

9. SUMMARY OF OPERATIONAL TEST RESULTS OF THE 2010 MOD-MSA: READING

This section presents both the raw score and scaled score summaries for the Mod-MSA: Reading by grade. Table 9.1.1 presents the raw score summary by grade. Table 9.2.1 presents the scale score summary by grade. Table 9.3.1 presents the percentage of students in each of the proficiency levels by grade. In addition, Appendix C provides frequency distributions and histograms of the scale scores of the 2010 Mod-MSA: Reading. For grades 6-8, the 2009 results are also depicted in the above mentioned tables.

9.1 Classical Descriptive Test Statistics

Table 9.1.1 contains the classical descriptive statistics of each form for each grade and includes:

- Numbers of students (based on a whole population)
- Numbers of items
- Minimum and maximum points achievable on the test
- Means and standard deviations of raw scores
- Test reliability (KR_{20})
- Standard error of measurement (SEM)

Table 9.1.1 Classical Descriptive Statistics for the 2010 Mod-MSA: Reading: Grades 3 through 8

Grade	Total # of Items	Min. Point	Max. Point	2009					2010				
				N	Mean	SD	Reliability	SEM	N	Mean	SD	Reliability	SEM
3	45	0	45	N/A	N/A	N/A	N/A	N/A	1058	17.53	6.45	0.82	3.08
4	45	0	45	N/A	N/A	N/A	N/A	N/A	1335	24.83	7.04	0.81	3.06
5	45	0	45	N/A	N/A	N/A	N/A	N/A	1459	24.68	6.65	0.79	3.04
6	45	0	45	1274	24.13	6.34	0.75	3.16	1356	25.22	6.92	0.81	3.03
7	45	0	45	1630	24.43	6.66	0.78	3.09	1702	25.97	6.54	0.79	3.02
8	45	0	45	1792	26.68	6.71	0.79	3.06	1969	27.23	6.89	0.81	3.04

Note. Analyses were conducted with the statewide population after applying equating exclusion criteria.

9.2 Scale Score Descriptive Statistics

Table 8.2.1 provides information about scale score descriptive statistics of the test by grade and years and includes:

- Numbers of students
- Mean and standard deviation of scale scores
- 10% quantile (P10), 25% quantile (Q1), median (P50), 75% quantile (Q3), 90% quantile, and IQR (Interquantile Range=Q3-Q1)

Table 9.2.1. Scale Score Descriptive Statistics Across Years for Mod-MSA, Reading: Grades 3-8

Grade	Year	<i>N</i>	<i>M</i>	<i>SD</i>	<i>P10</i>	<i>Q1</i>	<i>Mdn</i>	<i>Q3</i>	<i>P90</i>	<i>IQR</i>
3	2010	1058	50.69	12.16	37	42	50	58	66	16
4	2010	1335	50.73	11.84	37	43	49	58	67	15
5	2010	1459	50.65	11.44	37	43	51	57	66	14
6	2009	1274	49.81	12.02	35	41	49	58	66	17
	2010	1356	52.14	13.69	35	42	51	60	69	18
7	2009	1630	49.75	12.38	35	42	49	57	65	15
	2010	1702	50.48	12.27	35	42	50	59	67	17
8	2009	1792	50.04	12.17	35	42	50	59	65	17
	2010	1969	51.68	12.57	36	43	52	59	68	16

Note: 1. Grades 3-5 have no history since 2010 was the first year of their administration.
 2. Analyses were conducted with the statewide population after applying equating exclusion criteria.

9.3 Frequency of Students at Each Proficiency Level by Grade

Table 9.3.1 contains the pass rate at each performance level based on the cut scores shown in Table 4.7.1.

Table 9.3.1. Percentage of Students in Each Proficiency Level by Grade

Grade	<i>N</i>	Percentage of Performance Level		
		Basic	Proficient	Advanced
3	1058	62.19	25.43	12.38
4	1335	60.82	26.67	12.51
5	1459	61.34	32.69	5.96
6	1356	56.27	28.32	15.41
7	1702	68.80	26.26	4.94
8	1969	54.75	33.21	12.04

Note. 1. Percentages may not add up to 100% due to rounding.
 2. Analyses were conducted with the statewide population after applying equating exclusion criteria.

9.4. Test Characteristic Curves, Test Information Functions, and Conditional Standard Errors of Measurement

Test Characteristic Curves

In IRT models, an item characteristic curve (ICC) permits us to see how the probability of answering correctly depends on the latent trait, i.e., the proficiency of the students. The most common shape of the ICC in practice is the S-shaped curve, which increases monotonically from left to right with the lower asymptote approaching 0 and the upper asymptote approaching 1. Since the logistic model ICC is the probabilistic curve for the item, the score on the test can be presented by the summation of the probabilistic scores of each item plus the error of measurement, i.e.:

$$X = \sum_g P_g(\theta) + E,$$

where g is the item number and E is the standard error of measurement. The regression formula shown above for predicting X from θ scores is known as the test characteristic curve (TCC). The TCC for each grade 3 through 8 are provided in Figure 8.4.1a, 8.4.2a, 8.4.3a, 8.4.4a, 8.4.5a, and 8.4.6a, respectively. For grades 6 to 8, a comparison is made across years in these figures.

Test Information Functions

On a standardized achievement test, items could be too hard for the low- proficiency examinee. Similarly some items may be too easy for everyone and may not help in providing any discrimination for these students. These types of items provide little or no information at the cut scores where it really counts. In most testing situations it becomes necessary for us to understand the information provided by each item across the spectrum of different proficiency levels. Mathematically,

$$I_g(\theta) = \frac{[P'_g(\theta)]^2}{[P_g(\theta)][Q_g(\theta)]}, \text{ where}$$

g is the number of the item, $Q_g(\theta) = 1 - P_g(\theta)$, and the numerator is the first derivative of $P_g(\theta)$, which for the Rasch model $= P_g(\theta) Q_g(\theta)$.

The test information function is the sum of the item information functions for all items on the test and is useful in examining the total information provided by the test across the proficiency levels. Symbolically, the test information function at a particular proficiency level can be depicted as:

$$I(\theta) = \sum_g I_g(\theta)$$

The test information functions for each grade are provided in Figures 8.4.1b, 8.4.2b, 4.3b, 8.4.4b, 8.4.5b, and 8.4.6b. For grades 6 to 8, a comparison is made across years in these figures.

