Appendix 2-J: Pilot Development Partner Report: Pilot Phase 2 Test Results and Item Statistics

Test and Item Statistics

Item Statistics

All Pilot Phase 2 test items underwent comprehensive statistical analyses to provide NCSC with the information necessary to make informed decisions about the likelihood of each item providing reliable information that could be used in drawing valid inferences concerning student performance. The test development partner completed analyses using Pilot Phase 2 results to evaluate the performance of the developed identified (flagged) weaknesses of items for their classical difficulty or discrimination.

The conducted analyses were to answer critical questions about the developed test content and to inform item data reviews. First, Pilot Phase 2 analyses aimed to evaluate the appropriateness and accessibility of developed items and test forms to a sample of the student population. Second, analyses aimed to identify (flag) weaknesses of items for their difficulty or discrimination in advance of an item data review (e.g., incorrect keys, unmanageable administration characteristics, otherwise problematic design) and to finalize the operational blueprint. Third, analyses sought to provide item difficulty estimates to build the NCSC item bank and test scales for the operational test. The following five sections describe the scoring rules and analyses and a summary of results: Analysis Sample Descriptive Statistics, Mathematics Pilot Phase 2 Test Results, Mathematics Pilot Phase 2 IRT Fit Results, Reading Pilot Phase 2 Test Results and Reading IRT Fit Results.

Methodology for Flagging of Items

The applied Item flagging criteria identified items that may have not been functioning as intended due to errors in content or rendering. Final item flagging criteria were the same as in Pilot Phase 1, and were developed based on input from the TAC and NCSC partners.

- Difficult item: Low p-value < 0.50 Tier 1 (2 answer choice options)

  For Tier 1 items in both reading, mathematics, and writing, there are only two answer choices. If the Tier 1 p-value was less than 0.50, the item was flagged and reviewed at the IDR.

- Difficult item: Low p-value <0.40 Tiers 2--4 (3 answer choice options)

  For Tiers 2--4 items in both reading, mathematics, and writing, there are three answer choices. If the p-value was less than 0.40 for Tiers 2–4 items, the item was flagged and reviewed at the IDR.

- Easy item: High p-value > 0.90

  If the p-value was greater than 0.90 for an item in reading, mathematics, and writing (SR only), the item was flagged and reviewed at IDR.

- Low biserial correlation < 0.00

When a student's performance on an item did not correlate with the student's performance on the rest of the test, the item was flagged for a low biserial correlation and was to be reviewed at IDR.

- Tier reversal: Tier 1 items harder than Tier 4

  An evaluation of item tiers was conducted for the Pilot Phase 2 item data review, which included ordering items by difficulty and evaluating the order in light of items' tier design. A tier reversal flag indicated when p-values were not in descending order from Tier 1 to Tier 4. In reading, where passages are written in families, there are four variations of each passage, one for each of Tier 1 to Tier 4. Each tier of a passage has an associated set of items. The number of items associated with a passage varies by tier as well as by passage. Because there is not a direct one-to-one correspondence of items across the tiers, mean p-values were calculated in order to identify instances of tier reversal in reading. It was expected that Tier 1 would have the highest p-values and Tier 4 would have the lowest p-values. Items were flagged for a tier reversal and was to be reviewed at IDR if the Tier 4 p-value was greater than Tier 1.

- Distractor analysis: Proportion selecting distractor greater than proportion selecting key

  Items were flagged for a distractor and reviewed at IDR when statistics for the answer choices revealed that students were being drawn to a distractor(s) more often than to the correct response. Items with two possible correct responses were flagged when the p-value of a distractor was similar or higher than the p-value of the correct response. This could mean that there was a miskey (correct response not correctly noted), a second possible correct response, or a distractor with elements of a correct response.

During IDR, items will be reviewed based on their statistical properties, as well as qualitatively observable problems (e.g., long scroll downs).

A summary of mathematics, reading, and writing items flagged by type of flag in each grade is provided in the next sections of this document.

Test Statistics

IRT

The test development partner conducted Rasch IRT analysis for Pilot Phase 2. In the Pilot Phase 2 test design phase, a minimum sample size of 100 per item was determined necessary for calibration. Nearly all forms reached the minimum target sample size of 100 per item. Concurrent calibration placed items on the same scale. Session 1 items were common across all forms, and they were used to anchor the Session 2 items to a common scale. For ELA, the test developer calibrated concurrently all reading forms and writing selected-response items at each grade. For mathematics, the test developer partner developed calibrated concurrently all forms at each grade. The foundational items were calibrated with verbal and nonverbal responses together. The test developer partner applied the Rasch partial credit model for 2-part and 3-part selected-response items in ELA.

Model fit

Exhibit 1 presents the test development partner's applied default criteria for convergence and model fit. Model-data fit was monitored using Mean Square (MS) infit and MS outfit statistics, which indicate the

degree of accuracy and predictability with which the data fit the IRT model (Linacre, 2002). Infit is sensitive to misfit on items targeted at the ability level of the person, whereas outfit is sensitive to misfit on items with difficulty far from the ability of the person (Linacre, 2002). Typically, values less than 0.7 or greater than 1.3 for MS infit indicate misfit, and values greater than 1.3 for MS outfit indicate misfit (Wright & Linacre, 1994). The infit and outfit statistics can also be expressed as z-scores to facilitate a statistical test of model-data fit. However, an item is more likely to be flagged for minor model-data fit issues when the analyses are based on very large sample sizes, when in fact the fit of the model to the data is acceptable. Given that the MS infit statistic best addresses item misfit for items that are well-targeted to the ability level of the student, the test development partner traditionally prioritized this statistic in evaluating the fit of the Rasch model. That is, model-data fit provides an evaluation as to whether the use of the Rasch model is an appropriate and coherent framework for all scaling, equating, and score reporting activities.

Exhibit 1. Convergence Criteria

Criteria	Value	Explanation
converge	L	Convergence is controlled by logit change size
lconv	0.001	Iterations stop when the largest logit change is less than or equal to 0.001
rconv	NA	NA
mjmle	0	No maximum number of JMLE iterations
mprox	10	The maximum number of PROX (normal approximation) iterations is 10, which provides starting values for the JMLE algorithm
stbias	No	Does not apply a correction for JMLE item bias
extrsc	0.3	This value is subtracted from maximum possible scores and added to minimum possible scores

Mathematics Pilot Phase 2 Test Results

The test development partner converted all item responses from the student's chosen answer into a scored data file. For selected-response items, the test development partners converted student responses to 0 if incorrect and 1 if correct.

Exhibit 2 contains form summary statistics for Mathematics. The first column of the Exhibit indicates the grade, and the second column indicates the form. The next column contains the total number of items on each form. Each item is worth one score point, meaning the total number of points and number of items is equal. The remaining columns of the Exhibit indicate the number of students taking each form, the mean raw score and standard deviation, the minimum and maximum raw score, mean p-value and standard deviation, and Cronbach's alpha. Overall, the mathematics forms were relatively difficult, with mean raw scores ranging from 43 percent to 53 percent of the total points/form.

Cronbach's alpha reliability coefficient is a frequently used measure of internal consistency over the responses to a set of items measuring an underlying, unidimensional trait. Reliability coefficient alpha expresses the consistency of test scores as the ratio of true score variance to total score (observed) variance (true score variance plus error variance). Clearly, a larger index would indicate that test scores were influenced less by random sources of error. The reliability coefficient is a "unitless" index, which can be compared from test to test and ranges from 0 to 1, where 0.80 is typically considered the minimally acceptable level of reliability for assessments like the CRCT. While sensitive to random error

associated with content sampling variability, the index is not sensitive to other types of errors, such as temporal stability or variability in performance that might occur across different testing occasions. Cronbach's alpha is computed as (Crocker & Algina, 1986)

$$\hat{\alpha} = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_j^2}{\sigma_X^2}\right),$$

where k = number of items, σ_X^2 = the total score variance, and σ_j^2 = the variance of item j.

Cronbach's alpha coefficients range from 0.66 to 0.88 across grades for mathematics. If 0.90 is considered a conservative criterion for an acceptable level of reliability, reliability for all grades fall below this criterion. Each grade has at least one form that meets the minimally acceptable level of 0.80.

Exhibit 2. Mathematics Form Summary Statistics

Grade	Form	Total (Points)	Summary Statistics							
			N	Mean Raw Score	SD Raw Score	Min Raw Score	Max Raw Score	Mean p-value	SD p-value	Alpha
3	1	40	178	19.66	7.49	1	40	0.51	0.12	0.85
	2	40	108	21.00	7.94	5	38	0.52	0.14	0.88
	3	40	100	18.60	7.63	0	40	0.49	0.14	0.86
	4	40	107	20.51	7.53	0	37	0.51	0.13	0.86
4	1	40	205	17.30	6.69	2	39	0.45	0.14	0.82
	2	40	91	17.77	6.07	0	31	0.45	0.15	0.78
	3	40	119	18.85	5.25	5	33	0.47	0.16	0.69
	4	40	127	16.80	5.46	2	32	0.44	0.16	0.72
5	1	40	189	18.17	5.09	1	35	0.47	0.15	0.66
	2	40	78	18.03	7.14	1	40	0.47	0.16	0.85
	3	40	95	17.60	5.08	3	31	0.45	0.15	0.67
	4	40	101	18.06	5.58	6	36	0.46	0.16	0.73
6	1	40	187	20.02	6.42	4	40	0.50	0.14	0.80
	2	40	102	19.61	6.72	1	36	0.50	0.14	0.82
	3	40	90	19.77	6.69	6	38	0.50	0.13	0.81
	4	40	109	19.70	6.30	7	38	0.49	0.14	0.79
7	1	40	175	18.86	5.72	6	37	0.48	0.16	0.75
	2	40	109	20.26	7.67	0	37	0.51	0.15	0.87
	3	40	103	20.70	6.92	1	38	0.51	0.16	0.83
	4	40	154	19.25	6.17	8	40	0.50	0.15	0.78
8	1	40	203	20.51	7.38	1	39	0.52	0.12	0.85
	2	40	108	20.16	6.40	2	36	0.51	0.13	0.79
	3	40	77	17.62	6.62	6	40	0.47	0.14	0.81
	4	40	123	17.11	7.45	0	39	0.48	0.13	0.85
11	1	40	210	19.25	7.15	2	39	0.49	0.12	0.84
	2	40	102	17.51	6.35	3	34	0.47	0.12	0.79
	3	40	60	19.13	6.04	9	37	0.50	0.12	0.76
	4	40	124	19.33	7.97	3	39	0.49	0.12	0.87

Exhibits 3 and 13 for mathematics and reading, respectively, provide a summary of the number of items flagged by type of flag in each grade. Exhibits 4 and 14 contain a summary of item flags by tier and Core Content Connector (CCC) or content type. Exhibits 5 and 15 contain a summary of item flags by form. Exhibits 6 and 16 contain a summary of p-values by tier.

Overall, 33.9 percent of all items were flagged for at least one reason in mathematics. The majority of items were flagged for a low p-value or a high distractor p-value. The number of items flagged per form was relatively similar across forms, with the exception of Grade 8, Form 4. Of the items flagged for omit, one grade 5 item, 20 grade 8 items, and one grade 11 item were flagged as a result of a not reached rate greater than 5 percent. The grade 4 item and seven of the grade 5 items were flagged due to a combined omit and not reached rate greater than 5 percent. In evaluating the number of items flagged and the adequacy of the flagging criteria, it is important to consider that all items are two-option or three-option selected-response or one-point constructed-response items.

Exhibit 3. Number of Items Flagged in Mathematics by Grade

Grade	Items	Type of Flag						
		Low p-value	High p-value	Low biserial	Distractor p-value	Tier reversal	Omit	At least one flag
3	100	23	0	0	15	0	0	23
4	100	45	0	13	28	0	1	49
5	100	45	0	10	32	0	8	49
6	100	23	0	6	20	0	0	24
7	100	26	0	5	21	0	0	26
8	100	30	0	4	15	2	21	40
11	100	25	0	3	14	0	2	26

Exhibit 4. Number of Items Flagged in Mathematics by Tier and CCC

Grade	CCC	Tier 1	Tier 2	Tier 3	Tier 4
3	MA_3_DPS_1G1				
	MA_3_GM_1I1				
	MA_3_ME_1D2			1	1
	MA_3_NO_1J3		2		
	MA_3_NO_1L3		2	2	
	MA_3_NO_2C1		3	2	
	MA_3_NO_2D3			1	1
	MA_3_NO_2E1			3	1
	MA_3_PRF_2D1			2	
	MA_3_SE_1G1		1	1	
4	MA_4_DPS_1G3		4	3	1
	MA_4_GM_1H2				1
	MA_4_ME_1G2				
	MA_4_NO_1J5		1	2	
	MA_4_NO_1M1		3	4	1
	MA_4_NO_1N2	1	3	2	1
	MA_4_NO_2D7		1	1	1
	MA_4_NO_2E2		4	3	1
	MA_4_PRF_1E3		1	2	1
	MA_4_SE_1G2		3	3	1
5	MA_5_GM_1C3		2	3	1
	MA_5_ME_1B2	1		2	1
	MA_5_ME_2A1		2	4	
	MA_5_NO_1B1				
	MA_5_NO_1B4		1	1	1
	MA_5_NO_2A5		3	4	
	MA_5_NO_2C1			4	1
	MA_5_NO_2C2	1	1	4	1
	MA_5_PRF_1A1		2	1	
	MA_5_PRF_2B1	1	2	4	1
6	MA_6_DPS_1D3		3	2	1
	MA_6_GM_1D1		1	1	1
	MA_6_ME_2A2			4	1
	MA_6_NO_1D2				
	MA_6_NO_1D4			1	1
	MA_6_NO_1F1				1
	MA_6_NO_2A6		3	2	1
	MA_6_NO_2C3				
	MA_6_PRF_1C1		1		
	MA_6_PRF_1D1				

Exhibit 4. Number of Items Flagged in Mathematics by Tier and CCC (cont.)

Grade	CCC	Tier 1	Tier 2	Tier 3	Tier 4
7	MA_7_DPS_1K1				1
	MA_7_GM_1H2			2	1
	MA_7_ME_2D1			1	1
	MA_7_NO_2F1			1	1
	MA_7_NO_2F2			2	1
	MA_7_NO_2F6			2	2
	MA_7_NO_2I1		1	2	
	MA_7_NO_2I2		3	1	
	MA_7_PRF_1F1		1	2	1
	MA_7_PRF_1G2				
8	MA_8_DPS_1H1		2	1	
	MA_8_DPS_1K2		1		1
	MA_8_GM_1G1		1		1
	MA_8_ME_1E1		1	2	1
	MA_8_ME_2D2	1	1		
	MA_8_NO_1K3		3	2	1
	MA_8_PRF_1E2		2	2	1
	MA_8_PRF_1F2		2	2	
	MA_8_PRF_1G3	2	2	2	1
	MA_8_PRF_2E2	1		3	1
11	MA_H_DPS_1B1		2	1	
	MA_H_DPS_1C1			1	1
	MA_H_GM_1B1		1	2	1
	MA_H_ME_1A2			1	1
	MA_H_ME_1B2			1	1
	MA_H_NO_1A1	1		4	
	MA_H_PRF_1C1	1			1
	MA_H_PRF_2B1		1		
	MA_H_PRF_2B2				
	MA_H_PRF_2C1		2	2	1

Exhibit 5. Number of Items Flagged in Mathematics by Form

Grade	Form	Number of Items	Items with at least one flag
3	1	40	6
	2	40	7
	3	40	10
	4	40	9
4	1	40	16
	2	40	17
	3	40	17
	4	40	20
5	1	40	14
	2	40	21
	3	40	18
	4	40	17
6	1	40	9
	2	40	8
	3	40	9
	4	40	10
7	1	40	14
	2	40	9
	3	40	10
	4	40	11
8	1	40	8
	2	40	9
	3	40	15
	4	40	26
11	1	40	10
	2	40	14
	3	40	10
	4	40	10

Exhibit 6 contains a summary of mean p-values by grade and tier for mathematics. The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the tier. The next column contains the total number of items. Each item is worth one score point, meaning the total number of score points is the same as the number of items. The last two columns of the Exhibit indicate the mean p-value and standard deviation of p-value for each tier. Following the design of the items and tiers, it would be expect to see the highest p-values at Tier 1 and the lowest p-values at Tier 4. As expected, the mean p-values for Tier 1 were always highest (easier) as compared to the other tiers. In general, mean p-values for the other tiers followed the expected pattern; however, the means were very similar across Tiers 2–4.

