

APPENDIX C: YEAR 2006 MSA-MATH RECALIBRATION RESULTS FROM 3PL IRT TO THE RASCH MODEL USING EQUIPERCENTILE METHOD

It was required to replace the original calibration and equating IRT model (e.g., the 3PL) due to a change in the administrative structure of a program. Replacing the original model was undertaken with an eye that takes into account the inherent differences that exist between any two IRT models with an effort at preserving, at a minimum, the distribution of the performance classifications of the original model.

Because the data sets were originally run by the 3PL equating model, the 3PL scale scores were considered to be the base or the original scores. The Rasch model was then run to generate the new ability estimates. The equipercentile equating method was applied to link the two types of ability estimates, and the new Rasch ability estimates were linearly transformed to the new reporting scale scores. First, the distribution characteristics of the new scale scores were investigated. Other measures were also calculated to assess the consistency of performance classifications between the two models. These measures include correlation coefficients, kappa indices, overall performance level results, and overall raw score agreement indices.

The goal of equipercentile equating is to have at least some of the same score distribution characteristics in a population of examinees (Kolen & Brennan, 1995) when two tests are placed on the established scale. The equipercentile equating principle is applied in the following manner: For a given Form X score, the percentage of examinees earning scores at or below that Form X score is obtained. Next, the Form Y score that has the same percentage of examinees at or below that observed on Form X is obtained. The scores on Form X and Form Y that provide the same percent of students at or below their respective scores are considered to be equivalent, and Forms X and Y are equated. Thus, the distribution of scores on Form X converted to the Form Y would be equal to the distribution of scores on Form Y in the population at particular score points because the equipercentile function is developed by identifying scores on Form X that have the same percentile ranks as scores on Form Y (Kolen & Brennan, 1995).

The test of each grade had two operational forms and composed of five content standards across all grades. The number of items and score points for each standard were identical between the two operational forms within each grade. Tables C.1 through C.6 show the number of items that were included in each operational form with respect to content standards. Specifically, Table C.7 indicates how many common items appeared on both operational test forms. These common items were used for the purpose of form-to-form calibration and equating.

Each mixed-format operational form with SR, SPR, BCR and ECR within each grade was recalibrated with the dichotomous Rasch (Rasch, 1960) and the Rasch Partial Credit (Masters, 1982) models for the SR and SPR and the BCR and ECR items respectively. Form A of each grade was chosen as a base form, and the common items which appeared across two forms were screened using robust z and Rasch difficulty plots (“b-plots”) (SCDE, 2001) for determining their use as linking items. In addition, correlation coefficients as well as standard deviation ratio were also used for the purpose of the screening. Tables C.8 through C.13 contain more information on robust z values and correlations, and screening guidelines can be obtained from section 1.10, Linking, Equating, and Scaling Procedures. Once the useable linking items were identified from the list of common items, the two operational forms were equated using a fixed item parameter method. The result of this procedure put the two forms within each grade on the same scale.

Now that each form within grades was on the same scale, the Rasch ability estimate for each student was obtained, which in turn had to be equated with their previously estimated ability estimate based on the 3PL model.

Since ability estimates are seldom, if ever, reported directly to the examinees, the new ability estimates are linearly transformed by the use of a multiplicative and additive scaling constant so that they can be used as reporting scale scores. The new reporting scale scores have the same meaning of the original scale scores in terms of the performance cut scores and levels.

Equipercenile equating principle was applied to link and equate the two types of ability estimates. First, the percent of students at or below the two scale score proficient cuts, Basic/Proficient and Proficient/Advanced for the 3PL model were obtained. The theta location of these cuts were matched against their respective scale scores defined as $SS(B/P)$ and $SS(P/A)$ for the Basic/Proficient and Proficient/Advanced., respectively. Next, the Rasch ability estimates (defined as $\Theta(B/P)$ and $\Theta(P/A)$ for the Basic/Proficient and Proficient/Advanced cuts respectively) that had the same percentage of examinees at or below the cuts obtained from the 3PL model were obtained.

Given these two sets of cuts, the slope and the intercept were calculated such that

$$SS(B/P) = slope \times \Theta(B/P) + intercept$$

and

$$SS(P/A) = slope \times \Theta(P/A) + intercept$$

The slope and intercept obtained from the two equations above were used to transform the Rasch ability estimate into a Rasch-based scale score for each student in the original data sets. Applying this process produced a Rasch-based scale score system that matched well with the 3PL results with respect to the distribution of students for the Basic, Proficient, and Advanced performance classification categories. Table C.14 shows the slope and intercept of each grade that were obtained for calculating the Rasch scale scores.

The equipercenile method discussed above ensured the similarity in student distribution by performance category classification when the the 3PL IRT model was replaced by the Rasch model. However, in order to establish the accuracy and stability of the model transformation, the central moments of the Rasch scale scores were compared with those of the original 3PL scale scores. As shown in Table C.16, the results indicate that the distribution characteristics of the new Rasch scale scores were very similar to those of the original 3PL scale scores.

To further compare the two types of scale scores, Tukey plots were used as per Huynh (2006). The plots depicted in Figures C1 through C12 compare the cumulative distribution functions (CDFs) for the 3PL and Rasch scale scores and examines the percent and the cumulative percent differences between the two CDFs. As shown in figures, the “smoothness” of the 3PL CDF due to the pattern scoring vs. the step function CDF of the Rasch CDF can be observed. In general, however, there were no real differences between the two CDFs except at the low scale scores for the cumulative percent differences in grades 4 through 8.

As seen from Table C.17, the Pearson-product correlation coefficients between the 3PL and the Rasch scale scores ranged from .98 to .99. The results clearly indicate an almost perfect liner correlation between the two types of scale scores.

One of the main purposes of this study was to investigate how consistently the Rasch model could preserve the original performance levels of the 3PL model. Table C.18 shows the performance classifications of each grade. The results document that the Rasch model preserved the original performance levels as closely as possible in spite of the slightly increasing passing rates for the Rasch model across grades.

The Kappa Index of Agreement (K) which measures the association between the two models and helps evaluate the accuracy of classification results, was also calculated. K values range from -1 to +1 after adjustment for chance agreement. If the two models are in perfect agreement (i.e., if no change occurs), K equals 1. If the two models are completely different, K would equal -1. If the change in the results of the two models occurred by chance, then Kappa would equal 0. As seen in Table C.19, Kappa indices for all grades indicate that the agreement rate between the 3PL and the Rasch models were in excess of 0.90 across all grades.

Table C.20 shows the overall raw agreement rate of each grade. The results indicated that the overall performance levels assigned to students based on the Rasch model matched well with those of the 3PL model across all grades (from 95% to 96%). Tables C.21 through C.23 show the raw agreement rate of each performance level between the 3PL and the Rasch models.

A comparison of scale score distributions, correlation coefficients between scale scores, kappa indices, overall performance level results, and overall raw score agreement indices documented that the distribution of student scores of the original 3PL equating model remained similar when the item and ability estimates were transferred to the Rasch model via equipercentile equating.

Table C.1 Year 2006 Grade 3 Item Type and Score Points Distribution

Form	# of TeraNova	# of CRT SR	# of CRT BCR		Points of TeraNova	Points of CRT SR	Points of CRT BCR		Total Score
			Pt A	Pt B			Pt A	Pt B	
Form 1	11	39	7	7	11	39	7	14	71
Form 2	11	39	7	7	11	39	7	14	71

Table C.2 Year 2006 Grade 4 Item Type and Score Point Distribution

Form	# of TeraNova	# of CRT SR	# of CRT BCR		Points of TeraNova	Points of CRT SR	Points of CRT BCR		Total Score
			Pt A	Pt B			Pt A	Pt B	
Form 1	10	39	7	7	10	39	7	14	70
Form 2	10	40	7	7	10	40	7	14	71

Table C.3 Year 2006 Grade 5 Item Type and Score Point Distribution

	# of TeraNova	# of CRT SR	# of CRT BCR		# of CRT ECR		Points of TeraNova	Points of CRT SR	Points of CRT BCR		Points of CRT ECR		Total Score
			Pt A	Pt B	Pt A	Pt B			Pt A	Pt B	Pt A	Pt B	
F 1	13	36	7	7	1	1	13	36	7	14	1	3	74
F 2	13	36	7	7	1	1	13	36	7	14	1	3	74

Table C.4 Year 2005 Grade 6 Item Type and Score Point Distribution

	# of TeraNova	# of CRT SR	# of CRT BCR		# of CRT ECR		Points of TeraNova	Points of CRT SR	Points of CRT BCR		Points of CRT ECR		Total Score
			Pt A	Pt B	Pt A	Pt B			Pt A	Pt B			
F 1	5	43	6	6	1	1	5	43	6	12	1	3	70
F 2	5	43	6	6	1	1	5	43	6	12	1	3	70

Table C.5 Year 2006 Grade 7 Item Type and Score Point Distribution

	# of TeraNova	# of CRT SR	# of CRT SPR	# of CRT BCR		# of CRT ECR		Points of TeraNova	Points of CRT SR	Points of CRT SPR	Points of CRT BCR		Points of CRT ECR		Total Score
				Pt A	Pt B	Pt A	Pt B				Pt A	Pt B			
F 1	6	30	12	4	4	3	3	6	30	12	4	8	3	9	72
F 2	6	30	12	4	4	3	3	6	30	12	4	8	3	9	72

Table C.6 Year 2006 Grade 8 Item Type and Score Point Distribution

	# of TeraNova	# of CRT SR	# of CRT SPR	# of CRT BCR		# of CRT ECR		Points of TeraNova	Points of CRT SR	Points of CRT SPR	Points of CRT BCR		Points of CRT ECR		Total Score
				Pt A	Pt B	Pt A	Pt B				Pt A	Pt B			
F 1	11	25	12	5	5	3	3	11	25	12	5	10	3	9	75
F 2	11	25	12	5	5	3	3	11	25	12	5	10	3	9	75

