WestEd
WestEd.org

mCLASS® Math K-5 Research Brief

2024 Beginning of Year Field Study

Submitted to Amplify
March 20, 2025

WestEd is a nonpartisan, nonprofit organization that aims to improve the lives of children and adults at all ages of learning and development. We do this by addressing challenges in education and human development, reducing opportunity gaps, and helping build communities where all can thrive. WestEd staff conduct and apply research, provide technical assistance, and support professional learning. We work with early learning educators, classroom teachers, local and state leaders, and policymakers at all levels of government.

For more information, visit WestEd.org. For regular updates on research, free resources, solutions, and job postings from WestEd, subscribe to the E-Bulletin, our semimonthly e-newsletter, at WestEd.org/subscribe.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Overview

The mCLASS Math K–5 assessment system is designed to provide educators with reliable and valid measures both to identify students needing additional support in mathematics and to inform instructional decisions. This technical brief presents evidence supporting the psychometric quality of the fall 2024 administration, addressing technical standards outlined by the National Center on Intensive Intervention (NCII) and state requirements for screening measures. Many analyses are complete, and for those that are not, we provide the details of the planned analyses.

Administration Context

The assessment was administered during the beginning-of-year (BOY) window under standardized conditions. Students in grades K–5 completed a 30-minute mCLASS Math session followed by a 30-minute STAR Math assessment. The STAR Math assessment provides data for validity analyses. Students in grades K–3 also completed 20 minutes of one-on-one testing with selected Woodcock-Johnson IV subtests to determine the threshold score on mCLASS Math for additional dyscalculia screening.

Sample Characteristics

The fall 2024 field study included 8,134 students in grades K–5, reflecting diverse student populations across 18 school districts that span four distinct census regions; thus ensuring national representation. Sample demographic groups, including gender and race/ethnicity, are provided in Table 1. This robust sample allows for analysis of the assessment's performance across diverse student groups and educational contexts.

Table 1. Demographic Characteristics for Gender and Race/Ethnicity From the BOY Administration

Grade	Race	Number	Proportion
K	**Overall**	1,463	
	Gender		
	Female	608	0.49
	Male	640	0.51
	Race/Ethnicity		
	American Indian or Alaska Native	56	0.04
	Asian	44	0.03
	Black or African American	94	0.07
	Hispanic/Latino	82	0.06
	More than one race identified	171	0.12
	Unknown	193	0.13
	White	882	0.61
1	**Overall**	1,404	
	Gender		
	Female	612	0.49
	Male	640	0.51
	Race/Ethnicity		
	American Indian or Alaska Native	60	0.04
	Asian	19	0.01
	Black or African American	82	0.06
	Hispanic/Latino	131	0.09
	More than one race identified	146	0.10
	Unknown	152	0.11
	White	942	0.67

Grade	Race	Number	Proportion
2	**Overall**	1,715	
	Gender		
	Female	779	0.49
	Male	800	0.51
	Race/Ethnicity		
	American Indian or Alaska Native	81	0.05
	Asian	45	0.03
	Black or African American	110	0.06
	Hispanic/Latino	142	0.08
	More than one race identified	274	0.16
	Unknown	135	0.08
	White	1,065	0.62
3	**Overall**	1,172	
	Gender		
	Female	534	0.50
	Male	542	0.50
	Race/Ethnicity		
	American Indian or Alaska Native	74	0.07
	Asian	36	0.03
	Black or African American	91	0.08
	Hispanic/Latino	108	0.10
	More than one race identified	214	0.19
	Unknown	29	0.03
	White	658	0.60
4	**Overall**	1,111	
	Gender		

Grade	Race	Number	Proportion
	Female	513	0.48
	Male	558	0.52
	Race/Ethnicity		
	American Indian or Alaska Native	78	0.07
	Asian	32	0.03
	Black or African American	63	0.06
	Hispanic/Latino	110	0.10
	More than one race identified	248	0.22
	Unknown	32	0.03
	White	652	0.59
5	**Overall**	1,269	
	Gender		
	Female	576	0.50
	Male	580	0.50
	Race/Ethnicity		
	American Indian or Alaska Native	64	0.05
	Asian	45	0.04
	Black or African American	98	0.08
	Hispanic/Latino	137	0.11
	More than one race identified	249	0.21
	Unknown	37	0.03
	White	699	0.59

Dyscalculia Screening

To establish a scientifically rigorous approach to dyscalculia screening, this study invested in individual, face-to-face administration of specific Woodcock-Johnson IV (WJIV) subtests: Applied Problems, Math Facts Fluency, and Calculation. Trained assessors individually administered these diagnostic measures to students, providing criterion data for establishing screening cuts. The scores from the WJIV subtests were entered into Riverside's proprietary scoring system to identify overall math, broad math, and math calculation skill clusters. The skill clusters are norm-referenced scores with a mean of 100 and standard deviation of 15. Scores that fall 1.5 standard deviations below the mean may indicate a skill deficit (Grant, 2021; Barrett & Cottrell, 2015; Swanson & Jerman, 2006). Cut score development will use receiver operating characteristic (ROC) curve analysis to optimize sensitivity and specificity with particular attention to the educational consequences of identification decisions (Zou et al., 2014). This investment in individual diagnostic testing helps ensure that students at risk for dyscalculia are appropriately identified for further evaluation.

Classification Accuracy

Classification accuracy is the degree to which an assessment correctly classifies individuals into categories, or performance levels, when compared against an established criterion measure. Using the STAR Math assessment's established performance standards as benchmarks, we will conduct ROC curve analyses to evaluate mCLASS Math classification accuracy. We expect the lower bound of the area under the curve (AUC) confidence interval to exceed 0.80, with sensitivity and specificity rates of 0.80 or higher. In grades K–3, ROC analyses will also be used versus the WJIV external criterion measure.

Fairness and Bias Analyses

A comprehensive evaluation of test fairness and potential bias provides evidence for comparability of mCLASS Math scores across all student populations. We examine through differential item functioning (DIF) analyses whether individual items perform differently for different student groups, after controlling for overall mathematics ability. Any items showing statistically significant DIF will undergo content review to identify potential bias that might disadvantage members of a particular student group. If the sample size is sufficient, multiple-group confirmatory factor analyses will investigate whether the underlying factor structure of the assessment is consistent across student groups, providing evidence that the test measures the same constructs in the same way across the population.

Reliability Evidence

Reliability indices demonstrate the consistency of mCLASS Math scores. Internal consistency reliability is estimated using coefficient alpha (Cronbach, 1951). In addition, we provide Rasch model–based standard errors of measurement (SEM) as evidence of measurement reliability. The results from fall 2024 are provided in Table 2. Cronbach's alpha exceeded .84 for Grades 1–5 and approached 0.8 for kindergarten. The unconditional (overall) SEM was about 0.02 points for all grades. All reliability indices from representative samples have confidence interval lower bounds exceeding 0.70. The conditional SEM plots are provided in the Appendix; see Figures A1 through A6 for validity evidence.

Table 2. Coefficient Alpha Estimates and Unconditional Standard Error of Measurement (SEM) by Grade of Administration

Grade	Alpha	Alpha (LB)	Alpha (UB)	SEM
K	0.79	0.77	0.81	0.02
1	0.84	0.83	0.86	0.02
2	0.88	0.87	0.89	0.02
3	0.89	0.88	0.90	0.02
4	0.91	0.91	0.92	0.02
5	0.90	0.89	0.91	0.02

mCLASS Math assessments are designed to evaluate student performance against grade-level expectations throughout the year. The following validity framework offers multiple sources of evidence to support this interpretation of mCLASS Math scores. Concurrent validity is supported by correlations with external measures. The correlations between mCLASS Math and STAR Math assessment are given in Table 3. Correlations exceeded 0.73 in grades 1–5, whereas grade K had a moderate correlation with STAR at 0.49. The jump from .49 to .73 in grade 1 suggests that early math skills might be conceptualized differently between the two tests. Also, kindergarten is a period of rapid, uneven development in math skill. The grades 1–5 correlation coefficients meet or exceed 0.60 at the lower bound of confidence intervals.

Table 3. Correlations Between Renaissance STAR Math and Summed mCLASS Math Total Scores by Grade

Grade	Correlation	Correlation (LB)	Correlation (UB)	N
K	0.49	0.42	0.55	560
1	0.73	0.69	0.77	516
2	0.81	0.78	0.83	703
3	0.79	0.75	0.82	534
4	0.82	0.79	0.84	532
5	0.80	0.77	0.83	584

Content validity is supported by a comprehensive standards alignment, conducted in spring 2024. This analysis evaluated each item's alignment to the Common Core State Standards for Mathematics (CCSSM), identifying both the comprehensiveness of standard coverage and any potential gaps. Content specialists verified alignments and provided detailed documentation of the relationship between item content and grade-level mathematical constructs.

