

4. SCALE CREATION, EQUATING AND RAW SCORES TO SCALE SCORES CONVERSION VIA ITEM RESPONSE THEORY PROCEDURES

For the 2010 administration, there was no equating for Grades 3 to 5 as this was the first year of implementation of the Mod-MSA examinations for these grades. However, grades 6 to 8 forms were linked together by the common items non-equivalent groups (CINEG, Kolen & Brennan, 2004) design.

The Rasch model (Rasch, 1960) was used to develop, calibrate, and scale the Mod-MSA: Reading. The Rasch measurement model is regularly used to construct test forms, for scaling and equating, and to develop and maintain large item banks. All item and test analyses, including item-fit analysis, scaling, diagnosis, and performance prediction were accomplished within this framework. The statistical software used to calibrate and scale the Mod-MSA: Reading was WINSTEPS Version 3.46 (Linacre & Wright, 2000).

The Rasch Model

The most basic expression of the Rasch model is in the item characteristic curve (ICC). It shows the probability of a correct response to an item as a function of the ability, i.e., the proficiency level. The probability of a correct response is bounded by 1 (certainty of a correct response) and 0 (certainty of an incorrect response).

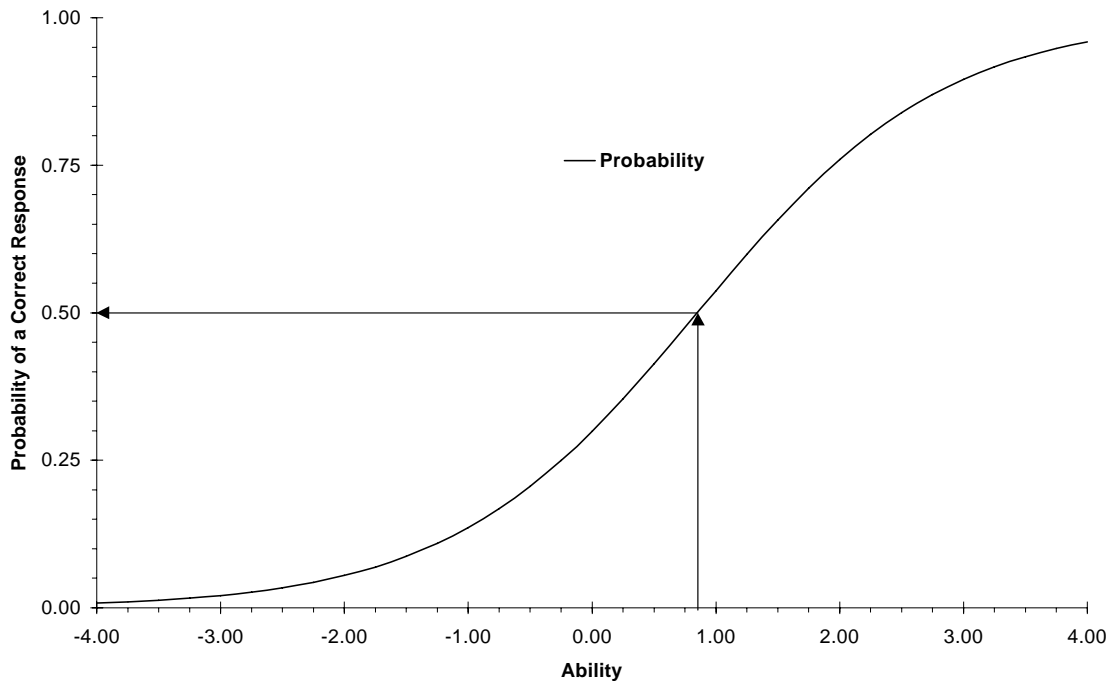


Figure 4.1 Item Characteristic Curve

As an example, consider Figure 4.1 which depicts an item that falls at approximately 0.85 on the ability, i.e., the proficiency (horizontal) scale. When a person answers an item at the same level as his or her proficiency, then that person has a probability of roughly 50% of answering the item correctly. Another way of expressing this is that if we have a group of 100 people, all of whom have a proficiency of 0.85, we would expect about 50% of them to answer the item correctly. A person whose proficiency was above 0.85 would have a higher probability of getting the item right, while a person whose proficiency is below 0.85 would have a lower probability of getting the item right. This makes intuitive sense and is the basic formulation of Rasch measurement for test items having only two possible categories (i.e., wrong or right).

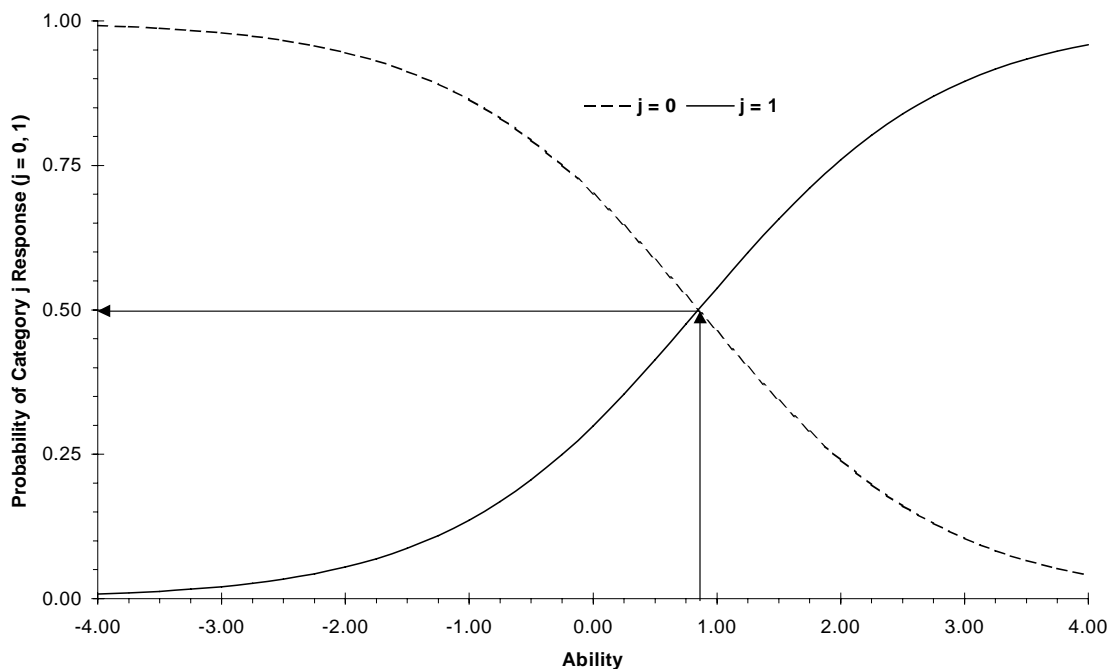


Figure 4.2 Category Response Curves for a One-Step Item

Figure 4.2 extends this formulation to show the probabilities of obtaining a wrong answer or a right answer. The curve on the left ($j = 0$) shows the probability of getting a score of “0” while the curve on the right ($j = 1$) shows the probability of getting a score of “1”. The point at which the two curves cross indicates the transition point on the proficiency scale where the most likely response changes from a “0” to a “1.” Here, the probability of answering the item correctly is 50%.