Conditional Standard Errors of Measurement (CSEM) of the Tests

The conditional standard error of measurement is the inverse of the information function. Under the Rasch (i.e., 1-PL IRT) model, the CSEM for each person is as follows (Andrich & Luo, 2004):

$$\sigma_{\hat{\beta}} = \frac{1}{\sqrt{\sum_{i=1}^L p_{vi}(1-p_{vi})}}$$

where

v = subscript for a person,

i = subscript for an item,

L = length of the test,

$\hat{\beta}$ = proficiency estimate, and

p_{vi} = the probability that a person answers an item correctly and defined as follows:

$$p_{vi} = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}} \text{ where } \beta_v \text{ is person's proficiency and } \delta_i \text{ is item's difficulty.}$$

A confidence band can be found for use in interpreting the proficiency estimate. For example, an approximate 68% confidence interval for $\hat{\beta}$ is given by

$$\hat{\beta} \pm SEM$$

For the item standard error, the above equation would be modified so that each item's difficulty estimate will be the summation over the different abilities in the test. Note that the standard error for item difficulty is smallest when the probability of passing is close to the probability of failing. That is, the standard error is small when an item is near the threshold level for many persons in the sample (Embretson & Reise, 2000). These stats are provided for each grade in Figures 8.4.1c, 8.4.2c, 8.4.3c, 8.4.4c, 8.4.5c, and 8.4.6c respectively. For grades 6 to 8, a comparison is made across years in these figures.