Exhibit 6. Summary of p-value by Tier in Mathematics

Grade	Tier	Items (Points)	Mean p-value	SD p-value
3	1	20	0.663	0.072
	2	35	0.497	0.114
	3	35	0.434	0.092
	4	10	0.398	0.102
4	1	20	0.653	0.119
	2	35	0.398	0.117
	3	35	0.393	0.095
	4	10	0.354	0.098
5	1	20	0.697	0.095
	2	35	0.446	0.086
	3	35	0.344	0.080
	4	10	0.375	0.093
6	1	20	0.698	0.070
	2	35	0.498	0.119
	3	35	0.440	0.094
	4	10	0.418	0.127
7	1	20	0.713	0.088
	2	35	0.491	0.112
	3	35	0.420	0.071
	4	10	0.358	0.097
8	1	20	0.662	0.142
	2	35	0.462	0.106
	3	35	0.452	0.104
	4	10	0.374	0.092
11	1	20	0.665	0.100
	2	35	0.456	0.077
	3	35	0.437	0.066
	4	10	0.394	0.078

Mathematics IRT Fit Results

Exhibit 7 contains a summary of the item difficulty estimates by grade for mathematics. The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the statistic (e.g., mean, median, standard deviation). The next six columns provide values for each of the statistics for item measure (difficulty), standard error, mean square infit, standardized infit, mean square outfit and standardized outfit. Exhibit 28 contains the same summary of the person ability estimates by grade for mathematics. The distribution of item measures is slightly negatively skewed, while the distribution of person measures is slightly positively skewed.

Exhibit 7. Mathematics Item Measure Summary

Grade	Statistic	Measure	SE	InMS	InZStd	OutMS	OutZStd
3	Items	100	100	100	100	100	100
	Mean	0.00	0.19	1.00	-0.08	1.00	-0.05
	Standard Deviation	0.65	0.05	0.12	1.96	0.18	1.77
	Min	-1.40	0.10	0.66	-6.21	0.59	-5.11
	10th	-0.84	0.10	0.84	-2.69	0.77	-2.22
	25th	-0.53	0.16	0.92	-1.19	0.86	-1.21
	50th	0.08	0.22	0.99	-0.06	0.99	-0.07
	75th	0.50	0.22	1.09	1.14	1.12	0.94
	90th	0.84	0.24	1.17	2.35	1.22	2.50
	Max	1.42	0.25	1.35	4.49	1.48	3.93
4	Items	100	100	100	100	100	100
	Mean	0.00	0.18	1.00	-0.10	1.00	-0.09
	Standard Deviation	0.71	0.05	0.11	1.89	0.15	1.77
	Min	-1.71	0.09	0.79	-5.39	0.74	-4.66
	10th	-0.99	0.09	0.86	-3.00	0.81	-2.57
	25th	-0.31	0.15	0.93	-1.27	0.89	-1.28
	50th	0.14	0.19	1.00	0.08	1.00	-0.04
	75th	0.42	0.22	1.07	0.99	1.09	1.11
	90th	0.74	0.24	1.15	2.07	1.20	1.96
	Max	1.60	0.30	1.30	5.26	1.48	4.64
5	Items	100	100	100	100	100	100
	Mean	0.00	0.20	1.00	-0.01	1.00	-0.04
	Standard Deviation	0.72	0.06	0.08	1.10	0.12	1.23
	Min	-1.74	0.10	0.82	-3.41	0.77	-3.04
	10th	-1.13	0.10	0.92	-1.32	0.86	-1.39
	25th	-0.34	0.15	0.94	-0.67	0.92	-0.85
	50th	0.11	0.22	0.99	-0.03	0.99	-0.09
	75th	0.54	0.24	1.04	0.69	1.06	0.67
	90th	0.83	0.26	1.09	1.41	1.14	1.38
	Max	1.19	0.31	1.41	2.97	1.50	3.91

Exhibit 7. Mathematics Item Measure Summary (cont.)

Grade	Statistic	Measure	SE	InMS	InZStd	OutMS	OutZStd
6	Items	100	100	100	100	100	100
	Mean	0.00	0.19	0.99	-0.18	0.98	-0.18
	Standard Deviation	0.68	0.05	0.11	1.69	0.15	1.55
	Min	-1.89	0.10	0.79	-3.54	0.72	-2.83
	10th	-0.95	0.10	0.87	-2.47	0.80	-2.15
	25th	-0.54	0.16	0.92	-1.34	0.87	-1.40
	50th	0.09	0.21	0.99	-0.18	0.98	-0.23
	75th	0.44	0.23	1.04	0.60	1.05	0.52
	90th	0.92	0.24	1.17	1.82	1.18	1.83
	Max	1.52	0.28	1.32	5.29	1.45	4.74
7	Items	100	100	100	100	100	100
	Mean	0.00	0.18	1.00	-0.12	0.99	-0.11
	Standard Deviation	0.72	0.05	0.10	1.53	0.16	1.54
	Min	-2.11	0.09	0.81	-2.85	0.74	-2.75
	10th	-1.10	0.09	0.89	-1.94	0.83	-1.87
	25th	-0.48	0.16	0.92	-1.19	0.88	-1.29
	50th	0.20	0.18	0.99	-0.10	0.97	-0.35
	75th	0.47	0.21	1.05	0.71	1.09	1.12
	90th	0.84	0.22	1.13	1.88	1.22	2.07
	Max	1.34	0.27	1.31	3.94	1.46	3.80
8	Items	100	100	100	100	100	100
	Mean	0.00	0.19	0.99	-0.10	0.99	-0.11
	Standard Deviation	0.70	0.06	0.08	1.35	0.16	1.25
	Min	-2.56	0.09	0.84	-3.28	0.73	-2.72
	10th	-0.86	0.10	0.91	-1.55	0.86	-1.40
	25th	-0.31	0.15	0.95	-1.06	0.91	-0.93
	50th	0.12	0.20	0.99	-0.24	0.96	-0.33
	75th	0.43	0.23	1.03	0.49	1.05	0.42
	90th	0.66	0.25	1.11	1.56	1.16	1.77
	Max	1.63	0.38	1.22	5.38	1.80	3.70
11	Items	100	100	100	100	100	100
	Mean	0.00	0.19	1.00	-0.09	0.99	-0.06
	Standard Deviation	0.57	0.06	0.10	1.64	0.15	1.51
	Min	-1.55	0.10	0.79	-3.70	0.71	-2.89
	10th	-0.95	0.10	0.89	-2.12	0.83	-1.86
	25th	-0.28	0.15	0.92	-1.08	0.88	-1.02
	50th	0.12	0.20	0.98	-0.24	0.97	-0.34
	75th	0.40	0.22	1.04	0.69	1.06	0.73
	90th	0.62	0.28	1.15	1.99	1.19	1.83
	Max	1.07	0.30	1.37	5.37	1.71	4.93

Exhibit 8. Mathematics Person Measure Summary

Grade	Variable	Measure	SE	InMS	InZStd	OutMS	OutZStd

3	Persons	493	493	493	493	493	493
	Mean	-0.03	0.38	1.00	0.00	1.00	0.00
	Standard Deviation	1.06	0.17	0.11	0.90	0.20	0.95
	Min	-3.98	0.33	0.73	-3.13	0.38	-2.98
	10th	-1.11	0.33	0.87	-1.05	0.81	-1.12
	25th	-0.71	0.33	0.93	-0.52	0.89	-0.58
	50th	-0.16	0.34	1.00	0.05	0.98	-0.05
	75th	0.52	0.37	1.07	0.51	1.09	0.61
	90th	1.34	0.43	1.14	1.08	1.23	1.23
	Max	5.03	1.91	1.30	2.85	2.67	2.82
4	Persons	542	542	542	542	542	542
	Mean	-0.30	0.36	1.00	-0.01	1.00	-0.01
	Standard Deviation	0.75	0.13	0.12	0.98	0.16	1.00
	Min	-3.18	0.33	0.72	-2.65	0.32	-2.37
	10th	-1.22	0.33	0.85	-1.33	0.80	-1.28
	25th	-0.78	0.34	0.92	-0.62	0.88	-0.73
	50th	-0.35	0.34	0.99	0.00	0.98	-0.03
	75th	0.15	0.36	1.08	0.64	1.10	0.67
	90th	0.61	0.39	1.17	1.30	1.23	1.37
	Max	3.82	2.18	1.36	2.61	1.54	2.98
5	Persons	463	463	463	463	463	463
	Mean	-0.24	0.36	1.00	-0.01	1.00	-0.03
	Standard Deviation	0.72	0.12	0.12	1.00	0.18	1.01
	Min	-3.50	0.33	0.64	-3.52	0.41	-3.38
	10th	-1.02	0.33	0.86	-1.29	0.82	-1.28
	25th	-0.65	0.33	0.91	-0.70	0.88	-0.74
	50th	-0.31	0.34	1.00	0.01	0.98	-0.06
	75th	0.06	0.35	1.08	0.66	1.10	0.66
	90th	0.51	0.38	1.17	1.27	1.21	1.32
	Max	5.07	1.99	1.44	3.22	2.83	3.09
6	Persons	488	488	488	488	488	488
	Mean	0.06	0.36	1.00	0.00	0.99	0.02
	Standard Deviation	0.84	0.09	0.12	1.05	0.17	1.07
	Min	-3.21	0.33	0.72	-3.29	0.22	-3.10
	10th	-0.77	0.33	0.85	-1.35	0.79	-1.31
	25th	-0.52	0.33	0.91	-0.65	0.88	-0.73
	50th	-0.07	0.34	0.99	-0.03	0.99	-0.03
	75th	0.48	0.36	1.08	0.71	1.10	0.77
	90th	1.21	0.39	1.16	1.36	1.22	1.46
	Max	5.10	1.83	1.43	3.04	1.61	3.36

Exhibit 8. Mathematics Person Measure Summary (cont.)

Grade	Variable	Measure	SE	InMS	InZStd	OutMS	OutZStd
7	Persons	541	541	541	541	541	541
	Mean	-0.02	0.37	1.00	-0.02	0.99	0.01

	Standard Deviation	0.86	0.14	0.13	1.05	0.20	1.08
	Min	-3.10	0.33	0.72	-2.90	0.32	-2.30
	10th	-0.89	0.34	0.84	-1.29	0.76	-1.28
	25th	-0.55	0.34	0.90	-0.70	0.85	-0.79
	50th	-0.20	0.34	0.98	-0.08	0.97	-0.12
	75th	0.39	0.36	1.08	0.68	1.12	0.80
	90th	1.14	0.39	1.19	1.43	1.26	1.48
	Max	5.03	2.18	1.45	3.95	1.86	4.50
8	Persons	511	511	511	511	511	511
	Mean	-0.04	0.38	1.00	0.03	1.00	0.04
	Standard Deviation	0.92	0.16	0.10	0.89	0.15	0.94
	Min	-3.90	0.33	0.73	-2.63	0.47	-2.34
	10th	-0.91	0.33	0.89	-0.95	0.84	-0.99
	25th	-0.60	0.33	0.93	-0.55	0.90	-0.60
	50th	-0.15	0.34	0.98	-0.05	0.98	-0.05
	75th	0.40	0.36	1.06	0.53	1.08	0.56
	90th	0.99	0.41	1.15	1.25	1.20	1.37
	Max	5.16	2.18	1.44	3.11	1.78	3.33
11	Persons	496	496	496	496	496	496
	Mean	-0.04	0.36	1.00	0.00	0.99	0.00
	Standard Deviation	0.92	0.08	0.09	0.81	0.14	0.86
	Min	-2.48	0.33	0.74	-2.67	0.43	-2.64
	10th	-0.96	0.33	0.89	-1.04	0.84	-1.12
	25th	-0.61	0.33	0.94	-0.48	0.90	-0.55
	50th	-0.25	0.34	0.99	0.04	0.99	-0.03
	75th	0.27	0.36	1.06	0.47	1.07	0.58
	90th	1.21	0.41	1.12	1.03	1.17	1.13
	Max	3.86	1.03	1.33	2.20	1.54	2.40

Exhibit 9 contains a summary of the principal component analysis of residual variation, as produced by WINSTEPS. The first column of Exhibit 9 indicates the grade. The second column of the Exhibit indicates the number of items, and the remaining three columns indicate the percent of variation associated with each dimension. As expected, the first dimension associated with the Rasch model accounts for the majority of score variation.

Exhibit 9. Principal Component Analysis in Mathematics

Grade	Number of Items	Percent First Dimension	Percent Second Dimension	Percent Third Dimension
3	100	33.3	2.5	1.8
4	100	27.0	3.0	2.1
5	100	23.5	3.7	2.8
6	100	27.5	3.0	2.2
7	100	32.6	3.3	2.5
8	100	32.7	2.1	1.6
11	100	32.8	2.2	1.8

Exhibit 10 contains a summary of the items flagged for misfit. The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the total number of items in each grade. The next three columns indicate the number of items flagged for infit, outfit, and the total number of items flagged by either infit or outfit. Overall, only a small number of items were flagged for misfit in each grade. In addition, the majority of these items were flagged for outfit, which is influenced by outliers. Exhibit 11 provides the statistics for each flagged item. Figures 1–5 contain item characteristic curves (ICCs) for each of the items flagged for infit.

Exhibit 10. Number of Mathematics Items Flagged for Misfit

Grade	Number of Items	Infit	Outfit	Total
3	100	2	5	6
4	100	0	2	2
5	100	1	2	2
6	100	1	5	5
7	100	1	5	5
8	100	0	5	5
11	100	1	3	3

Exhibit 11. Item Statistics for Misfitting Mathematics Items

Grade	Item	Measure	SE	InMS	InZstd	OutMS	OutZStd	Infit	Outfit
3	111434A	-1.34	0.11	0.97	-0.56	1.31	2.88		1
3	111399A	-0.66	0.22	1.10	1.31	1.48	3.37		1
3	112560A	0.32	0.22	1.26	2.85	1.44	3.39		1
3	112562A	0.50	0.22	1.35	3.47	1.44	3.23	1	1
3	122087A	0.18	0.21	0.66	-4.79	0.59	-4.08	1	
3	112561A	0.55	0.17	1.21	2.73	1.33	2.68		1
4	111703B	0.71	0.17	1.19	2.39	1.48	3.66		1
4	111718A	0.25	0.23	1.30	3.48	1.39	3.82		1
5	111304A	0.52	0.16	1.09	1.31	1.37	3.89		1
5	111264A	0.86	0.28	1.41	2.56	1.50	2.44	1	1
6	112667A	1.19	0.11	1.22	3.78	1.34	4.22		1
6	110984A	1.02	0.22	1.20	1.88	1.31	2.17		1
6	111489A	1.29	0.25	1.24	1.75	1.45	2.39		1
6	111462A	1.52	0.25	1.29	1.89	1.42	2.09		1
6	111466A	0.19	0.23	1.32	3.92	1.36	3.35	1	1
7	120525A	1.34	0.11	1.26	3.94	1.34	3.80		1
7	111751B	0.96	0.23	1.21	1.99	1.32	2.32		1
7	111781A	1.06	0.20	1.31	2.77	1.46	3.10	1	1
7	111786A	0.48	0.22	1.27	3.01	1.36	3.12		1
7	112854A	0.21	0.21	1.26	3.24	1.36	3.32		1
8	111253B	0.43	0.25	1.21	2.04	1.37	2.38		1
8	117075A	0.98	0.23	1.22	1.78	1.41	2.63		1
8	113977B	1.52	0.32	1.15	0.76	1.61	1.95		1
8	113976A	1.62	0.26	1.22	1.36	1.80	3.14		1
8	120571A	0.46	0.21	1.22	2.56	1.31	2.69		1
11	112734A	0.65	0.10	1.26	5.37	1.32	4.93		1
11	111559A	1.07	0.22	1.37	3.04	1.71	3.79	1	1
11	122021A	0.48	0.20	1.25	2.96	1.42	2.97		1