One important property of the Rasch model is its ability to separate the estimation of item/task parameters from the person parameters. With the Rasch model, the total score given by the sum of the categories in which a person responds is a sufficient statistic for estimating a person’s proficiency (i.e., no additional information need be estimated). The total number of responses

across examinees in a particular category is a sufficient statistic for estimating the step difficulty for that category. Thus with the Rasch model, the same total score will yield the same proficiency estimate for different examinees.

The parameters estimated by this model are (1) a proficiency estimate for each person, (2) m_i threshold (difficulty) estimate for each item. From these estimates, the conditional standard error estimates associated with proficiency and the standard error of the difficulty parameter estimates of each item can be calculated (See Section 8.4 for the derivation of the conditional standard error of measurement and the confidence interval set at each proficiency level).

4.1. Calibration and Scaling Procedures for Grades 3 to 5

For the 2010 administration, there was no equating for Grades 3 to 5 as this was the first year of implementation of the Mod-MSA examinations for these grades. However, for 2010, a new form of the test was created for Grades 6 to 8 and these forms were linked together by the *common items non-equivalent groups* (CINEG, Kolen & Brennan, 2004) design.

The calibration of the spring 2010 administration of the Mod-MSA: Reading was used to establish the base scale for the assessment in the area of reading at grades 3–5. Item parameters were calibrated using the Rasch measurement model, which placed all items on a common scale. Although the Rasch model is fairly robust, when setting the base scale for an assessment program it is desirable to minimize as many sources of error as practical during the calibration process. This calibration was, therefore, conducted using a two-phase approach. In the first phase only items with acceptable classical item statistics (i.e., non-negative point biserial correlations) and IRT model fit were included. This phase of calibration established the base scale. During the second phase of calibration the items excluded from phase one were placed on the established base scale. This was accomplished by anchoring the parameters obtained for the items included in phase one to their base scale values and only allowing the parameters of the items with less acceptable classical stats (those excluded from phase one) to be freely estimated. This method placed the parameters of the poorly functioning items on the base scale (thereby allowing these items to be selected for operational scoring if necessary) while ensuring that these items did not unduly influence the parameters of those items with acceptable statistics.

Following calibration, all items were sent to data review. Those items not selected as operational items, but not labeled as “do not use” (DNU) during data review, were archived in the item bank for possible future use. RS to SS tables were then created using the established scale parameters of the items selected for operational scoring.

4.2. Specifics for Creating the Base Scale for the Mod-MSA: Reading Grades 3-5

The base scale was created for each grade 3 to 5 and content area based on the strength of the items’ classical statistics. Items that had poor classical statistics were not included in the creation of the base scale for each grade and content area (for the purposes of this calibration poor item statistics means a negative point biserial correlation).

Items selected from above were calibrated using the Rasch model. From these items, all items showing poor infit and outfit stats (>2.00 and < 0.5) were dropped from the creation of the base scale.

All the items that were excluded from the creation of the base scale were placed on this scale by floating them (keeping their calibration values unanchored) while anchoring the base-scale items to their established calibrated values.

Operational item calibration took place after an identification of these items from data review. The operational form item calibrations remained the same as those established on the above scale for the creation of the RS to SS tables. The non-operational items with their respective calibrations were banked as FT items.

The specific steps in the process were as follows:

1. Conduct classical item analysis of all items on a test.
2. Conduct Rasch calibration of all items on a test that do *not* have negative point biserial correlations (based on results of Step 1).
3. Conduct Rasch calibration of all items used in Step 2 that show acceptable infit and outfit (≤ 2.00 and > 0.5) – this step establishes the base scale for the test.
4. Place the items excluded at Steps 2 and 3 on the base scale by conducting a Rasch calibration with all items used in Step 3 anchored to their base scale values.
5. Submit items for data review with their respective calibrations obtained as outlined above.
6. Create RS to SS scales (for total scores and strand scores), using base scale parameters of the items selected for operational scoring by data review members.

4.3. Calibration Equating the 2010 Mod-MSA: Reading Grades 6-8

The base scale for the Mod-MSA: Reading Grades 6 to 8 had been created in 2009. The procedures followed in creating the base scale were the same as those explained above in creating the scale for Grades 3 to 5 in 2010.

For Grades 6 there were 23 common items for use as linking items in the equating process while Grades 7 and 8 had 25 such common items. Items in these grades were placed on the 2009 established scale through the equating process. The calibrations of these items were then sent to data review and the same process was followed as in the Grade 3 to 5 calibration process described above to select operational items, and archiving non-operational items into the item bank.

4.4. Specifics of Linking and Equating the 2010 Mod-MSA Grades 6-8: Reading

The 2010 Mod-MSA was calibrated and equated by fixing item parameters of common linking items. To select unstable common items (outliers) from being linking items, the Robust Z procedure was used.

4.4.1. Generalized Robust Z Procedure

Generalized robust z values were calculated by the following procedures:

- Calculate the mean and standard deviation of the linking pool's structure measure parameters (D_{ij}) for the 2010 form
- Obtain the ratio of the standard deviations between form 2009 and form 2010
- Obtain the correlation between form 2009 and form 2010 structure measure parameters

- Calculate the difference between form 2009 and 2010 structure measure parameters for each item in the linking pool
- Calculate the mean of the differences calculated above
- Calculate the median of the differences
- Calculate the interquartile range of the differences
- Calculate the robust z for each structure measure parameter in the linking pool where the robust z is defined as (the difference between form 2009 and form 2010 item measure parameters minus the median of the differences) / (interquartile range multiplied by 0.74)
- Calculate the absolute z value of each item measure parameter

4.4.2. Guidelines for Selecting Linking Items

Once the above calculations are made, the following guidelines will determine possible sets of common items to be used for the Rasch equating (SCDE, 2001):

- Try not to include those items with an averaged absolute robust z exceeding 1.645
- Consider that the ratio of the standard deviations of form 2009 and form 2010 item measure parameters should be in the 90 to 110 percent range
- The correlation coefficient of form 2009 and 2010 should be greater than .95
- Do not eliminate more than 20 percent of total score point of the linking pool items

4.4.3. Step-by-step Procedure for Selecting Linking Items

1. Calculate robust Z for all items, the correlation between the fixed Rasch difficulties and the estimated Rasch difficulties, and the ratio of the standard deviations for the fixed and estimated Rasch difficulties. .
2. Check the correlation and ratio of SD of fixed and estimated Rasch parameters. If correlation is greater than 0.95 and ratio is between 0.9 and 1.1 then stop.
3. Choose the item with the largest absolute value of robust Z that is greater than 1.645 and drop from linking set. If no items have a robust Z with an absolute value greater than 1.645 then stop.
4. If the deletion of one more item from the linking set would result in 20% or more of the linking set items being dropped, then stop.
5. Recalculate correlation and SD ratio for remaining items and return to step 1. Do NOT recalculate robust Z values.