Grade 3

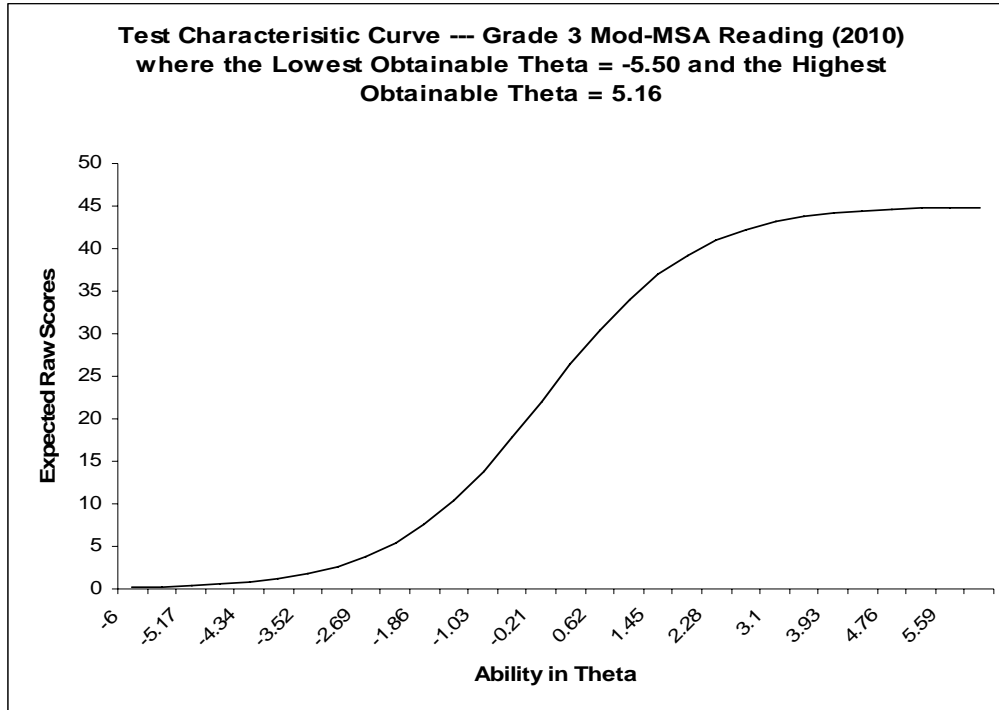


Figure 9.4.1a. TCC for Grade 3 Mod-MSA: Reading

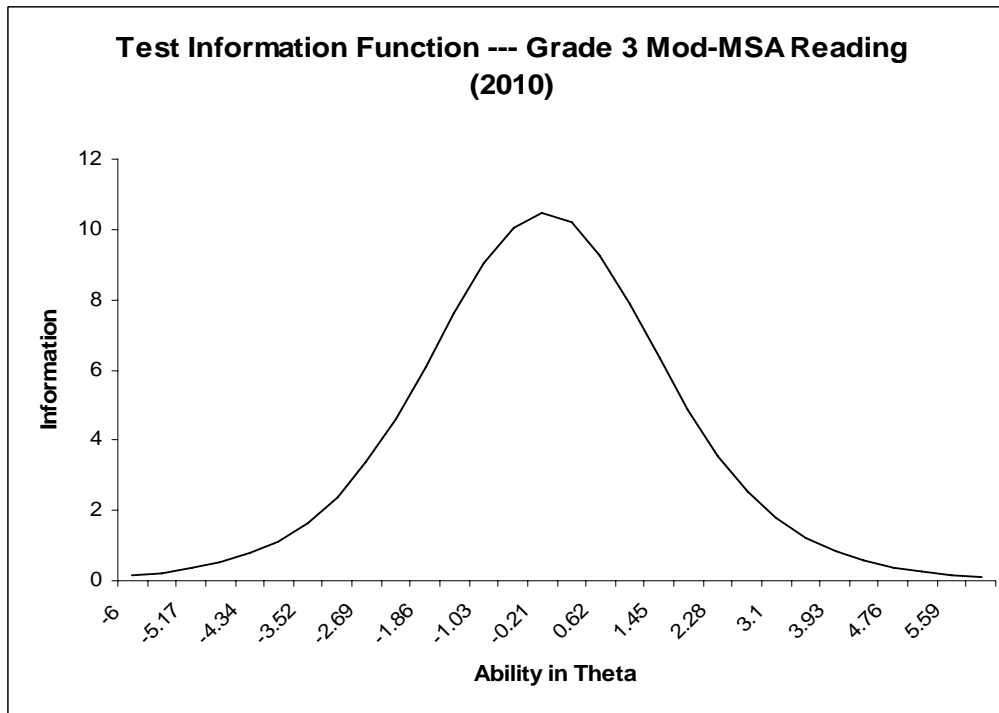


Figure 9.4.1b. TIF for Grade 3 Mod-MSA: Reading

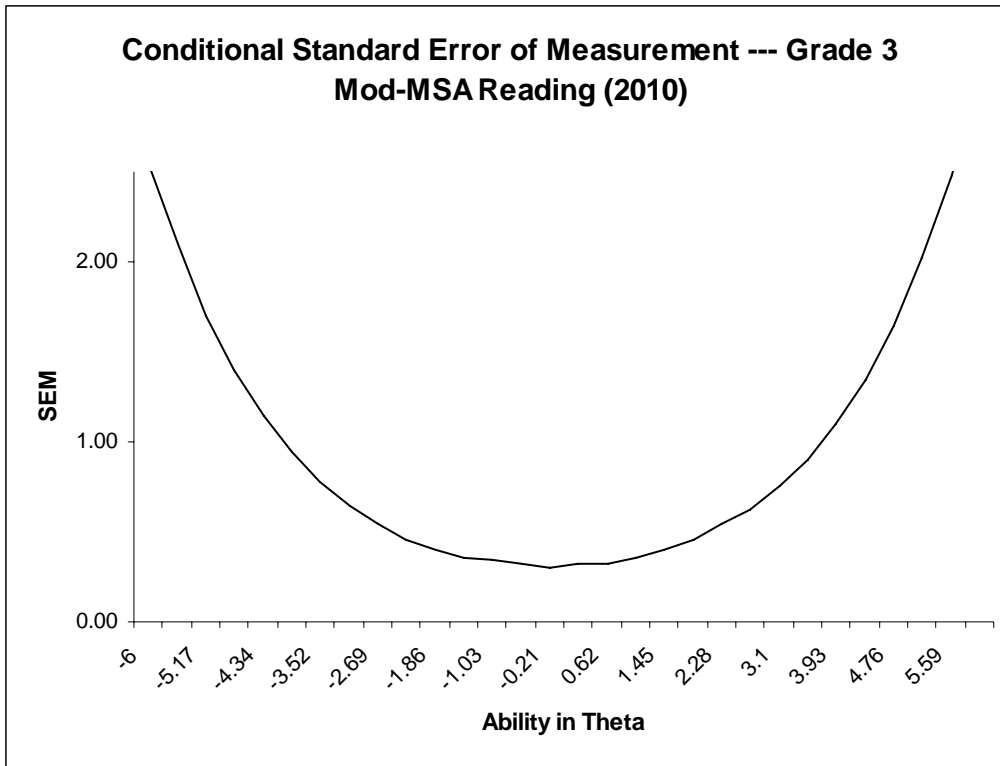


Figure 9.4.1c. CSEM for Grade 3 Mod-MSA: Reading