Internal structure validity is evaluated through confirmatory factor analyses (CFA; Brown, 2015), examining whether the internal structure aligns with the theoretical framework of mathematical ability. All factor analysis models were fit within a grade and the items were treated as categorical. All grade-level factor analyses suggested that a one-factor model fit the data. In all cases, model fit statistics were higher than commonly referenced thresholds, including CFI > 0.95 and RMSEA < 0.05 (Hu & Bentler, 1999). The CFA model fit statistics are shown in Table 4. Additionally, Rasch/IRT analyses provide evidence of construct validity through the examination of Wright maps (Boone et al., 2014) and construct representation across the score scale. These analyses examine whether item difficulties align with the mathematical difficulty spectrum. The Wright maps are shown in the Appendix; see Figures A7 through A12. Overall, the Wright maps show evidence of items across the ability scale for all grades.

Table 4. Model Fit Statistics for the Confirmatory Factor Analysis Model

Grade	N	CFI	RMSEA
K	993	0.973	0.026 [0.021, 0.032]
1	1,012	0.982	0.029 [0.024, 0.034]
2	1,371	0.957	0.042 [0.039, 0.045]
3	885	0.968	0.032 [0.028, 0.035]
4	929	0.955	0.049 [0.046, 0.051]
5	1,073	0.975	0.032 [0.029, 0.035]

Response process validity includes considering how students engage with the items. While cognitive labs have not been conducted to date, we examined the analysis of response patterns, distractor analyses, and item-total correlations. These analyses support that items measure ability as intended (see Tables A1 to A6 in the Appendix). Most items have item-total correlations greater than 0.2.

Consequential validity examines both the intended and unintended consequences of assessment use in educational settings. The mCLASS Math assessment is designed to provide teachers with actionable information about student mathematical understanding throughout the school year, enabling early identification of students who may need additional support. To evaluate this intended use, we analyzed score distributions by student groups to ensure score interpretations are fair across diverse student populations. Where sample sizes permit, we present disaggregated analyses in Tables 5 through 7. Analyses were reported if samples sizes within a group were greater than 10.

Table 5. Summary Statistics of the Summed Total mCLASS Math Scores

Grade	Avg. Total Score	SD Total Score	Min Total Score	Max Total Score	Max Possible Score
K	11.88	4.20	0	20	20
1	10.32	4.83	0	20	20
2	14.44	6.08	0	25	25
3	15.24	7.13	0	30	30
4	14.40	7.56	0	30	30
5	12.44	7.00	0	29	30

Table 6. Summary Statistics of the Summed Total mCLASS Math Scores by Gender

Grade	Gender	Avg. Total Score	SD Total Score	Min Total Score	Max Total Score	Number
K	Female	12.13	3.96	0.00	20.00	608
K	Male	11.97	4.18	0.00	20.00	640
1	Female	10.10	4.67	0.00	20.00	612
1	Male	10.69	4.85	0.00	20.00	640
2	Female	14.30	5.85	0.00	25.00	779
2	Male	14.96	6.08	0.00	25.00	800
3	Female	14.61	6.96	0.00	30.00	534
3	Male	15.82	7.20	0.00	29.00	542
4	Female	13.26	7.27	0.00	30.00	513
4	Male	15.40	7.75	0.00	30.00	558
5	Female	11.49	6.36	0.00	29.00	576
5	Male	13.27	7.57	0.00	29.00	580

Table 7. Summary Statistics of the Summed Total mCLASS Math Scores by Race and Ethnicity

Grade	Race/Ethnicity	Avg. Total Score	SD Total Score	Min Total Score	Max Total Score	Number
K	American Indian or Alaska Native	10.81	4.25	1.00	19.00	56
K	Asian	12.24	4.18	0.00	18.00	44
K	Black or African American	10.17	3.97	2.00	18.00	94
K	Hispanic/Latino	10.59	4.60	1.00	19.00	82
K	More than one race identified	12.38	3.92	0.00	20.00	171
K	Unknown	10.81	4.81	0.00	19.00	193
K	White	12.24	4.03	0.00	20.00	882
1	American Indian or Alaska Native	7.50	3.86	0.00	16.00	60
1	Asian	10.32	4.67	0.00	17.00	19
1	Black or African American	7.53	4.95	0.00	19.00	82
1	Hispanic/Latino	9.48	4.33	0.00	19.00	131
1	More than one race identified	9.48	4.70	0.00	19.00	146
1	Unknown	9.66	5.22	1.00	20.00	152
1	White	10.96	4.65	0.00	20.00	942
2	American Indian or Alaska Native	10.36	5.62	0.00	24.00	81

Grade	Race/Ethnicity	Avg. Total Score	SD Total Score	Min Total Score	Max Total Score	Number
2	Asian	17.48	5.42	0.00	25.00	45
2	Black or African American	10.59	6.63	0.00	24.00	110
2	Hispanic/Latino	12.27	6.04	0.00	25.00	142
2	More than one race identified	13.94	6.14	1.00	25.00	274
2	Unknown	12.15	6.73	0.00	25.00	135
2	White	15.43	5.54	0.00	25.00	1,065
3	American Indian or Alaska Native	9.03	5.84	0.00	25.00	74
3	Asian	19.50	6.55	3.00	30.00	36
3	Black or African American	10.11	5.94	0.00	29.00	91
3	Hispanic/Latino	15.02	7.06	2.00	29.00	108
3	More than one race identified	14.63	7.00	1.00	27.00	214
3	Unknown	16.00	7.81	2.00	27.00	29
3	White	16.55	6.68	0.00	29.00	658
4	American Indian or Alaska Native	7.90	5.62	0.00	21.00	78
4	Asian	19.09	5.97	7.00	28.00	32
4	Black or African American	9.57	6.88	0.00	25.00	63

Grade	Race/Ethnicity	Avg. Total Score	SD Total Score	Min Total Score	Max Total Score	Number
4	Hispanic/Latino	14.38	7.26	0.00	28.00	110
4	More than one race identified	14.53	7.87	0.00	28.00	248
4	Unknown	15.44	5.87	2.00	27.00	32
4	White	15.28	7.25	0.00	30.00	652
5	American Indian or Alaska Native	5.79	4.42	0.00	22.00	64
5	Asian	16.86	6.69	0.00	28.00	45
5	Black or African American	7.58	5.79	0.00	23.00	98
5	Hispanic/Latino	11.25	6.38	0.00	28.00	137
5	More than one race identified	12.40	7.30	0.00	29.00	249
5	Unknown	14.03	5.04	6.00	26.00	37
5	White	13.32	6.67	0.00	29.00	699

Item Analysis

Comprehensive item analyses combine classical-test theory and item-response theory approaches. Classical analyses include *p*-values (item difficulty); corrected item-total correlations (discrimination); and distractor analysis for multiple-choice items. Results from our item analyses are provided in the Appendix; see Tables A1 through A6. Items spanned the difficulty spectrum for each grade; showed evidence of only having one correct answer; and had corrected item-total correlations above 0.2 in most cases.

For dichotomous items, we employ the Rasch model (De Ayala, 2013). Item fit is evaluated using outfit and infit mean-squares; values below 0.5 or above 1.5 are flagged for review. Across the 155 items, seven items were flagged for content review. Differential Item Functioning (DIF) analyses using logistic regression (Swaminathan & Rogers, 1990) examine item fairness across student groups, requiring minimum sample sizes of 100 students in each group. We used the logistic regression approach to identifying DIF across gender and race categories in Table 1 and 2. Due to sample size constraints, we explored evidence of DIF for the following five groups: White; Black; more than one race identified; American Indian and/or Alaskan native; and other.

Logistic regression differential item functioning (LR DIF) analyses were conducted for race and gender separately, with each model including group membership as a categorical predictor; mCLASS Math total score as a continuous predictor; and the interaction between the continuous and categorical predictor. Items were flagged with *p*-values below 0.001 to account for multiple testing. For the 155 items across all grade levels and across all DIF analyses, two items were flagged for further content review. Grade 3 item 30022 and Grade 4 item 40006 showed C-level DIF for one of the race/ethnicity pairwise comparisons.

Note

This BOY research brief addresses fall 2024 administration data. Subsequent reports will incorporate winter and spring administrations, providing additional validity evidence for the tests' intended uses and score interpretations.

References

Barrett, C. A., & Cottrell, J. M. (2015). Defining the undefinable: Operationalization of methods to identify specific learning disabilities among practicing school psychologists. Psychology in the Schools, 53(2), 143-157. https://doi.org/10.1002/pits.21892

Boone, W.J., Staver, J.R., Yale, M.S. (2014). Wright Maps: First Steps. In: Rasch Analysis in the Human Sciences. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-6857-4_6

Brown, T. A. (2015). Confirmatory factor analysis for applied research. Guilford publications.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. psychometrika, 16(3), 297-334.