Figure 1. ICCs for Grade 3 Math Items Flagged for Infit

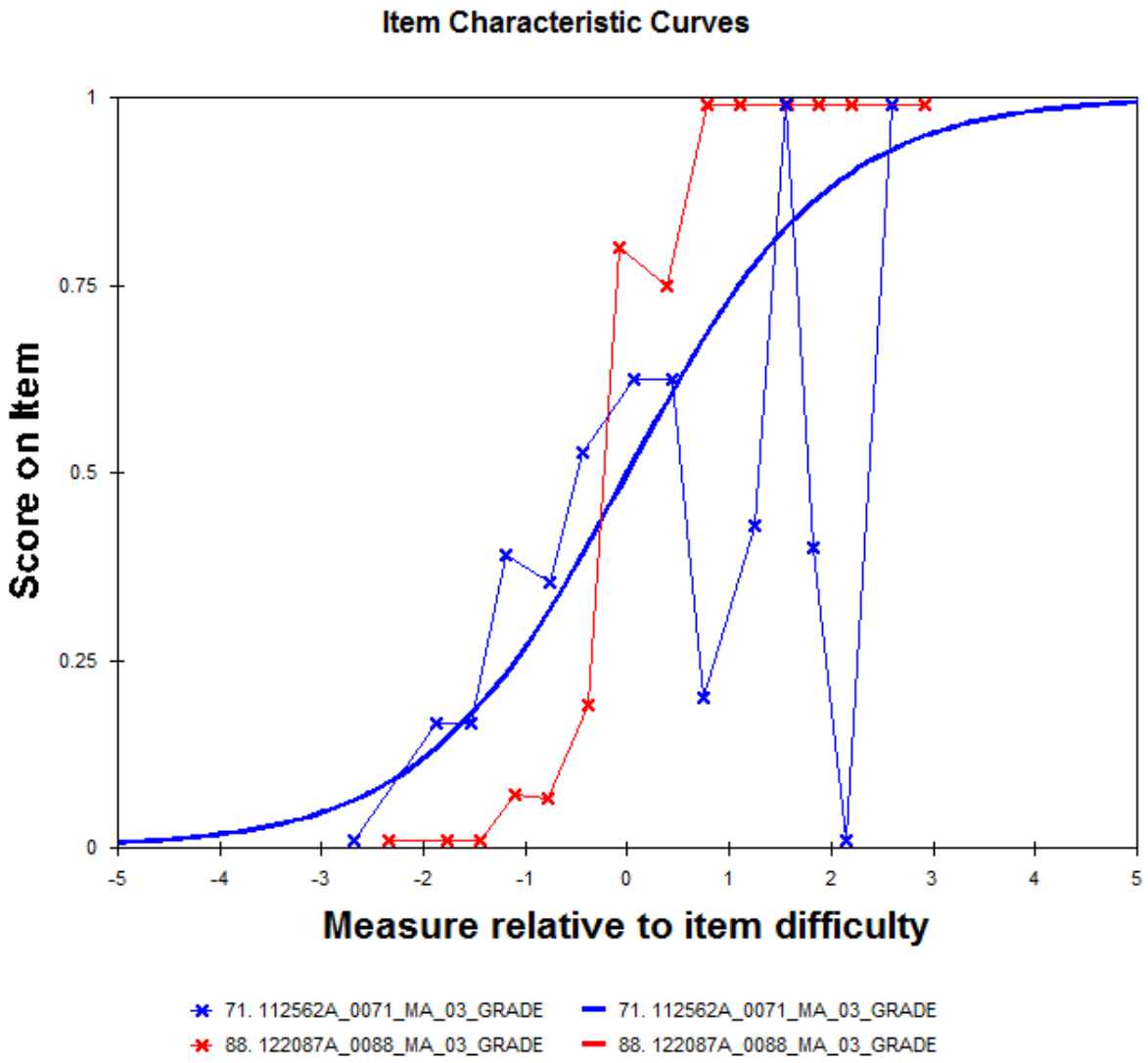


Figure 2. ICC for Grade 5 Math Item Flagged for Infit

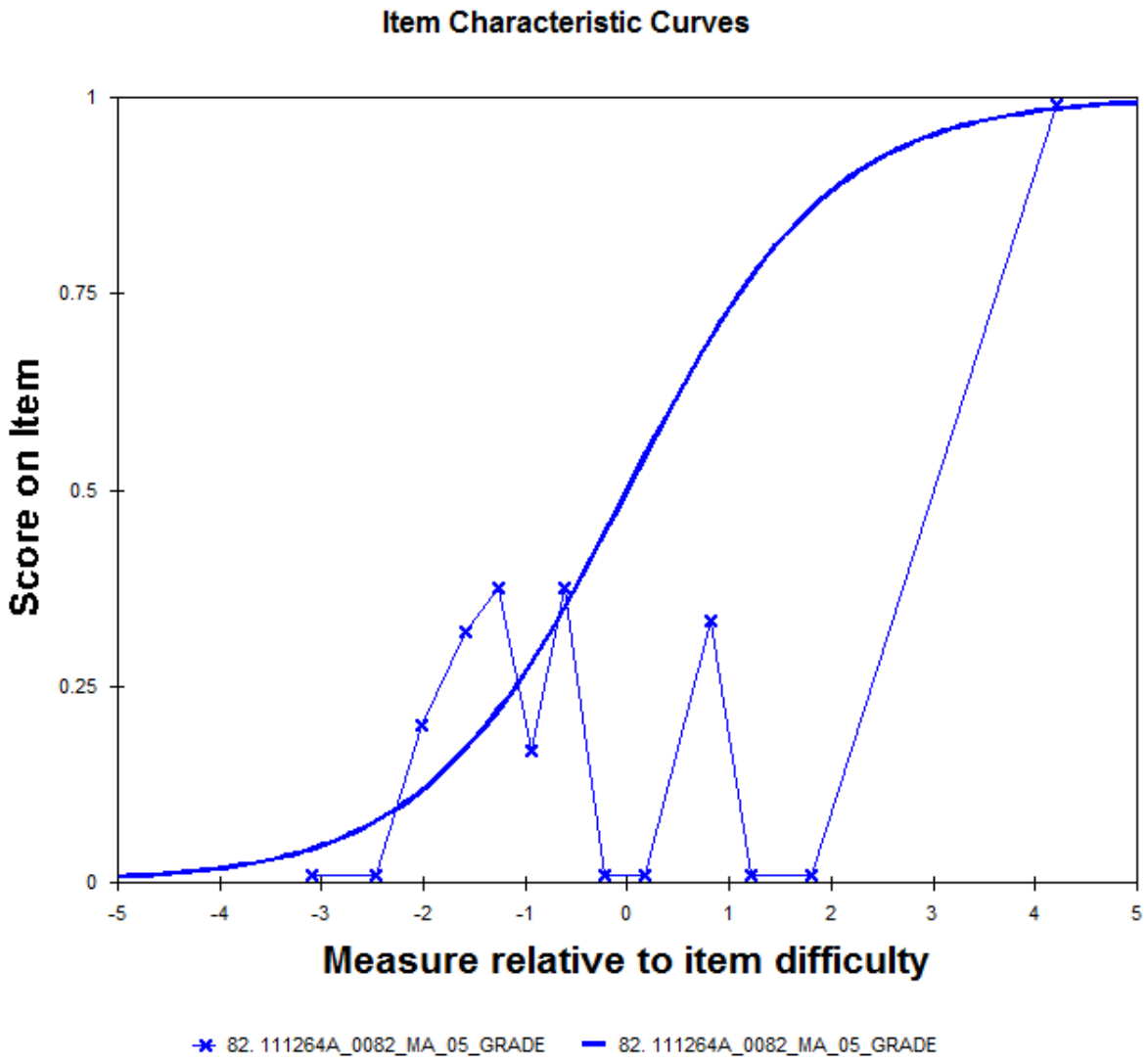


Figure 3. ICC for Grade 6 Math Item Flagged for Infit

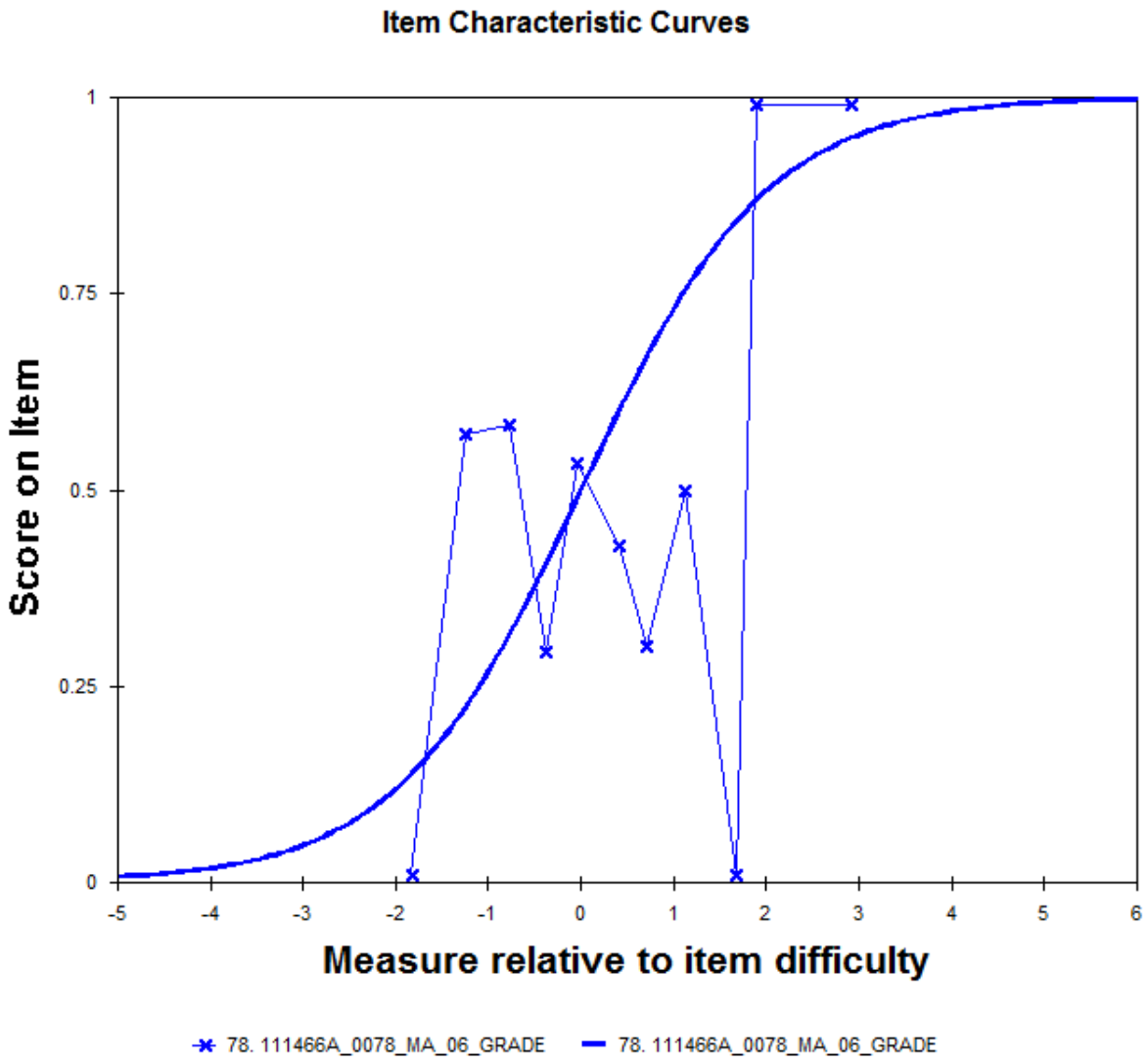


Figure 4. ICC for Grade 7 Math Item Flagged for Infit

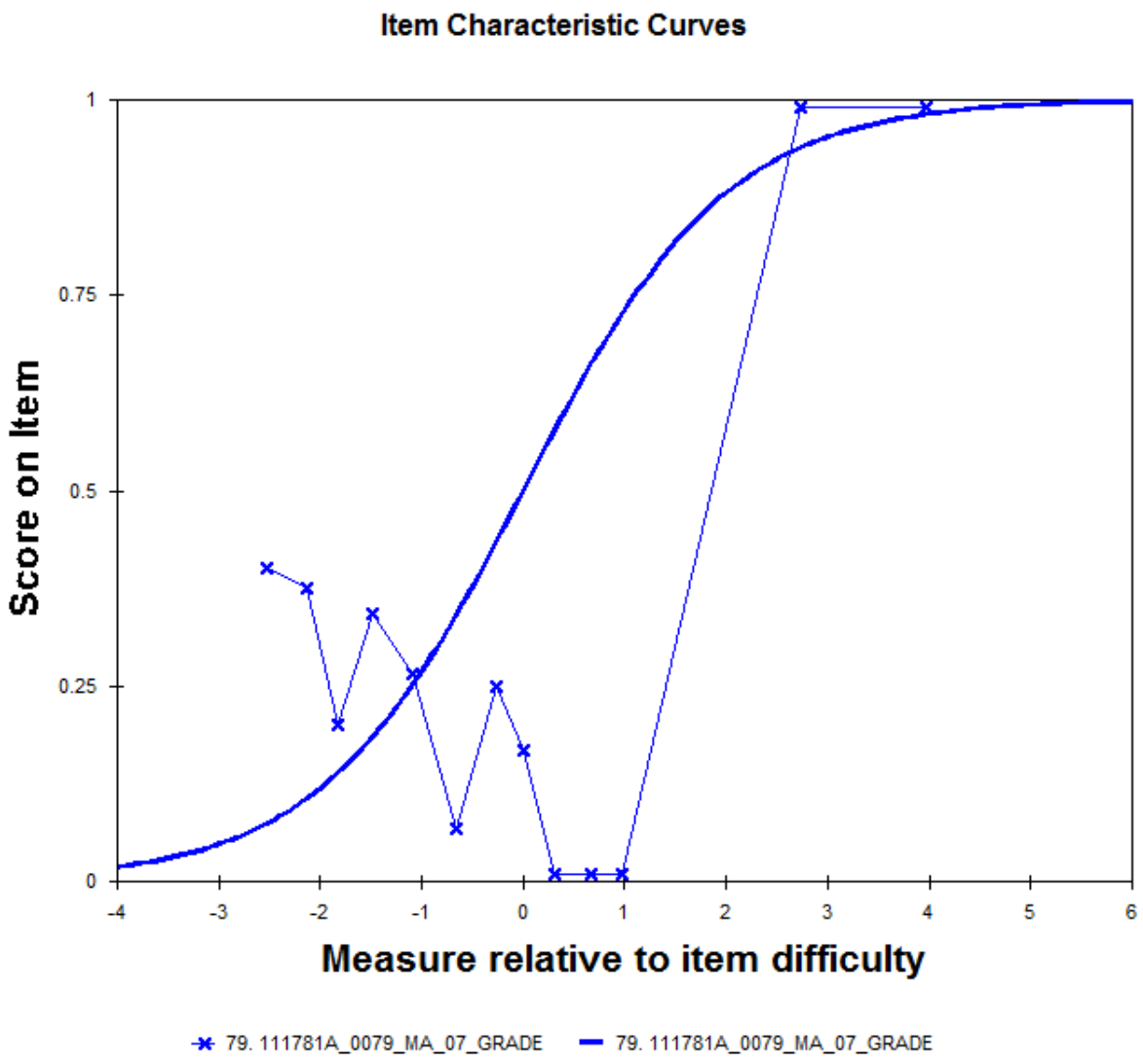
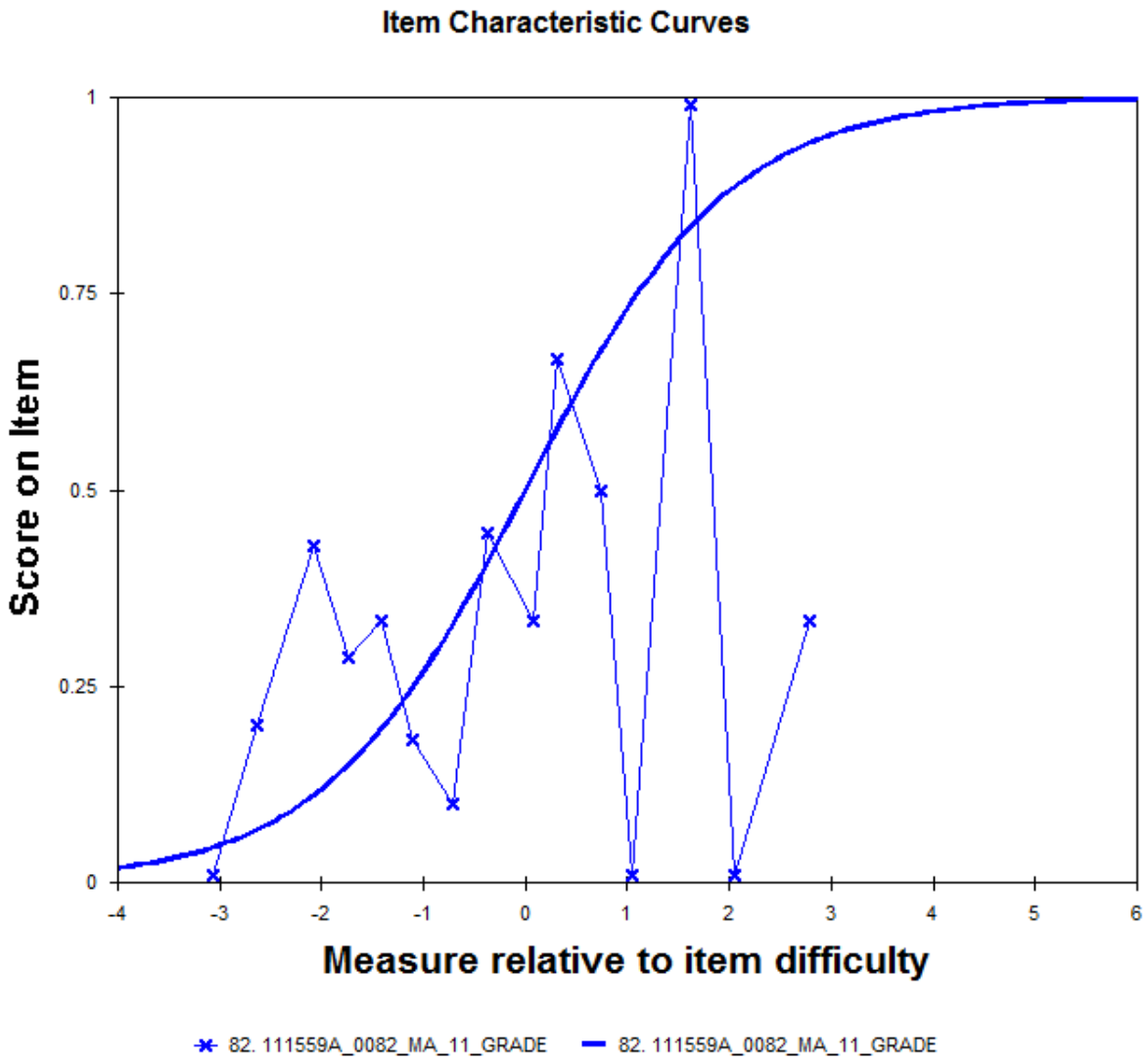


Figure 5. ICC for Grade 11 Math Item Flagged for Infit



Reading Pilot Phase 2 Test Results

The test development partner converted all item responses from the student's chosen answer into a scored data file. For selected-response items, the test development partners converted student responses to 0 if incorrect and 1 if correct.