Table C.7 Year-to-Year Common and Unique Items of Two Operational Forms

Grade	Form	Terra Nova	MD Common	Total Common	Unique Item	Total Items
3	1	11	27	38	26	64
	2	11	27	38	26	64
4	1	10	22	32	31	63
	2	10	22	32	32	64
5	1	13	27	40	25	65
	2	13	27	40	25	65
6	1	5	26	31	31	62
	2	5	26	31	31	62
7	1	6	28	34	28	62
	2	6	28	34	28	62
8	1	11	27	38	26	64
	2	11	27	38	26	64

Table C.8 Free Calibration Item Difficulties of Linking Items and Robust Z Values: Grade 3

Item Sequential Number	Y06 Form 1	Y06 Form 2	Item Sequential Number	Item Type	11	12	Robust Z
1	-0.5774	-0.538	1	SR	0.00	0.04	.6047
2	-1.3414	-1.3852	2	SR	0.00	-0.04	-.7495
3	-0.989	-0.9818	3	SR	0.00	0.01	.0806
4	-1.0159	-1.0036	4	SR	0.00	0.01	.1636
5	-1.0182	-1.0667	5	SR	0.00	-0.05	-.8260
6	-0.7445	-0.7199	6	SR	0.00	0.02	.3638
7	-1.006	-1.0167	7	SR	0.00	-0.01	-.2108
8	-2.384	-2.4273	8	SR	0.00	-0.04	-.7414
9	-1.231	-1.2693	9	SR	0.00	-0.04	-.6600
10	-2.6951	-2.7146	10	SR	0.00	-0.02	-.3540
11	-2.8	-2.923	11	SR	0.00	-0.12	-2.0386
12	0.9627	1.0445	12	SR	0.00	0.08	1.2948
13	0.7154	0.7913	13	SR	0.00	0.08	1.1988
15	-1.3766	-1.4465	15	SR	0.00	-0.07	-1.1743
16	1.8411	1.9914	16	SR	0.00	0.15	2.4097
17	-0.3242	-0.3574	17	SR	0.00	-0.03	-.5770
18	-1.3667	-1.4108	18	SR	0.00	-0.04	-.7544
19	-0.036	0.0342	19	SR	0.00	0.07	1.1060
20	-0.7332	-0.7336	20	SR	0.00	0.00	-.0431
25	1.2257	1.2649	46	SR	0.00	0.04	.6014
26	0.069	0.1579	26	SR	0.00	0.09	1.4104
28	0.2953	0.2867	30	SR	0.00	-0.01	-.1766
29	-0.1123	-0.1629	33	SR	0.00	-0.05	-.8602
31	-0.5906	-0.6251	31	SR	0.00	-0.03	-.5982
32	-1.3693	-1.7309	32	SR	0.00	-0.36	-5.9222
34	-0.6165	-0.605	34	SR	0.00	0.01	.1506
35	-1.819	-1.8221	35	SR	0.00	0.00	-.0871
36	0.0444	0.0604	36	SR	0.00	0.02	.2238
37	-0.5231	-0.3197	37	SR	0.00	0.20	3.2740
38	1.4814	1.6202	40	SR	0.00	0.14	2.2225
39	-0.2691	-0.2642	38	SR	0.00	0.00	.0431
42	-0.3652	-0.3302	42	SR	0.00	0.04	.5331
43	0.4861	0.5486	43	SR	0.00	0.06	.9807
44	1.3184	1.0151	44	SR	0.00	-0.30	-4.9733
46	0.0425	-0.0727	47	SR	0.00	-0.12	-1.9117
48	2.8084	2.9233	27	SR	0.00	0.11	1.8335
49	-2.6459	-2.8129	49	SR	0.00	-0.17	-2.7548
50	0.9317	0.9462	14	SR	0.00	0.01	.1994

Form Statistics

Mean	-.414	-.423
SD	1.270	1.319

Comparison of Each Form with Base Form (Form 1)

Correlation with Base	1.000	.997
SD ratio	100%	104%
Mean Diff	.000	-.009
Median Diff	.000	.002
IQR Diff	.000	.083

Rasch Item Difficulties of Common Items: Grade 3

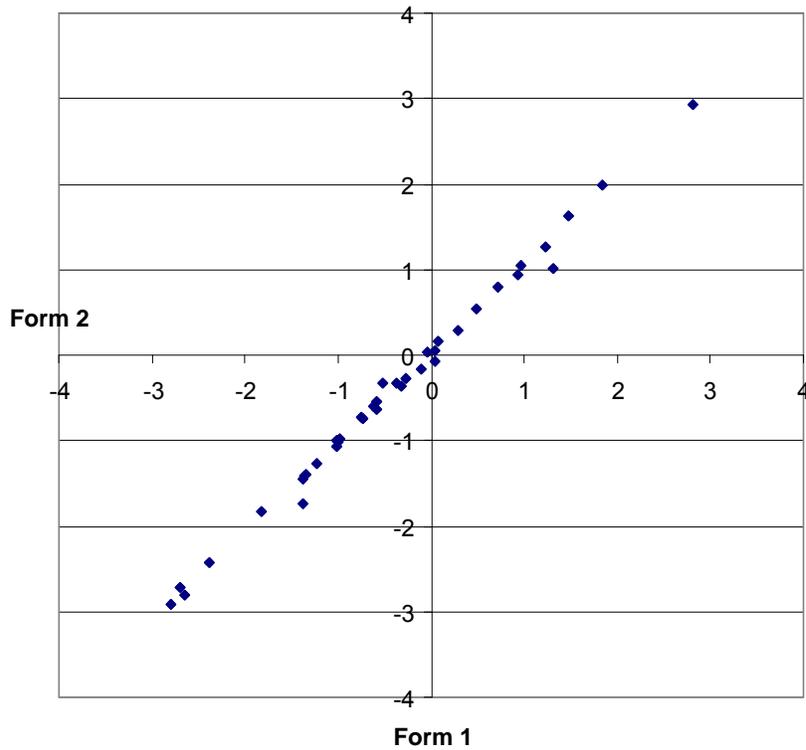


Table C.9 Free Calibration Item Difficulties of Linking Items and Robust Z Values: Grade 4

Item Sequential Number	Y06 Form 1	Y06 Form 2	Item Sequential Number	Item Type	11	12	Robust Z
1	-1.0694	-1.1474	1	SR	0.00	-0.08	.0863
2	-0.0237	-0.0964	2	SR	0.00	-0.07	.1686
3	-1.2632	-1.3609	3	SR	0.00	-0.10	-.2199
4	0.3659	0.2488	4	SR	0.00	-0.12	-.5214
5	0.3731	0.328	5	SR	0.00	-0.05	.5976
6	-2.0134	-2.1005	6	SR	0.00	-0.09	-.0552
7	0.5759	0.4523	7	SR	0.00	-0.12	-.6224
8	-0.3652	-0.4619	8	SR	0.00	-0.10	-.2044
9	0.2603	0.163	9	SR	0.00	-0.10	-.2137
10	0.8463	0.7501	10	SR	0.00	-0.10	-.1966
12	-0.799	-0.8578	11	SR	0.00	-0.06	.3847
13	0.1763	0.0483	14	SR	0.00	-0.13	-.6908
14	-1.055	-1.1127	16	SR	0.00	-0.06	.4018
18	0.7782	0.6761	17	SR	0.00	-0.10	-.2883
19	0.5403	0.3789	18	SR	0.00	-0.16	-1.2099
21	-1.7288	-1.8064	22	SR	0.00	-0.08	.0925
24	-0.7475	-0.6248	25	SR	0.00	0.12	3.2055
25	-2.1248	-2.1129	26	SR	0.00	0.01	1.4835
28	-0.9767	-1.0475	28	SR	0.00	-0.07	.1982
30	-1.7626	-1.7783	29	SR	0.00	-0.02	1.0545
31	0.7468	0.6104	30	SR	0.00	-0.14	-.8214
34	-0.3554	0.1357	35	SR	0.00	0.49	8.9310
35	-1.2169	-1.3526	36	SR	0.00	-0.14	-.8105
39	-0.2743	-0.4401	39	SR	0.00	-0.17	-1.2783
40	-0.8464	-0.7931	41	SR	0.00	0.05	2.1269
41	-0.0497	-0.1297	42	SR	0.00	-0.08	.0552
44	0.8666	0.8699	45	SR	0.00	0.00	1.3498
45	-0.9395	-0.9391	46	SR	0.00	0.00	1.3047
48	-0.1077	-0.5185	49	SR	0.00	-0.41	-5.0860
49	0.5508	0.6046	50	SR	0.00	0.05	2.1347
52	-0.5937	-0.9446	53	CR	0.00	-0.35	-4.1551
53	1.9494	1.8006	54	CR	0.00	-0.15	-1.0141

Form Statistics

Mean	-.321	-.392
SD	.965	.955

Comparison of Each Form with Base Form (Form 1)

Correlation with Base	1.000	.989
SD ratio	100%	99%
Mean Diff	.000	-.071
Median Diff	.000	-.084
IQR Diff	.000	.087