De Ayala, R. J. (2013). The theory and practice of item response theory. Guilford Publications.

Grant, G. (2021, November 8). Assessment plan: A guide to evaluating for dyscalculia. Riverside Insights Blog. https://blog.riversideinsights.com/evaluating-dyscalculia-woodcock-johnson-assessment-plan

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal, 6(1), 1-55.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. Journal of Educational measurement, 27(4), 361-370.

Swanson, H. L., & Jerman, O. (2006). Math Disabilities: A Selective Meta-Analysis of the Literature. Review of Educational Research, 76(2), 249-274. https://doi.org/10.3102/00346543076002249.

Zou, K. H., Liu, A., Bandos, A. I., Ohno-Machado, L., & Rockette, H. E. (2011). Statistical evaluation of diagnostic performance: topics in ROC analysis. CRC press.

Appendix

Table A1. Item Statistics for Kindergarten Items

Order	Grade	Item	Domain	*p*	itc	Number Missing	Number Response	Percent Missing
1.00	K	00001	CC	0.67	0.35	106.00	1,357.00	7.20
2.00	K	00003	CC	0.89	0.27	204.00	1,259.00	13.90
3.00	K	00011	MD	0.61	0.44	67.00	1,390.00	4.60
4.00	K	00002	CC	0.69	0.49	42.00	1,414.00	2.90
5.00	K	00005	CC	0.69	0.42	40.00	1,416.00	2.70
6.00	K	00010	MD	0.91	0.35	165.00	1,298.00	11.30
7.00	K	00018	MD	0.49	0.51	97.00	1,354.00	6.70
8.00	K	00012	G	0.77	0.36	32.00	1,418.00	2.20
9.00	K	00009	OA	0.65	0.42	36.00	1,411.00	2.50
10.00	K	00017	OA	0.34	0.35	70.00	1,377.00	4.80
11.00	K	00013	CC	0.87	0.33	100.00	1,363.00	6.80
12.00	K	00014	CC	0.88	0.24	100.00	1,363.00	6.80
13.00	K	00004	CC	0.33	0.30	60.00	1,380.00	4.20
14.00	K	00008	OA	0.77	0.45	30.00	1,409.00	2.10
15.00	K	00007	OA	0.22	0.34	39.00	1,399.00	2.70
16.00	K	00019	G	0.61	0.37	55.00	1,408.00	3.80
17.00	K	00015	CC	0.79	0.30	83.00	1,380.00	5.70
18.00	K	00021	OA	0.66	0.46	24.00	1,412.00	1.70
19.00	K	00016	OA	0.07	0.06	67.00	1,369.00	4.70
20.00	K	00020	G	0.68	0.09	24.00	1,411.00	1.70

Table A2. Item Statistics for Grade 1 Items

Order	Grade	Item	Domain	*p*	itc	Number Missing	Number Response	Percent Missing
1.00	1	10001	CC	0.75	0.28	12.00	1,392.00	0.90
2.00	1	10002	OA	0.65	0.53	29.00	1,374.00	2.10
3.00	1	10023	OA	0.44	0.35	53.00	1,349.00	3.80
4.00	1	10016	OA	0.31	0.10	72.00	1,332.00	5.10
5.00	1	10021	MD	0.59	0.44	38.00	1,364.00	2.70
6.00	1	10008	NBT	0.52	0.33	37.00	1,364.00	2.60
7.00	1	10010	MD	0.71	0.39	181.00	1,220.00	12.90
8.00	1	10009	NBT	0.49	0.56	36.00	1,365.00	2.60
9.00	1	10003	OA	0.75	0.40	32.00	1,369.00	2.30
10.00	1	10019	NBT	0.23	0.49	50.00	1,351.00	3.60
11.00	1	10015	MD	0.15	0.41	43.00	1,358.00	3.10
12.00	1	10022	G	0.53	0.42	37.00	1,364.00	2.60
13.00	1	10005	OA	0.53	0.61	42.00	1,359.00	3.00
14.00	1	10018	NBT	0.40	0.53	69.00	1,332.00	4.90
15.00	1	10017	NBT	0.58	0.60	45.00	1,356.00	3.20
16.00	1	10020	MD	0.46	0.34	132.00	1,269.00	9.40
17.00	1	10014	OA	0.47	0.57	52.00	1,349.00	3.70
18.00	1	10011	MD	0.83	0.41	51.00	1,350.00	3.60
19.00	1	10004	OA	0.54	0.49	64.00	1,337.00	4.60
20.00	1	10013	G	0.79	0.21	21.00	1,380.00	1.50

Table A3. Item Statistics for Grade 2 Items

Order	Grade	Item	Domain	*p*	itc	Number Missing	Number Response	Percent Missing
1.00	2	20004	OA	0.73	0.48	11.00	1,704.00	0.60
2.00	2	20023	MD	0.47	0.51	26.00	1,688.00	1.50
3.00	2	20012	G	0.62	0.39	16.00	1,698.00	0.90
4.00	2	20009	NBT	0.19	0.27	49.00	1,664.00	2.90
5.00	2	20014	OA	0.75	0.39	45.00	1,670.00	2.60
6.00	2	20018	NBT	0.70	0.54	40.00	1,672.00	2.30
7.00	2	20021	MD	0.44	0.55	36.00	1,676.00	2.10
8.00	2	20008	NBT	0.74	0.35	74.00	1,637.00	4.30
9.00	2	20001	OA	0.86	0.39	28.00	1,683.00	1.60
10.00	2	20005	OA	0.51	0.53	49.00	1,663.00	2.90
11.00	2	20020	NBT	0.73	0.28	36.00	1,679.00	2.10
12.00	2	20011	MD	0.56	0.64	39.00	1,671.00	2.30
13.00	2	20027	NBT	0.63	0.52	30.00	1,685.00	1.70
14.00	2	20007	OA	0.80	0.51	40.00	1,669.00	2.30
15.00	2	20026	OA	0.51	0.36	62.00	1,653.00	3.60
16.00	2	20022	MD	0.67	0.42	44.00	1,663.00	2.60
17.00	2	20019	NBT	0.62	0.56	49.00	1,656.00	2.90
18.00	2	20013	MD	0.47	0.59	45.00	1,660.00	2.60
19.00	2	20003	OA	0.27	0.45	48.00	1,656.00	2.80
20.00	2	20006	OA	0.52	0.62	52.00	1,651.00	3.10
21.00	2	20025	G	0.95	0.22	142.00	1,573.00	8.30
22.00	2	20002	OA	0.63	0.44	43.00	1,659.00	2.50
23.00	2	20010	MD	0.29	0.29	52.00	1,650.00	3.10
24.00	2	20017	NBT	0.69	0.51	40.00	1,662.00	2.40
25.00	2	20024	MD	0.48	0.35	49.00	1,666.00	2.90

Table A4. Item Statistics for Grade 3 Items

Order	Grade	Item	Domain	*p*	itc	Number Missing	Number Response	Percent Missing
1.00	3	30002	OA	0.64	0.50	5.00	1,167.00	0.40
2.00	3	30006	NBT	0.29	0.26	36.00	1,136.00	3.10
3.00	3	30009	NF	0.67	0.35	18.00	1,154.00	1.50
4.00	3	30008	NBT	0.67	0.48	14.00	1,158.00	1.20
5.00	3	30018	OA	0.70	0.49	9.00	1,163.00	0.80
6.00	3	30025	MD	0.47	0.28	28.00	1,144.00	2.40
7.00	3	30019	OA	0.44	0.44	23.00	1,149.00	2.00
8.00	3	30003	OA	0.25	0.49	35.00	1,137.00	3.00
10.00	3	30015	G	0.54	0.46	11.00	1,161.00	0.90
11.00	3	30024	NBT	0.34	0.56	34.00	1,138.00	2.90
12.00	3	30022	NBT	0.65	0.49	17.00	1,153.00	1.50
13.00	3	30005	OA	0.83	0.21	11.00	1,160.00	0.90
14.00	3	30012	MD	0.53	0.51	35.00	1,136.00	3.00
15.00	3	30021	NBT	0.64	0.52	23.00	1,149.00	2.00
16.00	3	30010	MD	0.51	0.21	45.00	1,127.00	3.80
18.00	3	30031	G	0.21	0.20	16.00	1,148.00	1.40
19.00	3	30023	NBT	0.53	0.45	82.00	1,079.00	7.10
20.00	3	30029	MD	0.64	0.52	29.00	1,130.00	2.50
21.00	3	30014	G	0.95	0.26	97.00	1,075.00	8.30
22.00	3	30032	G	0.69	0.37	28.00	1,126.00	2.40
23.00	3	30016	OA	0.41	0.47	44.00	1,108.00	3.80
24.00	3	30028	MD	0.66	0.43	52.00	1,120.00	4.40
25.00	3	30027	NBT	0.32	0.47	87.00	1,062.00	7.60
26.00	3	30030	MD	0.46	0.52	0.00	1,145.00	0.00
27.00	3	30033	NBT	0.23	0.52	60.00	1,078.00	5.30
28.00	3	30007	NBT	0.47	0.63	56.00	1,079.00	4.90
29.00	3	30020	OA	0.44	0.50	76.00	1,096.00	6.50
30.00	3	30017	OA	0.39	0.58	58.00	1,073.00	5.10
31.00	3	30026	NBT	0.51	0.50	56.00	1,074.00	5.00
32.00	3	30013	MD	0.67	0.35	40.00	1,089.00	3.50