Grade 4

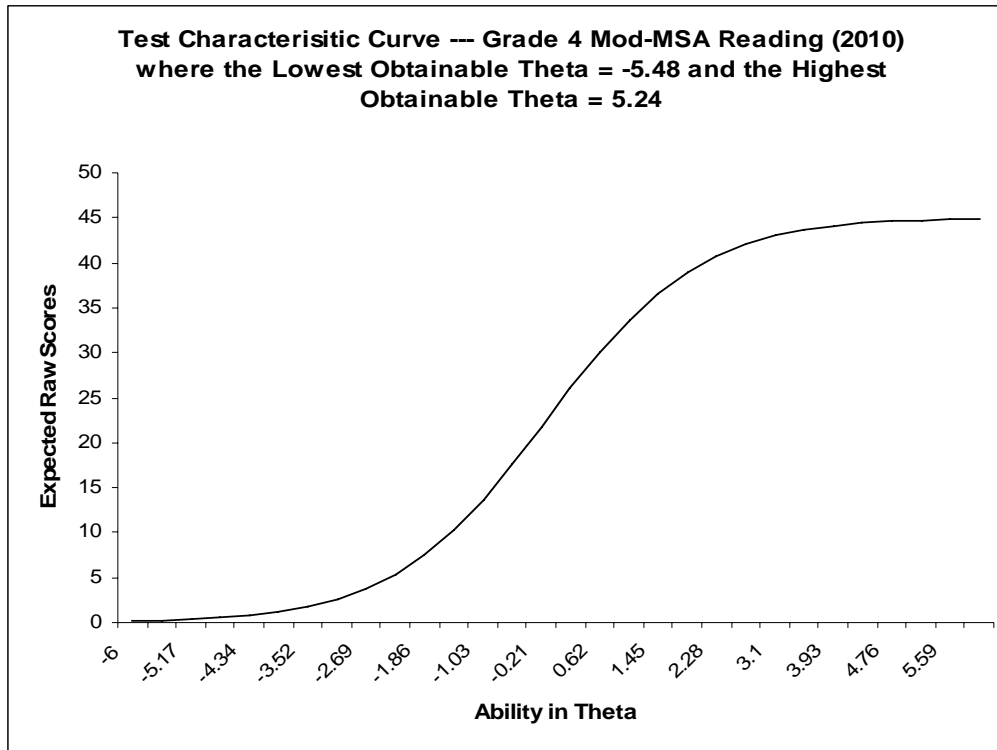


Figure 9.4.2a. TCC for Grade 4 Mod-MSA: Reading

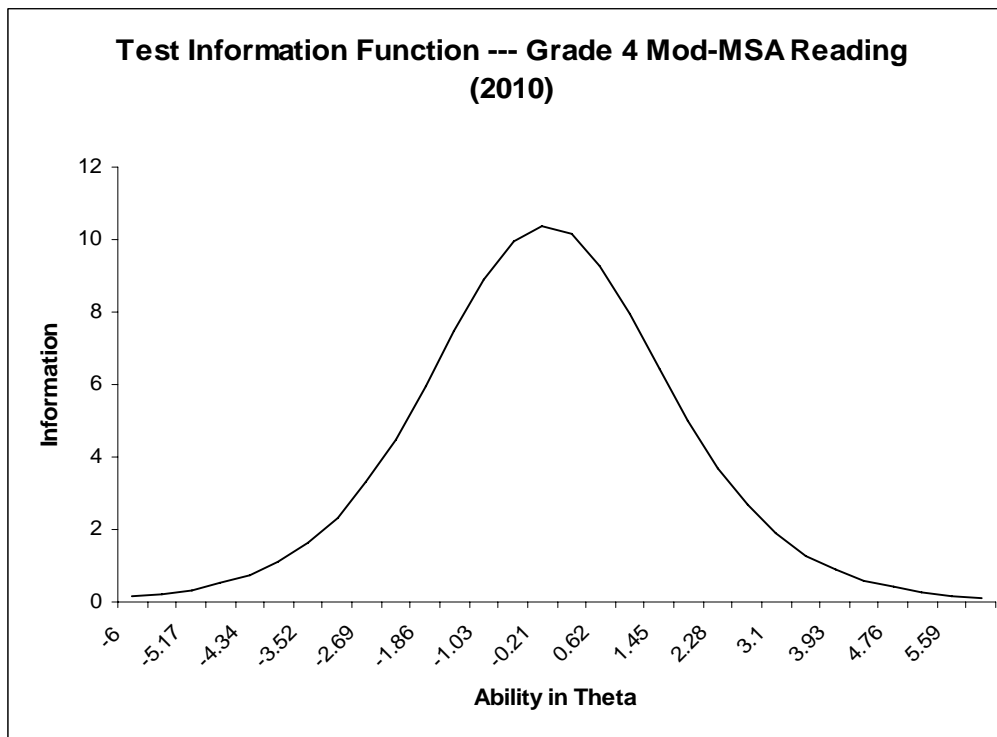


Figure 9.4.2b. TIF for Grade 4 Mod-MSA: Reading

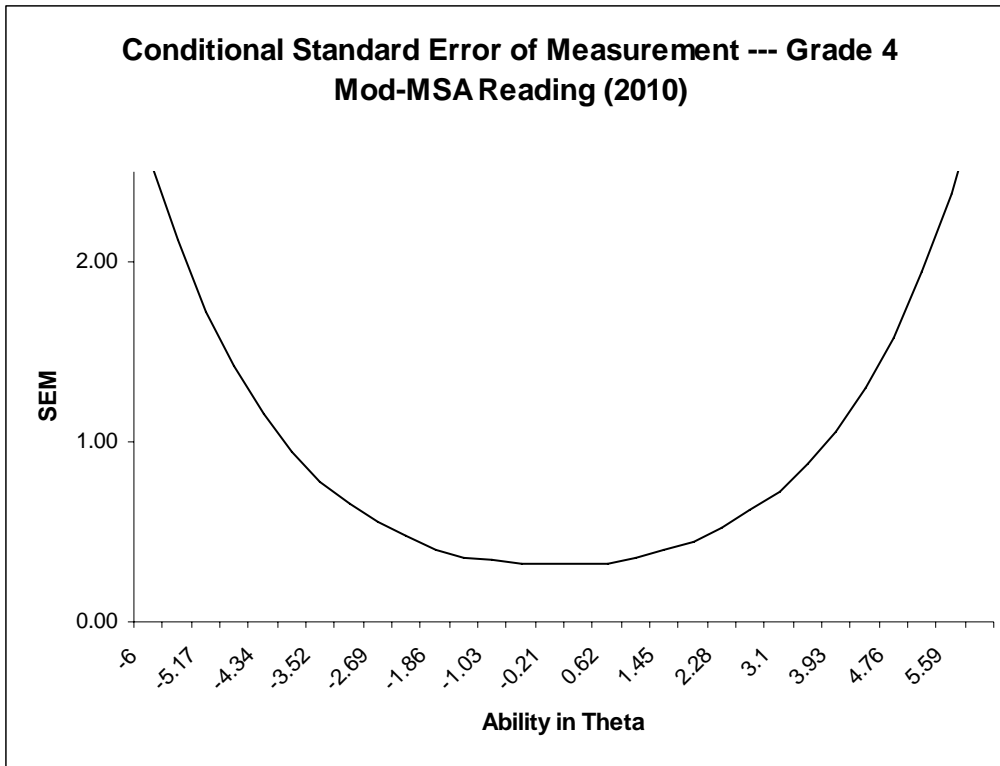