In grades 3 and 4, reading foundational items were scored such that a set of 3 to 5 items was worth a maximum of one point. Initially, analyses were conducted such that a student would receive a score of 1 if they answered all of the items correctly and a score of 0 if any of the foundational items were answered incorrectly. However, the foundational items scored in this manner were very difficult. Therefore, a new scoring rule was applied. At Tier 1, if students answered all three items in the set correctly, they received a score of 1. They received a score of 0 if they answered any incorrectly. At Tiers

2–4, if students answered at least 4 out of 5 items correctly, they received a score of 1. If they answered less than 4 correctly, they received a score of 0.

For 2-part and 3-part selected-response items in reading, the desired scoring was for items to be scored as dependent items, where students received partial credit for a correct answer on any of the items. In order to account for the dependency of the items while maintaining integer scoring, the test development partner analyzed each set of items as a single polytomous item, with a maximum of 2 or 3 score points.

Exhibit 12 contains form summary statistics for reading. The first column of the Exhibit indicates the grade, and the second column indicates the form. The next column contains the total number of items and points on each form. Given that the test development partner evaluated some items as a single polytomous item, with a maximum of 2 or 3 score points, in column three, there are more points possible than there are total "items." The remaining columns of the Exhibit indicate the number of students taking each form, the mean raw score and standard deviation, the minimum and maximum raw score, mean p-value and standard deviation, and Cronbach's alpha. Overall, the reading forms were moderately difficult, with mean raw scores ranging from 53percent to 65 percent of the total points/form. Cronbach's alpha ranged from 0.76 to 0.88 across grades. All forms fell below the criterion of 0.90, but all except three forms met the minimally acceptable level of 0.80.

When reviewing the reading results, it is important to note that one item was suppressed in reading grade 4, following initial key check reviews. This item is included in the counts of flagged items, only. The remaining items in the passage were included in the analyses.

Exhibit 12. Reading Form Summary Statistics

Grade	Form	ItemTotal (Points)	Summary Statistics							
			N	Mean Raw Score	SD Raw Score	Min Raw Score	Max Raw Score	Mean p-value	SD p-value	Alpha
3	1	28 (31)	178	17.92	5.78	0	29	0.63	0.19	0.84
	2	29 (32)	98	18.65	5.74	5	29	0.62	0.15	0.83
	3	30 (32)	115	18.91	5.21	3	28	0.62	0.22	0.81
	4	29 (32)	113	18.00	6.31	0	30	0.60	0.18	0.86
4	1	28 (32)	192	18.65	6.33	0	30	0.62	0.16	0.86
	2	28 (33)	129	17.35	6.48	0	30	0.58	0.15	0.85
	3	28 (33)	110	19.15	6.76	4	30	0.62	0.17	0.87
	4	28 (34)	122	18.65	6.82	2	32	0.59	0.12	0.86
5	1	26 (30)	154	16.59	6.12	0	28	0.62	0.17	0.85
	2	26 (32)	99	19.73	6.32	3	31	0.62	0.13	0.85
	3	26 (31)	88	18.01	6.37	0	29	0.61	0.14	0.84
	4	25 (31)	127	17.43	5.88	1	29	0.58	0.14	0.81
6	1	24 (29)	180	18.97	5.64	4	29	0.68	0.16	0.84
	2	24 (29)	103	16.85	5.31	5	27	0.59	0.18	0.80
	3	24 (29)	116	18.19	5.44	6	29	0.68	0.16	0.82
	4	24 (29)	98	17.92	4.85	7	29	0.61	0.18	0.76
7	1	24 (29)	162	18.04	6.46	0	29	0.67	0.14	0.88
	2	24 (31)	121	17.74	5.94	1	29	0.59	0.14	0.82
	3	24 (29)	126	18.69	5.55	1	29	0.67	0.14	0.83
	4	24 (31)	141	17.29	5.38	1	29	0.57	0.14	0.78

Exhibit 12. Reading Form Summary Statistics (cont.)

Grade	Form	Total (Points)	Summary Statistics							
			N	Mean Raw Score	SD Raw Score	Min Raw Score	Max Raw Score	Mean p-value	SD p-value	Alpha
8	1	24 (29)	173	17.97	5.63	0	29	0.66	0.14	0.83
	2	24 (31)	116	18.55	6.06	5	31	0.62	0.13	0.83
	3	24 (29)	104	17.82	5.27	4	28	0.61	0.17	0.79
	4	24 (31)	114	17.39	5.92	4	30	0.57	0.14	0.81
11	1	26 (31)	168	20.12	6.46	1	31	0.68	0.16	0.87
	2	26 (32)	122	20.61	6.00	3	31	0.66	0.15	0.84
	3	26 (29)	109	18.30	5.32	9	28	0.68	0.16	0.82
	4	26 (32)	96	19.90	6.74	0	32	0.65	0.15	0.88

Exhibit 13 contains a summary of the total number of items flagged by type of flag as well as overall in each grade in reading. Exhibit 14 contains the number of items flagged by Tier and content type in reading. Exhibit 15 contains the number of items flagged by form in reading. Overall, 10.3 percent of all items were flagged for at least one reason in reading. The majority of items were flagged for a low p-value or a high distractor p-value. The number of items flagged per form was relatively similar across forms. Of the items flagged for omit, both grade 3 items were flagged as a result of an omit rate greater than 5 percent. One grade 5 item was flagged based on an omit rate greater than 5 percent, two were flagged based on a not reached rate greater than 5 percent, and two were flagged due to a combined omit and not reached rate greater than 5 percent. In evaluating the number of items flagged and adequacy of the flagging criteria, it is important to consider that all items are two-option or three-option selected-response or one-point constructed-response items. The test developer suppressed one reading item in grade 4 following initial key check reviews; however, it is included in the flagged item counts.

Exhibit 13. Number of Items Flagged in Reading by Grade

Grade	Items	Type of Flag						
		Low p-value	High p-value	Low biserial	Distractor p-value	Tier reversal	Omit	At least one flag
3	65	10	0	2	2	0	2	14
4	62	8	0	1	3	0	0	8
5	55	4	0	1	2	0	5	7
6	54	6	0	1	4	0	0	7
7	60	2	0	0	1	0	0	2
8	60	4	0	0	2	0	0	4
11	53	2	0	0	0	0	0	2

Exhibit 14. Number of Items Flagged in Reading by Tier and Content Type

Grade	Content Type	Tier 1	Tier 2	Tier 3	Tier 4
3	Foundational – Non-Verbal		2	1	1
	Foundational – Verbal		2	1	1
	Informational Passage		2	4	
	Literary Passage				
4	Foundational – Non-Verbal	1		1	1
	Foundational – Verbal		1		1
	Informational Passage			2	
	Literary Passage		1		
5	Informational Passage		3	2	
	Literary Passage		1		1
6	Informational Passage			3	4
	Literary Passage				
7	Informational Passage				2
	Literary Passage				
8	Informational Passage			1	3
	Literary Passage				
11	Informational Passage			2	
	Literary Passage				

Exhibit 15. Number of Items Flagged in Reading by Form

Grade	Form	Total (Points)	At least one flag
3	1	28 (31)	6
	2	29 (32)	7
	3	30 (32)	9
	4	29 (32)	6
4	1	28 (32)	4
	2	29 (34)	7
	3	28 (33)	4
	4	28 (34)	4
5	1	26 (30)	5
	2	26 (32)	4
	3	26 (31)	3
	4	25 (31)	4
6	1	24 (29)	1
	2	24 (29)	3
	3	24 (29)	1
	4	24 (29)	5
7	1	24 (29)	0
	2	24 (31)	0
	3	24 (29)	0
	4	24 (31)	2
8	1	24 (29)	0
	2	24 (31)	0
	3	24 (29)	1
	4	24 (31)	3
11	1	26 (31)	2
	2	26 (32)	2
	3	26 (29)	2
	4	26 (32)	2

Exhibit 16 contains a summary of mean p-values by grade and tier for reading. The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the tier. The next three columns contain the total number of items, mean p-value and standard deviation of p-value for each tier. Following the design of the items and tiers, it would be expected to see the highest p-values at Tier 1 and the lowest p-values at Tier 4. In general, mean p-values for the other tiers followed the expected pattern; however, the means were very similar across Tiers 2–4.

Exhibit 16. Summary of p-value by Tier in Reading

Grade	Tier	Number of Items	Summary Statistics	
			Mean p-value	SD p-value
3	1	17	0.81	0.07
	2	19	0.65	0.10
	3	22	0.54	0.10
	4	7	0.56	0.10
4	1	17	0.78	0.10
	2	17	0.55	0.17
	3	20	0.57	0.16
	4	7	0.49	0.22
5	1	15	0.74	0.15
	2	15	0.54	0.10
	3	20	0.51	0.10
	4	5	0.46	0.15
6	1	17	0.77	0.07
	2	12	0.58	0.12
	3	19	0.53	0.11
	4	6	0.53	0.13
7	1	20	0.82	0.07
	2	12	0.63	0.11
	3	20	0.51	0.11
	4	8	0.43	0.12
8	1	16	0.77	0.07
	2	16	0.61	0.10
	3	20	0.52	0.08
	4	8	0.45	0.06
11	1	17	0.77	0.12
	2	13	0.63	0.10
	3	18	0.52	0.11
	4	5	0.43	0.07

Reading IRT Fit Results

Exhibit 17 contains a summary of the item difficulty estimates by grade for reading and ELA. Since the test developer calibrated items concurrently, the summaries are based on both reading forms and ELA Session 1. The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the statistic (e.g., mean, median, standard deviation). The next six columns provide values for each of the statistics for item measure (difficulty), standard error, mean square infit, standardized infit, mean square outfit and standardized outfit. Exhibit 18 contains the same summary of the person ability estimates by grade for reading. The distribution of item measures tends to be slightly negatively skewed or to exhibit very little skew, while the distribution of person measures is slightly positively skewed.

Exhibit 17. Reading and ELA Session 1 Item Measure Summary

Grade	Statistic	Measure	SE	InMS	InZStd	OutMS	OutZStd
3	Items	65	65	65	65	65	65
	Mean	0.00	0.18	1.00	0.09	0.95	-0.12
	Standard Deviation	0.95	0.07	0.17	2.40	0.29	2.57
	Min	-1.74	0.06	0.77	-5.52	0.56	-5.28
	10th	-1.35	0.08	0.83	-2.07	0.60	-2.45
	25th	-0.66	0.12	0.87	-1.43	0.72	-1.73
	50th	-0.01	0.21	0.98	-0.11	0.94	-0.39
	75th	0.57	0.24	1.08	0.91	1.14	1.03
	90th	1.40	0.24	1.21	2.53	1.32	2.87
	Max	2.77	0.29	1.62	9.90	1.96	9.90
4	Items	61	61	61	61	61	61
	Mean	0.00	0.16	1.02	0.01	0.98	-0.06
	Standard Deviation	0.81	0.06	0.20	2.81	0.27	2.64
	Min	-1.72	0.05	0.69	-6.84	0.49	-5.76
	10th	-1.24	0.08	0.81	-3.57	0.65	-3.05
	25th	-0.56	0.09	0.88	-1.81	0.77	-1.74
	50th	0.02	0.17	0.98	-0.21	1.00	0.00
	75th	0.51	0.21	1.14	1.70	1.15	1.40
	90th	0.97	0.23	1.28	3.76	1.43	3.51
	Max	2.40	0.29	1.58	9.90	1.69	9.90

Exhibit 17. Reading and ELA Session 1 Item Measure Summary (cont.)

Grade	Statistic	Measure	SE	InMS	InZStd	OutMS	OutZStd
5	Items	60	60	60	60	60	60
	Mean	0.00	0.17	0.98	-0.19	0.93	-0.32
	Standard Deviation	0.81	0.07	0.13	2.27	0.21	2.29
	Min	-1.85	0.05	0.75	-4.15	0.62	-4.42
	10th	-1.3	0.08	0.84	-2.53	0.7	-2.57
	25th	-0.44	0.10	0.90	-1.51	0.76	-1.64
	50th	0.02	0.19	0.94	-0.72	0.89	-0.90
	75th	0.62	0.23	1.04	0.49	1.03	0.33
	90th	1.01	0.24	1.15	2.19	1.26	2.53
	Max	1.58	0.31	1.41	9.90	1.57	9.15
6	Items	59	59	59	59	59	59
	Mean	0.00	0.18	1.01	0.18	0.97	0.00
	Standard Deviation	0.99	0.06	0.14	1.83	0.27	2.05
	Min	-1.89	0.05	0.74	-3.68	0.53	-4.03
	10th	-1.47	0.08	0.85	-1.85	0.63	-2.29
	25th	-0.51	0.12	0.9	-1.1	0.75	-1.47
	50th	0.06	0.20	1.00	0.00	0.98	-0.26
	75th	0.69	0.23	1.10	1.21	1.13	1.43
	90th	1.11	0.25	1.20	2.56	1.30	2.13
	Max	2.04	0.27	1.38	5.89	1.66	5.79
7	Items	65	65	65	65	65	65
	Mean	0.00	0.17	1.00	0.03	0.94	-0.19
	Standard Deviation	0.89	0.06	0.16	2.27	0.25	2.35
	Min	-2.47	0.06	0.7	-5.67	0.55	-4.98
	10th	-1.18	0.08	0.81	-2.63	0.6	-2.83
	25th	-0.72	0.13	0.86	-1.65	0.71	-1.73
	50th	0.08	0.18	1.00	0.00	0.99	-0.14
	75th	0.64	0.20	1.08	1.04	1.07	0.82
	90th	0.96	0.22	1.24	3.19	1.28	2.67
	Max	2.24	0.31	1.44	7.65	1.52	8.45
8	Items	65	65	65	65	65	65
	Mean	0.00	0.17	1.00	0.20	0.98	0.03
	Standard Deviation	0.87	0.07	0.13	1.93	0.21	1.90
	Min	-2.56	0.05	0.66	-4.24	0.51	-4.20
	10th	-1.42	0.07	0.86	-1.70	0.68	-1.83
	25th	-0.47	0.13	0.91	-1.00	0.81	-1.21
	50th	0.13	0.18	0.99	-0.08	0.99	-0.11
	75th	0.64	0.22	1.07	1.31	1.09	1.22
	90th	1.03	0.24	1.22	2.79	1.31	2.28
	Max	1.55	0.33	1.32	8.84	1.52	6.96
11	Items	57	57	57	57	57	57
	Mean	0.00	0.16	0.99	-0.09	0.93	-0.28
	Standard Deviation	0.89	0.07	0.16	2.56	0.27	2.54
	Min	-2.23	0.06	0.74	-5.31	0.51	-4.58
	10th	-1.22	0.08	0.84	-2.23	0.62	-2.52
	25th	-0.45	0.10	0.89	-1.37	0.73	-1.56
	50th	0.00	0.16	0.96	-0.54	0.87	-0.83
	75th	0.67	0.22	1.08	0.80	1.07	0.49
	90th	1.05	0.26	1.21	2.03	1.31	2.29
	Max	1.86	0.30	1.57	9.90	1.72	9.90

Exhibit 18. Reading and ELA Session 1 Person Measure Summary

Grade	Statistic	Measure	SE	InMS	InZStd	OutMS	OutZStd
3	Persons	859	859	859	859	859	859
	Mean	0.79	0.52	1.00	0.10	0.93	-0.01
	Standard Deviation	1.25	0.17	0.27	1.04	0.39	0.99
	Min	-4.91	0.40	0.46	-2.83	0.14	-2.58
	10th	-0.78	0.41	0.67	-1.22	0.47	-1.16
	25th	-0.12	0.43	0.79	-0.61	0.62	-0.70
	50th	0.73	0.47	0.98	0.00	0.87	-0.16
	75th	1.54	0.53	1.18	0.75	1.17	0.67
	90th	2.26	0.67	1.36	1.43	1.45	1.30
	Max	5.22	1.86	2.14	4.16	2.82	4.12
4	Persons	964	964	964	964	964	964
	Mean	0.53	0.49	1.03	0.19	0.95	0.02
	Standard Deviation	1.25	0.16	0.23	0.92	0.34	0.92
	Min	-4.02	0.38	0.48	-2.54	0.17	-2.29
	10th	-0.94	0.39	0.73	-1.02	0.55	-1.12
	25th	-0.46	0.41	0.87	-0.37	0.73	-0.62
	50th	0.51	0.45	1.01	0.18	0.94	-0.05
	75th	1.40	0.50	1.18	0.79	1.14	0.62
	90th	2.12	0.63	1.33	1.34	1.33	1.25
	Max	5.03	1.90	1.89	3.27	5.75	3.43
5	Persons	765	765	765	765	765	765
	Mean	0.49	0.46	1.03	0.15	0.96	0.02
	Standard Deviation	1.04	0.13	0.27	0.98	0.31	1.04
	Min	-3.86	0.37	0.43	-2.38	0.25	-2.64
	10th	-0.71	0.38	0.71	-1.11	0.58	-1.26
	25th	-0.22	0.40	0.84	-0.52	0.73	-0.72
	50th	0.46	0.44	1.01	0.12	0.94	-0.12
	75th	1.28	0.47	1.19	0.78	1.17	0.75
	90th	1.87	0.56	1.38	1.41	1.37	1.48
	Max	3.90	2.18	2.51	4.49	2.30	3.12
6	Persons	777	777	777	777	777	777
	Mean	0.75	0.52	1.00	0.06	0.93	-0.02
	Standard Deviation	1.25	0.19	0.29	0.96	0.44	1.01
	Min	-3.08	0.41	0.41	-2.60	0.17	-2.45
	10th	-0.76	0.42	0.66	-1.10	0.51	-1.13
	25th	-0.15	0.43	0.81	-0.53	0.64	-0.65
	50th	0.60	0.47	0.98	0.02	0.84	-0.18
	75th	1.46	0.53	1.14	0.57	1.10	0.44
	90th	2.38	0.66	1.39	1.24	1.43	1.43
	Max	5.20	1.86	2.42	3.81	5.50	3.25

Exhibit 18. Reading and ELA Session 1 Person Measure Summary (cont.)