Rasch Item Difficulties of Common Items: Grade 4

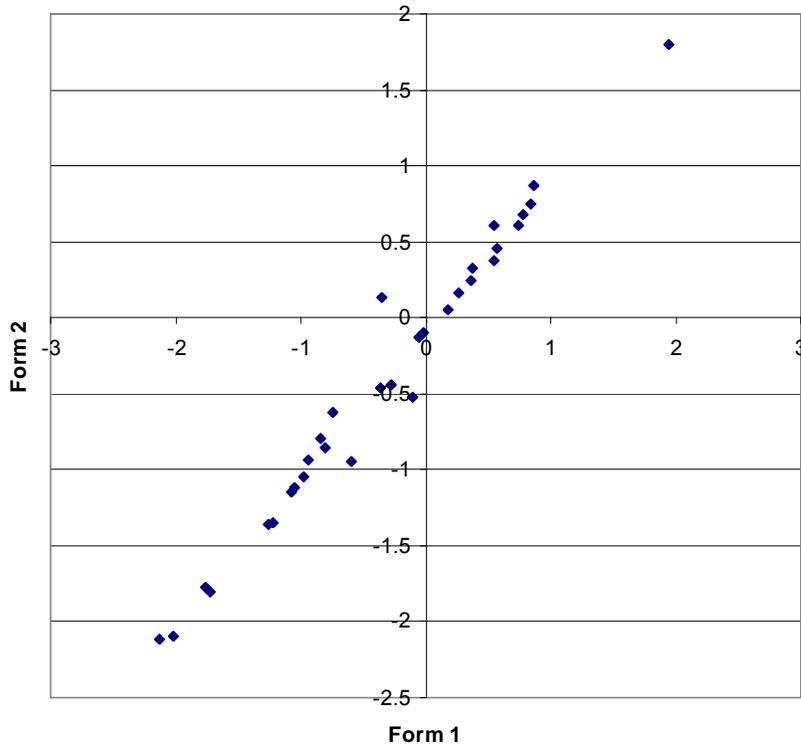


Table C.10 Free Calibration Item Difficulties of Linking Items and Robust Z Values: Grade 5

Item Sequential Number	Y06 Form 1	Y06 Form 2	Item Sequential Number	Item Type	11	12	Robust Z
1	-0.6672	-0.7886	1	SR	0.00	-0.12	.1533
2	-0.2842	-0.3872	2	SR	0.00	-0.10	.4870
3	-1.1777	-1.3356	3	SR	0.00	-0.16	-.5088
4	-0.0008	-0.1369	4	SR	0.00	-0.14	-.1134
5	-0.6322	-0.7773	5	SR	0.00	-0.15	-.2766
6	-0.9107	-1.0586	6	SR	0.00	-0.15	-.3274
7	0.077	-0.0283	7	SR	0.00	-0.11	.4453
8	-0.2025	-0.3463	8	SR	0.00	-0.14	-.2530
9	0.4557	0.2954	9	SR	0.00	-0.16	-.5523
10	-0.1595	-0.334	10	SR	0.00	-0.17	-.8099
11	-0.0496	-0.1994	11	SR	0.00	-0.15	-.3619
12	0.2015	0.0432	12	SR	0.00	-0.16	-.5161
13	-1.5434	-1.706	13	SR	0.00	-0.16	-.5941
16	0.203	0.1804	17	SR	0.00	-0.02	1.9454
17	0.3214	0.1003	16	SR	0.00	-0.22	-1.6552
19	-0.331	-0.4414	18	SR	0.00	-0.11	.3528
20	0.0148	-0.1637	20	SR	0.00	-0.18	-.8825
21	-1.0845	-1.1458	21	SR	0.00	-0.06	1.2434
22	1.5483	1.4255	22	SR	0.00	-0.12	.1279
23	1.5795	1.3911	23	SR	0.00	-0.19	-1.0620
24	-1.4191	-1.6077	24	SR	0.00	-0.19	-1.0657
25	0.6342	0.4653	25	SR	0.00	-0.17	-.7083
27	-1.6886	-1.6946	27	SR	0.00	-0.01	2.2465
28	0.8118	0.7498	28	SR	0.00	-0.06	1.2307
32	1.0449	0.9124	33	SR	0.00	-0.13	-.0481
33	-1.1516	-1.1424	37	SR	0.00	0.01	2.5222
34	-0.0507	-0.2289	36	SR	0.00	-0.18	-.8770
37	-0.5779	-0.7973	35	SR	0.00	-0.22	-1.6243
38	0.5383	0.5367	38	SR	0.00	0.00	2.3263
39	-0.6839	-0.7642	39	SR	0.00	-0.08	.8988

Item Sequential Number	Y06 Form 1	Y06 Form 2	Item Sequential Number	Item Type	11	12	Robust Z
41	-0.9093	-1.1286	41	SR	0.00	-0.22	-1.6225
42	-0.1826	0.001	42	SR	0.00	0.18	5.6857
43	-0.6898	-0.8144	43	SR	0.00	-0.12	.0952
44	0.6218	0.4527	44	SR	0.00	-0.17	-.7120
46	0.1746	0.0818	46	SR	0.00	-0.09	.6720
47	-1.255	-1.2204	47	SR	0.00	0.03	2.9829
48	-1.1293	-1.2424	48	SR	0.00	-0.11	.3038
49	0.2895	0.1785	49	SR	0.00	-0.11	.3419
62	1.7699	1.6427	62	CR	0.00	-0.13	.0481
63	2.2928	2.3586	63	CR	0.00	0.07	3.5489

Form Statistics

Mean	-.105	-.217
SD	.937	.942

Comparison of Each Form with Base Form (Form 1)

Correlation with Base	1.000	.996
SD ratio	100%	101%
Mean Diff	.000	-.112
Median Diff	.000	-.130
IQR Diff	.000	.074

Rasch Item Difficulties of Common Items: Grade 5

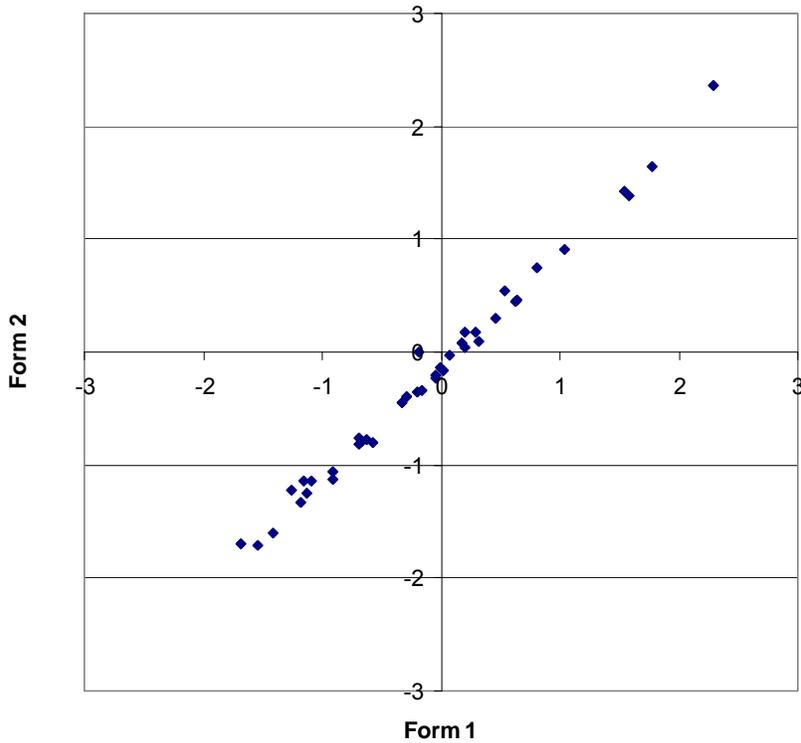


Table C.11 Free Calibration Item Difficulties of Linking Items and Robust Z Values: Grade 6

Item Sequential Number	Y06 Form 1	Y06 Form 2	Item Sequential Number	Item Type	11	12	Robust Z
1	-0.8419	-0.8551	1	SR	0.00	-0.01	-.1482
2	-1.3621	-1.3902	2	SR	0.00	-0.03	-.3805
3	-0.9964	-1.0001	3	SR	0.00	0.00	.0000
4	-1.0753	-1.0614	4	SR	0.00	0.01	.2745
5	-0.6919	-0.7209	5	SR	0.00	-0.03	-.3946
6	0.2409	0.2378	6	SR	0.00	0.00	.0094
8	1.2969	1.1553	8	SR	0.00	-0.14	-2.1506
9	-0.2844	-0.3255	10	SR	0.00	-0.04	-.5833
11	0.3674	0.339	11	SR	0.00	-0.03	-.3852
12	-0.7278	-0.7284	12	SR	0.00	0.00	.0483
14	-0.4703	-0.5421	14	SR	0.00	-0.07	-1.0621
15	0.135	0.1819	15	SR	0.00	0.05	.7891
19	0.6666	0.5973	16	SR	0.00	-0.07	-1.0231
20	0.8563	0.8737	20	SR	0.00	0.02	.3291
24	0.6406	0.7963	27	SR	0.00	0.16	2.4859
25	1.0083	0.9732	25	SR	0.00	-0.04	-.4897
26	0.1004	-0.0817	26	SR	0.00	-0.18	-2.7822
30	-0.4092	-0.4184	31	SR	0.00	-0.01	-.0858
31	0.658	0.6275	32	SR	0.00	-0.03	-.4180
32	-0.2581	-0.0766	35	SR	0.00	0.18	2.8883
35	-1.3362	-1.2695	33	SR	0.00	0.07	1.0979
36	-1.8302	-1.6454	37	SR	0.00	0.18	2.9398
37	-1.6189	-1.6172	36	SR	0.00	0.00	.0842
38	-0.0894	-0.0286	38	SR	0.00	0.06	1.0059
39	-0.7001	-0.5618	39	SR	0.00	0.14	2.2146
40	0.5144	0.279	40	SR	0.00	-0.24	-3.6135
43	0.5885	0.3203	44	SR	0.00	-0.27	-4.1250
44	0.4777	0.4634	43	SR	0.00	-0.01	-.1653
47	-0.7843	-0.7179	48	SR	0.00	0.07	1.0932
57	0.9049	1.0565	57	CR	0.00	0.15	2.4220
58	0.1675	0.1783	58	CR	0.00	0.01	.2261

Form Statistics

Mean	-.157	-.160
SD	.837	.808

Comparison of Each Form with Base Form (Form 1)

Correlation with Base	1.000	.992
SD ratio	100%	97%
Mean Diff	.000	-.003
Median Diff	.000	-.004
IQR Diff	.000	.087