Figure 9.4.2c. CSEM for Grade 4 Mod-MSA: Reading

Grade 5

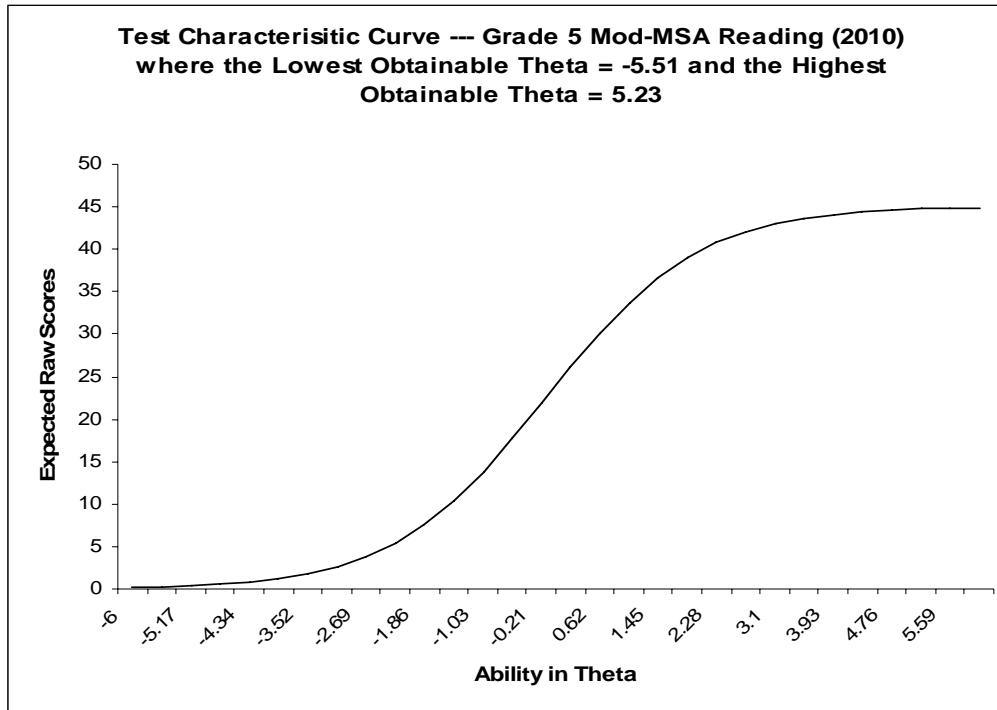


Figure 9.4.3a. TCC for Grade 5 Mod-MSA: Reading

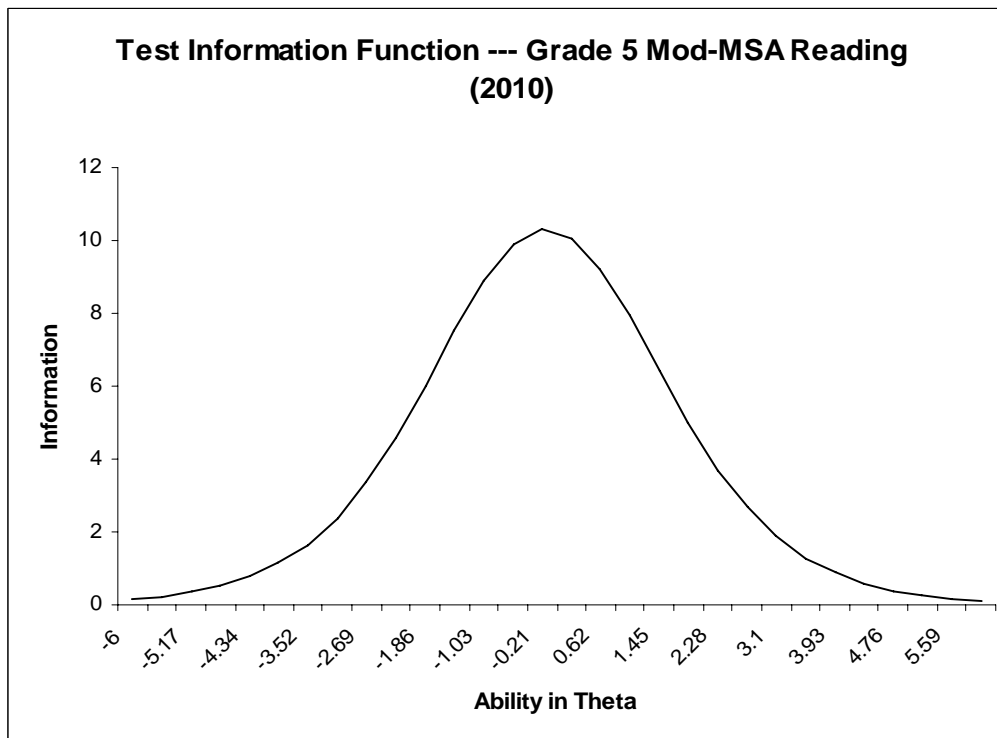


Figure 9.4.3b. TIF for Grade 5 Mod-MSA: Reading

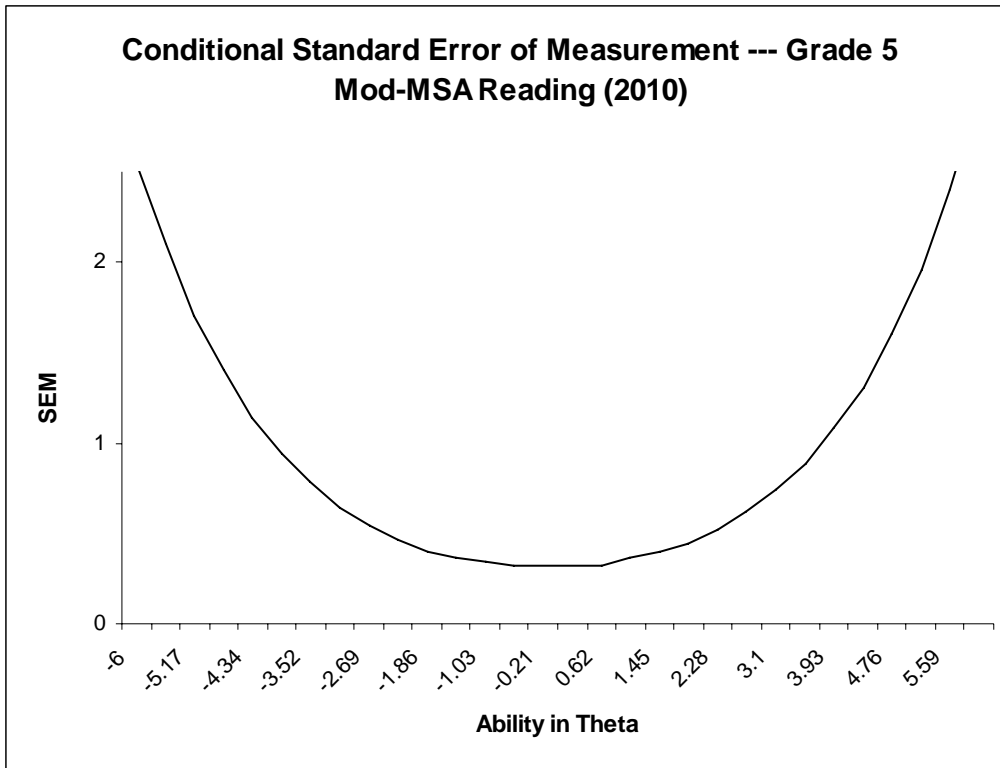