Grade	Statistic	Measure	SE	InMS	InZStd	OutMS	OutZStd
7	Persons	873	873	873	873	873	873
	Mean	0.62	0.51	1.02	0.12	0.93	-0.03
	Standard Deviation	1.19	0.18	0.30	1.10	0.39	1.15
	Min	-5.27	0.38	0.40	-3.58	0.24	-3.26
	10th	-0.81	0.39	0.67	-1.21	0.52	-1.21
	25th	-0.17	0.41	0.81	-0.52	0.67	-0.76
	50th	0.69	0.48	0.99	0.10	0.85	-0.27
	75th	1.33	0.54	1.19	0.74	1.14	0.54
	90th	2.08	0.64	1.44	1.52	1.46	1.58
	Max	4.77	1.90	2.03	3.39	2.83	4.01
8	Persons	936	936	936	936	936	936
	Mean	0.60	0.51	0.99	0.03	0.96	0.02
	Standard Deviation	1.09	0.20	0.23	0.87	0.33	0.87
	Min	-2.97	0.37	0.40	-2.23	0.21	-2.14
	10th	-0.59	0.38	0.73	-1.01	0.64	-0.95
	25th	-0.13	0.40	0.83	-0.50	0.75	-0.56
	50th	0.55	0.47	0.97	0.00	0.90	-0.12
	75th	1.18	0.52	1.10	0.47	1.11	0.49
	90th	1.93	0.64	1.30	1.15	1.33	1.25
	Max	4.80	2.06	2.03	4.10	5.20	4.30
11	Persons	879	879	879	879	879	879
	Mean	0.92	0.53	1.01	0.13	0.93	0.05
	Standard Deviation	1.30	0.22	0.26	0.97	0.40	1.01
	Min	-3.05	0.39	0.45	-2.41	0.10	-2.15
	10th	-0.56	0.40	0.69	-1.02	0.42	-1.07
	25th	-0.01	0.42	0.82	-0.47	0.66	-0.68
	50th	0.73	0.46	1.00	0.06	0.91	-0.06
	75th	1.78	0.56	1.15	0.60	1.15	0.63
	90th	2.69	0.68	1.35	1.30	1.43	1.39
	Max	5.55	2.03	2.05	3.62	3.34	3.92

Exhibit 19 contains a summary of the principal component analysis (PCA) of residual variation, as produced by WINSTEPS. The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the number of items, and the remaining three columns indicate the percent of variation associated with each dimension. As expected, the first dimension associated with the Rasch model accounts for the majority of score variation. Since the test development partner calibrated items concurrently, the PCA is based on both reading forms and ELA Session 1.

Exhibit 19. Principal Components Analysis in Reading and ELA

Grade	Number of Items	Percent First Dimension	Percent Second Dimension	Percent Third Dimension
3	65	56.7	1.5	1.3
4	61	49.8	2.1	1.7
5	60	42.3	2.7	2.0
6	59	56.7	1.6	1.5
7	65	52.6	2.5	1.9
8	65	50.2	1.8	1.6
11	57	59.1	1.6	1.4

Exhibit 20 contains a summary of the reading items flagged for misfit. The test developer partners provided a summary of the Writing items in a separate memo along with the writing constructed-response items. The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the total number of items. The next three columns indicate the number of items flagged for infit, outfit, and the total number of items flagged by either infit or outfit. Overall, a larger number of items were flagged for misfit in each grade as compared to mathematics. However, the total number of flagged items for infit is still relatively low. Exhibit 21 provides the statistics for each flagged item. Figures 6-18 contain item characteristic curves (ICCs) for each of the items flagged for infit. For polytomously scored items, category probability curves are presented.

Exhibit 20. Number of Items Flagged for Misfit in Reading

Grade	Total Number of Items	Infit	Outfit	Total
3	65	4	7	7
4	61	6	9	10
5	60	2	5	5
6	59	2	5	5
7	65	4	5	5
8	65	3	7	8
11	57	3	8	8

Exhibit 21. Item Statistics for Misfitting Items in Reading

Grade	Item	Measure	SE	Parm 2	SE 2	InMS	InZstd	OutMS	OutZStd	Infit	Outfit
3	114011A	0.94	0.08			1.40	9.90	1.59	9.90	1	1
3	300003P	0.18	0.08	2.62	0.11	1.42	8.67	1.48	9.53	1	1
3	117688B	0.88	0.22			1.30	3.11	1.40	2.87		1
3	114725A	0.93	0.23			1.33	3.39	1.42	2.93	1	1
3	115987A	-0.24	0.18			1.21	2.53	1.32	1.91		1
3	114726A	-0.03	0.24			1.22	1.95	1.75	3.45		1
3	117671B	0.76	0.21			1.62	6.38	1.96	6.49	1	1
4	113100A	-0.87	0.19			0.69	-3.57	0.49	-3.00	1	
4	116618A	-0.02	0.21			1.28	3.15	1.43	2.44		1
4	116619A	0.70	0.21			1.26	2.87	1.31	2.32		1
4	400004P	-0.71	0.09	1.20	0.08	1.51	9.90	1.54	9.90	1	1
4	400016P	0.37	0.22	1.21	0.25	1.49	3.74	1.52	3.26	1	1
4	117323A	0.90	0.17			1.30	3.98	1.44	3.75		1
4	400013P	-0.68	0.28	0.45	0.23	1.55	3.76	1.47	2.40	1	1
4	116606A	0.64	0.20			1.32	3.45	1.55	3.97	1	1
4	400006P	-0.05	0.18	1.52	0.19	1.58	5.54	1.69	5.80	1	1
4	116607A	0.97	0.21			1.22	2.27	1.36	2.44		1
5	119972B	1.27	0.08			1.41	9.90	1.57	9.15	1	1
5	500003P	-0.46	0.20	1.54	0.22	1.37	3.41	1.39	3.44	1	1
5	117110B	1.58	0.20			1.19	1.76	1.42	2.39		1
5	121730A	0.59	0.23			1.13	1.54	1.37	2.53		1
5	114323A	-0.25	0.25			1.09	0.82	1.31	1.40		1
6	119997B	1.58	0.23			1.20	2.04	1.33	2.12		1
6	120011B	1.68	0.23			1.34	2.87	1.66	3.39	1	1
6	120000A	1.01	0.22			1.38	4.37	1.66	4.88	1	1
6	120013B	2.02	0.25			1.15	1.17	1.48	2.10		1
6	121349B	2.04	0.25			1.14	1.13	1.32	1.60		1
7	700003P	-0.69	0.09	1.72	0.09	1.36	7.65	1.41	8.45	1	1
7	120467A	0.06	0.20			1.30	3.41	1.40	2.67		1
7	700008P	0.28	0.22	0.87	0.22	1.44	3.70	1.52	3.50	1	1
7	700009P	-0.08	0.21			1.40	3.17	1.40	3.21	1	1
7	121503A	1.09	0.21	2.37	0.28	1.31	3.49	1.51	3.94	1	1
8	800012P	-0.45	0.30			0.66	-2.75	0.51	-2.40	1	
8	118801A	0.56	0.21			1.22	2.79	1.32	2.50		1
8	121040A	1.28	0.22	-0.76	0.23	1.32	3.10	1.46	3.06	1	1
8	121206A	-0.85	0.23			0.88	-1.10	1.37	1.43		1
8	121043B	1.18	0.22			1.24	2.53	1.31	2.25		1
8	121160A	1.55	0.23			1.20	1.82	1.36	2.11		1
8	121165A	0.42	0.21			1.27	3.49	1.52	4.21		1
8	800009P	-0.11	0.23			1.31	2.61	1.31	2.28	1	1
11	113727A	1.22	0.24			1.40	3.42	1.57	3.53	1	1
11	121718A	-1.05	0.26			1.09	0.73	1.31	1.05		1
11	121065A	1.61	0.08	1.22	0.24	1.37	8.98	1.63	9.11	1	1
11	110003P	0.77	0.08			1.57	9.90	1.72	9.90	1	1
11	110013P	0.45	0.25			1.28	2.03	1.31	2.20		1
11	124319A	-0.23	0.12			1.26	4.47	1.52	4.16		1
11	110010P	0.12	0.22			1.29	2.42	1.31	2.37		1
11	113711A	0.51	0.22	2.94	0.12	1.15	1.62	1.32	2.29		1