Table A5. Item Statistics for Grade 4 Items

Order	Grade	Item	Domain	*p*	itc	Number Missing	Number Response	Percent Missing
1.00	4	40006	NBT	0.74	0.41	8.00	1,103.00	0.70
2.00	4	40012	NF	0.61	0.57	9.00	1,102.00	0.80
3.00	4	40036	NBT	0.25	0.48	9.00	1,102.00	0.80
4.00	4	40003	OA	0.39	0.53	15.00	1,095.00	1.40
6.00	4	40028	NF	0.72	0.51	11.00	1,098.00	1.00
7.00	4	40014	MD	0.21	0.43	14.00	1,094.00	1.30
8.00	4	40023	NBT	0.56	0.47	37.00	1,070.00	3.30
9.00	4	40035	G	0.05	0.23	13.00	1,094.00	1.20
10.00	4	40005	NBT	0.88	0.30	21.00	1,090.00	1.90
11.00	4	40017	OA	0.59	0.57	25.00	1,079.00	2.30
12.00	4	40008	NBT	0.79	0.50	15.00	1,088.00	1.40
13.00	4	40030	MD	0.59	0.64	22.00	1,080.00	2.00
14.00	4	40007	OA	0.52	0.61	17.00	1,084.00	1.50
15.00	4	40011	NF	0.24	0.48	45.00	1,055.00	4.10
16.00	4	40024	NBT	0.77	0.48	38.00	1,073.00	3.40
17.00	4	40015	MD	0.53	0.52	34.00	1,060.00	3.10
18.00	4	40002	OA	0.36	0.59	28.00	1,063.00	2.60
19.00	4	40010	NF	0.60	0.52	24.00	1,063.00	2.20
20.00	4	40018	OA	0.36	0.55	24.00	1,057.00	2.20
21.00	4	40020	OA	0.45	0.51	54.00	1,057.00	4.90
22.00	4	40025	NBT	0.64	0.18	57.00	1,054.00	5.10
23.00	4	40022	NBT	0.60	0.57	25.00	1,046.00	2.30
24.00	4	40009	NF	0.28	0.53	34.00	1,035.00	3.20
25.00	4	40032	MD	0.33	0.43	27.00	1,040.00	2.50
27.00	4	40033	G	0.41	0.16	73.00	1,038.00	6.60
28.00	4	40026	NF	0.35	0.50	23.00	1,039.00	2.20
29.00	4	40029	MD	0.39	0.46	28.00	1,031.00	2.60
30.00	4	40019	OA	0.64	0.61	26.00	1,030.00	2.50
31.00	4	40027	NF	0.40	0.56	29.00	1,025.00	2.80
32.00	4	40031	MD	0.57	0.59	24.00	1,028.00	2.30

Table A6. Item Statistics for Grade 5 Items

Order	Grade	Item	Domain	*p*	itc	Number Missing	Number Response	Percent Missing
1.00	5	50015	MD	0.55	0.48	4.00	1,265.00	0.30
2.00	5	50034	NBT	0.37	0.53	11.00	1,258.00	0.90
3.00	5	50005	NBT	0.68	0.40	11.00	1,257.00	0.90
4.00	5	50028	MD	0.55	0.57	16.00	1,252.00	1.30
5.00	5	50011	NF	0.59	0.39	13.00	1,255.00	1.00
6.00	5	50032	MD	0.26	0.50	19.00	1,249.00	1.50
7.00	5	50021	NBT	0.88	0.38	21.00	1,248.00	1.70
8.00	5	50002	OA	0.35	0.53	17.00	1,249.00	1.30
10.00	5	50025	NF	0.29	0.07	27.00	1,242.00	2.10
11.00	5	50030	G	0.37	0.26	20.00	1,249.00	1.60
12.00	5	50008	NF	0.71	0.35	13.00	1,250.00	1.00
13.00	5	50026	MD	0.12	0.47	24.00	1,237.00	1.90
14.00	5	50018	NBT	0.33	0.41	17.00	1,239.00	1.40
15.00	5	50009	NF	0.39	0.64	26.00	1,228.00	2.10
16.00	5	50022	NBT	0.67	0.40	29.00	1,224.00	2.30
17.00	5	50031	NF	0.35	0.43	30.00	1,221.00	2.40
18.00	5	50004	OA	0.35	0.48	41.00	1,206.00	3.30
19.00	5	50013	MD	0.42	0.38	23.00	1,219.00	1.90
20.00	5	50007	NBT	0.47	0.50	29.00	1,210.00	2.30
21.00	5	50033	NF	0.27	0.48	18.00	1,218.00	1.50
22.00	5	50029	MD	0.18	0.49	48.00	1,187.00	3.90
23.00	5	50017	OA	0.75	0.52	29.00	1,201.00	2.40
24.00	5	50016	G	0.15	0.42	30.00	1,196.00	2.40
25.00	5	50006	NBT	0.26	0.49	29.00	1,194.00	2.40
26.00	5	50024	NF	0.10	0.43	42.00	1,177.00	3.40
27.00	5	50020	NBT	0.61	0.44	79.00	1,189.00	6.20
28.00	5	50027	MD	0.33	0.59	35.00	1,180.00	2.90
29.00	5	50010	NF	0.56	0.34	26.00	1,184.00	2.10
31.00	5	50019	NBT	0.36	0.56	36.00	1,172.00	3.00
32.00	5	50023	NF	0.36	0.64	31.00	1,176.00	2.60

Figure A1. Information and Conditional Standard Error of Measurement (SEM) for Kindergarten

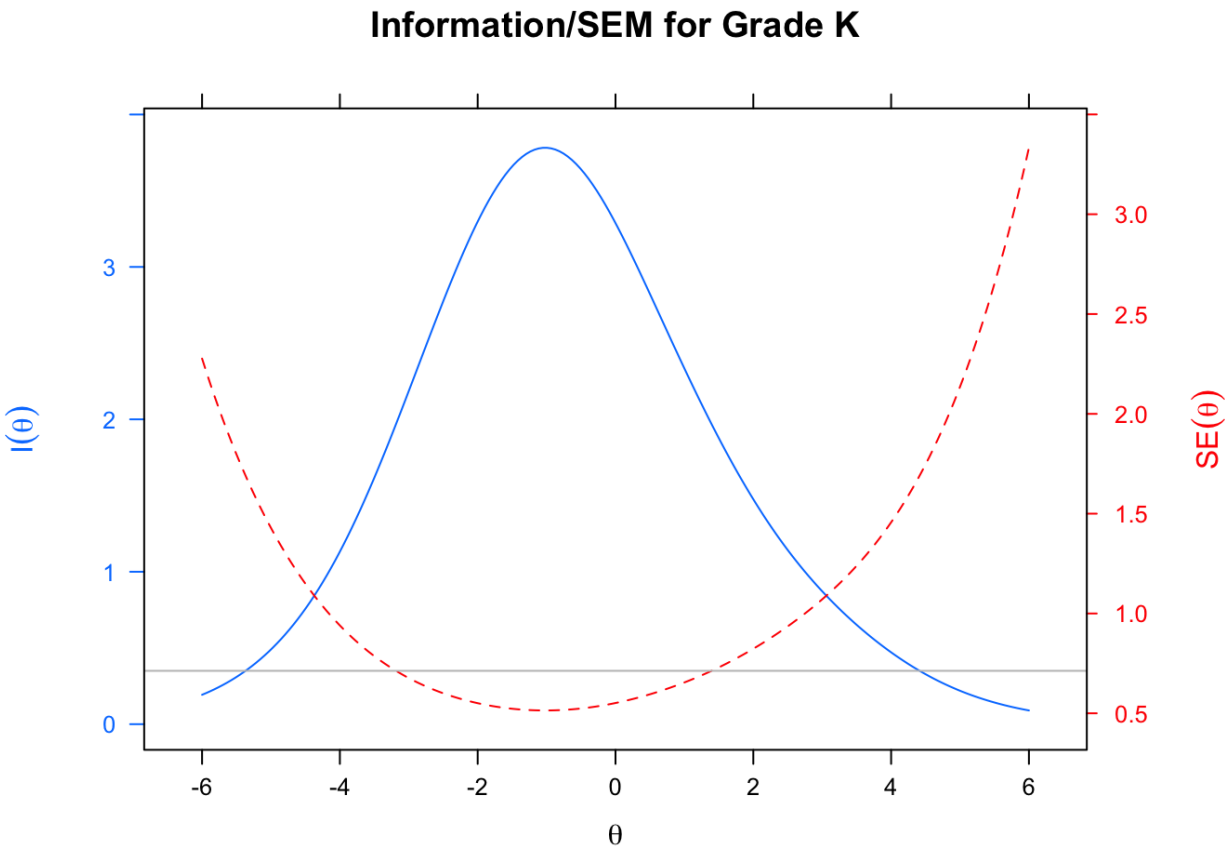

Figure A2. Information and Conditional Standard Error of Measurement (SEM) for Grade 1