Figure 9.4.3c. CSEM for Grade 5 Mod-MSA: Reading

Grade 6

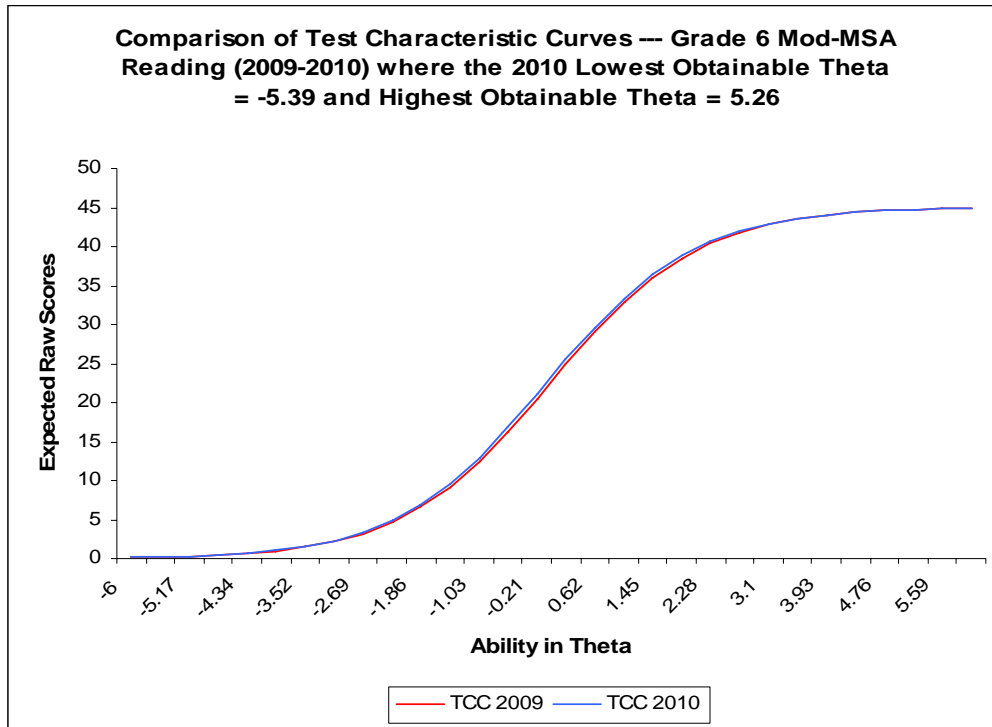


Figure 9.4.4a. TCC Comparison for Grade 6 Mod-MSA: Reading

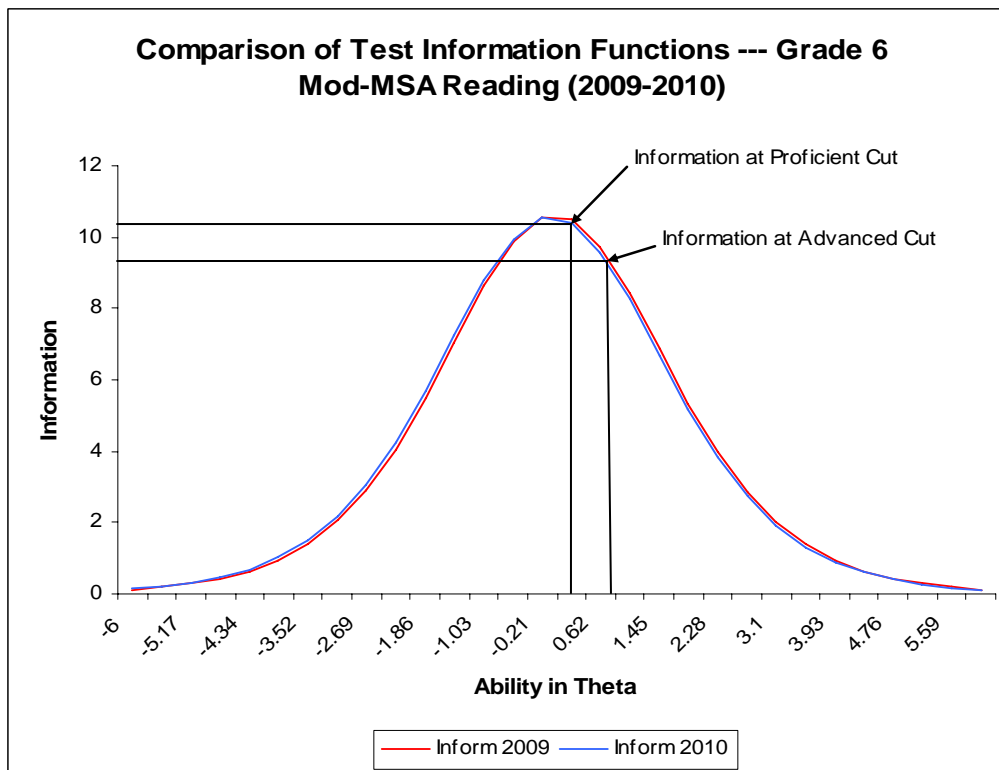


Figure 9.4.4b. TIF Comparison for Grade 6 Mod-MSA: Reading (2009 with 2010)

