2	233100	205	6	Conventions-Spelling
3	10030500	195	5	Conventions-Capitalization
4	10029500	185	4	Conventions-Punctuation
5	10025920	185	3	Grammar and Usage
6	10029040	187	3	Grammar and Usage
7	10021900	194	6	Conventions-Spelling
8	10029010	187	4	Conventions-Punctuation
9	630190	185	3	Grammar and Usage
10	652640	194	1	Composition and the Writing Process
11	260420	188	5	Conventions-Capitalizaiton
12	10030280	188	1	Composition and the Writing Process
13	231220	190	1	Composition and the Writing Process
14	231140	190	6	Conventions-Spelling
15	10026130	201	2	Composition and Structure
16	10029150	193	6	Conventions-Spelling
17	10029490	194	5	Conventions-Capitalization
18	10022400	196	4	Conventions-Punctuation
19	10029000	194	3	Grammar and Usage
20	10029750	194	1	Composition and the Writing Process
21	10026250	203	4	Conventions-Punctuation
22	10026080	199	2	Composition and Structure
23	10022430	204	4	Conventions-Punctuation
24	231330	201	5	Conventions-Capitalization
25	10021800	204	1	Composition and the Writing Process
26	656390	206	3	Grammar and Usage
27	10029990	207	2	Composition and Structure
28	10024970	207	1	Composition and the Writing Process
29	608360	207	2	Composition and Structure
30	235290	207	2	Composition and Structure
31	232780	209	5	Conventions-Capitalization
32	10022170	209	5	Conventions-Capitalization
33	232970	211	6	Conventions-Spelling
34	10030010	213	4	Conventions-Punctuation
35	235630	213	6	Conventions-Spelling
36	10021810	215	1	Composition and the Writing Process
37	600110	214	5	Conventions-Capitalization
38	10025160	215	3	Grammar and Usage
39	556540	214	2	Composition and Structure
40	10022340	215	4	Conventions-Punctuation
41	236040	216	3	Grammar and Usage
42	236700	218	6	Conventions-Spelling

ISAT Grade 8 Core Language Usage Items - Spring 2003

Seq	Item ID	RIT	Goal ID	Goal Area
1	243640	192	5	Conventions-Capitalization
2	10029760	200	5	Conventions-Capitalization
3	10024160	201	6	Conventions-Spelling
4	10026120	201	2	Composition and Structure
5	10030090	201	1	Composition and the Writing Process
6	10028550	203	6	Conventions-Spelling
7	10021990	203	4	Conventions-Punctuation
8	251640	204	1	Composition and the Writing Process
9	10026300	206	3	Grammar and Usage
10	10029980	207	2	Composition and Structure
11	10022120	209	2	Composition and Structure
12	612201	209	3	Grammar and Usage
13	620770	210	5	Conventions-Capitalization
14	10025180	211	6	Conventions-Spelling
15	10021830	211	2	Composition and Structure
16	10030340	212	2	Composition and Structure
17	10024860	213	1	Composition and the Writing Process
18	625820	213	4	Conventions-Punctuation
19	625650	219	1	Composition and the Writing Process
20	625590	214	6	Conventions-Spelling
21	10020560	215	1	Composition and the Writing Process
22	232181	215	4	Conventions-Punctuation
23	625610	216	6	Conventions-Spelling
24	628370	219	5	Conventions-Capitalization
25	600320	216	5	Conventions-Capitalization
26	600350	219	4	Conventions-Punctuation
27	235770	223	3	Grammar and Usage
28	633040	224	4	Conventions-Punctuation
29	10025170	224	6	Conventions-Spelling
30	615550	224	3	Grammar and Usage
31	240820	224	6	Conventions-Spelling
32	610660	225	5	Conventions-Capitalization
33	10022480	226	2	Composition and Structure
34	10022160	228	4	Conventions-Punctuation
35	10028910	228	1	Composition and the Writing Process
36	250250	228	3	Grammar and Usage
37	637320	229	5	Conventions-Capitalization
38	10024930	230	2	Composition and Structure
39	10022530	231	4	Conventions-Punctuation
40	618200	232	3	Grammar and Usage
41	10031120	238	1	Composition and the Writing Process
42	236860	239	3	Grammar and Usage