Figure 6. ICCs for Grade 3 Reading Items Flagged for Infit

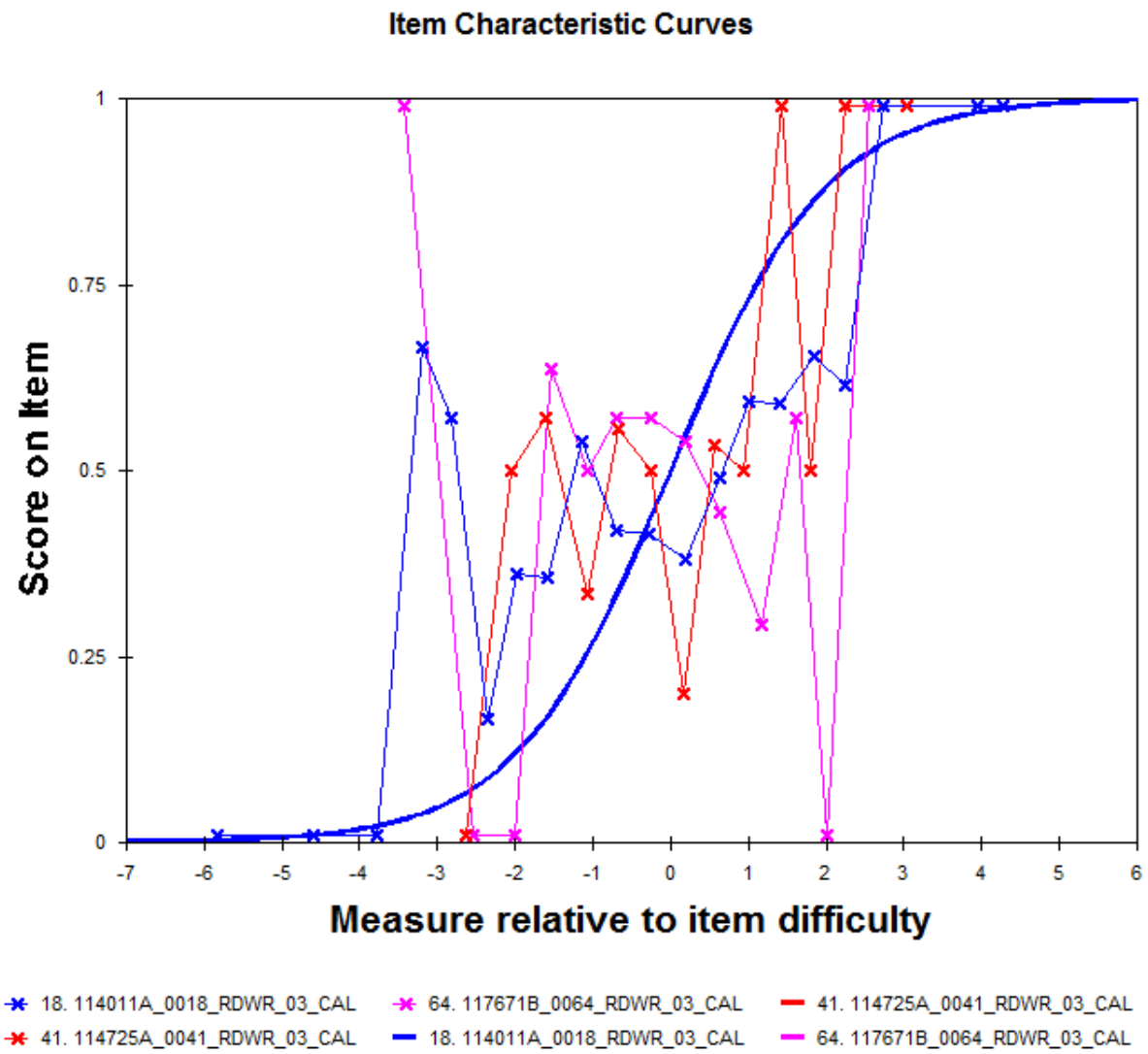


Figure 7. Probability Curves for Grade 3 Reading Items Flagged for Infit

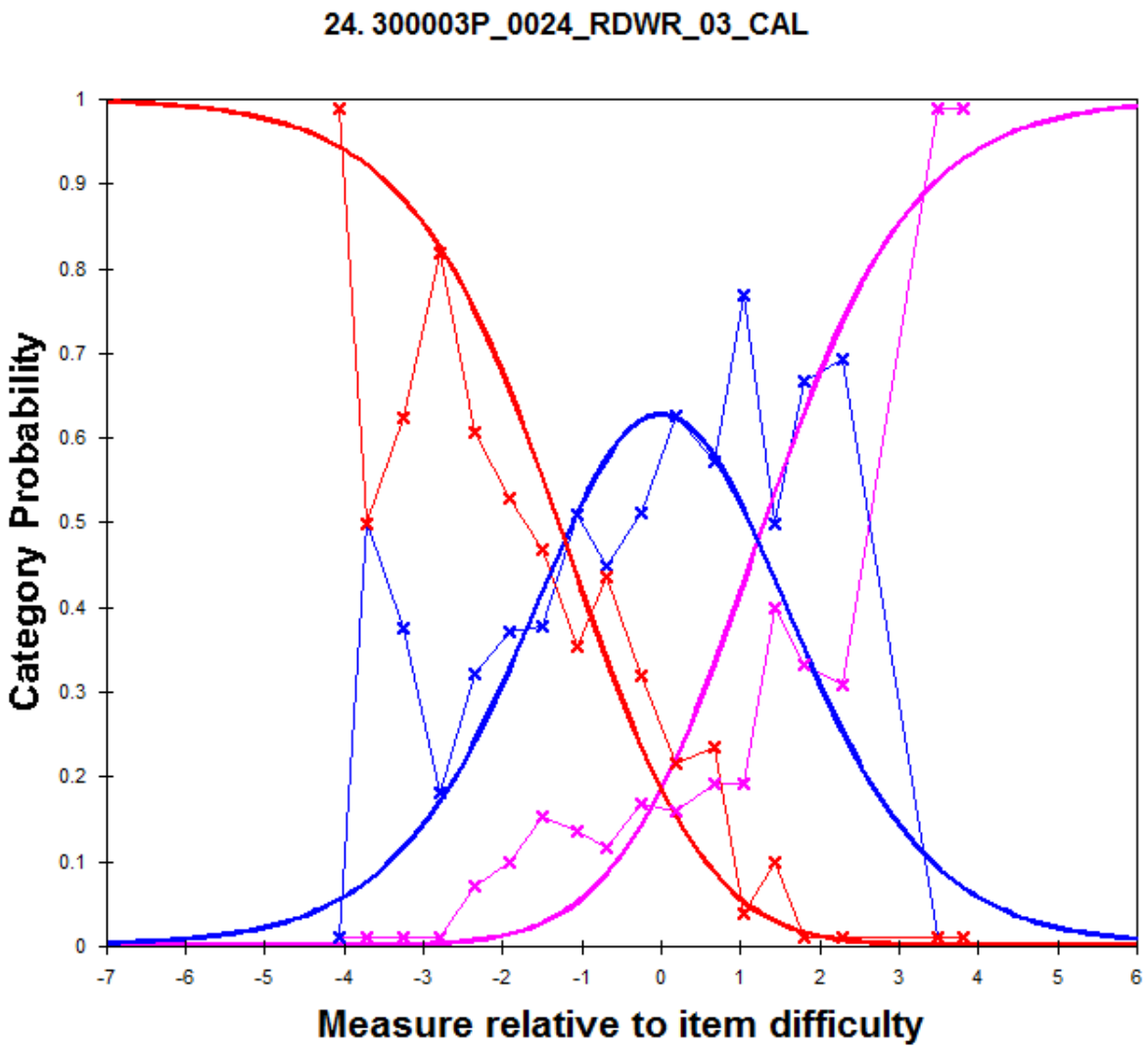


Figure 8. ICCs for Grade 4 Reading Items Flagged for Infit

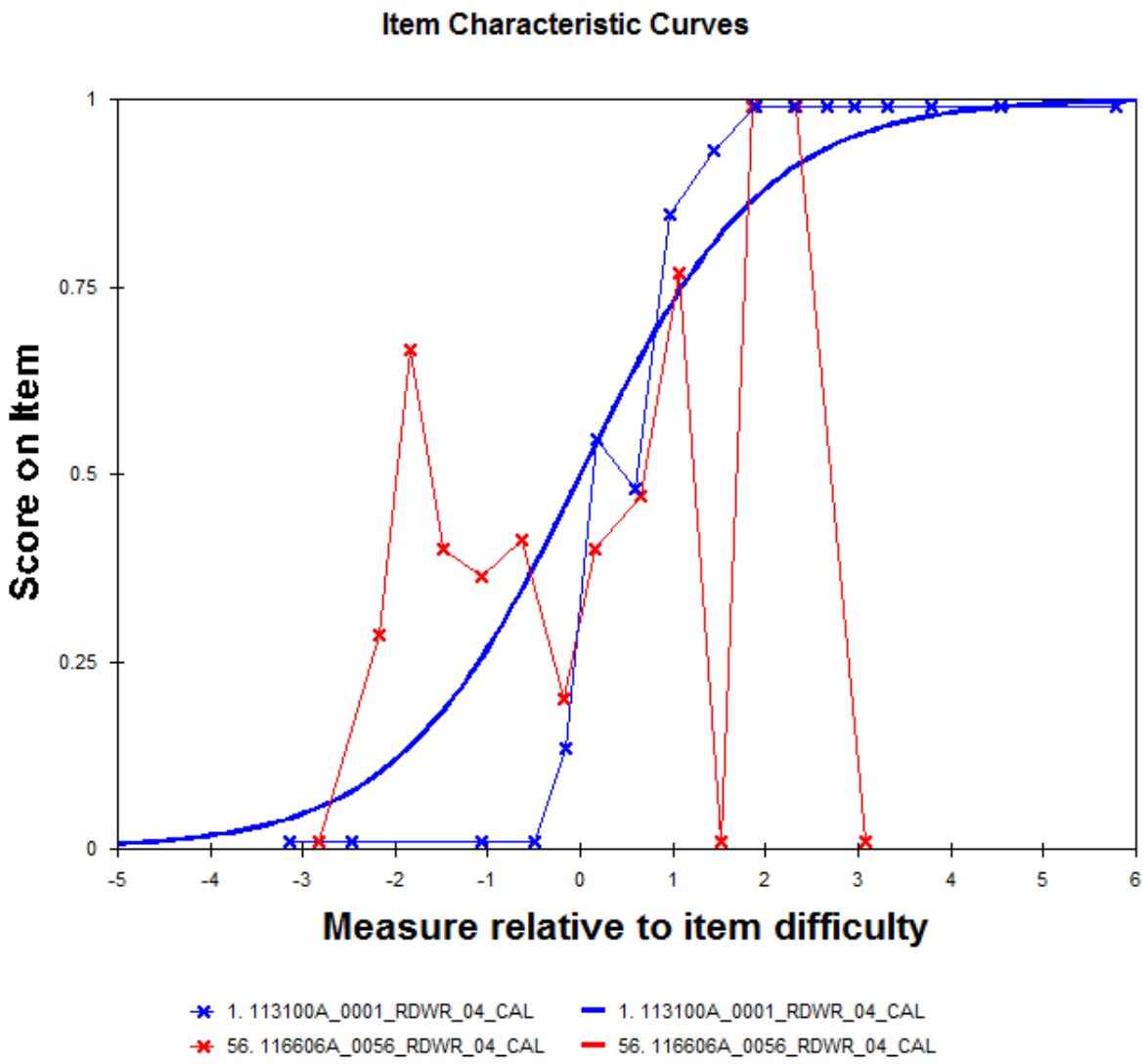


Figure 9. Probability Curves for Grade 4 Reading Items Flagged for Infit

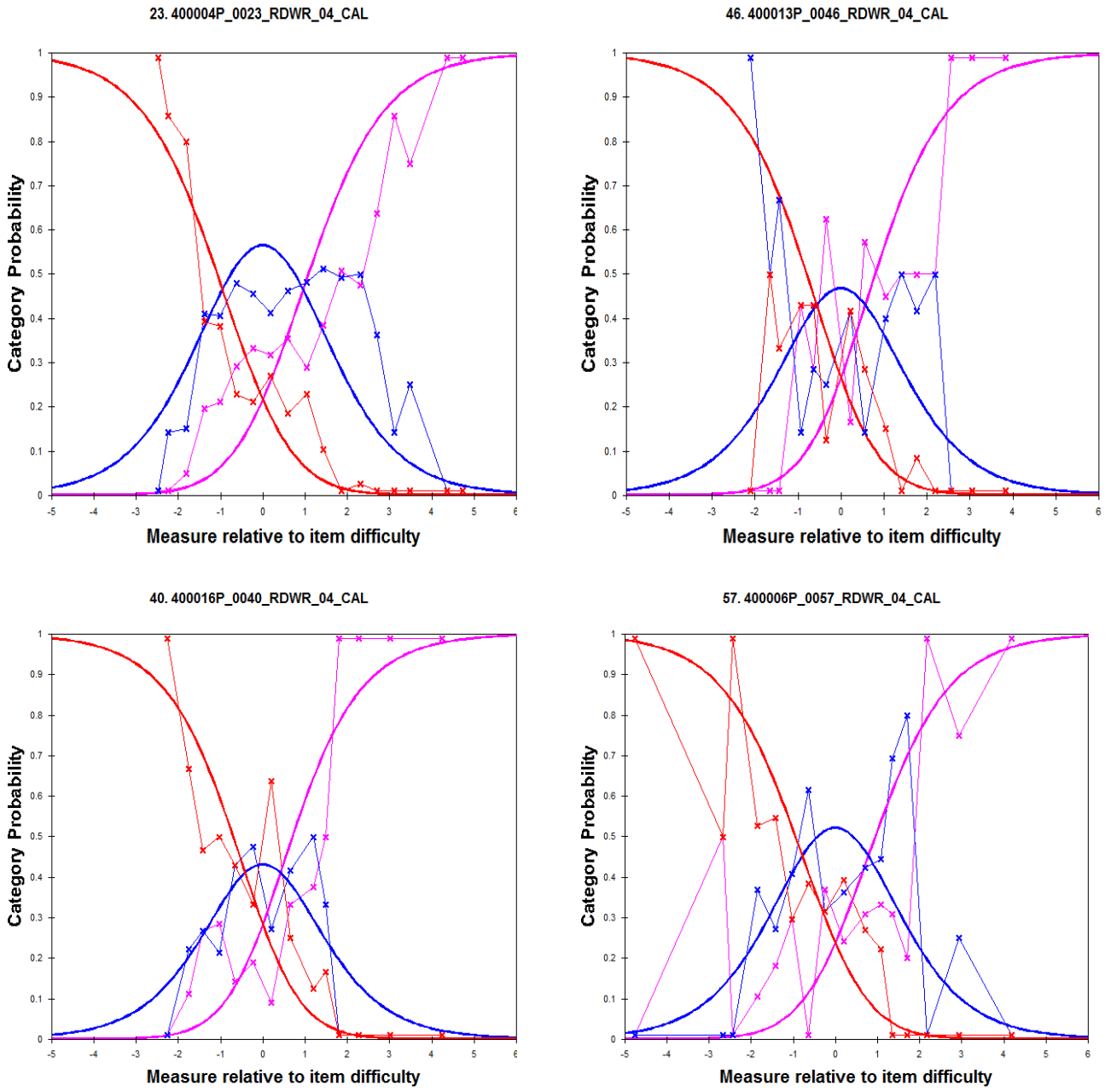


Figure 10. ICCs for Grade 5 Reading Items Flagged for Infit

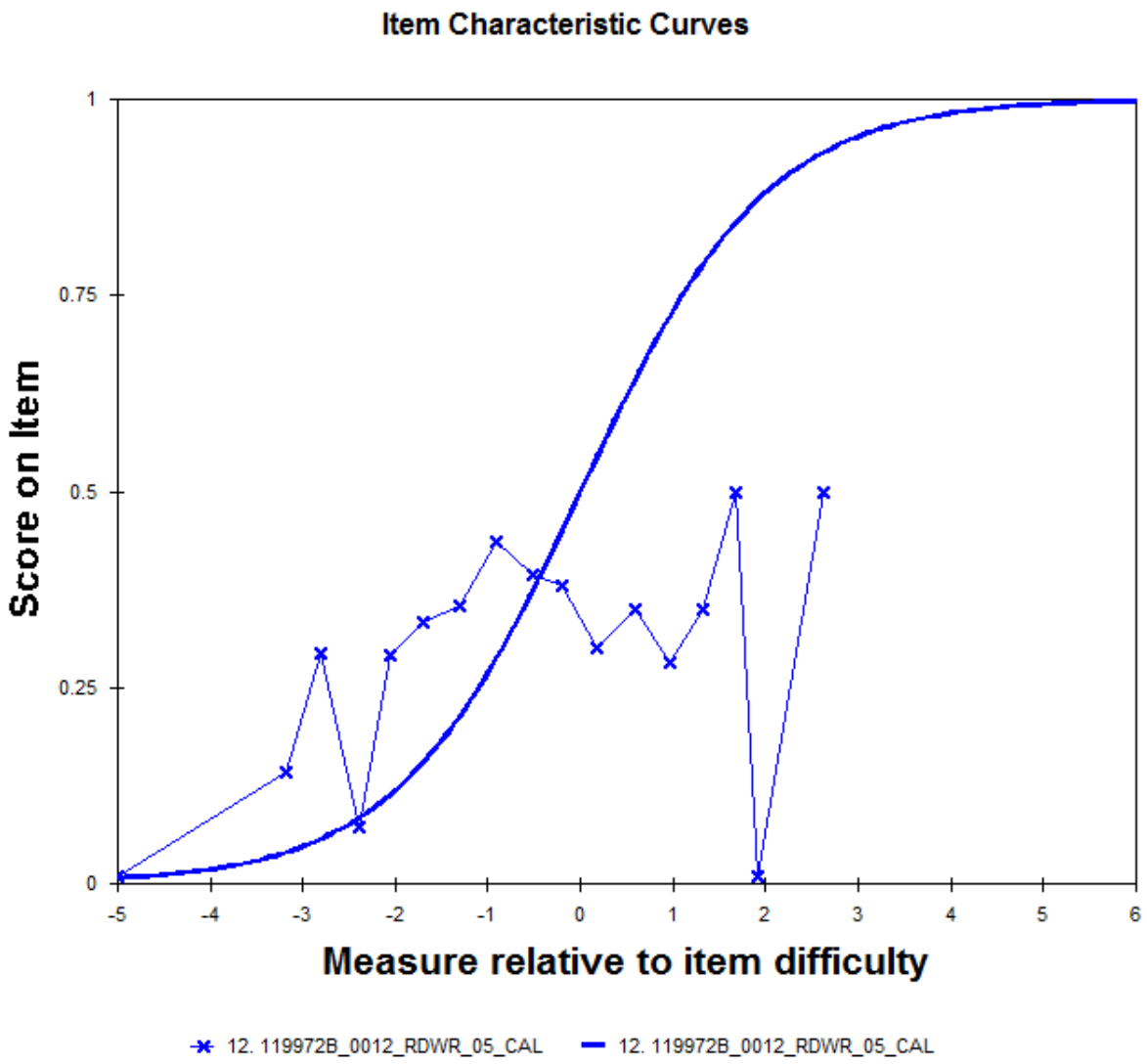


Figure 11. Probability Curves for Grade 5 Reading Items Flagged for Infit

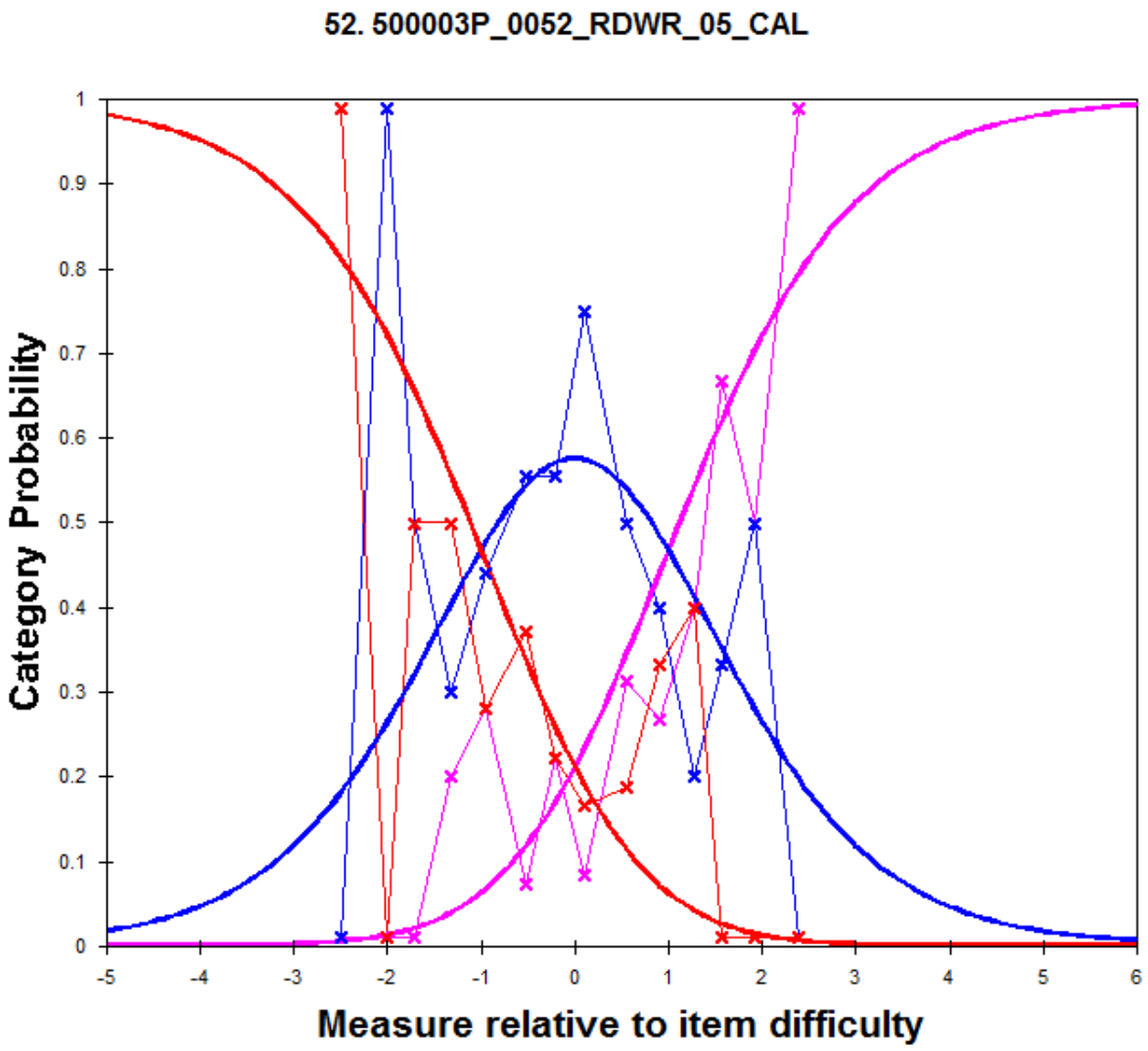