The step-by-step procedure is graphically displayed in Figure 4.4.1, below. Tables 4.4.1 to 4.4.3 provide the unequated Rasch item difficulty comparison of the core linking items between 2009 and 2010 for grades 6 to 8 together with their robust z values.

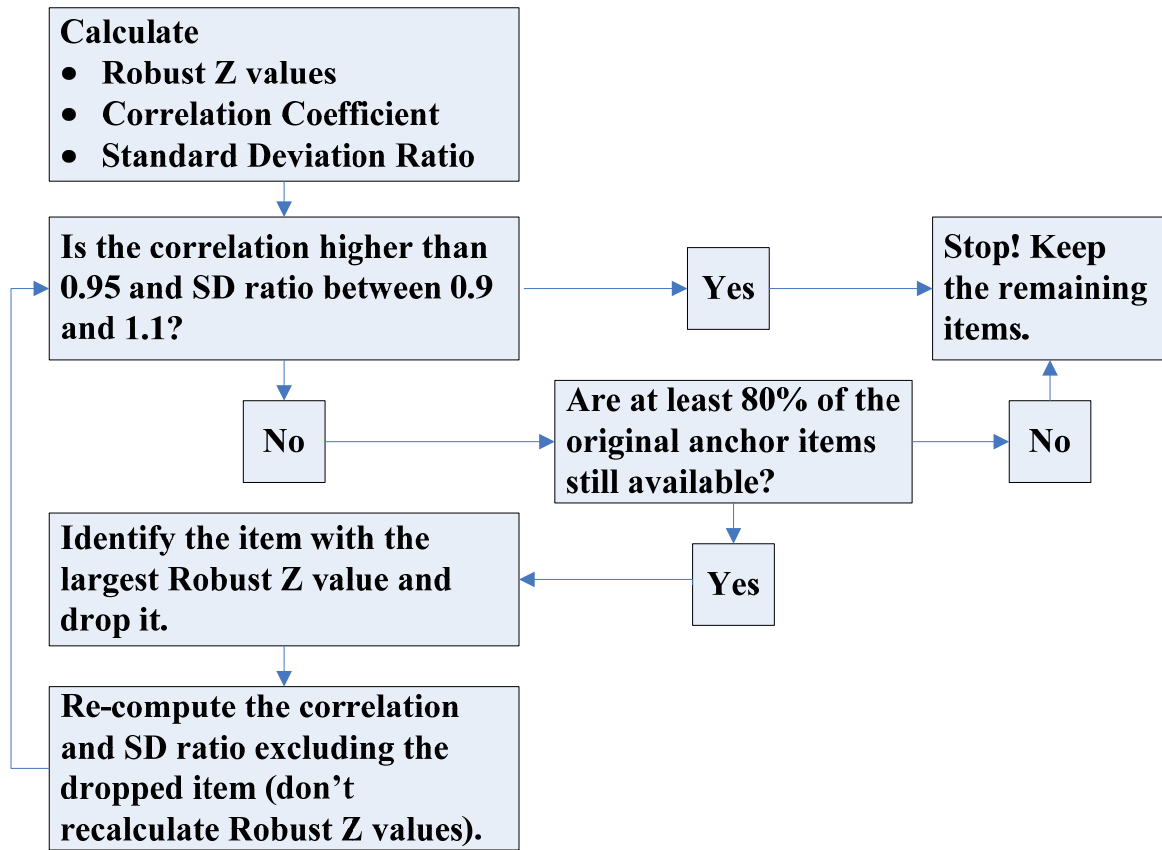


Figure 4.3.1. Anchor Evaluation Steps Chart for Mod-MSA

**Table 4.4.1. Unequated Core Linking Item Difficulties of Previous Year vs. Year 2010:
Grade 6**

Item No.	Item Seq. No.	Rasch Diff. 2009	Rasch Diff.2010	Robust Z*
1	1	-0.7867	-0.6640	-0.67
2	2	-1.2831	-1.0811	-1.32
3	3	-0.3530	0.0364	<u>-2.86</u>
4	4	-0.4888	-0.4854	0.32
5	5	-0.0595	-0.1022	0.70
6	7	-0.5508	-0.6483	1.15
7	8	1.3285	1.4436	-0.60
8	9	-0.8831	-1.1560	<u>2.59</u>
9	10	0.1881	0.1052	1.03
10	12	-0.4133	-0.2714	-0.82
11	13	-0.1792	0.0364	-1.43
12	14	-0.6591	-0.6172	0.00
13	15	-0.3109	-0.3000	0.26
14	16	0.2119	0.4491	-1.61
15	17	0.0695	0.1464	-0.29
16	18	0.8270	0.9642	-0.78
17	19	-0.4815	-0.4457	0.05
18	20	0.4870	0.4908	0.31
19	21	0.3073	0.1738	1.44
20	50	-0.5472	-0.7221	<u>1.78</u>
21	51	0.0153	0.0088	0.40
22	52	-0.3179	-0.2571	-0.16
23	53	-0.2517	-0.0327	-1.46

Note: Bold, underlined values are for Robust Z > 1.645

**Table 4.4.2. Unequated Core Linking Item Difficulties of Previous Year vs. Year 2010:
Grade 7**

Item No.	Item Seq. No.	Rasch Diff. 2009	Rasch Diff.2010	Robust Z*
1	1	-0.7780	-0.5090	-0.57
2	2	-1.3540	-1.1078	-0.03
3	3	0.8276	0.9922	<u>1.91</u>
4	5	-0.9143	-0.8588	<u>4.51</u>
5	6	-0.8469	-0.5615	-0.96
6	22	0.0908	0.3745	-0.92
7	23	0.7546	1.0260	-0.63
8	24	0.4769	0.7183	0.08
9	25	0.0775	0.4595	<u>-3.26</u>
10	26	0.4687	0.7064	0.17
11	27	-0.1197	-0.0485	<u>4.13</u>
12	28	0.4089	0.6273	0.63
13	29	0.0297	0.2819	-0.17
14	30	0.4035	0.6628	-0.34
15	31	0.6886	1.1033	<u>-4.04</u>
16	39	-1.6963	-1.4900	0.92
17	40	-0.1924	-0.0761	<u>3.06</u>
18	41	-1.1534	-0.9085	0.00
19	42	0.0244	0.2742	-0.12
20	43	-1.3660	-1.2209	<u>2.37</u>
21	44	0.1918	0.2973	<u>3.32</u>
22	45	-0.2739	-0.0642	0.84
23	46	0.1732	0.4208	-0.06
24	47	-0.4429	-0.1798	-0.43
25	48	-0.5024	-0.2937	0.86