Rasch Item Difficulties of Common Items: Grade 6

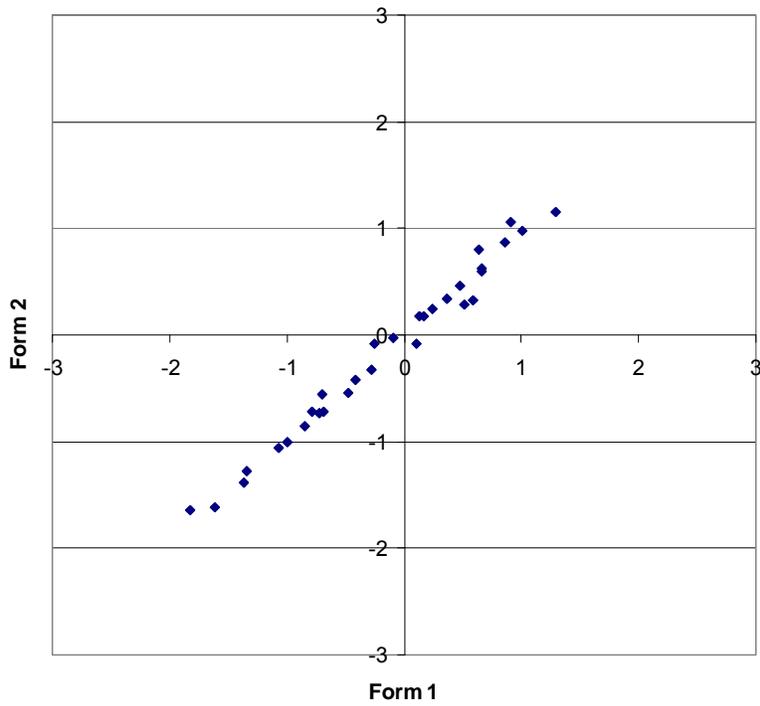


Table C.12 Free Calibration Item Difficulties of Linking Items and Robust Z Values: Grade 7

Item Sequential Number	Y06 Form 1	Y06 Form 2	Item Sequential Number	Item Type	11	12	Robust Z
1	-0.1196	-0.0841	1	SR	0.00	0.04	.3998
2	-1.3674	-1.3401	2	SR	0.00	0.03	.2429
3	-0.6924	-0.6801	3	SR	0.00	0.01	-.0440
4	-2.7317	-2.7636	4	SR	0.00	-0.03	-.8894
5	-0.4273	-0.4097	5	SR	0.00	0.02	.0574
6	-0.0772	-0.0868	6	SR	0.00	-0.01	-.4629
8	1.0539	1.1384	9	SR	0.00	0.08	1.3370
9	0.1508	0.1851	7	SR	0.00	0.03	.3768
10	-0.642	-0.6049	8	SR	0.00	0.04	.4304
12	-0.4706	-0.5243	16	SR	0.00	-0.05	-1.3064
13	-1.1551	-0.9829	10	SR	0.00	0.17	3.0145
15	-0.6035	-0.7209	12	SR	0.00	-0.12	-2.5248
16	-0.6621	-0.6575	11	SR	0.00	0.00	-.1913
17	-0.4683	-0.4628	19	SR	0.00	0.01	-.1741
18	-0.6359	-0.5132	20	SR	0.00	0.12	2.0677
21	0.1104	0.1878	14	SR	0.00	0.08	1.2012
23	0.9745	1.0655	22	SR	0.00	0.09	1.4613
25	-0.0583	-0.0755	25	SR	0.00	-0.02	-.6083
26	-1.4991	-1.5078	26	SR	0.00	-0.01	-.4457
27	-1.2172	-1.1718	27	SR	0.00	0.05	.5891
28	-1.2028	-1.1998	28	SR	0.00	0.00	-.2219
30	-0.7302	-0.8046	30	SR	0.00	-0.07	-1.7023
31	0.5663	0.5356	31	SR	0.00	-0.03	-.8665
32	0.0092	0.0321	32	SR	0.00	0.02	.1588
33	-0.4333	-0.4929	33	SR	0.00	-0.06	-1.4193
34	-0.2963	-0.4138	29	SR	0.00	-0.12	-2.5267
35	0.5231	0.4806	35	SR	0.00	-0.04	-1.0922
49	0.0932	0.1913	49	CR	0.00	0.10	1.5971
50	-0.22	-0.1841	50	CR	0.00	0.04	.4074
51	-0.6284	-0.6736	51	SPR	0.00	-0.05	-1.1438
53	0.2605	0.4123	53	SPR	0.00	0.15	2.6243
58	0.5245	0.5235	55	SPR	0.00	0.00	-.2984
59	1.7931	1.81	59	SPR	0.00	0.02	.0440
62	1.3895	1.4393	60	SPR	0.00	0.05	.6733

Form Statistics

Mean	-.261	-.246
SD	.883	.897

Comparison of Each Form with Base Form (Form 1)

Correlation with Base	1.000	.997
SD ratio	100%	102%
Mean Diff	.000	.016
Median Diff	.000	.015
IQR Diff	.000	.071

Rasch Item Difficulties of Common Items: Grade 7

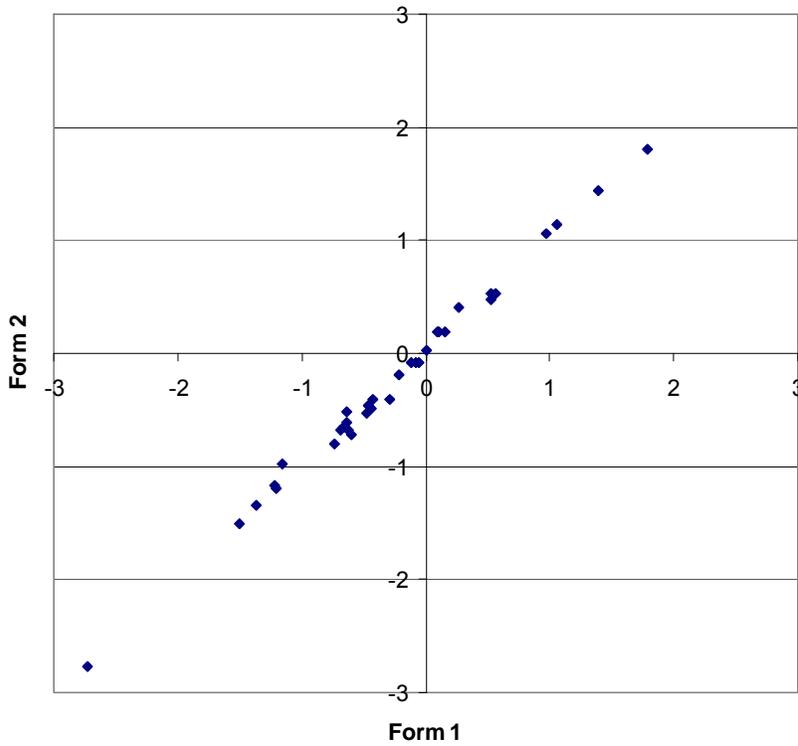


Table C.13 Free Calibration Item Difficulties of Linking Items and Robust Z Values: Grade 8

Item Sequential Number	Y06 Form 1	Y06 Form 2	Item Sequential Number	Item Type	11	12	Robust Z
1	-0.6252	-0.6178	1	SR	0.00	0.01	.7611
2	0.8284	0.7995	2	SR	0.00	-0.03	.0581
3	-0.9965	-1.0054	3	SR	0.00	-0.01	.4454
4	-0.8823	-0.9227	4	SR	0.00	-0.04	-.1646
5	-2.8306	-2.8601	5	SR	0.00	-0.03	.0465
6	-2.5699	-2.6687	6	SR	0.00	-0.10	-1.2957
7	0.2295	0.1833	7	SR	0.00	-0.05	-.2770
8	-1.5057	-1.5467	8	SR	0.00	-0.04	-.1762
9	0.3465	0.3217	9	SR	0.00	-0.02	.1375
10	-0.7124	-0.802	10	SR	0.00	-0.09	-1.1175
11	-0.4901	-0.4764	11	SR	0.00	0.01	.8831
12	-0.2177	-0.2203	12	SR	0.00	0.00	.5675
13	1.4965	1.4771	13	SR	0.00	-0.02	.2421
14	-1.3613	-1.4092	14	SR	0.00	-0.05	-.3099
15	-0.1452	-0.1581	18	SR	0.00	-0.01	.3680
16	-0.0881	-0.109	19	SR	0.00	-0.02	.2130
17	-0.1085	-0.1173	16	SR	0.00	-0.01	.4474
18	-1.2003	-1.2808	17	SR	0.00	-0.08	-.9412
19	-0.2581	-0.4144	15	SR	0.00	-0.16	-2.4093
20	-1.4852	-1.555	24	SR	0.00	-0.07	-.7340
22	1.0306	0.7796	20	SR	0.00	-0.25	-4.2434
23	-0.5815	-0.64	22	SR	0.00	-0.06	-.5152
24	0.5139	0.4796	21	SR	0.00	-0.03	-.0465
26	-0.4061	-0.464	27	SR	0.00	-0.06	-.5035
27	0.3257	0.3438	29	SR	0.00	0.02	.9684
28	-0.6275	-0.7134	26	SR	0.00	-0.09	-1.0458
29	0.1649	0.1207	28	SR	0.00	-0.04	-.2382
30	0.2379	0.2952	25	SR	0.00	0.06	1.7276
32	1.2102	1.0762	31	SR	0.00	-0.13	-1.9774
33	-1.0918	-1.0232	34	SR	0.00	0.07	1.9464
34	-0.4851	-0.4309	33	SR	0.00	0.05	1.6675
35	-0.533	-0.501	36	SR	0.00	0.03	1.2376

Item Sequential Number	Y06 Form 1	Y06 Form 2	Item Sequential Number	Item Type	11	12	Robust Z
51	1.0668	1.1663	51	CR	0.00	0.10	2.5449
52	0.7711	0.8069	52	CR	0.00	0.04	1.3112
59	1.6966	1.6296	58	SPR	0.00	-0.07	-0.6798
60	-0.3965	-0.5906	60	SPR	0.00	-0.19	-3.1414
61	0.4163	0.3718	62	SPR	0.00	-0.04	-0.2440
63	0.2569	0.3	64	SPR	0.00	0.04	1.4525