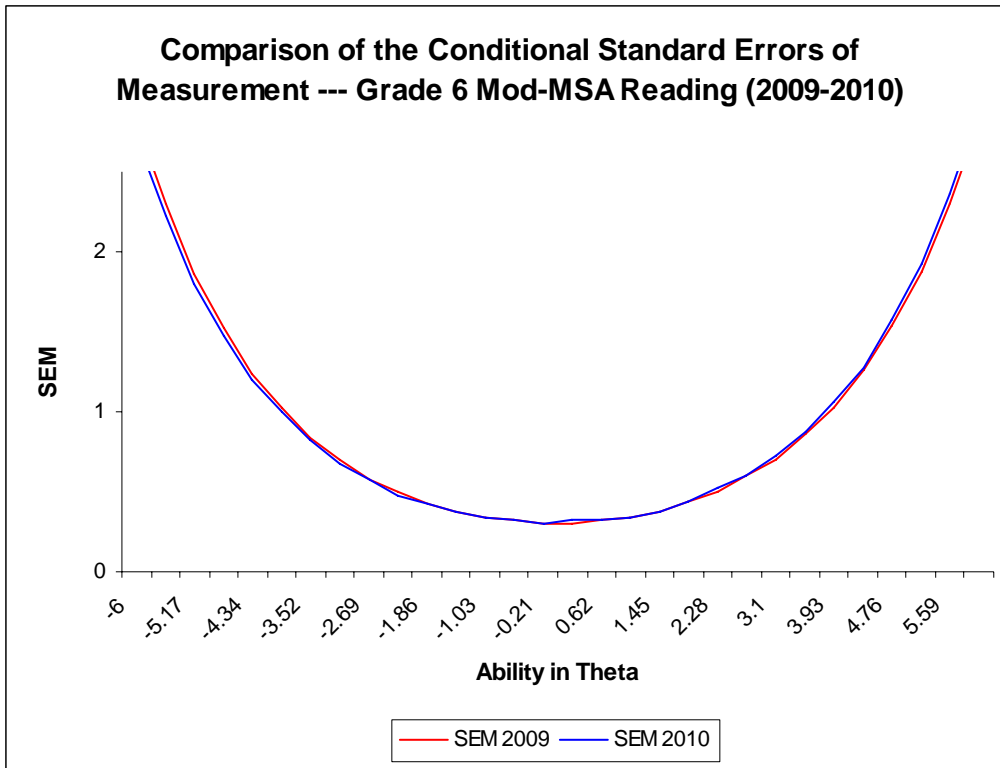


Figure 9.4.4c. CSEM Comparison for Grade 6 Mod-MSA: Reading

Grade 7

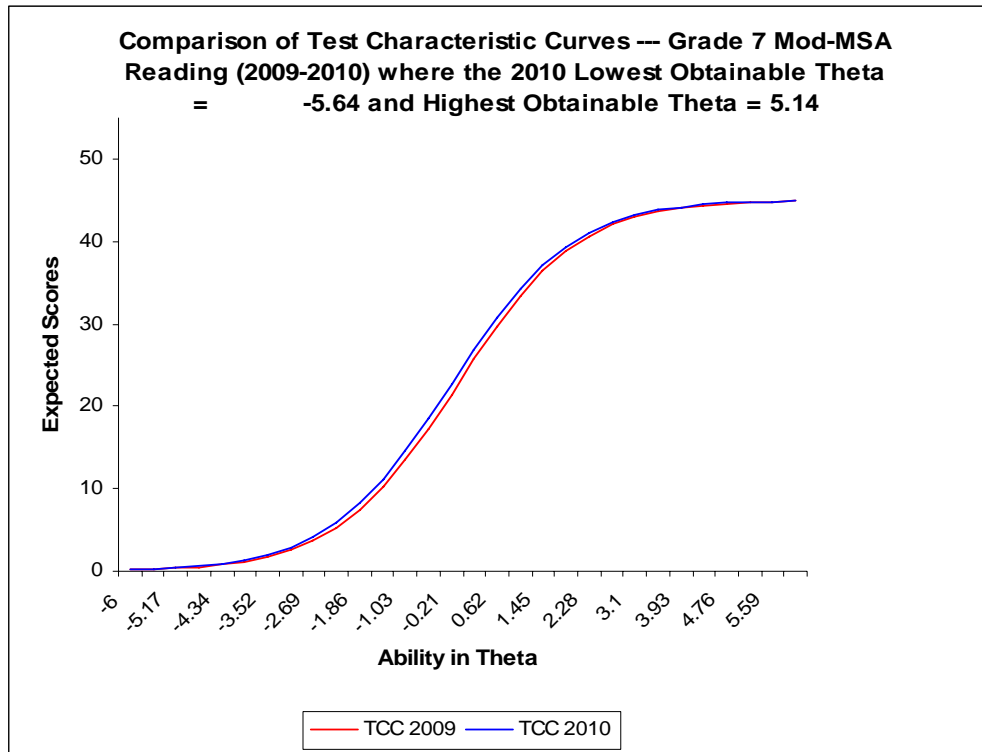


Figure 9.4.5a. TCC Comparison for Grade 7 Mod-MSA: Reading

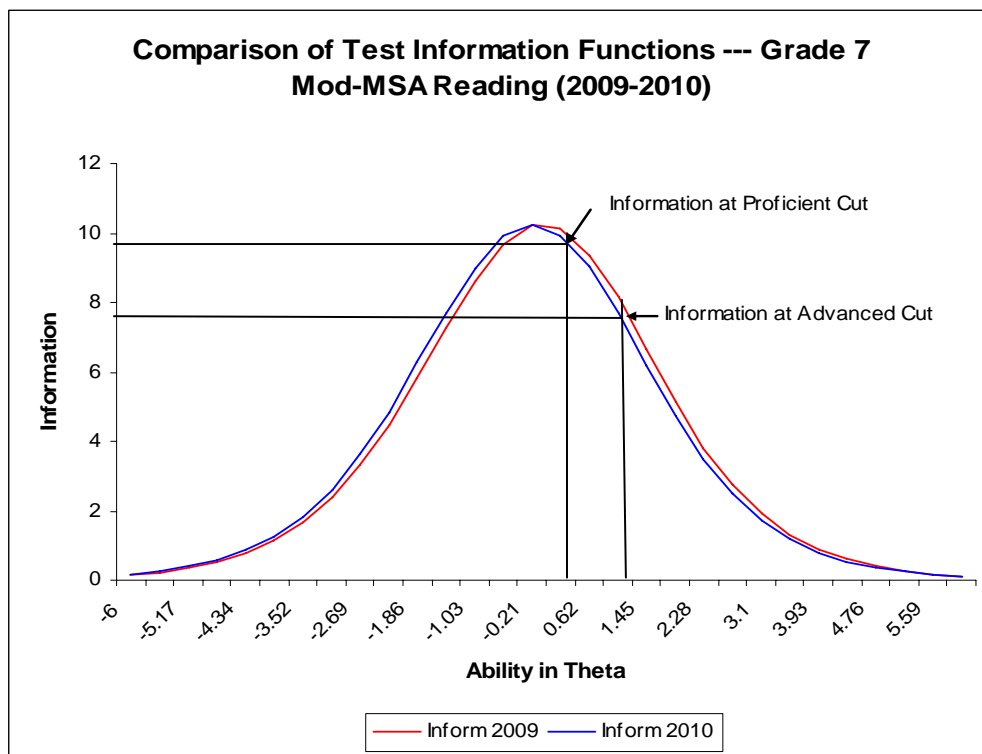


Figure 9.5.5b. TIF Comparison for Grade 7 Mod-MSA: Reading

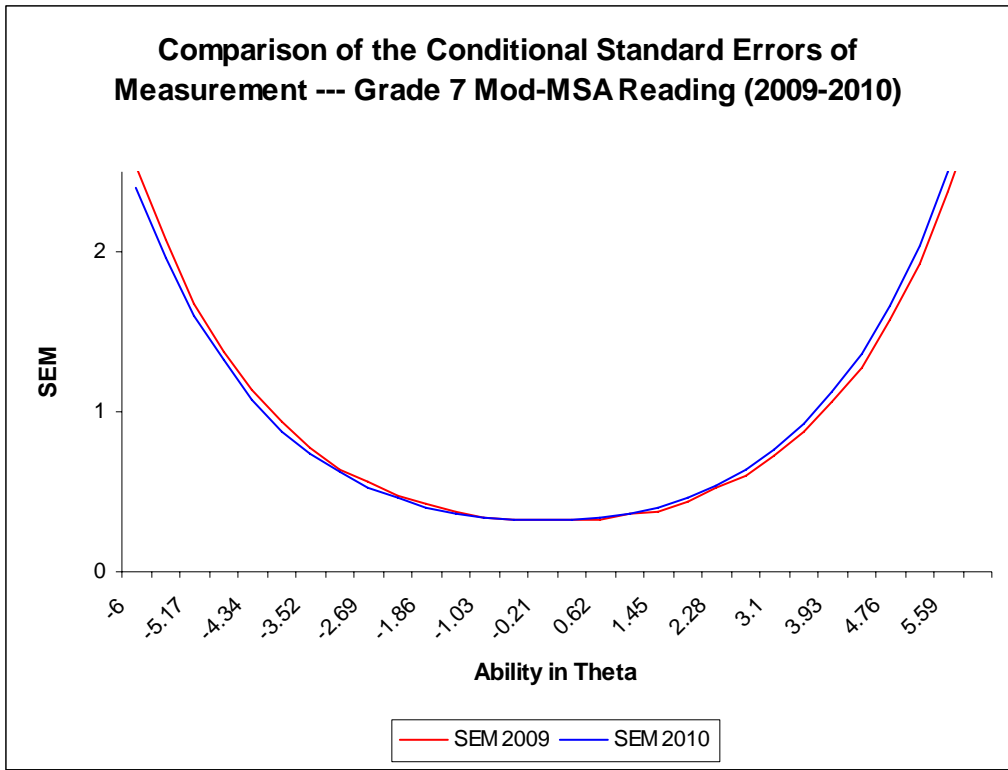