Note: Bold, underlined values are for Robust Z > 1.645

Table 4.4.3. Unequated Core Linking Item Difficulties of Previous Year vs. Year 2010: Grade 8

Item No.	Item Seq. No.	Rasch Diff. 2009	Rasch Diff.2010	Robust Z*
1	1	-1.1648	-1.3041	0.58
2	2	-0.4197	-0.3775	-1.13
3	4	0.4446	0.3926	-0.24
4	5	0.0045	-0.0457	-0.26
5	6	-0.1989	-0.1547	-1.15
6	15	-1.3174	-1.5258	1.22
7	16	0.2613	0.1423	0.38
8	17	0.0618	-0.0170	0.01
9	18	0.3175	0.5589	<u>-3.00</u>
10	19	-1.1686	-1.4334	<u>1.75</u>
11	27	-0.2773	-0.3581	0.03
12	28	0.2393	0.1213	0.38
13	29	-0.5397	-0.5208	-0.91
14	30	0.2686	0.1913	-0.01
15	31	-0.0760	-0.1806	0.25
16	39	-0.9453	-0.9953	-0.26
17	40	-0.2879	-0.2740	-0.86
18	41	0.2760	0.3822	<u>-1.73</u>
19	42	0.0370	-0.1072	0.62
20	43	-0.8923	-0.8741	-0.90
21	51	0.2711	0.1002	0.87
22	52	-0.2458	-0.3158	-0.08
23	53	0.4618	0.2783	0.99
24	54	-0.3358	-0.4206	0.06
25	55	-0.1834	-0.3043	0.40

Note: Bold, underlined values are for Robust Z > 1.645

4.5. Reporting Scale Scores for the 2010 Mod-MSA: Reading

The Mod-MSA reports student scores on the total performance of students on the reading examination (total score) as well as the reporting of their strand scores outlined in Section 2.3.

In order to facilitate the use and interpretation of the results of the 2010 Mod-MSA Reading, a scale score was created for each point on the raw score tables (total scores as well as strand scores) that had a mean = 50; a standard deviation = 12; and the lowest and highest obtainable scale scores (LOSS and HOSS) as 2 and 98, respectively. As is the case with standard MSA, the lowest obtainable raw score (zero) was automatically set to the LOSS and the highest obtainable raw score (51) set to the HOSS in the event that the actual scale score associated with these raw scores fell above or below these values, respectively.

Once RS to Theta tables were produced by the WINSTEPS 3.46 program after data review, theta to scale score constants were calculated using the following formula:

$$SS = Slope \times Theta + Intercept$$

$$SEM_{CSS} = Slope \times SEM_{CT}$$

where

Slope = 12 / the standard deviation of the theta values, and

Intercept = 50 – slope × mean of the theta values

Theta = the *IRT* proficiency estimate at a particular raw score on the scoring continuum

SEM_{CSS} = the standard error of the scale score, and

SEM_{CT} = the standard error conditional on proficiency (theta) estimates

Table 4.5.1 depicts the slope and intercept that were used for each grade. It should be noted that the same slopes and intercepts were used for Grades 6 to 8 as those used in 2009. Similarly, the same slopes and intercept for each of the grades 3 to 8 will be used for future administrations. Total raw score to scale score conversion tables for Grades 3-8 are provided in Tables 4.5.2 to 4.5.7, while strand level RS to SS are provided in Tables 4.5.8 to Tables 4.5.13.

Each student's total raw score for the strands was a summation of the individual item score within a strand level. The strand levels were classified as stated in section 2.3 and the item parameters within each strand was obtained using the Winsteps program in the same manner as those obtained for the total test. Once the item parameters were available, thetas (student proficiency scores) were calculated for each raw score point that could be obtained within each strand. The thetas were transferred to scale scores, using the same slope and intercept as that which were applied for the total reading test score.

Table 4.5.1. The 2010 Mod-MSA, Reading Slope and Intercept for the Transfer of RS to SS Across Grades

Grade	Slope	Intercept
3	13.8375	47.9876
4	15.3069	48.7765
5	15.8311	49.1418
6	18.4057	48.0880
7	17.1743	47.9523
8	16.5287	43.8652

Raw Score to Scale Score Conversion Tables for the Total Score

Table 4.5.2. The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion Table: Grade 3

Raw Score	Proficiency Estimates	SE	SS ¹	SE(SS)	SS-1SE (SS) ²	SS+1SE (SS) ²
0	-5.5040	2.0073	2	28	-	-
1	-4.0956	1.0147	2	14	-	-
2	-3.3724	0.7282	2	10	-	-
3	-2.9365	0.6034	7	8	-	-
4	-2.6179	0.5303	12	7	5	19
5	-2.3632	0.4815	15	7	8	22
6	-2.1488	0.4462	18	6	12	24
7	-1.9620	0.4194	21	6	15	27
8	-1.7951	0.3984	23	6	17	29
9	-1.6432	0.3815	25	5	20	30
10	-1.5030	0.3677	27	5	22	32
11	-1.3721	0.3563	29	5	24	34
12	-1.2486	0.3467	31	5	26	36
13	-1.1312	0.3387	32	5	27	37
14	-1.0189	0.3320	34	5	29	39
15	-0.9107	0.3263	35	5	30	40
16	-0.8057	0.3216	37	4	33	41
17	-0.7035	0.3178	38	4	34	42
18	-0.6036	0.3147	40	4	36	44
19	-0.5054	0.3123	41	4	37	45
20	-0.4084	0.3105	42	4	38	46
21	-0.3124	0.3093	44	4	40	48
22	-0.2169	0.3087	45	4	41	49
23	-0.1216	0.3087	46	4	42	50
24	-0.0262	0.3093	48	4	44	52
25	0.0698	0.3104	49	4	45	53
26	0.1666	0.3122	50	4	46	54
27	0.2649	0.3146	52	4	48	56
28	0.3647	0.3177	53	4	49	57
29	0.4669	0.3215	54	4	50	58
30	0.5717	0.3262	56	5	51	61
31	0.6799	0.3318	57	5	52	62
32	0.7922	0.3385	59	5	54	64
33	0.9094	0.3465	61	5	56	66
34	1.0328	0.3561	62	5	57	67
35	1.1635	0.3675	64	5	59	69
36	1.3036	0.3813	66	5	61	71
37	1.4553	0.3982	68	6	62	74
38	1.6220	0.4192	70	6	64	76
39	1.8087	0.4460	73	6	67	79
40	2.0231	0.4813	76	7	69	83
41	2.2776	0.5302	80	7	73	87
42	2.5962	0.6033	84	8	76	92
43	3.0319	0.7280	90	10	-	-
44	3.7549	1.0147	98	14	-	-
45	5.1633	2.0072	98	28	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors may not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