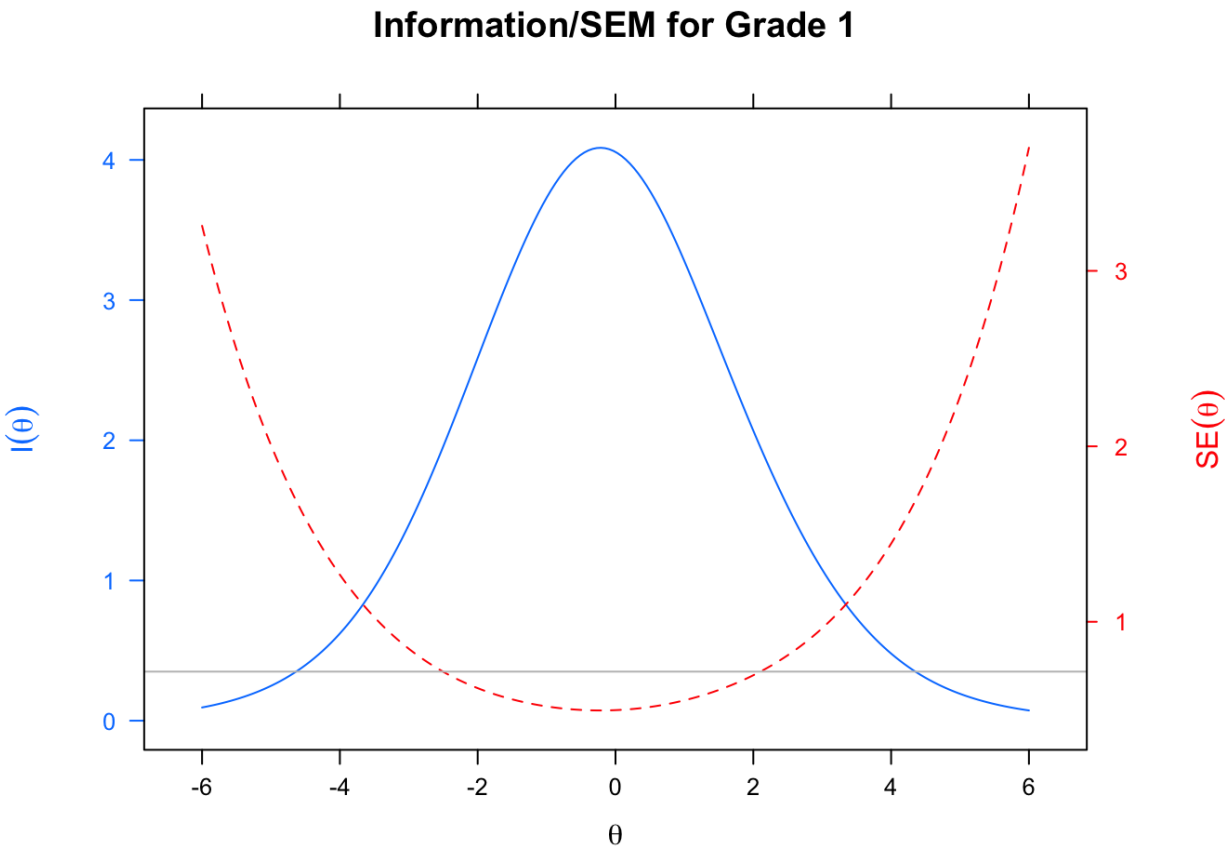

Figure A3. Information and Conditional Standard of Error of Measurement (SEM) for Grade 2

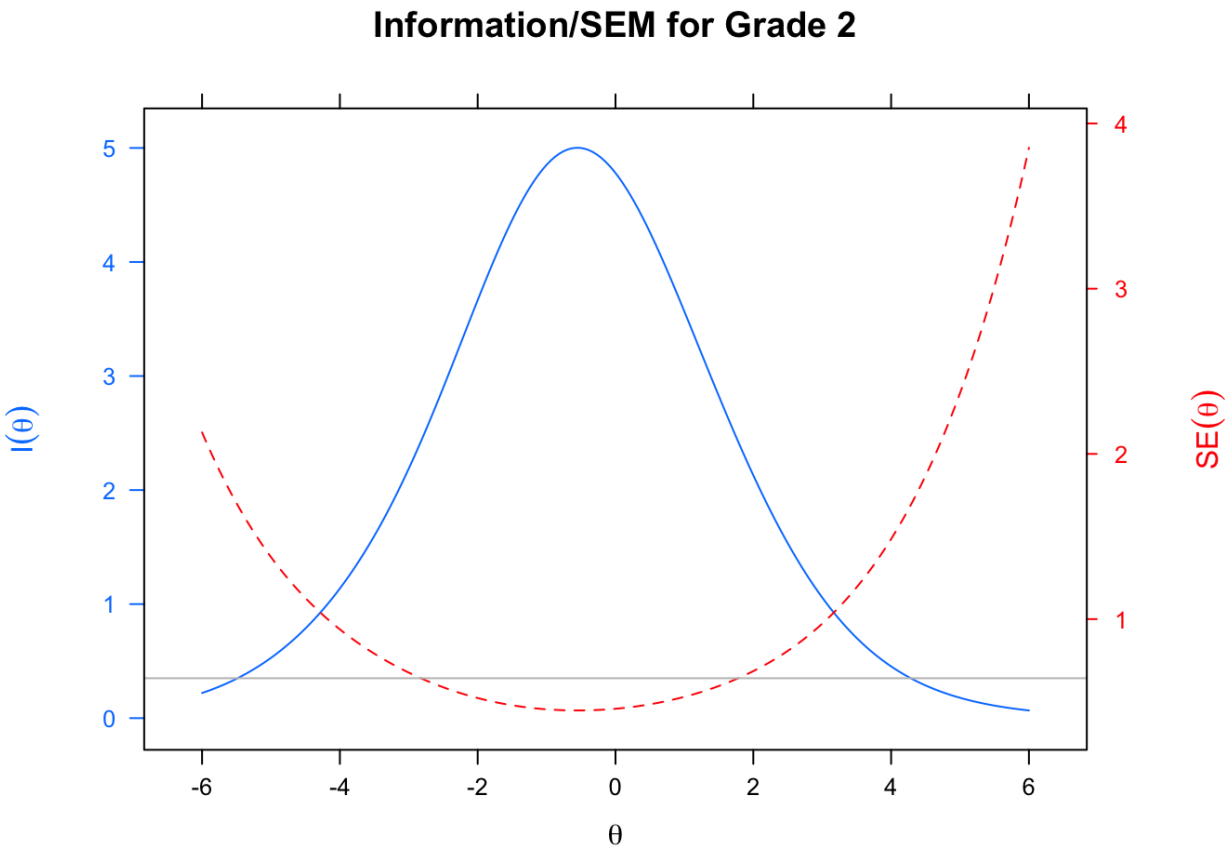

Figure A4. Information and Conditional Standard Error of Measurement (SEM) for Grade 3