ISAT Grade 10 Core Language Usage Items - Spring 2003

Seq	Item ID	RIT	Goal ID	Goal Area
1	243640	192	1	Composition and the Writing Process
2	251220	205	5	Convention-Capitalization
3	612861	205	3	Grammar and Usage
4	602901	206	6	Conventions-Spelling
5	232390	207	5	Convention-Capitalization
6	10024900	207	3	Grammar and Usage
7	637570	228	1	Composition and the Writing Process
8	10025010	210	6	Conventions-Spelling
9	630190	185	1	Composition and the Writing Process
10	602811	210	1	Composition and the Writing Process
11	260830	211	5	Convention-Capitalization
12	10022240	214	6	Conventions-Spelling
13	609300	212	4	Conventions-Punctuation
14	10030470	213	2	Composition and Structure
15	252290	213	2	Composition and Structure
16	10022000	214	4	Conventions-Punctuation
17	10027510	214	1	Composition and the Writing Process
18	236660	216	6	Conventions-Spelling
19	232270	216	5	Convention-Capitalization
20	230220	216	4	Conventions-Punctuation
21	10027400	217	1	Composition and the Writing Process
22	230080	218	3	Grammar and Usage
23	251710	218	5	Convention-Capitalization
24	10025060	218	2	Composition and Structure
25	611070	220	6	Conventions-Spelling
26	10031130	221	1	Composition and the Writing Process
27	238540	222	4	Conventions-Punctuation
28	10022230	223	5	Convention-Capitalization
29	245730	223	2	Composition and Structure
30	243720	224	2	Composition and Structure
31	10019050	224	6	Conventions-Spelling
32	621770	225	4	Conventions-Punctuation
33	10022200	226	3	Grammar and Usage
34	10022290	225	4	Conventions-Punctuation
35	10018890	225	3	Grammar and Usage
36	10030170	226	6	Conventions-Spelling
37	10022460	226	3	Grammar and Usage
38	610561	226	1	Composition and the Writing Process
39	10030150	227	6	Conventions-Spelling
40	614440	227	1	Composition and the Writing Process
41	553220	227	2	Composition and Structure
42	620890	228	5	Convention-Capitalization
43	622280	228	2	Composition and Structure
44	232040	229	4	Conventions-Punctuation
45	643750	230	4	Conventions-Punctuation
46	10025110	231	3	Grammar and Usage
47	618580	231	3	Grammar and Usage
48	630970	231	4	Conventions-Punctuation
49	235780	231	3	Grammar and Usage
50	637040	233	5	Convention-Capitalization
51	619431	233	3	Grammar and Usage
52	618540	233	6	Conventions-Spelling
53	253160	235	5	Convention-Capitalization
54	10024920	236	1	Composition and the Writing Process
55	653020	237	2	Composition and Structure
56	10031100	240	2	Composition and Structure

ISAT Grade 4 Core Mathematics Items - Spring 2003

Seq	Item ID	RIT	Goal ID	Goal Area
1	315220	183	6	Geometry
2	10022970	183	1	Number Sense
3	15400	185	2	Estimation & Computations
4	271400	188	7	Data Analysis, Probability & Stats
5	276050	188	5	Algebra, Functions & Mathematical Models
6	213440	189	4	Measurement
7	303250	192	4	Measurement
8	272780	192	7	Data Analysis, Probability & Stats
9	340900	195	2	Estimation & Computations
10	10022900	192	1	Number Sense
11	302140	195	5	Algebra, Functions & Mathematical Models
12	40730	195	3	Mathematical Reasoning & Problem Solving
13	331260	226	6	Geometry
14	33999	197	1	Number Sense
15	35600	197	3	Mathematical Reasoning & Problem Solving
16	10023340	198	5	Algebra, Functions & Mathematical Models
17	212200	200	4	Measurement
18	202560	200	3	Mathematical Reasoning & Problem Solving
19	10022790	200	1	Number Sense
20	203800	201	6	Geometry
21	187200	203	7	Data Analysis, Probability & Stats
22	218800	204	4	Measurement
23	188600	204	7	Data Analysis, Probability & Stats
24	9430	205	2	Estimation & Computations
25	347071	206	6	Geometry
26	340	207	5	Algebra, Functions & Mathematical Models
27	300760	207	1	Number Sense
28	325680	210	2	Estimation & Computations
29	204280	208	3	Mathematical Reasoning & Problem Solving
30	277440	208	6	Geometry
31	303970	209	7	Data Analysis, Probability & Stats
32	278560	210	3	Mathematical Reasoning & Problem Solving
33	333650	210	2	Estimation & Computations
34	202410	212	4	Measurement
35	317600	219	7	Data Analysis, Probability & Stats
36	315540	215	6	Geometry
37	203250	216	5	Algebra, Functions & Mathematical Models
38	10023490	222	1	Number Sense
39	343510	224	3	Mathematical Reasoning & Problem Solving
40	210590	224	2	Estimation & Computations
41	308410	224	5	Algebra, Functions & Mathematical Models
42	312420	227	4	Measurement