Table 4.5.3. The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion
Table: Grade 4

Raw Score	Proficiency Estimates	SE	SS ¹	SE(SS)	SS-1SE (SS) ²	SS+1SE (SS) ²
0	-5.4793	2.0073	2	31	-	-
1	-4.0708	1.0148	2	16	-	-
2	-3.3476	0.7283	2	11	-	-
3	-2.9115	0.6035	4	9	-	-
4	-2.5927	0.5305	9	8	-	-
5	-2.3379	0.4817	13	7	6	20
6	-2.1233	0.4464	16	7	9	23
7	-1.9361	0.4197	19	6	13	25
8	-1.7689	0.3988	22	6	16	28
9	-1.6168	0.3820	24	6	18	30
10	-1.4762	0.3682	26	6	20	32
11	-1.3449	0.3568	28	5	23	33
12	-1.2210	0.3474	30	5	25	35
13	-1.1032	0.3394	32	5	27	37
14	-0.9903	0.3328	34	5	29	39
15	-0.8815	0.3272	35	5	30	40
16	-0.7759	0.3226	37	5	32	42
17	-0.6731	0.3188	38	5	33	43
18	-0.5725	0.3158	40	5	35	45
19	-0.4735	0.3135	42	5	37	47
20	-0.3758	0.3118	43	5	38	48
21	-0.2790	0.3107	45	5	40	50
22	-0.1827	0.3102	46	5	41	51
23	-0.0865	0.3103	47	5	42	52
24	0.0100	0.3109	49	5	44	54
25	0.1069	0.3121	50	5	45	55
26	0.2049	0.3140	52	5	47	57
27	0.3042	0.3165	53	5	48	58
28	0.4054	0.3196	55	5	50	60
29	0.5088	0.3236	57	5	52	62
30	0.6150	0.3283	58	5	53	63
31	0.7245	0.3340	60	5	55	65
32	0.8384	0.3408	62	5	57	67
33	0.9572	0.3489	63	5	58	68
34	1.0822	0.3585	65	5	60	70
35	1.2148	0.3700	67	6	61	73
36	1.3567	0.3839	70	6	64	76
37	1.5105	0.4008	72	6	66	78
38	1.6793	0.4218	74	6	68	80
39	1.8683	0.4486	77	7	70	84
40	2.0850	0.4839	81	7	74	88
41	2.3420	0.5326	85	8	77	93
42	2.6631	0.6055	90	9	-	-
43	3.1017	0.7300	96	11	-	-
44	3.8277	1.0162	98	16	-	-
45	5.2383	2.0080	98	31	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

**Table 4.5.4. The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion
Table: Grade 5**

Raw Score	Proficiency Estimates	SE	SS ¹	SE(SS)	SS-1SE (SS) ²	SS+1SE (SS) ²
0	-5.5057	2.0075	2	32	-	-
1	-4.0969	1.0151	2	16	-	-
2	-3.3732	0.7286	2	12	-	-
3	-2.9367	0.6039	3	10	-	-
4	-2.6173	0.5310	8	8	-	-
5	-2.3619	0.4823	12	8	4	20
6	-2.1466	0.4472	15	7	8	22
7	-1.9587	0.4206	18	7	11	25
8	-1.7908	0.3997	21	6	15	27
9	-1.6379	0.3829	23	6	17	29
10	-1.4966	0.3692	25	6	19	31
11	-1.3645	0.3579	28	6	22	34
12	-1.2399	0.3485	30	6	24	36
13	-1.1212	0.3406	31	5	26	36
14	-1.0076	0.3340	33	5	28	38
15	-0.8978	0.3285	35	5	30	40
16	-0.7915	0.3239	37	5	32	42
17	-0.6878	0.3202	38	5	33	43
18	-0.5862	0.3172	40	5	35	45
19	-0.4864	0.3148	41	5	36	46
20	-0.3880	0.3131	43	5	38	48
21	-0.2902	0.3120	45	5	40	50
22	-0.1931	0.3115	46	5	41	51
23	-0.0960	0.3115	48	5	43	53
24	0.0011	0.3122	49	5	44	54
25	0.0990	0.3133	51	5	46	56
26	0.1976	0.3151	52	5	47	57
27	0.2976	0.3175	54	5	49	59
28	0.3995	0.3206	55	5	50	60
29	0.5033	0.3244	57	5	52	62
30	0.6102	0.3291	59	5	54	64
31	0.7202	0.3347	61	5	56	66
32	0.8343	0.3413	62	5	57	67
33	0.9535	0.3493	64	6	58	70
34	1.0787	0.3587	66	6	60	72
35	1.2114	0.3701	68	6	62	74
36	1.3533	0.3838	71	6	65	77
37	1.5069	0.4005	73	6	67	79
38	1.6755	0.4214	76	7	69	83
39	1.8640	0.4480	79	7	72	86
40	2.0801	0.4831	82	8	74	90
41	2.3364	0.5318	86	8	78	94
42	2.6567	0.6047	91	10	-	-
43	3.0941	0.7292	98	12	-	-
44	3.8186	1.0155	98	16	-	-
45	5.2282	2.0076	98	32	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

**Table 4.5.5 The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion
Table: Grade 6**

Raw Score	Proficiency Estimates	SE	SS ¹	SE(SS)	SS-1SE (SS) ²	SS+1SE (SS) ²
0	-5.3858	2.0071	2	37	-	-
1	-3.9781	1.0143	2	19	-	-
2	-3.2560	0.7274	2	13	-	-
3	-2.8211	0.6025	2	11	-	-
4	-2.5036	0.5294	2	10	-	-
5	-2.2499	0.4804	7	9	-	-
6	-2.0366	0.4450	11	8	3	19
7	-1.8506	0.4182	14	8	6	22
8	-1.6848	0.3971	17	7	10	24
9	-1.5340	0.3802	20	7	13	27
10	-1.3947	0.3663	22	7	15	29
11	-1.2649	0.3549	25	7	18	32
12	-1.1423	0.3453	27	6	21	33
13	-1.0259	0.3373	29	6	23	35
14	-0.9144	0.3306	31	6	25	37
15	-0.8070	0.3250	33	6	27	39
16	-0.7030	0.3203	35	6	29	41
17	-0.6017	0.3165	37	6	31	43
18	-0.5025	0.3134	39	6	33	45
19	-0.4050	0.3110	41	6	35	47
20	-0.3089	0.3093	42	6	36	48
21	-0.2137	0.3081	44	6	38	50
22	-0.1189	0.3076	46	6	40	52
23	-0.0244	0.3076	48	6	42	54
24	0.0704	0.3082	49	6	43	55
25	0.1659	0.3094	51	6	45	57
26	0.2621	0.3113	53	6	47	59
27	0.3597	0.3137	55	6	49	61
28	0.4591	0.3168	57	6	51	63
29	0.5607	0.3207	58	6	52	64
30	0.6650	0.3255	60	6	54	66
31	0.7728	0.3312	62	6	56	68
32	0.8847	0.3379	64	6	58	70
33	1.0015	0.3460	67	6	61	73
34	1.1245	0.3556	69	7	62	76
35	1.2550	0.3671	71	7	64	78
36	1.3948	0.3810	74	7	67	81
37	1.5463	0.3980	77	7	70	84
38	1.7129	0.4191	80	8	72	88
39	1.8997	0.4460	83	8	75	91
40	2.1139	0.4813	87	9	78	96
41	2.3685	0.5303	92	10	-	-
42	2.6872	0.6034	98	11	-	-
43	3.1232	0.7283	98	13	-	-
44	3.8465	1.0149	98	19	-	-
45	5.2553	2.0074	98	37	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