Form Statistics

Mean	-.237	-.273
SD	1.000	1.004

Comparison of Each Form with Base Form (Form 1)

Correlation with Base	1.000	.998
SD ratio	100%	100%
Mean Diff	.000	-.036
Median Diff	.000	-.032
IQR Diff	.000	.070

Rasch Item Difficulties of Common Items: Grade 8

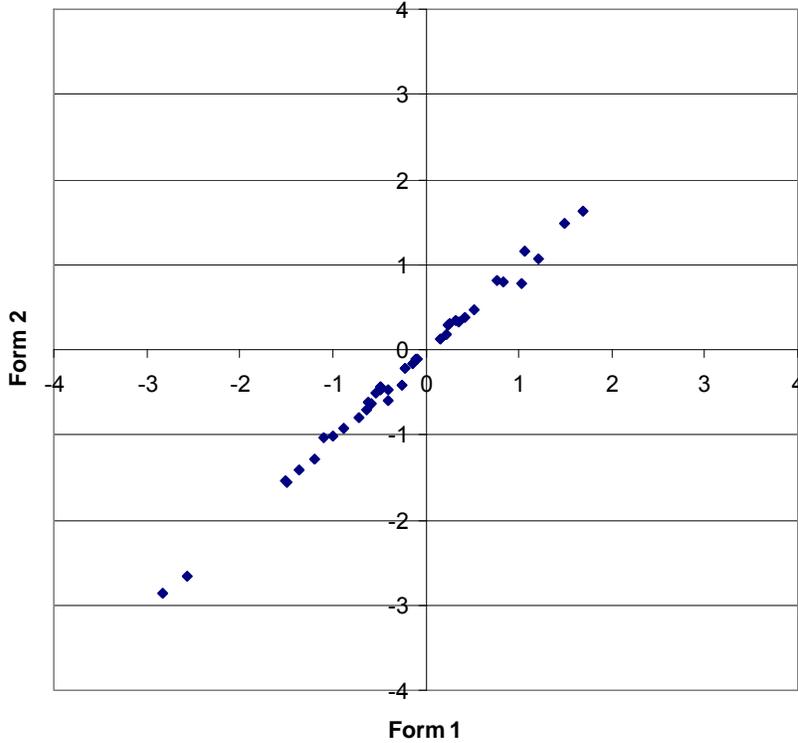


Table C.14 Rasch Equating Slope and Constant of 2006 MSA-Math

Grade	Slope	Intercept
3	32.6935	352.2959
4	32.8398	380.2954
5	30.7057	390.2866
6	29.6236	398.5595
7	28.1690	405.9549
8	28.3634	418.4843

Table C.15 Performance Level Cut Points of 2006 MSA-Math

Grade	Proficient	Advanced
3	379	441
4	374	433
5	392	453
6	396	447
7	396	451
8	407	444

Table C.16 Scale Score Moments between 3PL and 1PL of Each Grade

Grade	Model	M	SD	P10	Q1	Mdn	Q3	P90	IQR
3	3PL	411.06	43.64	356	384	413	440	463	56
	Rasch	411.57	42.40	357	385	414	441	463	56
4	3PL	410.47	43.54	355	385	414	440	462	55
	Rasch	412.83	40.46	359	386	413	441	465	55
5	3PL	414.91	45.14	360	389	418	445	468	56
	Rasch	417.96	38.63	370	390	417	443	469	53
6	3PL	406.27	48.39	349	383	412	439	460	56
	Rasch	411.44	38.36	364	385	411	439	460	54
7	3PL	402.02	50.92	338	374	408	438	461	64
	Rasch	408.17	41.85	357	378	406	438	464	60
8	3PL	408.10	47.74	352	383	412	440	464	57
	Rasch	414.78	39.63	369	388	411	440	468	52

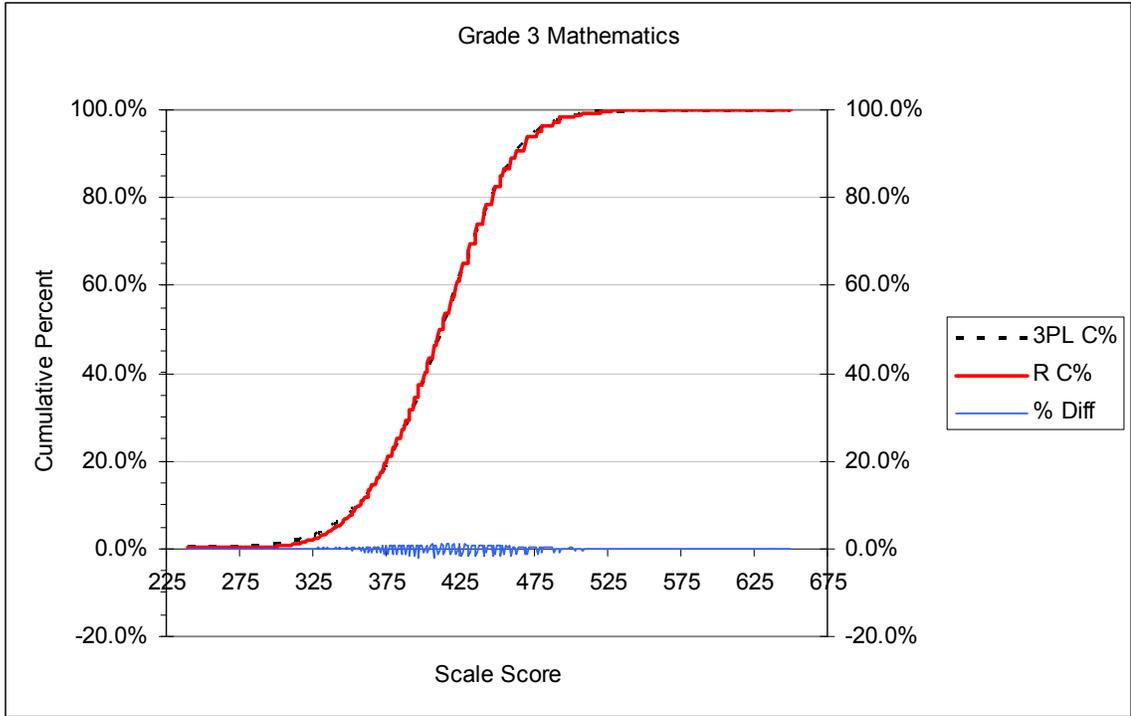


Figure C.1 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the percent differences between CDFs: Grade 3

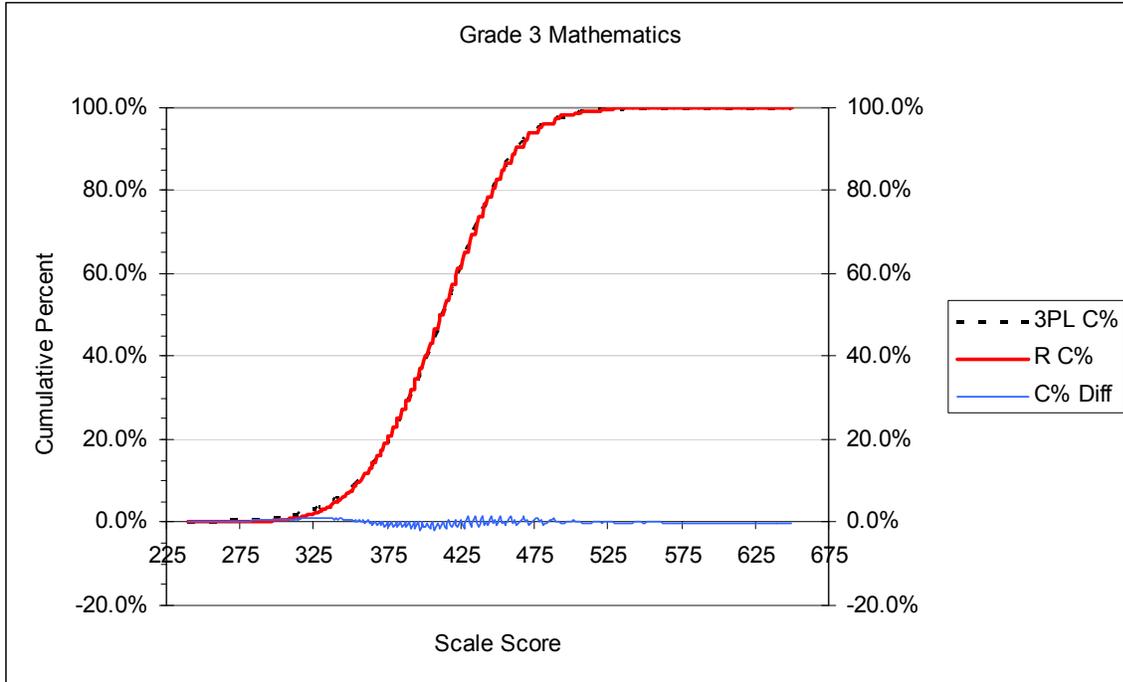


Figure C.2 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the cumulative percent differences between CDFs: Grade 3