Figure 12. ICCs for Grade 6 Reading Items Flagged for Infit

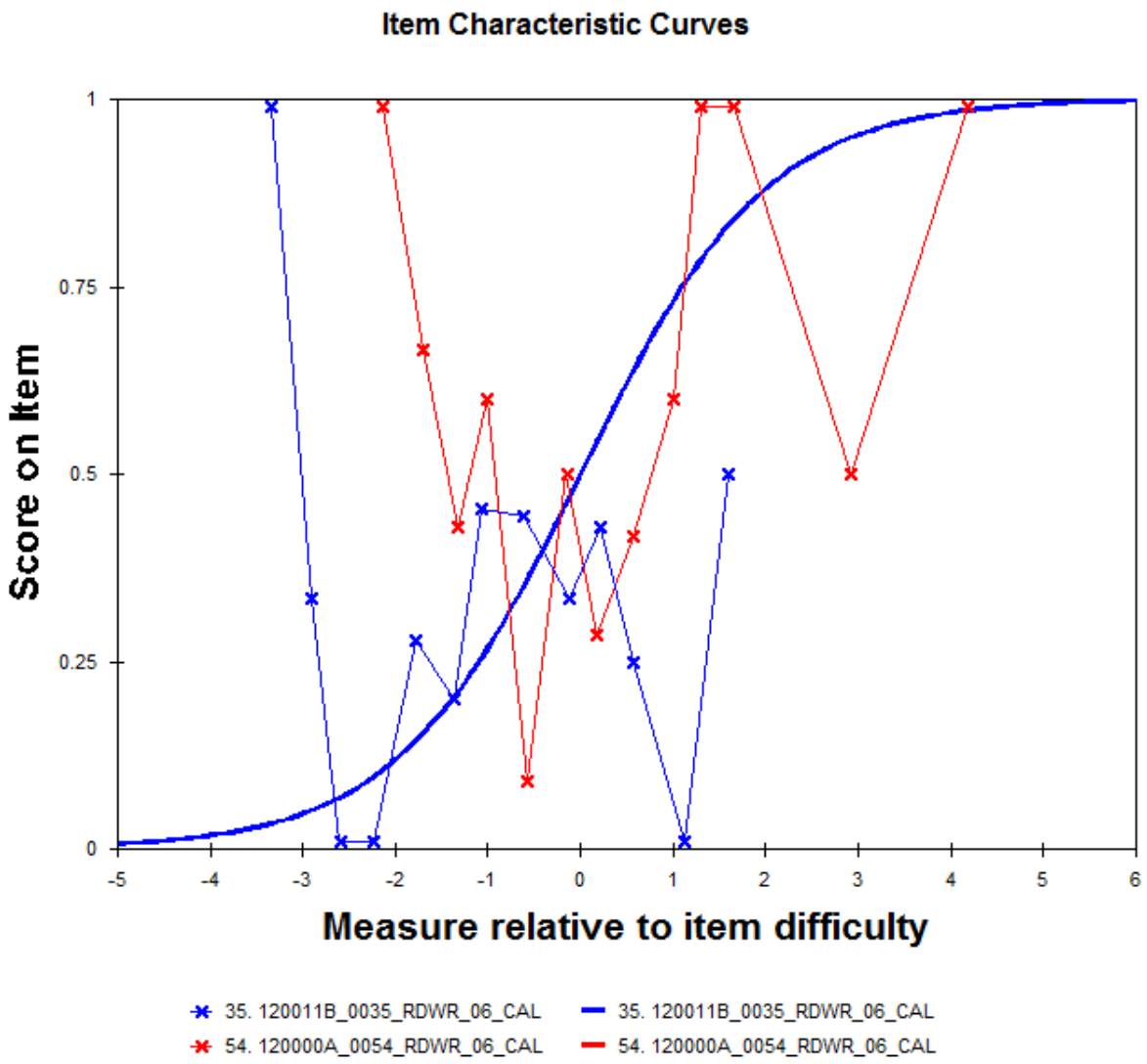


Figure 13. ICCs for Grade 7 Reading Items Flagged for Infit

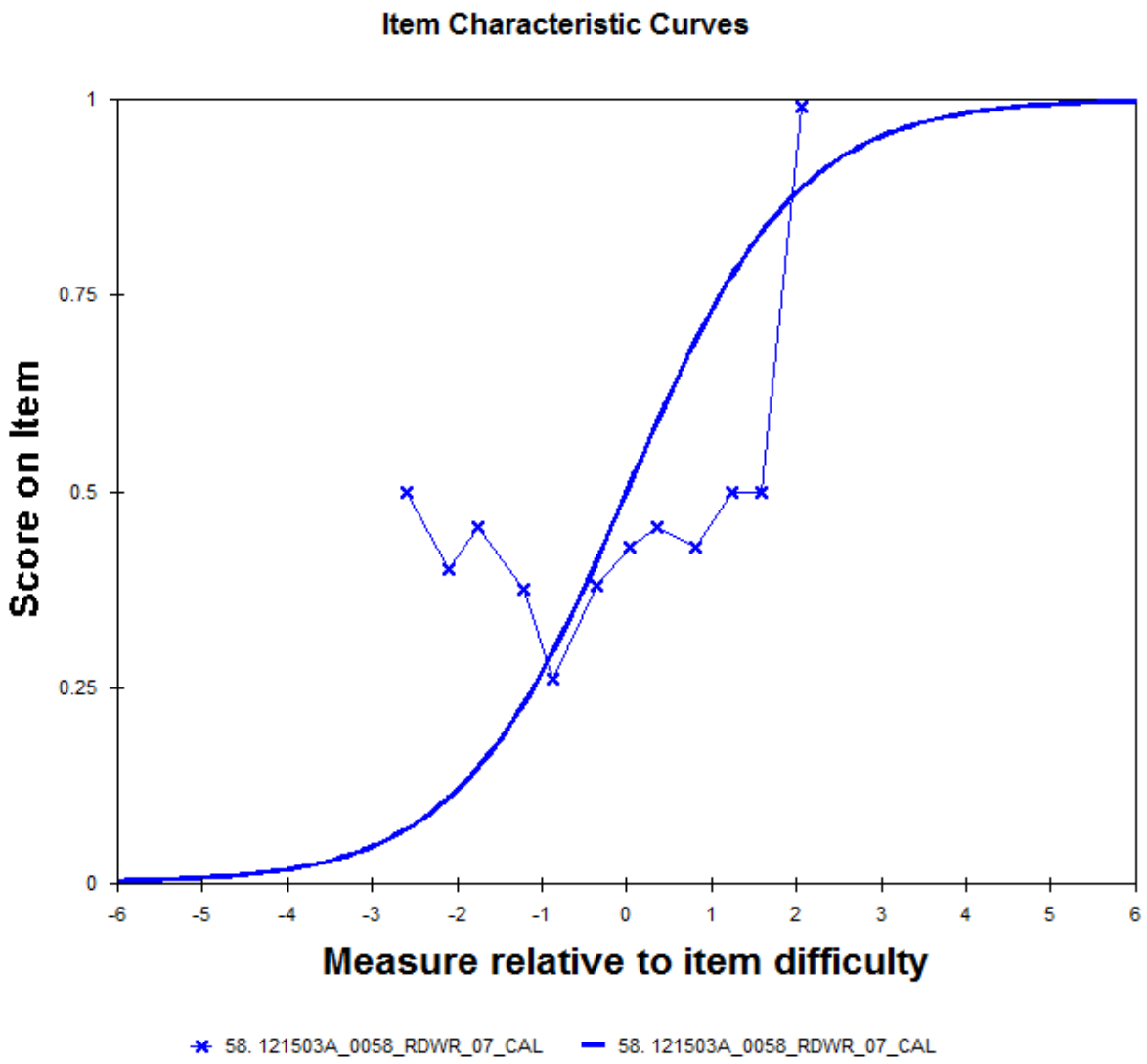
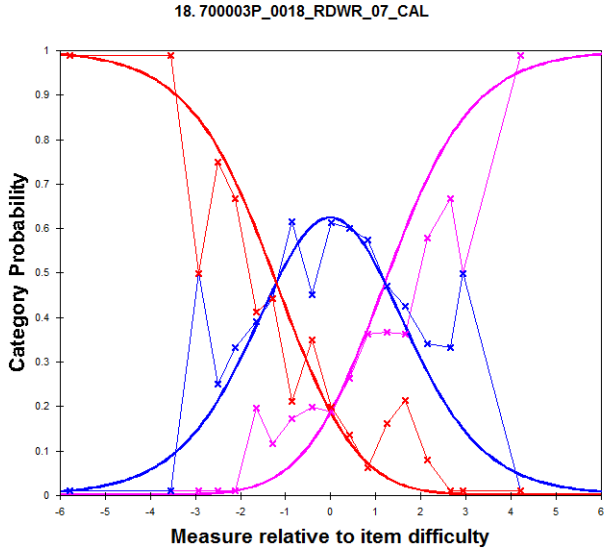
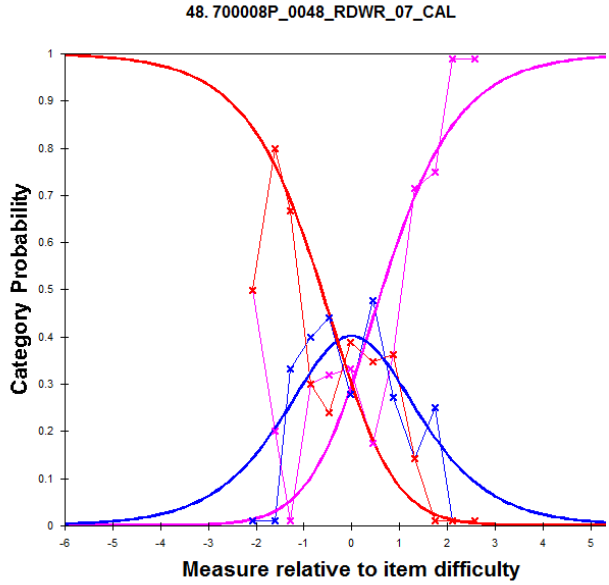


Figure 14. Probability Curves for Grade 7 Reading Items Flagged for Infit





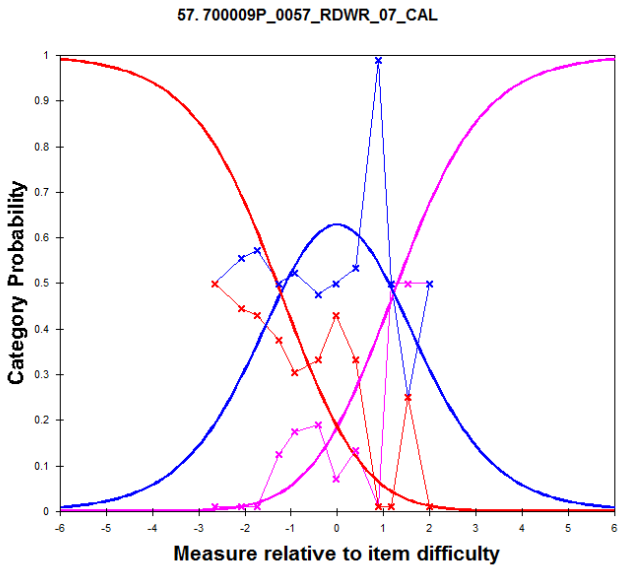

57. 700009P_0057_RDWR_07_CAL
Category Probability
Measure relative to item difficulty

Figure 15. ICCs for Grade 8 Reading Items Flagged for Infit

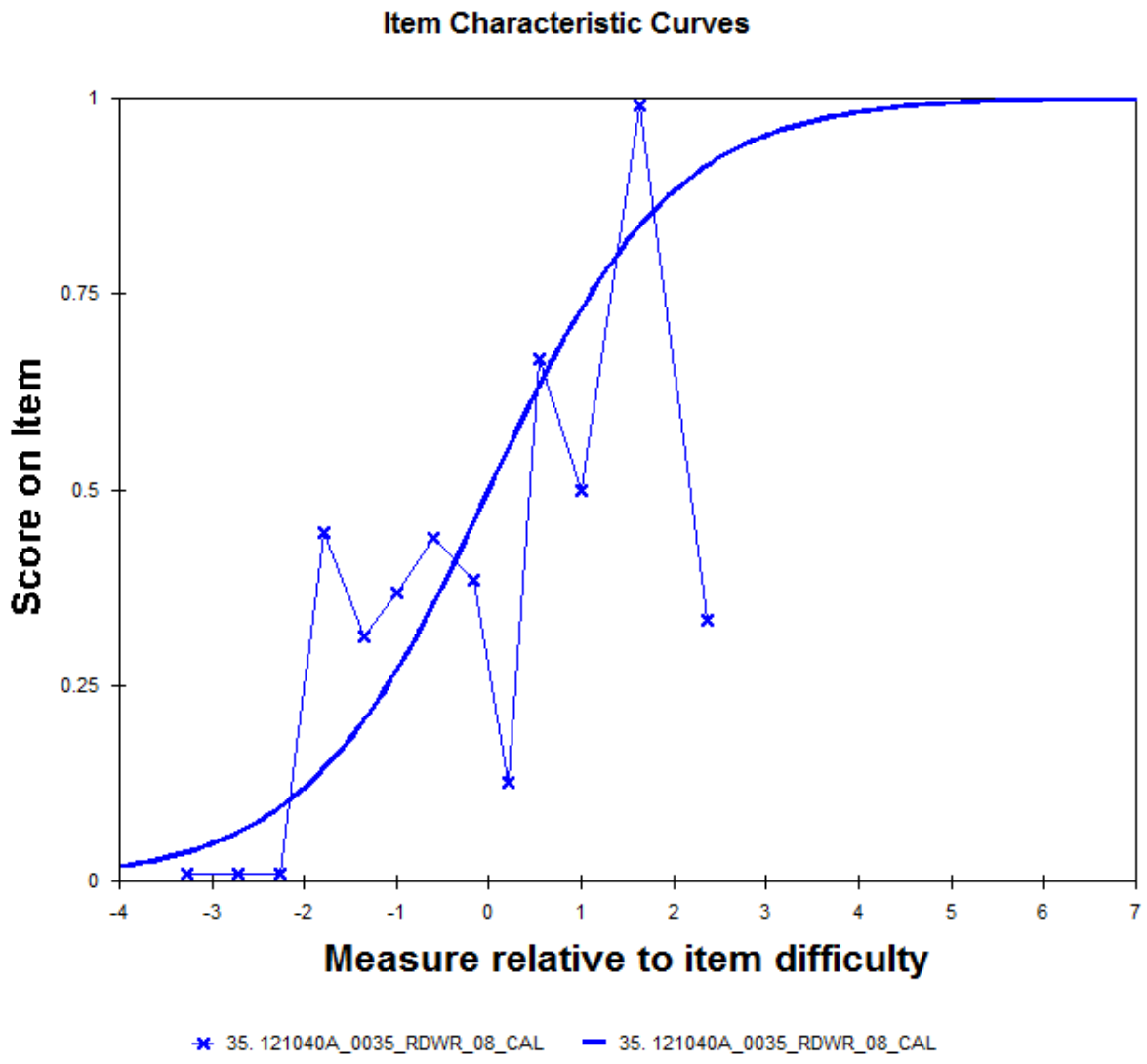
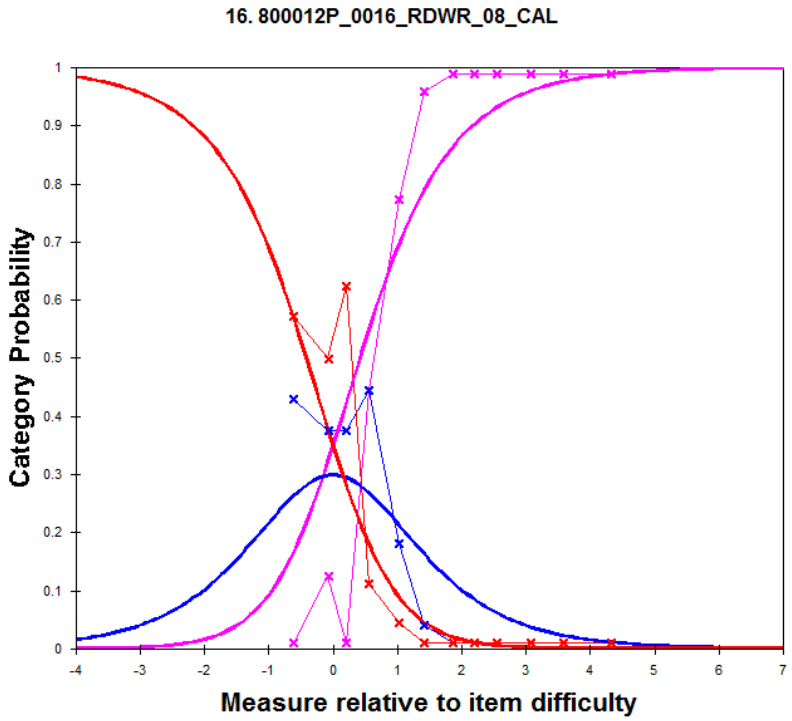