**Table 4.5.6. The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion
Table: Grade 7**

Raw Score	Proficiency Estimates	SE	SS ¹	SE(SS)	SS-1SE (SS) ²	SS+1SE (SS) ²
0	-5.6365	2.0084	2	34	-	-
1	-4.2246	1.0170	2	17	-	-
2	-3.4971	0.7312	2	13	-	-
3	-3.0568	0.6069	2	10	-	-
4	-2.7339	0.5343	2	9	-	-
5	-2.4750	0.4857	5	8	-	-
6	-2.2566	0.4507	9	8	-	-
7	-2.0657	0.4241	12	7	5	19
8	-1.8949	0.4032	15	7	8	22
9	-1.7394	0.3864	18	7	11	25
10	-1.5955	0.3726	21	6	15	27
11	-1.4610	0.3612	23	6	17	29
12	-1.3340	0.3516	25	6	19	31
13	-1.2132	0.3436	27	6	21	33
14	-1.0976	0.3368	29	6	23	35
15	-0.9861	0.3311	31	6	25	37
16	-0.8781	0.3263	33	6	27	39
17	-0.7729	0.3224	35	6	29	41
18	-0.6700	0.3192	36	5	31	41
19	-0.5690	0.3167	38	5	33	43
20	-0.4694	0.3147	40	5	35	45
21	-0.3707	0.3134	42	5	37	47
22	-0.2728	0.3127	43	5	38	48
23	-0.1750	0.3126	45	5	40	50
24	-0.0772	0.3130	47	5	42	52
25	0.0209	0.3139	48	5	43	53
26	0.1200	0.3155	50	5	45	55
27	0.2202	0.3178	52	5	47	57
28	0.3221	0.3207	53	6	47	59
29	0.4260	0.3243	55	6	49	61
30	0.5326	0.3288	57	6	51	63
31	0.6425	0.3343	59	6	53	65
32	0.7563	0.3408	61	6	55	67
33	0.8751	0.3486	63	6	57	69
34	0.9998	0.3580	65	6	59	71
35	1.1319	0.3692	67	6	61	73
36	1.2731	0.3828	70	7	63	77
37	1.4259	0.3995	72	7	65	79
38	1.5937	0.4204	75	7	68	82
39	1.7812	0.4470	79	8	71	87
40	1.9963	0.4821	82	8	74	90
41	2.2515	0.5308	87	9	78	96
42	2.5707	0.6037	92	10	-	-
43	3.0070	0.7284	98	13	-	-
44	3.7303	1.0149	98	17	-	-
45	5.1389	2.0073	98	34	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

**Table 4.5.7. The 2010 Mod-MSA, Reading: Total Raw Score to Scale Score Conversion
Table: Grade 8**

Raw Score	Proficiency Estimates	SE	SS ¹	SE(SS)	SS-1SE (SS) ²	SS+1SE (SS) ²
0	-5.3918	2.0078	2	33	-	-
1	-3.9819	1.0156	2	17	-	-
2	-3.2570	0.7294	2	12	-	-
3	-2.8193	0.6048	2	10	-	-
4	-2.4990	0.5319	3	9	-	-
5	-2.2427	0.4830	7	8	-	-
6	-2.0268	0.4478	10	7	3	17
7	-1.8385	0.4210	13	7	6	20
8	-1.6704	0.4000	16	7	9	23
9	-1.5173	0.3831	19	6	13	25
10	-1.3760	0.3692	21	6	15	27
11	-1.2440	0.3577	23	6	17	29
12	-1.1195	0.3481	25	6	19	31
13	-1.0012	0.3400	27	6	21	33
14	-0.8880	0.3332	29	6	23	35
15	-0.7790	0.3275	31	5	26	36
16	-0.6733	0.3227	33	5	28	38
17	-0.5705	0.3187	34	5	29	39
18	-0.4700	0.3155	36	5	31	41
19	-0.3712	0.3130	38	5	33	43
20	-0.2739	0.3112	39	5	34	44
21	-0.1774	0.3099	41	5	36	46
22	-0.0817	0.3092	43	5	38	48
23	0.0139	0.3091	44	5	39	49
24	0.1096	0.3096	46	5	41	51
25	0.2057	0.3107	47	5	42	52
26	0.3027	0.3123	49	5	44	54
27	0.4010	0.3147	50	5	45	55
28	0.5009	0.3177	52	5	47	57
29	0.6030	0.3214	54	5	49	59
30	0.7077	0.3260	56	5	51	61
31	0.8158	0.3316	57	5	52	62
32	0.9279	0.3382	59	6	53	65
33	1.0449	0.3462	61	6	55	67
34	1.1680	0.3557	63	6	57	69
35	1.2984	0.3671	65	6	59	71
36	1.4381	0.3808	68	6	62	74
37	1.5894	0.3977	70	7	63	77
38	1.7558	0.4187	73	7	66	80
39	1.9420	0.4454	76	7	69	83
40	2.1558	0.4808	79	8	71	87
41	2.4097	0.5296	84	9	75	93
42	2.7276	0.6027	89	10	-	-
43	3.1628	0.7276	96	12	-	-
44	3.8851	1.0144	98	17	-	-
45	5.2929	2.0071	98	33	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