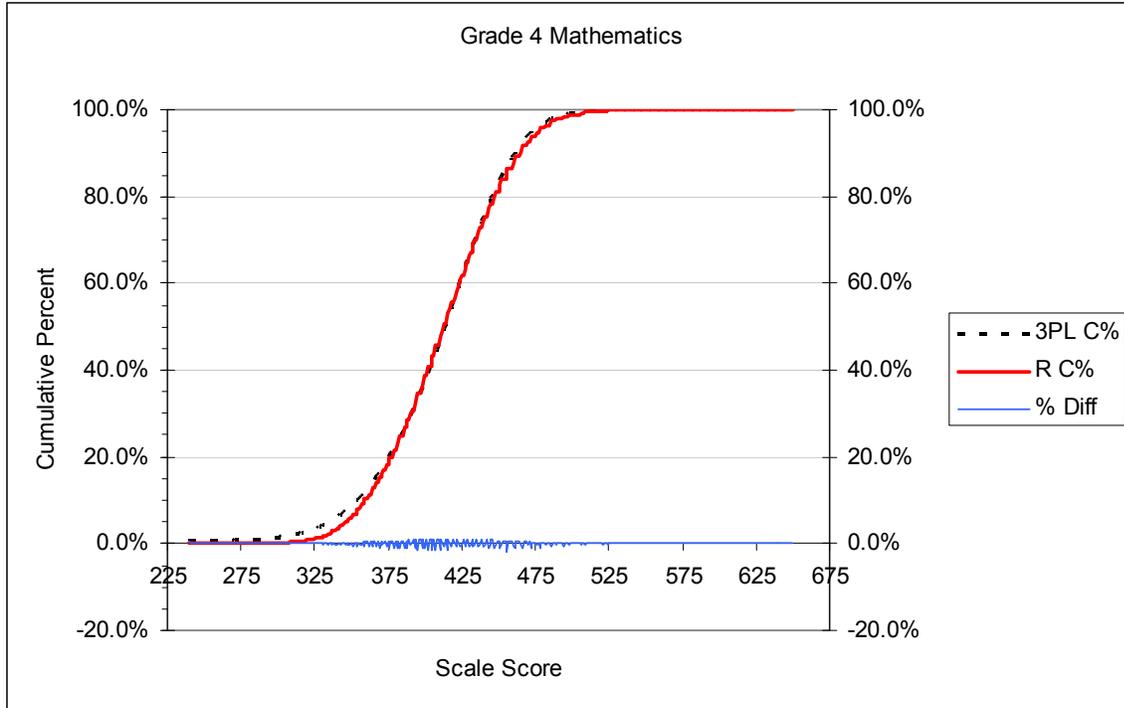


Figure C.3 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the percent differences between CDFs: Grade 4

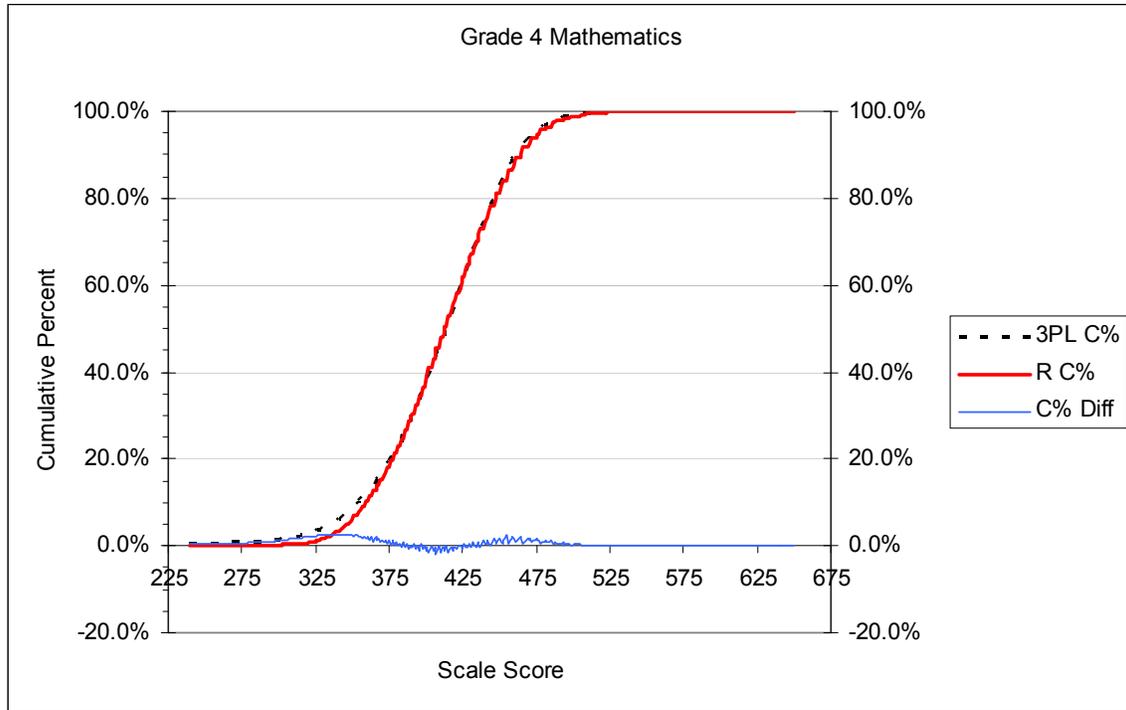


Figure C.4 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the cumulative percent differences between CDFs: Grade 4

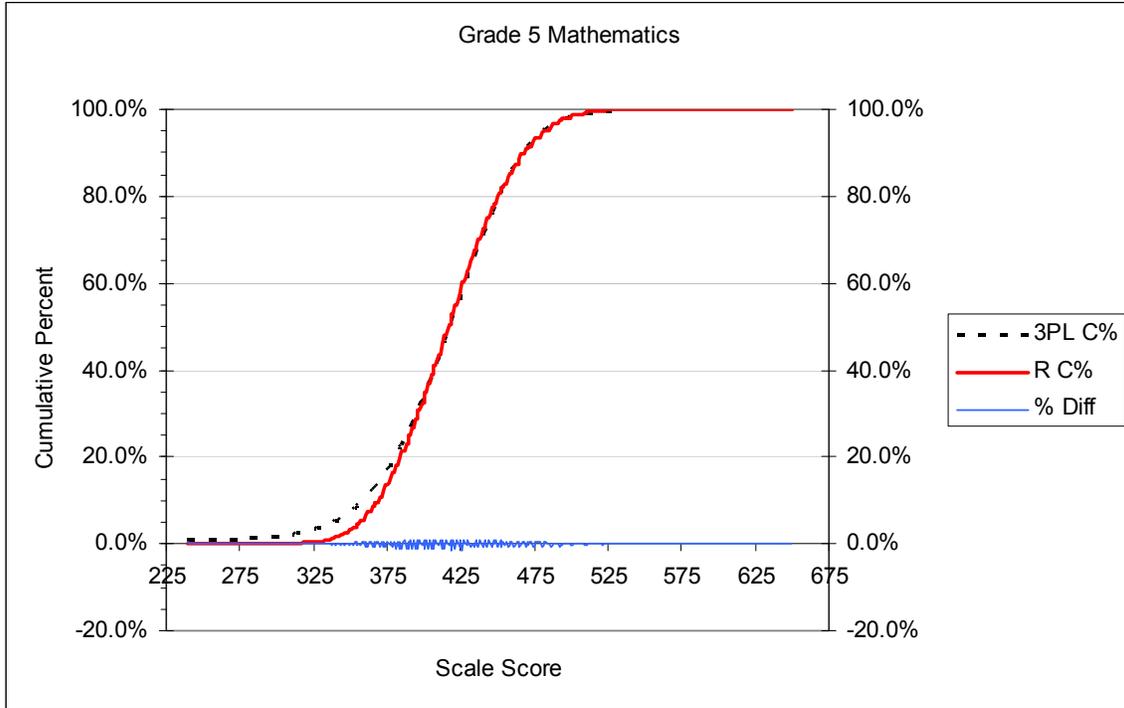


Figure C.5 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the percent differences between CDFs: Grade 5

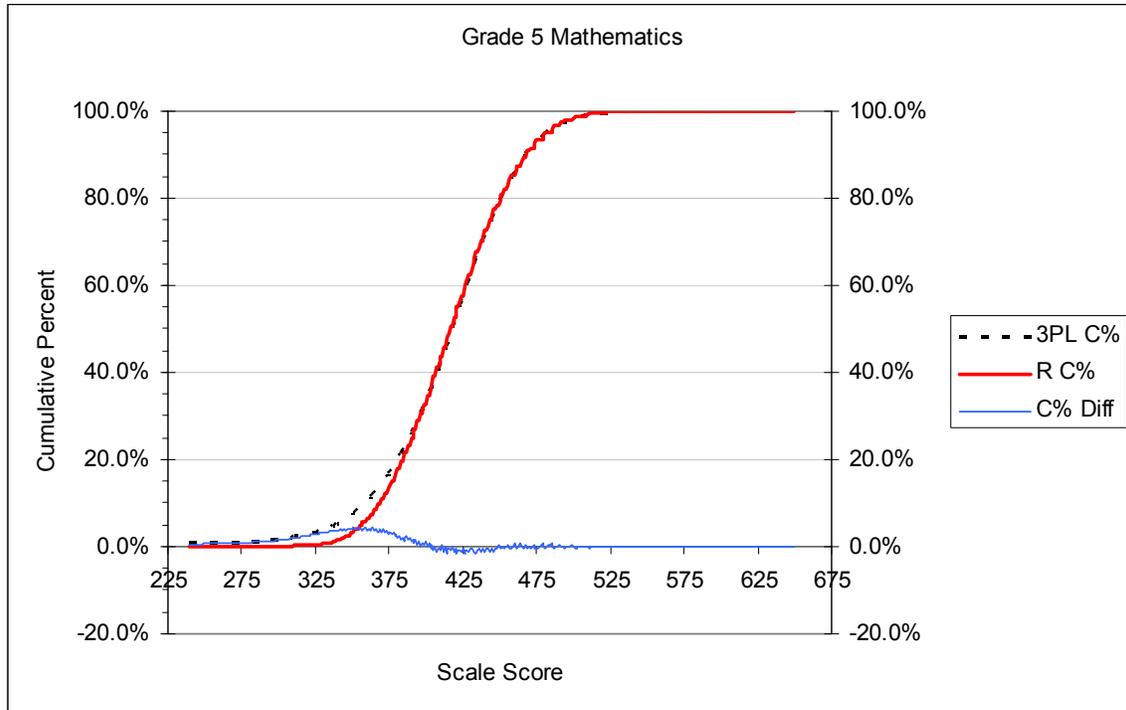


Figure C.6 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the cumulative percent differences between CDFs: Grade 5