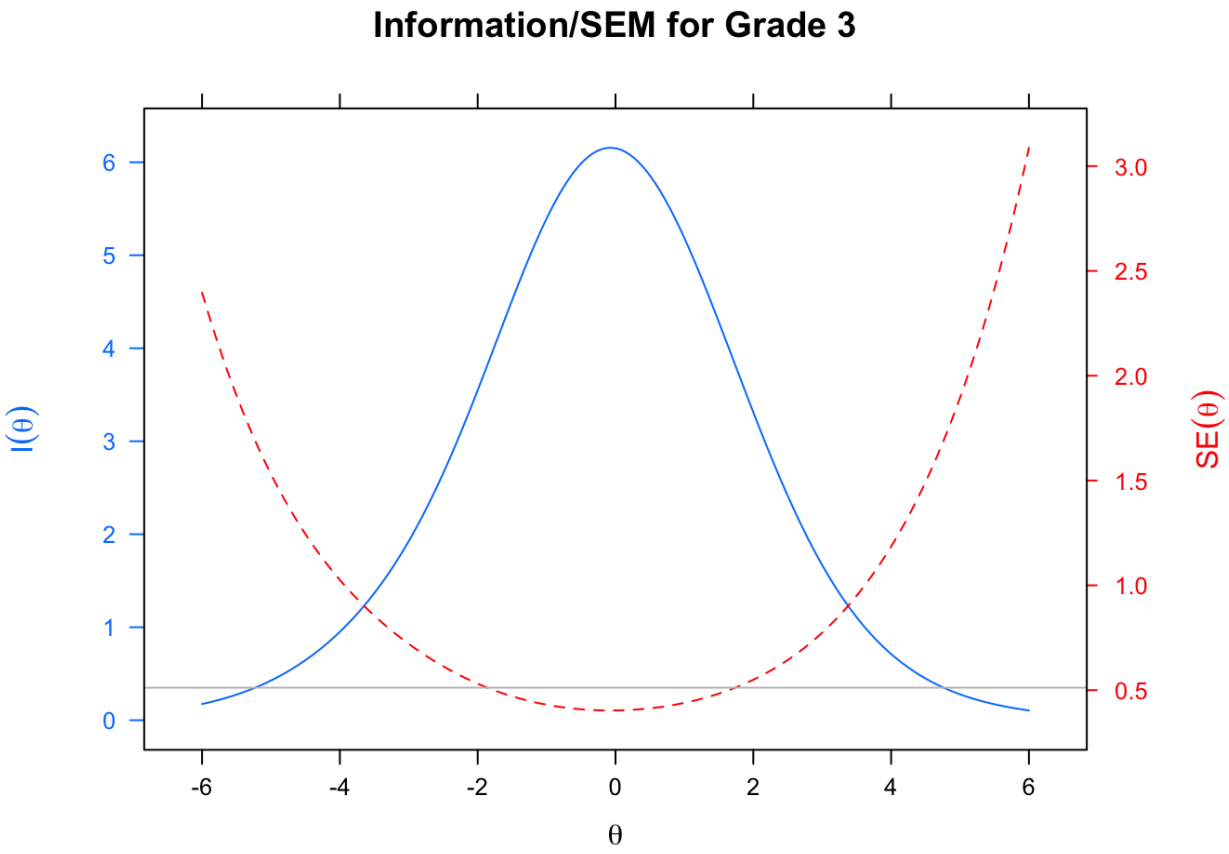

Figure A5. Information and Conditional Standard Error of Measurement (SEM) for Grade 4

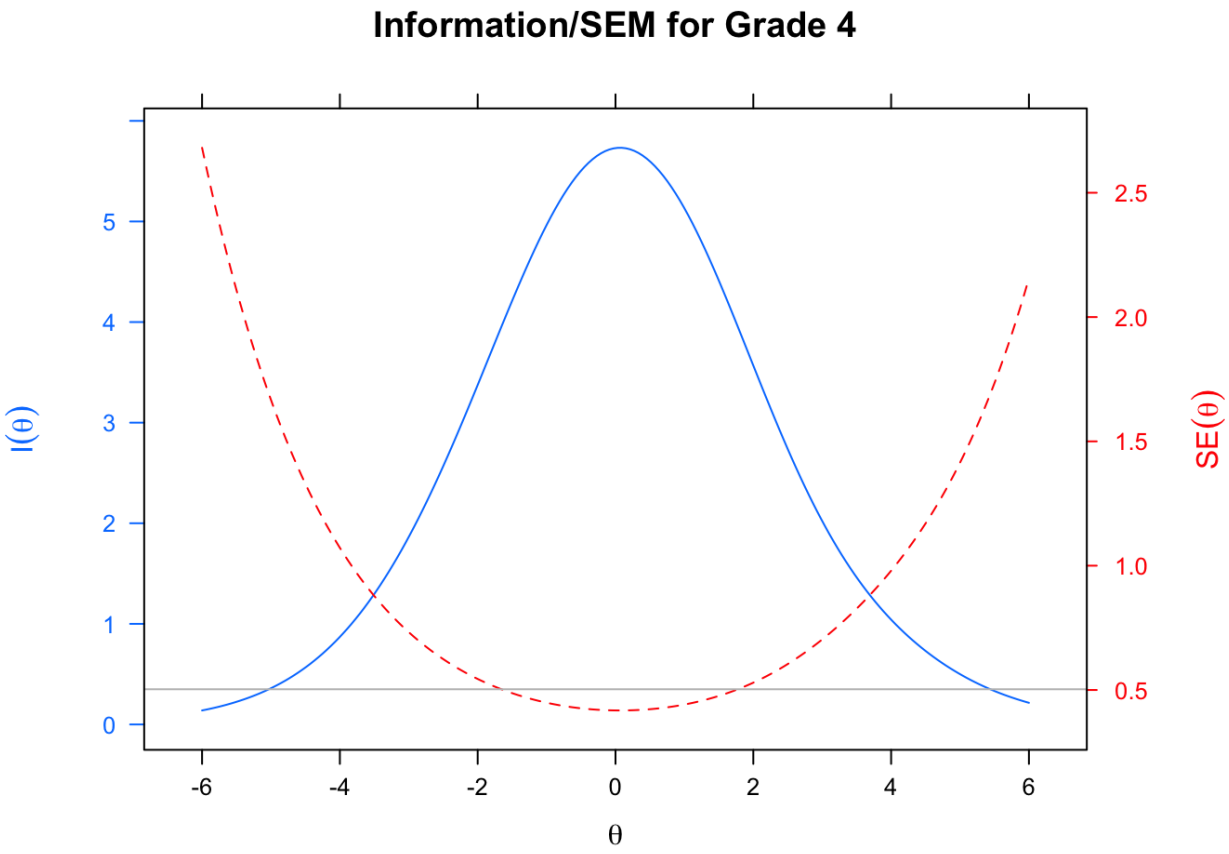

Figure A6. Information and Conditional Standard Error of Measurement (SEM) for Grade 5

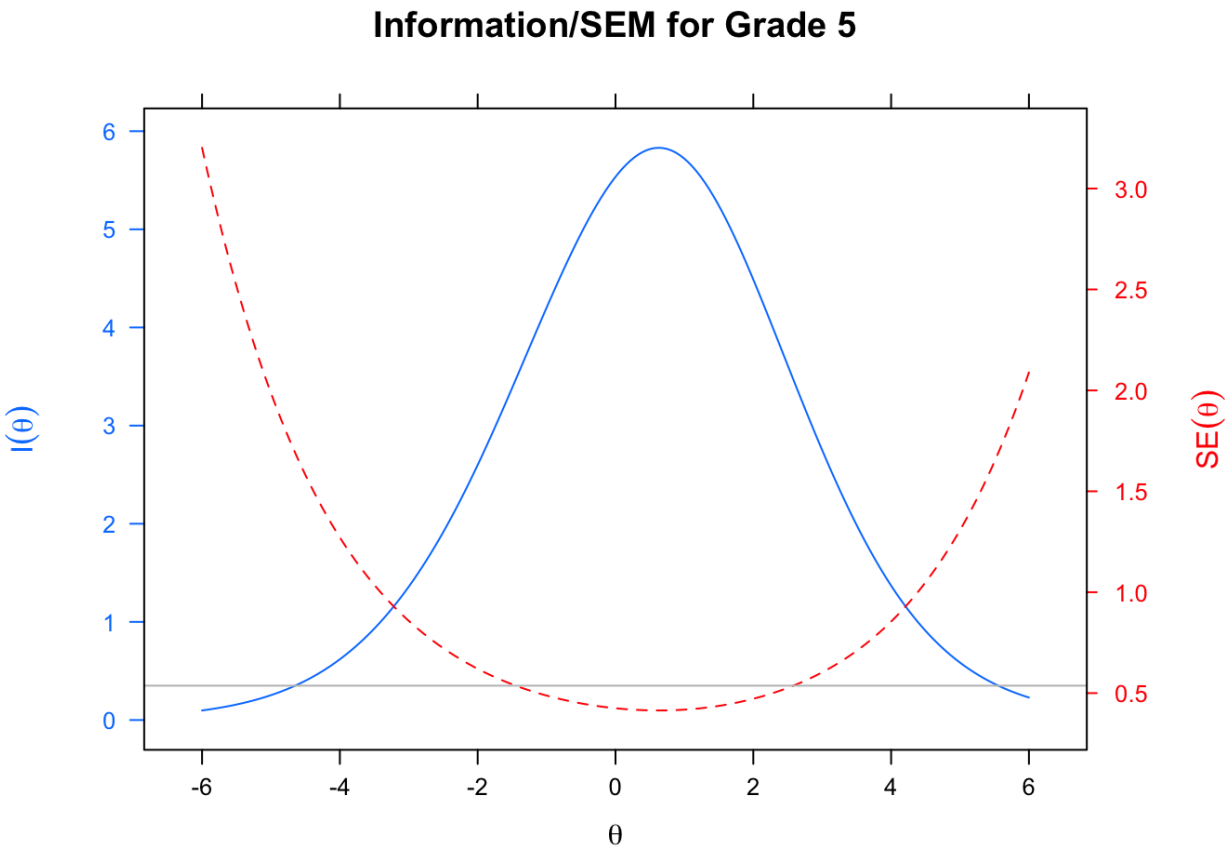

Figure A7. Kindergarten Wright Map Showing the Item and Person Estimates From the Rasch Model on a Similar Scale