ISAT Grade 8 Core Mathematics Items - Spring 2003

Seq	Item ID	RIT	Goal ID	Goal Area
1	10026450	210	6	Geometry
2	10023250	212	7	Data Analysis, Probability & Stats
3	10023160	213	1	Number Sense
4	10025260	215	6	Geometry
5	10020680	216	5	Algebra, Functions, & Mathematical Models
6	10032230	218	5	Algebra, Functions, & Mathematical Models
7	10019760	221	7	Data Analysis, Probability & Stats
8	10020050	222	5	Algebra, Functions, & Mathematical Models
9	10027690	225	7	Data Analysis, Probability & Stats
10	4910	226	3	Mathematical Reasoning & Problem Solving
11	770	226	4	Measurement
12	10027590	230	5	Algebra, Functions, & Mathematical Models
13	334590	230	6	Geometry
14	10020120	231	5	Algebra, Functions, & Mathematical Models
15	277780	241	6	Geometry
16	206190	233	4	Measurement
17	105630	235	1	Number Sense
18	278200	236	4	Measurement
19	312850	238	3	Mathematical Reasoning & Problem Solving
20	327750	238	7	Data Analysis, Probability & Stats
21	205620	239	4	Measurement
22	10025230	245	5	Algebra, Functions, & Mathematical Models
23	10019710	248	7	Data Analysis, Probability & Stats
24	10022650	209	1	Number Sense
25	10023030	210	2	Estimation & Computation
26	4570	216	3	Mathematical Reasoning & Problem Solving
27	202620	217	4	Measurement
28	16590	218	2	Estimation & Computation
29	14950	224	6	Geometry
30	16790	224	1	Number Sense
31	16780	239	2	Estimation & Computation
32	1020	225	2	Estimation & Computation
33	211570	229	3	Mathematical Reasoning & Problem Solving
34	355580	230	3	Mathematical Reasoning & Problem Solving
35	353360	231	1	Number Sense
36	10019300	242	1	Number Sense
37	9070	243	6	Geometry
38	211440	244	3	Mathematical Reasoning & Problem Solving
39	156600	244	2	Estimation & Computation
40	306970	245	2	Estimation & Computation
41	303170	246	1	Number Sense
42	313870	247	3	Mathematical Reasoning & Problem Solving

ISAT Grade 10 Core Mathematics Items - Spring 2003

Seq	Item ID	RIT	Goal ID	Goal Area
1	10024620	225	3	Measurements
2	105210	225	4	Algebra
3	205670	225	6	Data Analysis, Probability, & Stats
4	104440	227	5	Geometry
5	348270	228	1	Estimation & Computation
6	205640	230	6	Data Analysis, Probability, & Stats
7	317160	256	3	Measurements
8	195800	232	6	Data Analysis, Probability, & Stats
9	630190	185	7	Functions & Mathematical Models
10	161500	235	3	Measurements
11	328010	236	5	Geometry
12	213220	236	7	Functions & Mathematical Models
13	212600	238	3	Measurements
14	10001210	240	7	Functions & Mathematical Models
15	211590	242	2	Mathematical Reasoning & Problem Solving
16	349590	244	6	Data Analysis, Probability, & Stats
17	342940	245	7	Functions & Mathematical Models
18	105550	245	4	Algebra
19	101310	246	4	Algebra
20	211780	247	3	Measurements
21	10013870	247	7	Functions & Mathematical Models
22	10000300	248	5	Geometry
23	13200	248	2	Mathematical Reasoning & Problem Solving
24	10002680	248	6	Data Analysis, Probability, & Stats
25	205260	248	5	Geometry
26	100220	248	4	Algebra
27	101960	249	5	Geometry
28	105021	250	7	Functions & Mathematical Models
29	206020	250	6	Data Analysis, Probability, & Stats
30	100430	250	4	Algebra
31	100080	251	4	Algebra
32	105640	252	1	Estimation & Computation
33	100130	253	1	Estimation & Computation
34	10012270	253	7	Functions & Mathematical Models
35	101770	255	5	Geometry
36	104690	255	5	Geometry
37	101060	255	2	Mathematical Reasoning & Problem Solving
38	101160	255	7	Functions & Mathematical Models
39	100530	256	4	Algebra
40	105380	256	7	Functions & Mathematical Models
41	10002310	260	5	Geometry
42	10002490	260	2	Mathematical Reasoning & Problem Solving
43	350290	261	2	Mathematical Reasoning & Problem Solving
44	100270	262	4	Algebra
45	10011810	263	5	Geometry
46	211860	264	4	Algebra
47	101490	265	2	Mathematical Reasoning & Problem Solving
48	100600	266	1	Estimation & Computation
49	100580	268	4	Algebra
50	278710	270	3	Measurements
51	10032250	229	1	Estimation & Computation
52	205450	230	1	Estimation & Computation
53	205020	240	1	Estimation & Computation
54	341641	236	2	Mathematical Reasoning & Problem Solving
55	10019820	239	3	Measurements
56	313420	244	1	Estimation & Computation
57	309580	252	1	Estimation & Computation
58	154600	264	1	Estimation & Computation
59	328110	257	6	Data Analysis, Probability, & Stats
60	140000	275	1	Estimation & Computation