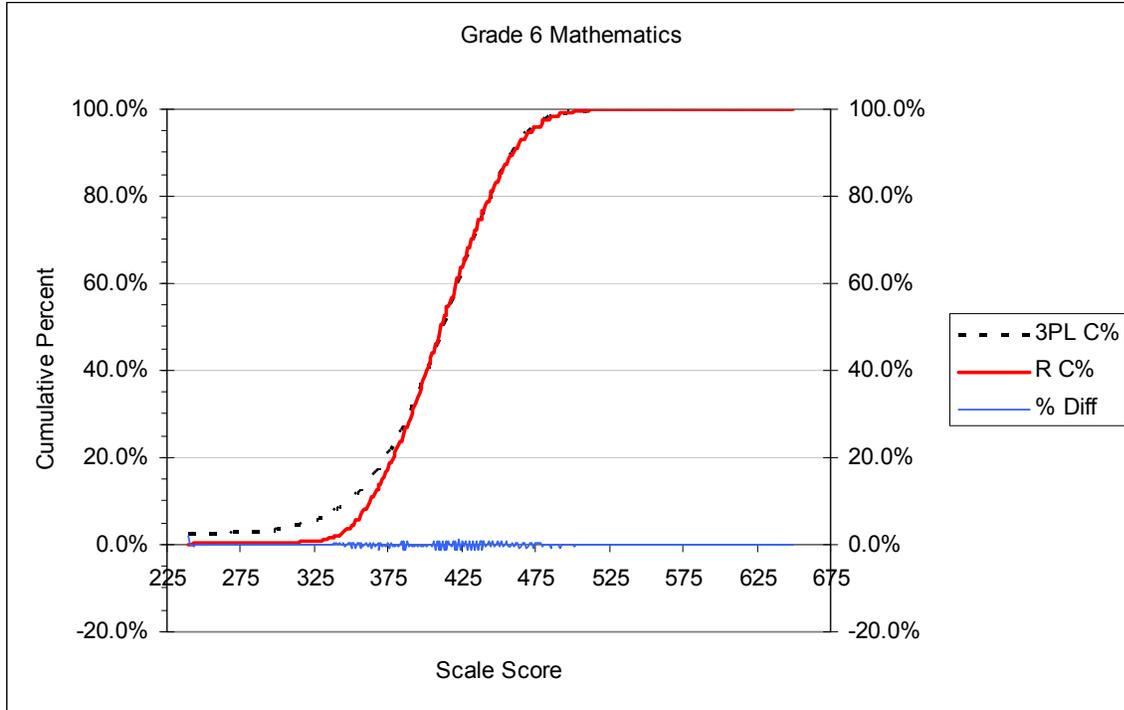


Figure C.7 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the percent differences between CDFs: Grade 6

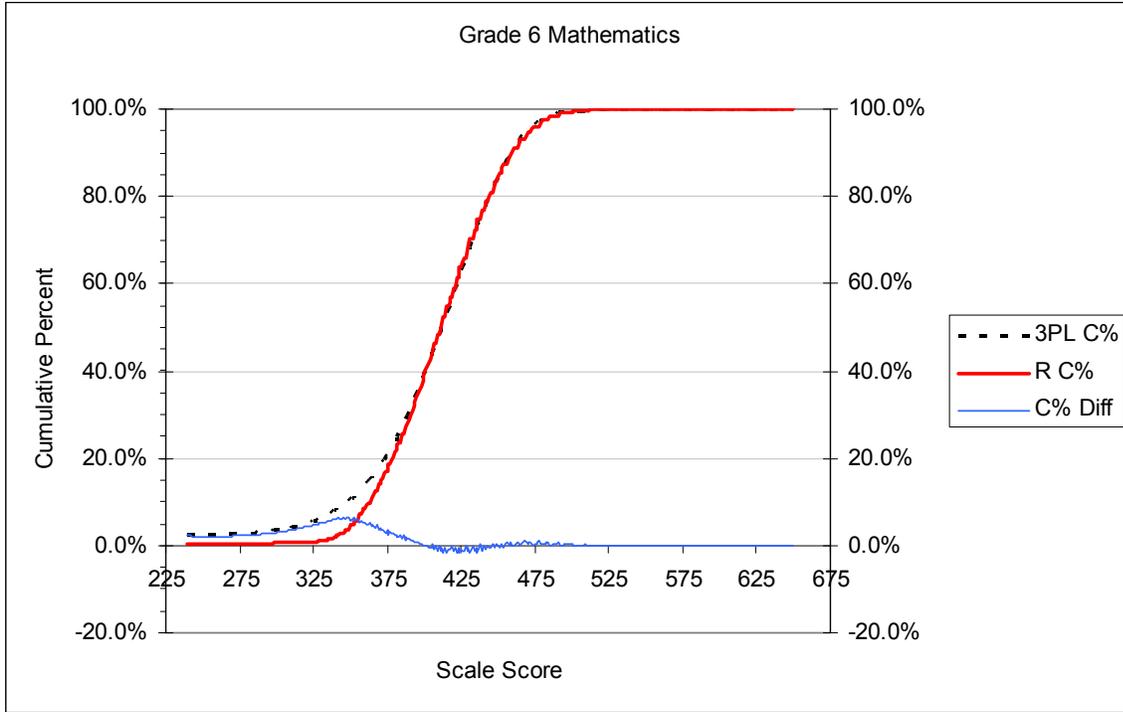


Figure C.8 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the cumulative percent differences between CDFs: Grade 6

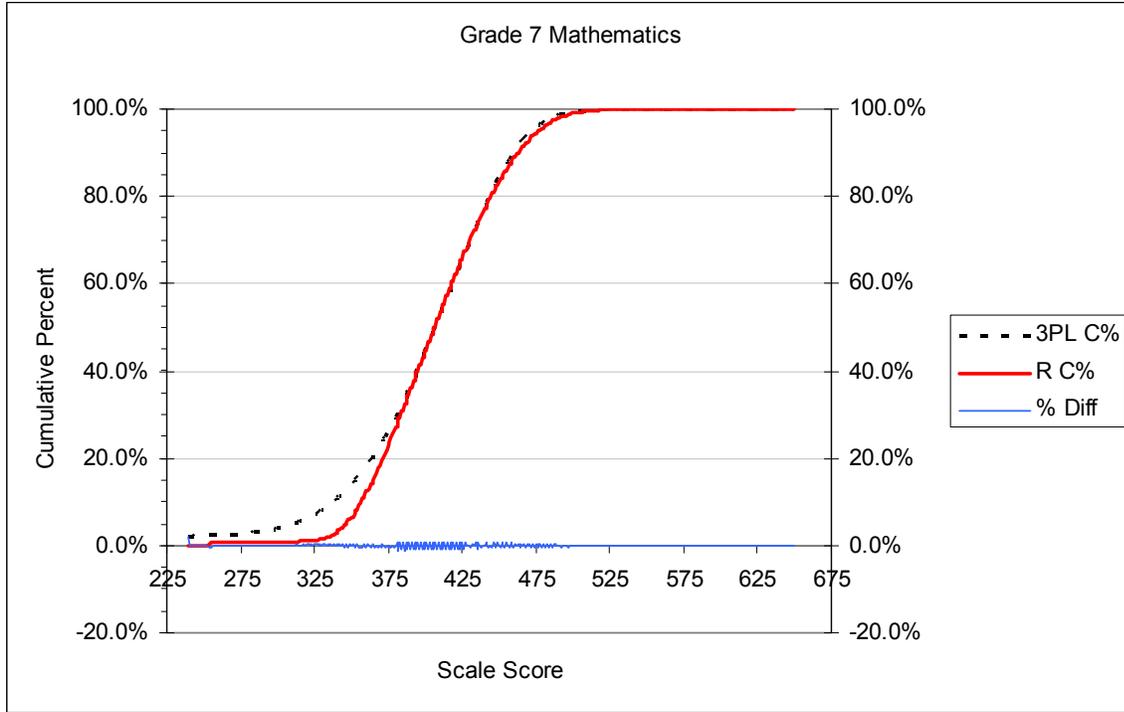


Figure C.9 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the percent differences between CDFs: Grade 7

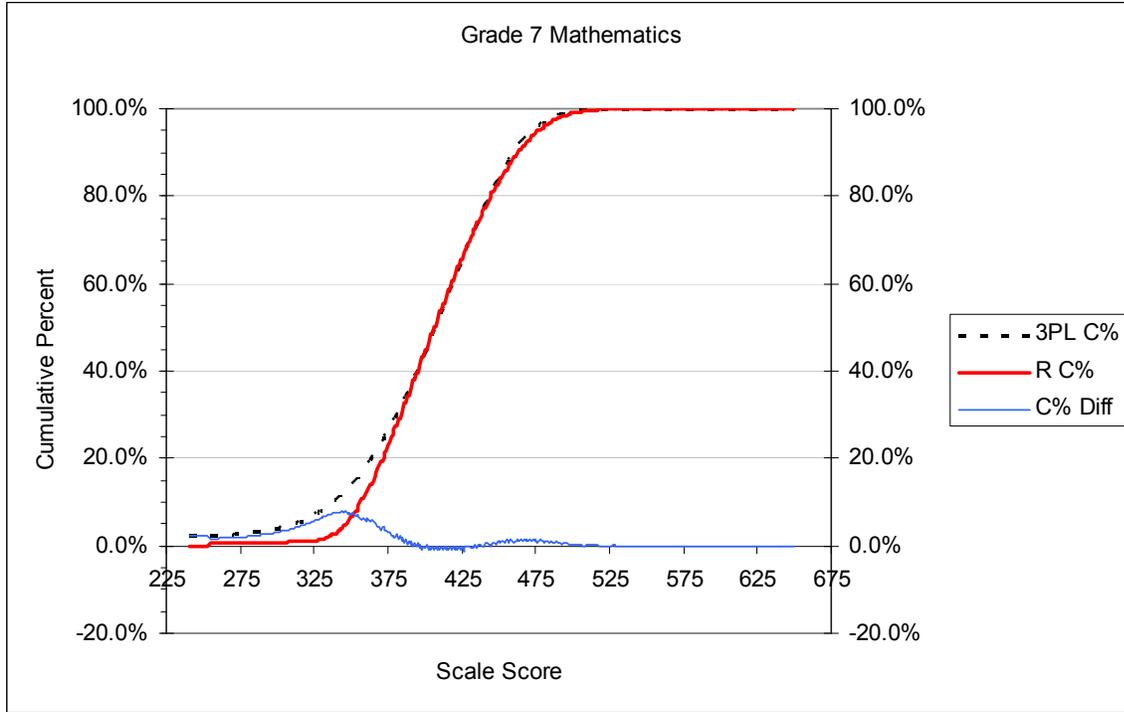


Figure C.10 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the cumulative percent differences between CDFs: Grade 7