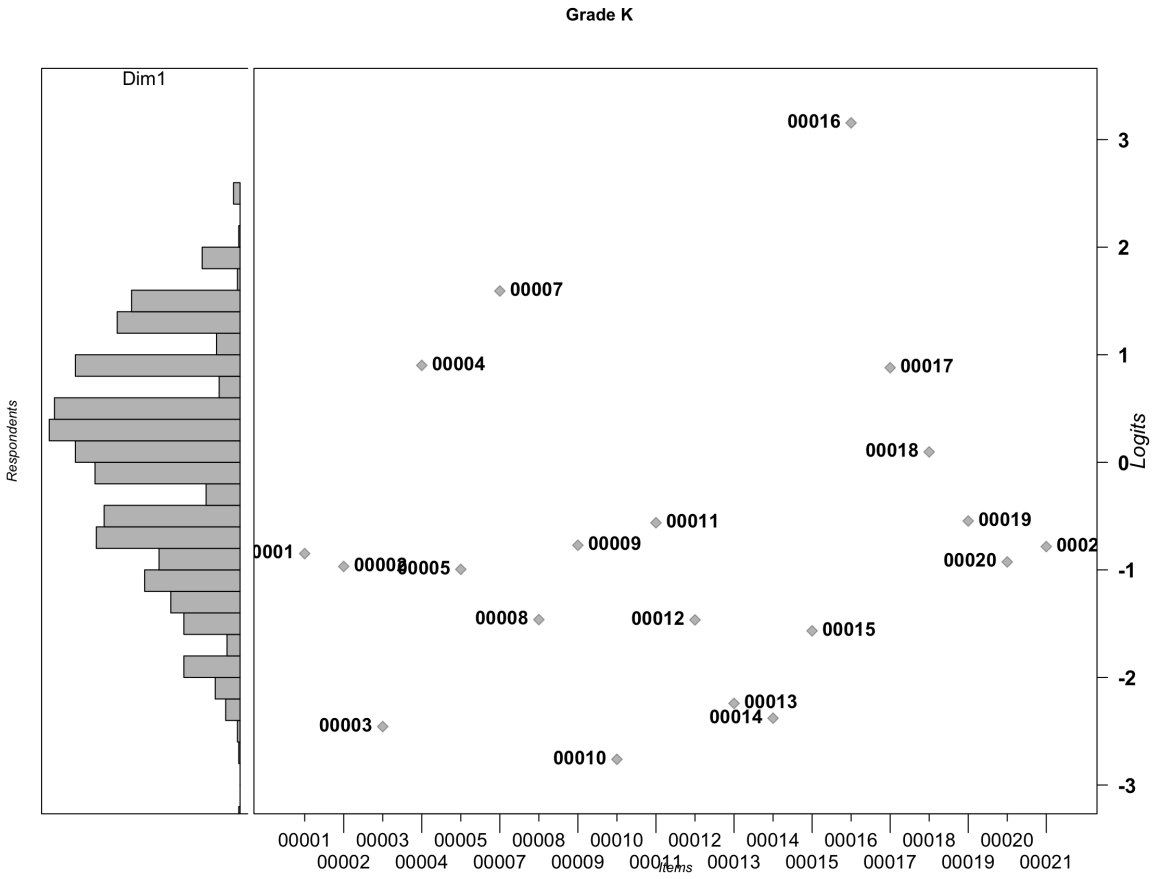

Figure A8. Grade 1 Wright Map Showing the Item and Person Estimates From the Rasch Model on a Similar Scale

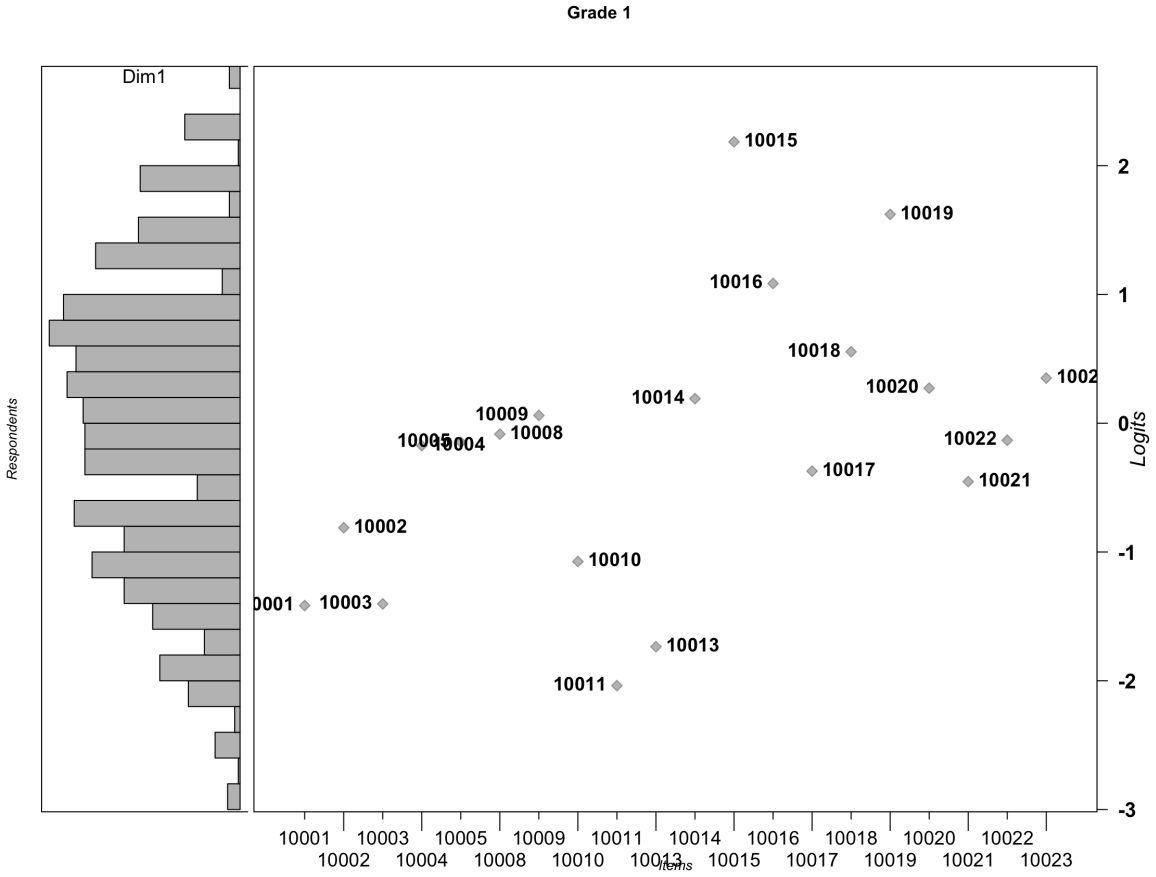

Figure A9. Grade 2 Wright Map Showing the Item and Person Estimates From the Rasch Model on a Similar Scale