Raw Score to Scale Score Conversion Tables for the Subscales

Table 4.5.8. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 3

Subscale Strand	Raw Score	Scale Score ¹ (SS)	Standard Error (SEM)	(SS – 1SEM) ²	(SS + 1SEM) ²
General Reading	0	2	28	-	-
General Reading	1	5	14	-	-
General Reading	2	15	11	4	26
General Reading	3	22	9	13	31
General Reading	4	28	8	20	36
General Reading	5	32	8	24	40
General Reading	6	36	7	29	43
General Reading	7	40	7	33	47
General Reading	8	44	7	37	51
General Reading	9	48	7	41	55
General Reading	10	52	7	45	59
General Reading	11	56	8	48	64
General Reading	12	60	8	52	68
General Reading	13	66	9	57	75
General Reading	14	73	11	62	84
General Reading	15	84	14	70	98
General Reading	16	98	28	-	-
Literary	0	2	28	-	-
Literary	1	7	14	-	-
Literary	2	18	11	7	29
Literary	3	25	9	16	34
Literary	4	30	8	22	38
Literary	5	35	8	27	43
Literary	6	40	8	32	48
Literary	7	44	8	36	52
Literary	8	48	8	40	56
Literary	9	53	8	45	61
Literary	10	57	8	49	65
Literary	11	63	9	54	72
Literary	12	70	11	59	81
Literary	13	81	15	66	96
Literary	14	98	28	-	-
Informational	0	2	28	-	-
Informational	1	11	14	-	-
Informational	2	22	11	11	33
Informational	3	29	9	20	38
Informational	4	34	8	26	42
Informational	5	39	8	31	47
Informational	6	43	7	36	50
Informational	7	47	7	40	54
Informational	8	51	7	44	58
Informational	9	55	7	48	62
Informational	10	59	8	51	67
Informational	11	64	8	56	72
Informational	12	69	9	60	78
Informational	13	76	11	65	87
Informational	14	87	14	-	-
Informational	15	98	28	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

Table 4.5.9. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 4

Subscale Strand	Raw Score	Scale Score ¹ (SS)	Standard Error (SEM)	(SS – 1SEM) ²	(SS + 1SEM) ²
General Reading	0	2	31	-	-
General Reading	1	2	16	-	-
General Reading	2	11	12	-	-
General Reading	3	19	10	9	29
General Reading	4	25	9	16	34
General Reading	5	30	9	21	39
General Reading	6	35	8	27	43
General Reading	7	39	8	31	47
General Reading	8	44	8	36	52
General Reading	9	48	8	40	56
General Reading	10	53	9	44	62
General Reading	11	58	9	49	67
General Reading	12	64	10	54	74
General Reading	13	72	12	60	84
General Reading	14	84	16	-	-
General Reading	15	98	31	-	-
Literary	0	2	31	-	-
Literary	1	7	16	-	-
Literary	2	19	12	7	31
Literary	3	27	10	17	37
Literary	4	33	9	24	42
Literary	5	38	9	29	47
Literary	6	42	8	34	50
Literary	7	47	8	39	55
Literary	8	51	8	43	59
Literary	9	55	8	47	63
Literary	10	60	9	51	69
Literary	11	65	9	56	74
Literary	12	71	10	61	81
Literary	13	79	12	67	91
Literary	14	90	16	-	-
Literary	15	98	31	-	-
Informational	0	2	31	-	-
Informational	1	7	16	-	-
Informational	2	19	12	7	31
Informational	3	27	10	17	37
Informational	4	33	9	24	42
Informational	5	38	9	29	47
Informational	6	43	8	35	51
Informational	7	48	8	40	56
Informational	8	52	8	44	60
Informational	9	57	8	49	65
Informational	10	62	9	53	71
Informational	11	67	9	58	76
Informational	12	73	10	63	83
Informational	13	81	12	69	93
Informational	14	93	16	-	-
Informational	15	98	31	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

Table 4.5.10. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 5

Subscale Strand	Raw Score	Scale Score ¹ (SS)	Standard Error (SEM)	(SS – 1SEM) ²	(SS + 1SEM) ²
General Reading	0	2	32	-	-
General Reading	1	2	16	-	-
General Reading	2	7	12	-	-
General Reading	3	15	10	5	25
General Reading	4	21	9	12	30
General Reading	5	26	9	17	35
General Reading	6	31	8	23	39
General Reading	7	35	8	27	43
General Reading	8	40	8	32	48
General Reading	9	44	8	36	52
General Reading	10	49	9	40	58
General Reading	11	54	9	45	63
General Reading	12	60	10	50	70
General Reading	13	68	12	56	80
General Reading	14	80	16	64	96
General Reading	15	98	32	-	-
Literary	0	2	32	-	-
Literary	1	6	16	-	-
Literary	2	18	12	6	30
Literary	3	26	10	16	36
Literary	4	32	9	23	41
Literary	5	38	9	29	47
Literary	6	42	9	33	51
Literary	7	47	8	39	55
Literary	8	51	8	43	59
Literary	9	56	9	47	65
Literary	10	61	9	52	70
Literary	11	66	9	57	75
Literary	12	72	10	62	82
Literary	13	80	12	68	92
Literary	14	93	17	-	-
Literary	15	98	32	-	-
Informational	0	2	32	-	-
Informational	1	11	16	-	-
Informational	2	23	12	11	35
Informational	3	31	10	21	41
Informational	4	37	9	28	46
Informational	5	43	9	34	52
Informational	6	47	9	38	56
Informational	7	52	8	44	60
Informational	8	56	8	48	64
Informational	9	61	9	52	70
Informational	10	66	9	57	75
Informational	11	71	9	62	80
Informational	12	77	10	67	87
Informational	13	85	12	73	97
Informational	14	97	16	-	-
Informational	15	98	32	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

Table 4.5.11. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 6

Subscale Strand	Raw Score	Scale Score ¹ (SS)	Standard Error (SEM)	(SS – 1SEM) ²	(SS + 1SEM) ²
General Reading	0	2	37	-	-
General Reading	1	2	19	-	-
General Reading	2	7	14	-	-
General Reading	3	16	12	4	28
General Reading	4	23	11	12	34
General Reading	5	29	10	19	39
General Reading	6	35	10	25	45
General Reading	7	40	10	30	50
General Reading	8	46	10	36	56
General Reading	9	51	10	41	61
General Reading	10	57	11	46	68
General Reading	11	63	11	52	74
General Reading	12	71	12	59	83
General Reading	13	80	14	66	94
General Reading	14	95	19	-	-
General Reading	15	98	37	-	-
Literary	0	2	37	-	-
Literary	1	2	19	-	-
Literary	2	12	14	-	-
Literary	3	21	12	9	33
Literary	4	28	11	17	39
Literary	5	34	10	24	44
Literary	6	40	10	30	50
Literary	7	45	10	35	55
Literary	8	50	10	40	60
Literary	9	55	10	45	65
Literary	10	60	10	50	70
Literary	11	67	11	56	78
Literary	12	74	12	62	86
Literary	13	83	14	69	97
Literary	14	97	19	-	-
Literary	15	98	37	-	-
Informational	0	2	37	-	-
Informational	1	2	19	-	-
Informational	2	14	14	-	-
Informational	3	23	12	11	35
Informational	4	30	11	19	41
Informational	5	36	10	26	46
Informational	6	42	10	32	52
Informational	7	47	10	37	57
Informational	8	52	10	42	62
Informational	9	58	10	48	68
Informational	10	63	10	53	73
Informational	11	69	11	58	80
Informational	12	77	12	65	89
Informational	13	86	14	-	-
Informational	14	98	19	-	-
Informational	15	98	37	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