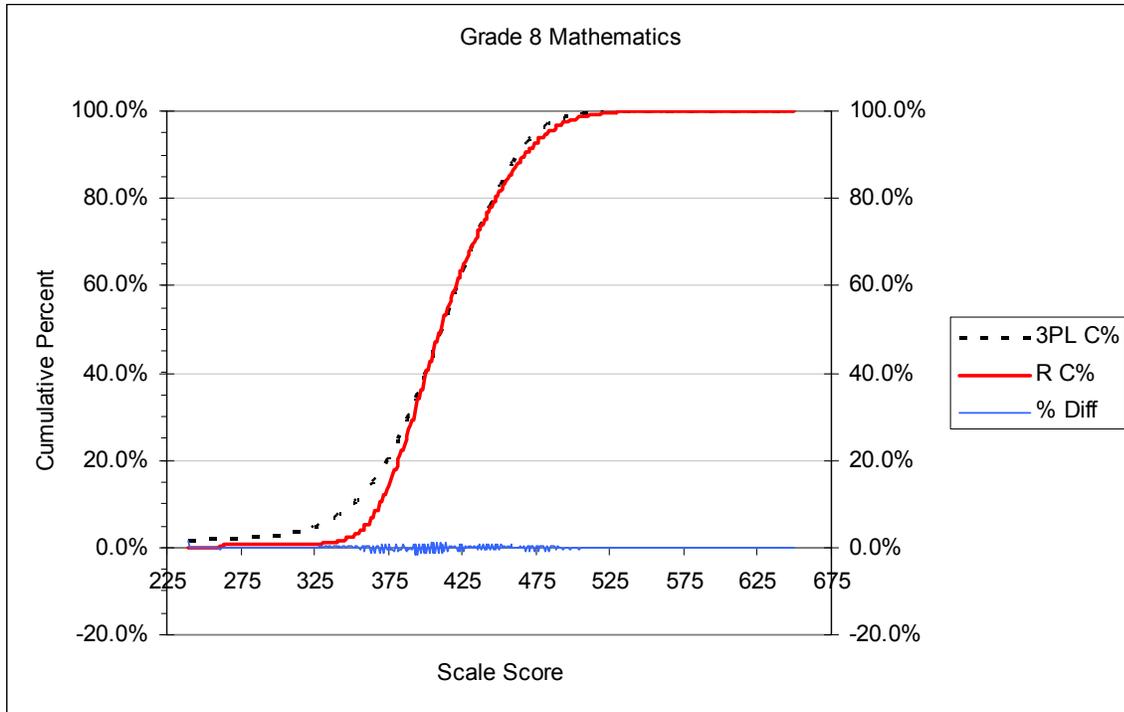


Figure C.11 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the percent differences between CDFs: Grade 8

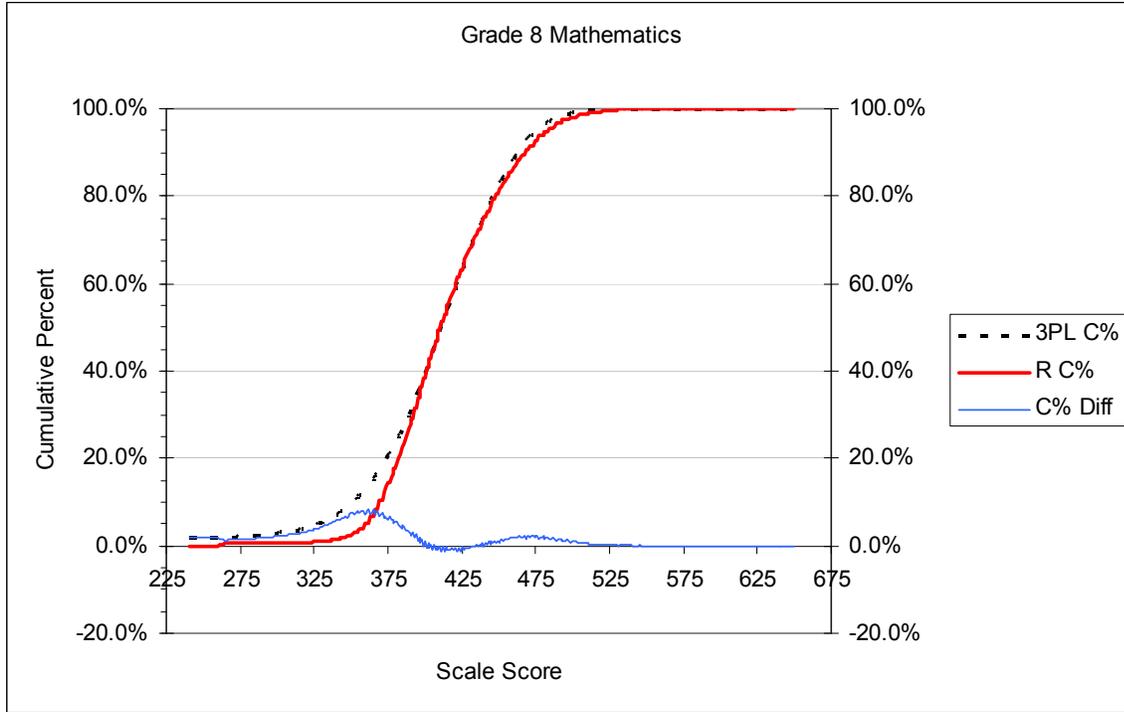


Figure C.12 Cumulative distribution functions (CDFs) for the 3PL and the Rasch scale scores with the cumulative percent differences between CDFs: Grade 8

Table C.17 Correlation between 3PL and Rasch Ability Estimates

Grade	Correlation Coefficient
3	0.99
4	0.98
5	0.98
6	0.95
7	0.96
8	0.95

Table C.18 Overall Performance Level Results of Each Grade

Grade	Model	3PL Vs. Rasch Performance Level			
		Below	Proficient	Advanced	Pass Rate
3	3PL	21.10%	54.17%	24.72%	78.89%
	Rasch	21.03%	52.84%	26.13%	78.97%
4	3PL	18.17%	49.79%	32.04%	81.83%
	Rasch	16.81%	50.46%	32.73%	83.19%
5	3PL	26.84%	54.00%	19.15%	73.15%
	Rasch	25.06%	55.59%	19.35%	74.94%
6	3PL	34.57%	46.77%	18.66%	65.43%
	Rasch	33.99%	47.10%	18.92%	66.02%
7	3PL	40.16%	44.01%	15.83%	59.84%
	Rasch	39.68%	43.60%	16.72%	60.32%
8	3PL	45.08%	32.46%	22.46%	54.92%
	Rasch	44.91%	31.73%	23.36%	55.09%

Table C.19 Kappa Indices for Classification Agreement between 3PL and 1PL: All Grades

Grade	Kappa
3	0.92
4	0.93
5	0.93
6	0.93
7	0.95
8	0.94

Table C.20 Overall Raw Agreement Index between 3PL and 1PL: All Grades

Grade	Consistent Classification	Inconsistent classification
3	95.17%	4.83%
4	95.73%	4.27%
5	95.99%	4.01%
6	95.85%	4.15%
7	96.67%	3.33%
8	96.34%	3.66%

Table C.21 Classification Consistency of Each Performance Level between 3PL and 1PL: Grade 3

	BL	PA	AD
B	20.12%	0.99%	0.00%
PA	0.91%	51.09%	2.17%
AD	0.00%	0.76%	23.96%

Note. B: Basic; PA: Proficient; AD: Advanced

Table C.22 Classification Consistency of Each Performance Level between 3PL and 1PL: Grade 4

	BL	PA	AD
B	16.51%	1.66%	0.00%
PA	0.30%	47.99%	1.50%
AD	0.00%	0.80%	31.23%

Table C.23 Classification Consistency of Each Performance Level between 3PL and 1PL: Grade 5

	BL	PA	AD
B	24.76%	2.09%	0.00%
PA	0.30%	52.79%	0.92%
AD	0.00%	0.71%	18.44%

Table C.24 Classification Consistency of Each Performance Level between 3PL and 1PL: Grade 6

	BL	PA	AD
B	33.07%	1.50%	0.00%
PA	0.91%	44.86%	1.00%
AD	0.00%	0.74%	17.92%

Note. B: Basic; PA: Proficient; AD: Advanced

Table C.25 Classification Consistency of Each Performance Level between 3PL and 1PL: Grade 7

	BL	PA	AD
B	38.99%	1.17%	0.00%
PA	0.69%	42.14%	1.18%
AD	0.00%	0.29%	15.54%

Table C.26 Classification Consistency of Each Performance Level between 3PL and 1PL: Grade 8

	BL	PA	AD
B	43.75%	1.33%	0.00%
PA	1.16%	30.26%	1.03%
AD	0.00%	0.13%	22.33%