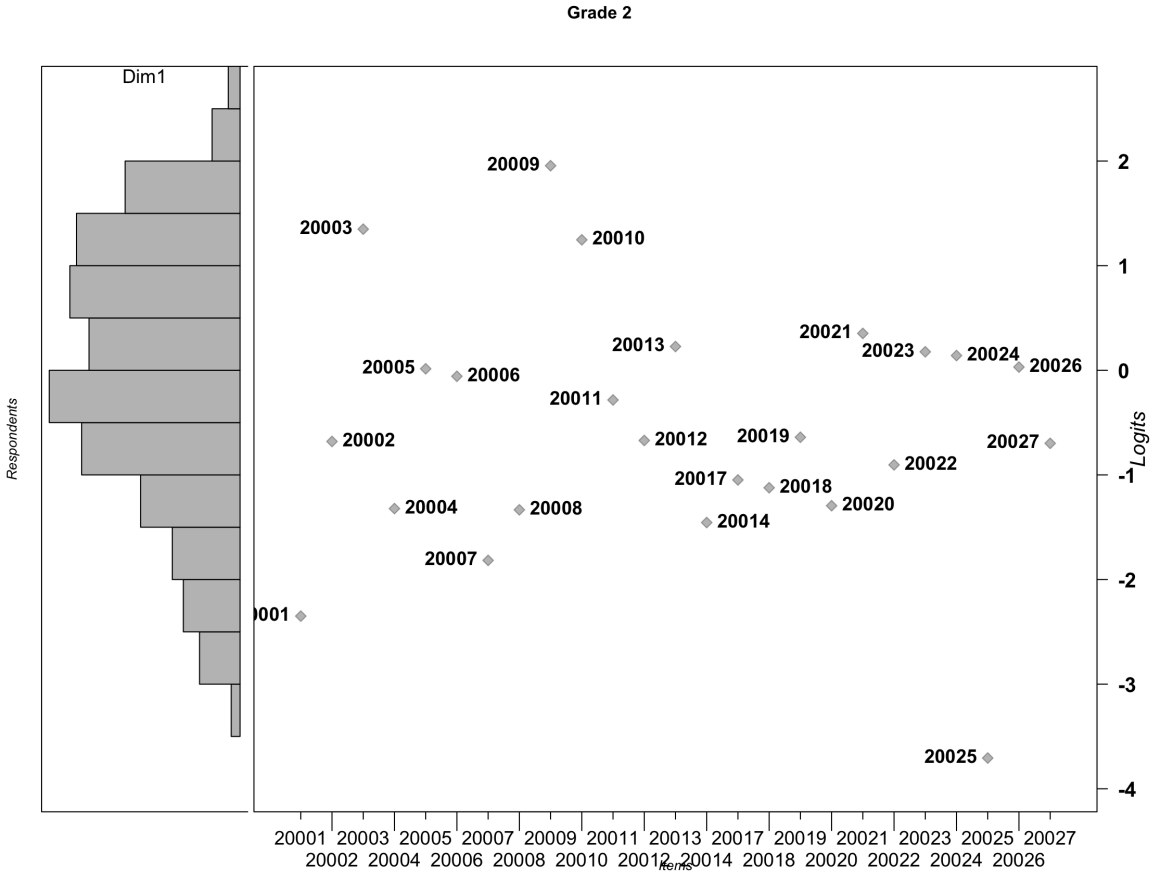

Figure A10. Grade 3 Wright Map Showing the Item and Person Estimates From the Rasch Model on a Similar Scale

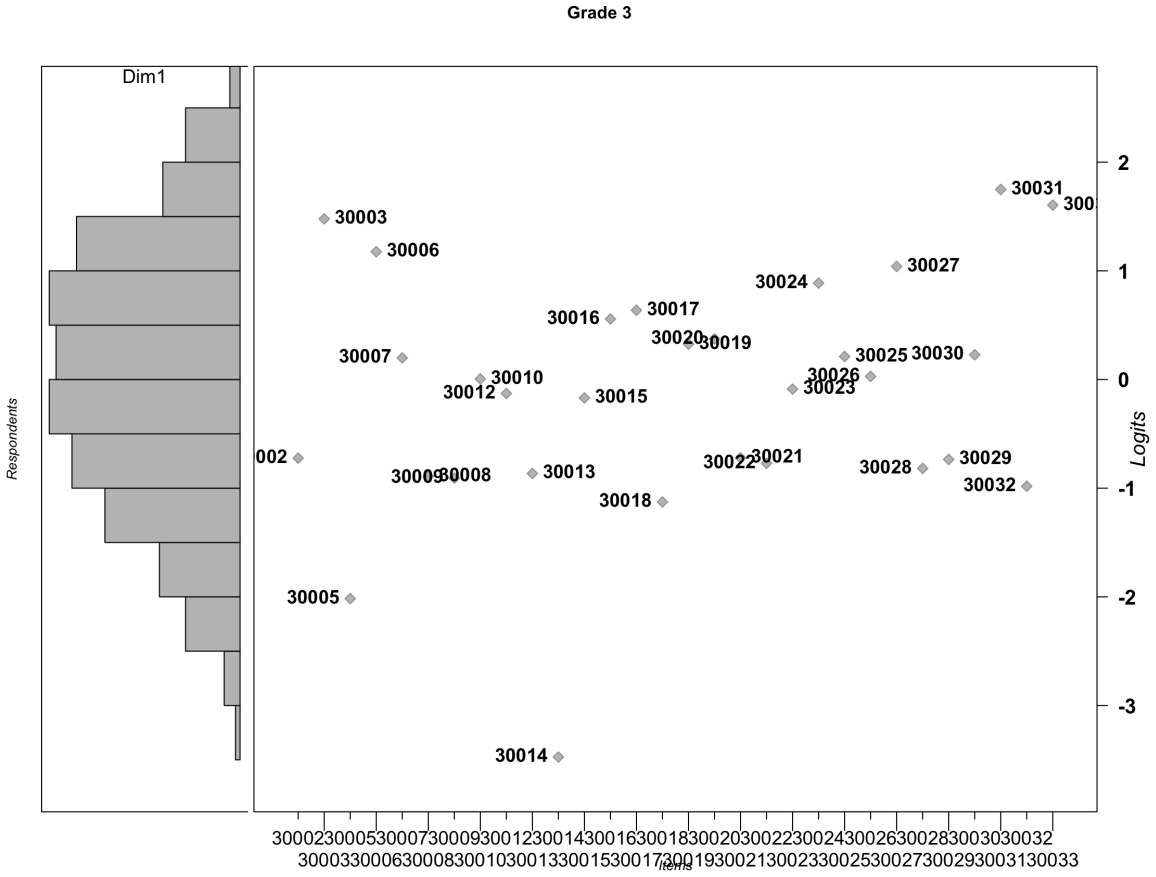

Figure A11. Grade 4 Wright Map Showing the Item and Person Estimates From the Rasch Model on a Similar Scale

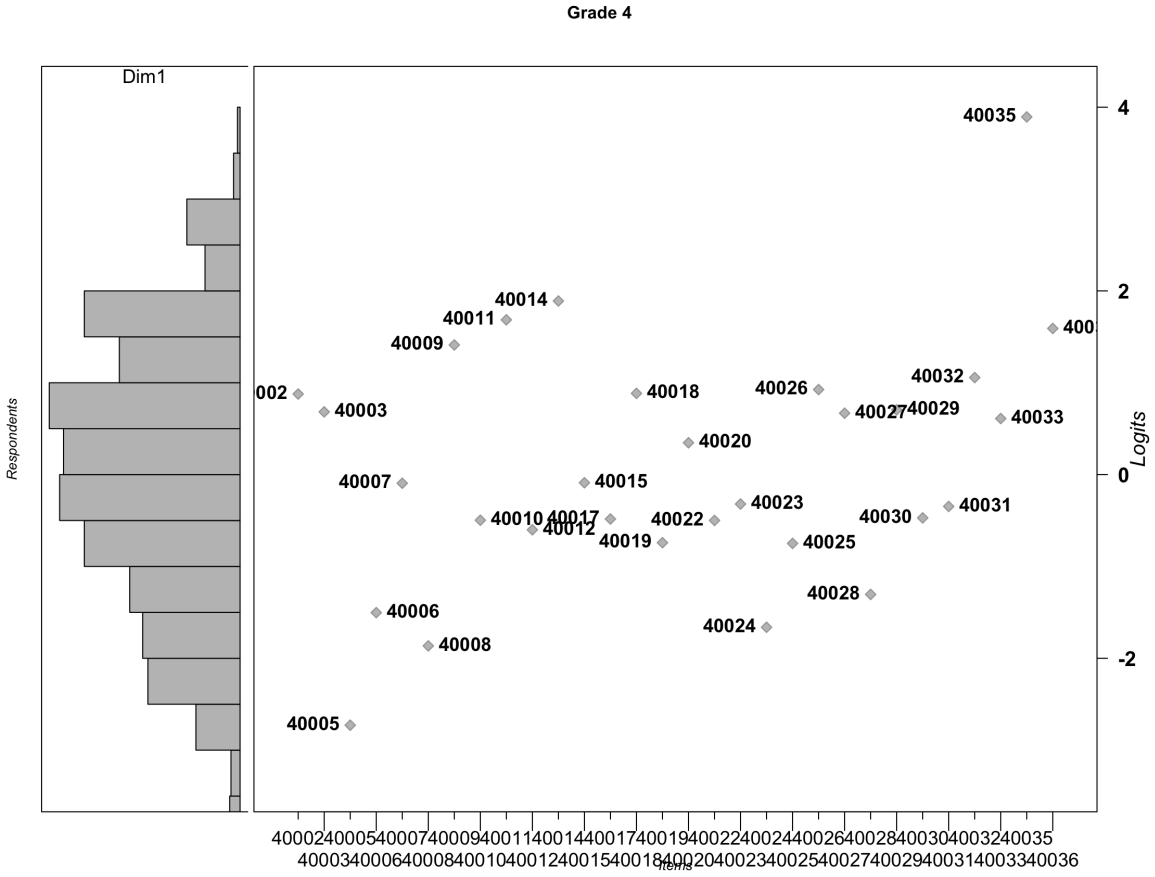

Figure A12. Grade 5 Wright Map Showing the Item and Person Estimates From the Rasch Model on a Similar Scale