Figure 9.5.5c. CSEM Comparison for Grade 7 Mod-MSA: Reading

Grade 8

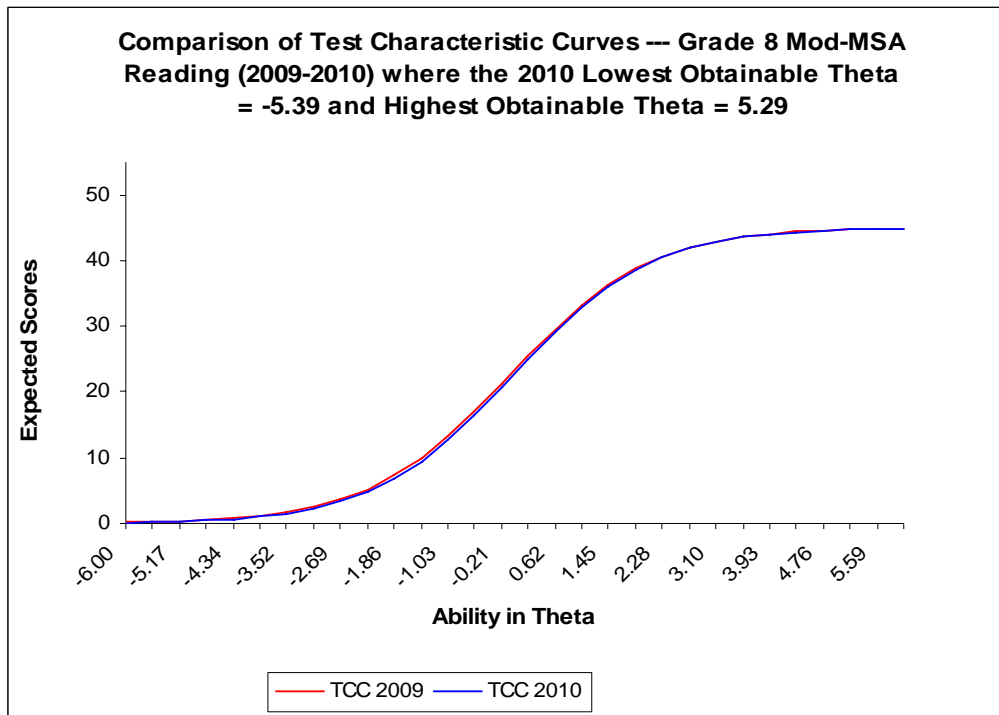


Figure 9.4.6a. TCC Comparison for Grade 8 Mod-MSA: Reading

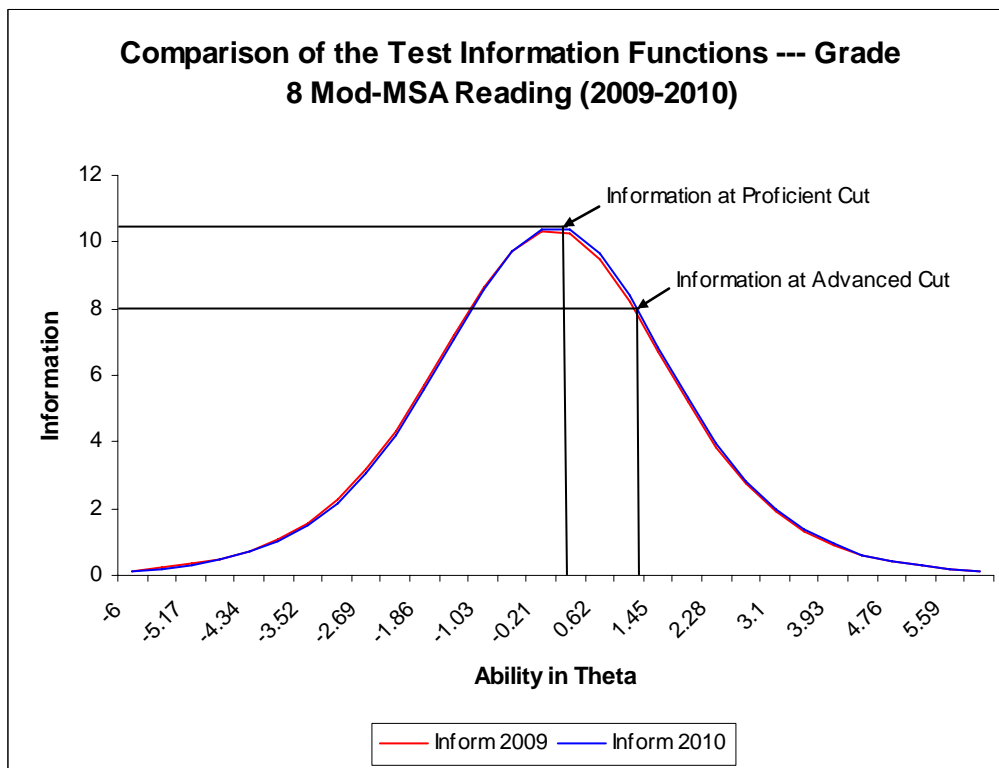


Figure 9.4.6b. TIF Comparison for Grade 8 Mod-MSA: Reading

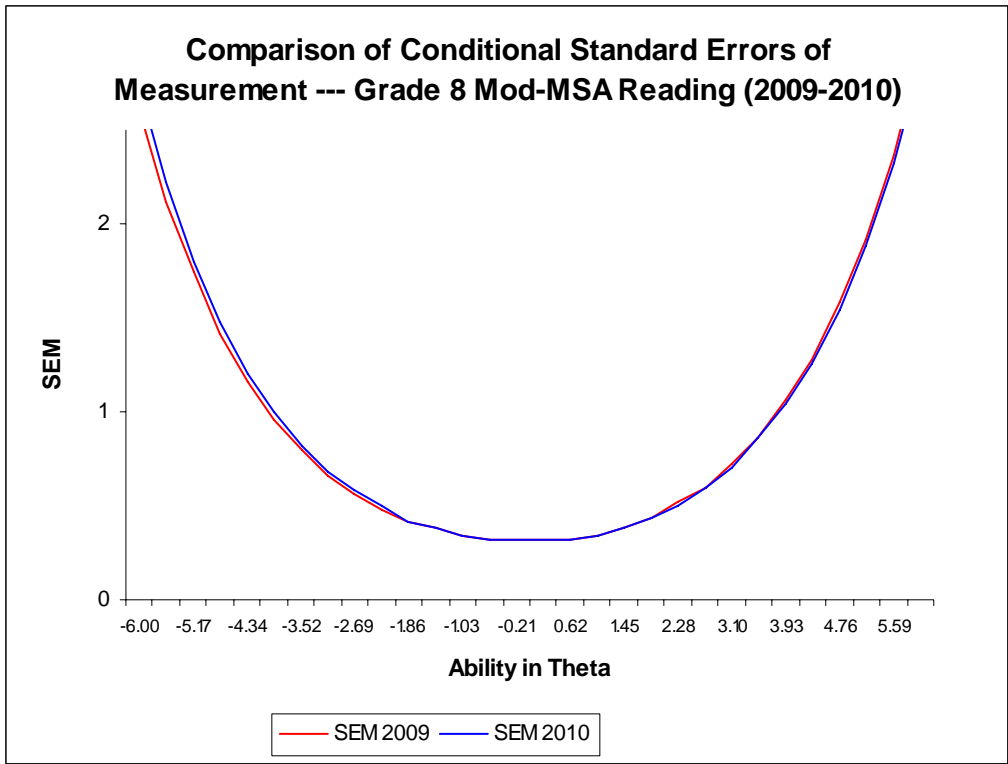


Figure 9.4.6c. CSEM Comparison for Grade 8 Mod-MSA: Reading