Table 4.5.12. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 7

Subscale Strand	Raw Score	Scale Score ¹ (SS)	Standard Error (SEM)	(SS – 1SEM) ²	(SS + 1SEM) ²
General Reading	0	2	35	-	-
General Reading	1	2	18	-	-
General Reading	2	3	13	-	-
General Reading	3	12	11	-	-
General Reading	4	19	10	9	29
General Reading	5	25	10	15	35
General Reading	6	30	10	20	40
General Reading	7	36	9	27	45
General Reading	8	41	9	32	50
General Reading	9	46	10	36	56
General Reading	10	51	10	41	61
General Reading	11	58	11	47	69
General Reading	12	65	12	53	77
General Reading	13	73	13	60	86
General Reading	14	87	18	-	-
General Reading	15	98	35	-	-
Literary	0	2	35	-	-
Literary	1	2	18	-	-
Literary	2	9	13	-	-
Literary	3	17	11	6	28
Literary	4	24	10	14	34
Literary	5	30	10	20	40
Literary	6	35	9	26	44
Literary	7	40	9	31	49
Literary	8	45	9	36	54
Literary	9	50	9	41	59
Literary	10	55	10	45	65
Literary	11	61	10	51	71
Literary	12	68	11	57	79
Literary	13	76	13	63	89
Literary	14	90	18	-	-
Literary	15	98	35	-	-
Informational	0	2	35	-	-
Informational	1	4	18	-	-
Informational	2	18	13	5	31
Informational	3	26	11	15	37
Informational	4	33	10	23	43
Informational	5	39	10	29	49
Informational	6	44	9	35	53
Informational	7	49	9	40	58
Informational	8	54	9	45	63
Informational	9	58	9	49	67
Informational	10	64	10	54	74
Informational	11	69	10	59	79
Informational	12	76	11	65	87
Informational	13	84	13	71	97
Informational	14	98	18	-	-
Informational	15	98	35	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

Table 4.5.13. The 2010 Mod-MSA, Reading: Subscale Total of Raw Score to Scale Score Conversion Table: Grade 8

Subscale Strand	Raw Score	Scale Score ¹ (SS)	Standard Error (SEM)	(SS – 1SEM) ²	(SS + 1SEM) ²
General Reading	0	2	33	-	-
General Reading	1	2	17	-	-
General Reading	2	8	13	-	-
General Reading	3	17	11	6	28
General Reading	4	23	10	13	33
General Reading	5	29	9	20	38
General Reading	6	34	9	25	43
General Reading	7	38	9	29	47
General Reading	8	43	9	34	52
General Reading	9	47	9	38	56
General Reading	10	52	9	43	61
General Reading	11	57	9	48	66
General Reading	12	62	10	52	72
General Reading	13	69	11	58	80
General Reading	14	77	13	64	90
General Reading	15	90	17	-	-
General Reading	16	98	33	-	-
Literary	0	2	33	-	-
Literary	1	2	17	-	-
Literary	2	12	13	-	-
Literary	3	20	11	9	31
Literary	4	27	10	17	37
Literary	5	32	9	23	41
Literary	6	37	9	28	46
Literary	7	42	9	33	51
Literary	8	47	9	38	56
Literary	9	52	9	43	61
Literary	10	58	10	48	68
Literary	11	64	11	53	75
Literary	12	72	13	59	85
Literary	13	85	17	-	-
Literary	14	98	33	-	-
Informational	0	2	34	-	-
Informational	1	2	17	-	-
Informational	2	11	13	-	-
Informational	3	20	11	9	31
Informational	4	26	10	16	36
Informational	5	32	10	22	42
Informational	6	38	9	29	47
Informational	7	43	9	34	52
Informational	8	48	9	39	57
Informational	9	52	9	43	61
Informational	10	58	9	49	67
Informational	11	63	10	53	73
Informational	12	70	11	59	81
Informational	13	78	13	65	91
Informational	14	91	17	-	-
Informational	15	98	33	-	-

Note. 1. LOSS was set to 2 while the HOSS was set at 98

2. Because of the ceiling effect set by the LOSS and HOSS, the confidence intervals set by the standard errors *may* not follow the expected pattern of equal or progressively larger bandwidth as one moves up and down the extreme ends of the scoring continuum. This would also be the case when the standard error is larger than the estimated scale score, and one would have to force the ceiling effect to counter negative score values at the lower end or higher than the ceiling values at the upper end of the bandwidth. These values are, therefore, left blank.

4.6. Score Interpretation

Interpretation of the 2010 Mod-MSA: Reading test scores depends primarily on the understanding of the scale score and the performance level descriptors.

Scale Scores

As explained in section 4.5, Reporting Scale Scores for the 2010 Mod-MSA: Reading, the tests produced scale scores that ranged between 2 and 98. These scale scores have the same meaning within the same grade, but are not comparable across grade levels.

It should be noted that for scale scores, a higher score simply means a higher performance on the reading tests. Performance levels and descriptions can then be used to give specific interpretation to the scale scores because they are developed to bring meaning to those scale scores.

Performance Level Descriptors

As explained previously, performance level descriptors provide specific information about students' performance levels and help interpret the 2010 Mod-MSA: Reading scale scores. They describe what students at a particular level generally know and can be applicable to all students within each grade level.

Maryland standards are divided into three levels of achievement (<http://mdk12.org/instruction/curriculum/index.html>):

- Advanced is a highly challenging and exemplary level of achievement indicating outstanding accomplishment in meeting students' needs.
- Proficient is a realistic and rigorous level of achievement indicating proficiency in meeting students' needs.
- Basic is a level of achievement indicating that more work is needed to attain proficiency in meeting students' needs.

The proficient levels described above were translated as classification scale score cuts through a standard setting procedure discussed in Appendix D.

4.7. Final Performance Level Cut Points for the Mod-MSA: Reading

For grade 3-5 a standard setting procedure was undertaken (see Appendix D) to obtain the cuts at the performance levels. The final cut points adopted by MSDE for the 2010 administration of the Mod-MSA: Reading test, grades 3-5 in raw score points, scale score, and theta metric were adjusted by the executive committee. There are two cut points that correspond to the three performance levels discussed above. Any score below the proficient cut point is the basic performance level.

Table 4.7.1 contains information about the cutoff scale score of each performance level. It should be noted that the same cutoff scores set by the standard setting procedure in 2009 for grades 6-8 were applied in 2010.

Table 4.7.1 Mod-MSA: Reading Scale Score Cut Scores: Grades 3 through 8

Grade	Cut Scores at Performance Levels	
	Proficient	Advanced
3	54	64
4	53	65
5	53	69
6	54	67
7	56	72
8	54	66