Figure 16. Probability Curves for Grade 8 Reading Items Flagged for Infit



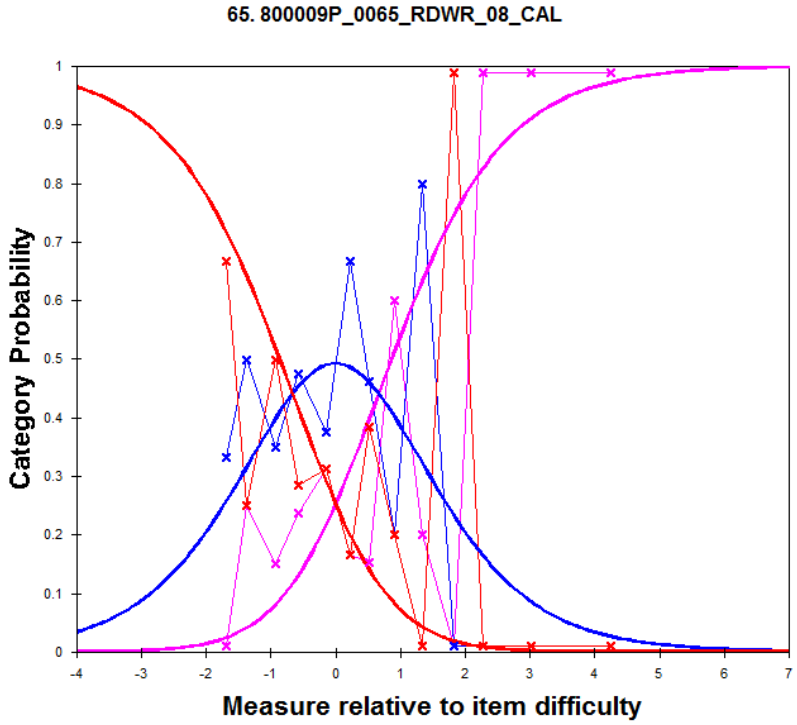


Figure 17. ICCs for Grade 11 Reading Items Flagged for Infit

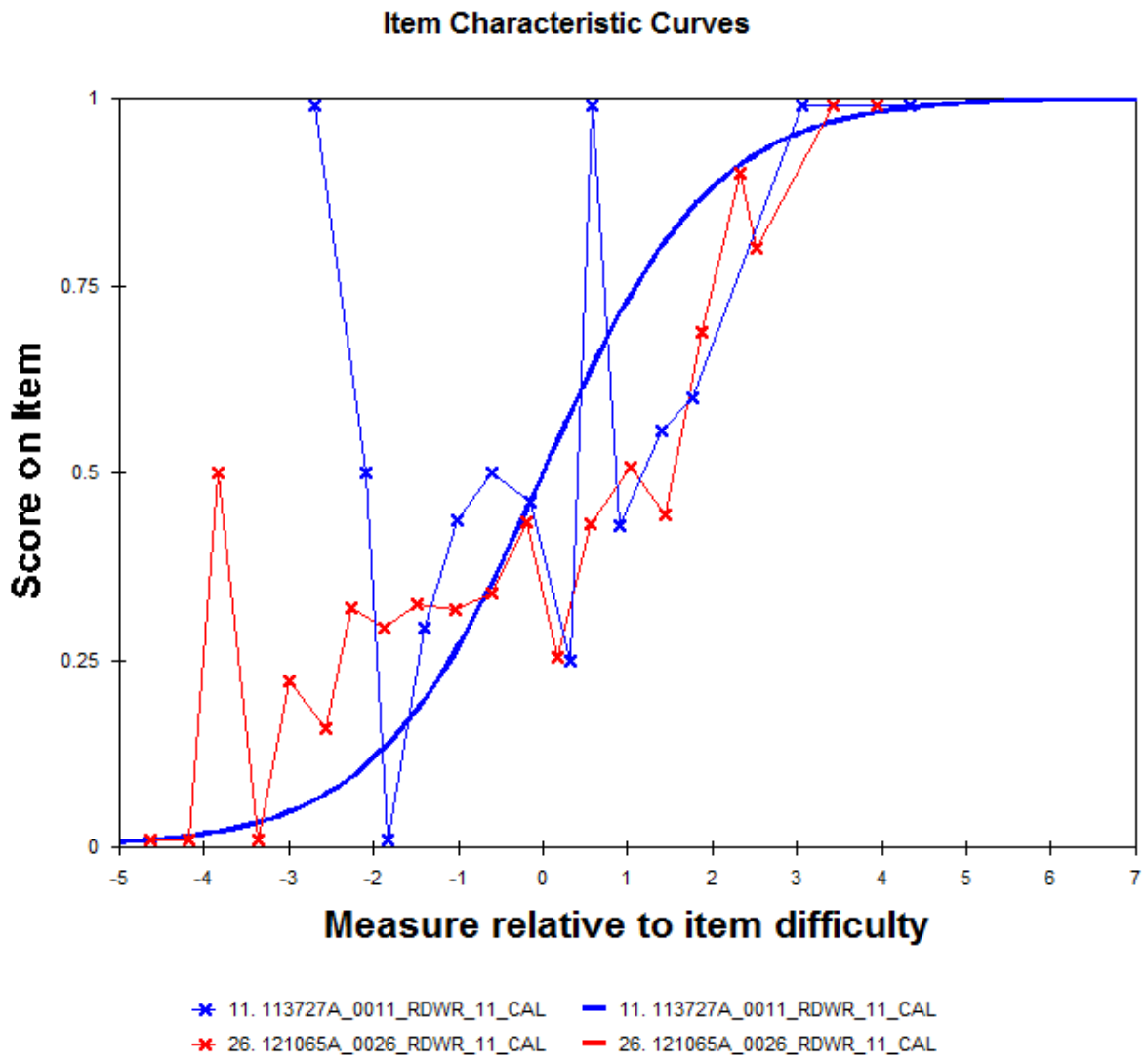
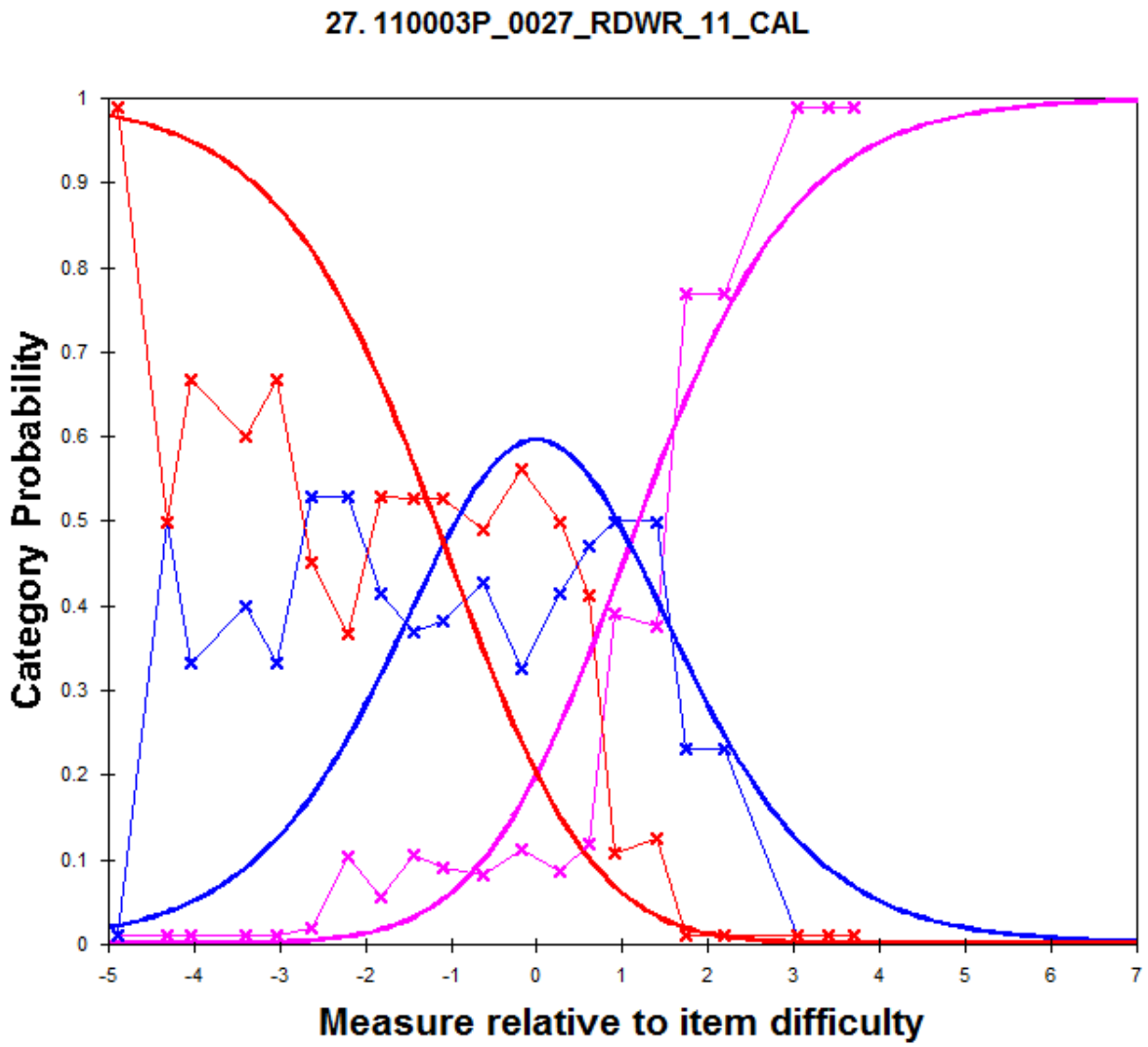


Figure 18. Probability Curves for Grade 11 Reading Items Flagged for Infit



ELA Selected-Response Pilot Phase 2 Test Results

ELA Classical Item Analysis Results

The test developer partner converted all item responses from the student's chosen answer into a scored data file. For selected-response items, the test developer converted all student responses to 0 if incorrect and 1 if correct.

In grades 3 and 4, the test developer partner scored the reading foundational items such that a set of 3 to 5 items was worth a maximum of one point. Initially, the test developer partner conducted analyses such that a student would receive a score of 1 if they answered all of the items correctly and a score of 0 if any of the foundational items were answered incorrectly. However, the foundational items scored in this manner were very difficult. Therefore, the test developer applied a new scoring rule. At Tier 1, if students answered all three items in the set correctly, they received a score of 1. They received a score of 0 if they answered any incorrectly. At Tiers 2–4, if students answered at least 4 out of 5 items correctly, they received a score of 1. If they answered less than 4 correctly, they received a score of 0.

For 2-part and 3-part selected-response items in reading, the desired scoring was for items to be scored as dependent items, where students received partial credit for a correct answer on any of the items. In order to account for the dependency of the items while maintaining integer scoring, the test developer partner analyzed each set of items as a single polytomous item, with a maximum of 2 or 3 score points.

Exhibit 22 contains form summary statistics for ELA Session 1. The ELA forms consisted of the same reading selections as Session 1 from the reading forms. In addition, each ELA form contained four stand-alone selected-response writing items, and one Tier 1 writing prompt item set A Tier 1 writing prompt consisted of a set of 4 – 6 selected-response items. However, for the Tier 1 writing prompt item set, the test developer partner converted student scores on the set of items to a total score of 0–2 points. The Tier 2 and Tier 3 writing constructed-response items were administered as part of Session 2. The test developer partner converted the Tier 2 items to a total score of 0–5 points, while Tier 3 items were converted to a total score of 0–7 points.

The results presented in Exhibit 26 are summary statistics for Session 1 of ELA, which includes all of the selected-response or dichotomously scored constructed-response reading and writing items, including the Tier 1 writing prompt item sets. The first column of the Exhibit indicates the grade, and the second column indicates the form. The next column contains the total number of items and points on each form. Given the inclusion of the foundational items, the polytomous items, and the Tier 1 writing prompt item sets, the total number of items will not match the total number of points possible. The remaining columns of the Exhibit indicate the number of students taking each form, the mean raw score and standard deviation, the minimum and maximum raw score, mean p-value and standard deviation, and Cronbach's alpha.

When reviewing the ELA results, the test developer partner suppressed one selected-response writing item in grade 11, following initial key check reviews. This item is included in the counts of flagged items only. It is also important to note that since ELA Session 1 was the same across forms, differences in form statistics are a result of the sample of students taking the form rather than the items. Lastly, as in reading and mathematics, the test developer partner calculated p-values for items by pooling across all student responses. Therefore, the mean and SD of p-values reflect the pooled p-values. Overall, the ELA forms were moderately difficult, as expected given reading results. Cronbach's alpha ranged from 0.74 to 0.83.

Exhibit 22. ELA Session 1 Form Summary Statistics

Grade	Form	Item Total (points)	N	Mean Raw Score	SD Raw Score	Min. Raw Score	Max Raw Score	Mean p-value	SD p-value	Alpha
3	1	22 (25)	170	15.49	4.69	4	23	0.68	0.16	0.81
	2	22 (25)	185	15.97	4.74	1	24	0.68	0.16	0.83
4	1	22 (26)	188	15.47	5.15	4	25	0.65	0.13	0.83
	2	22 (26)	223	15.06	5.24	1	24	0.65	0.13	0.83
5	1	21 (25)	147	13.81	4.35	1	23	0.63	0.15	0.73
	2	21 (25)	150	15.23	4.71	4	23	0.63	0.15	0.79
6	1	19 (23)	150	14.49	4.66	2	23	0.66	0.17	0.81
	2	19 (23)	130	15.44	4.46	5	23	0.66	0.17	0.81
7	1	17 (21)	157	13.25	4.27	0	20	0.63	0.18	0.80
	2	17 (21)	166	13.09	4.09	3	21	0.63	0.18	0.78
8	1	17 (21)	183	13.63	3.81	3	21	0.63	0.15	0.74
	2	17 (21)	246	13.61	3.89	1	21	0.63	0.15	0.74
11	1	21 (24)	182	16.13	4.56	0	24	0.70	0.14	0.82
	2	21 (24)	202	16.16	4.32	2	24	0.70	0.14	0.78

Exhibit 23 contains a summary of the number of selected-response items flagged by type of flag as well as overall in each grade in writing. Each ELA grade consisted of four stand-alone selected-response items, and one Tier 1 prompt item set, which was a compilation of selected-response items. Additional information is provided on the Tier 1 writing prompt item set in the section “Writing Constructed-Response Pilot Phase 2 Results”. Few of the selected-response writing items were flagged. Only two types of items were flagged: Tier 1 writing prompts and Tier 3 stand-alone selected-response items. As indicated below, some Tier 1 writing prompts were flagged for high p-values and Tier 3 stand-alone writing item for low p-values. Since ELA Session 1 was the same across both forms, there were no differences across forms.

Exhibit 23. Number of Items Flagged in Writing by Grade

Grade	Items	Low p-value	High p-value	Low biserial	Distractor p-value	Tier reversal
3	5	0	0	0	0	0
4	5	0	0	0	0	0
5	5	1	0	0	0	0
6	5	0	0	0	0	0
7	5	1	1	0	1	0
8	5	0	1	0	0	0
11	5	1	1	1	1	0

Exhibit 24 contains a summary of the p-values for each stand-alone selected-response writing item in ELA. Additional details for the Tier 1 writing prompt item sets are presented in section "Writing Constructed-Response Pilot Phase 2 Results". The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the tier. The next two columns contain the total number of items and p-value for the item. It is important to note that the writing items were selected from different item families. That is, in each grade, there was one writing item from each of four different families. Therefore, since the p-values are represented across item families, it is not strictly expected to see the highest p-values at Tier 1, and the lowest p-values at Tier 4.

Exhibit 24. Summary of p-value by Tier for Individual Writing Selected-Response Items

Grade	Tier	Number of Items	P-value
3	1	1	0.85
	2	1	0.79
	3	1	0.57
	4	1	0.58
4	1	1	0.68
	2	1	0.50
	3	1	0.57
	4	1	0.46
5	1	1	0.85
	2	1	0.69
	3	1	0.39
	4	1	0.47
6	1	1	0.70
	2	1	0.47
	3	1	0.64
	4	1	0.53
7	1	1	0.81
	2	1	0.61
	3	1	0.23
	4	1	0.50
8	1	1	0.86
	2	1	0.54
	3	1	0.64
	4	1	0.45
11	1	1	0.71
	2	1	0.73
	3**	1	0.30
	4	1	0.48

ELA IRT Fit Results

Exhibit 25 contains a summary of the item difficulty estimates by grade for reading and ELA. Since the test developer calibrated the items concurrently, the summaries are based on both reading forms and ELA Session 1. These are the same results presented in the reading and mathematics technical memo. The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the statistic (e.g., mean, median, standard deviation). The next six columns provide values for each of the statistics for item measure (difficulty), standard error, mean square infit, standardized infit, mean square outfit and standardized outfit. Exhibit 26 contains the same summary of the person ability

estimates by grade for reading. The distribution of item measures tends to be slightly negatively skewed or to exhibit very little skew, while the distribution of person measures is slightly positively skewed.

Exhibit 25. Reading and ELA Session 1 Item Measure Summary

Grade	Statistic	Measure	SE	InMS	InZStd	OutMS	OutZStd
3	Items	65	65	65	65	65	65
	Mean	0.00	0.18	1.00	0.09	0.95	-0.12
	Standard Deviation	0.95	0.07	0.17	2.40	0.29	2.57
	Min	-1.74	0.06	0.77	-5.52	0.56	-5.28
	10th	-1.35	0.08	0.83	-2.07	0.60	-2.45
	25th	-0.66	0.12	0.87	-1.43	0.72	-1.73
	50th	-0.01	0.21	0.98	-0.11	0.94	-0.39
	75th	0.57	0.24	1.08	0.91	1.14	1.03
	90th	1.40	0.24	1.21	2.53	1.32	2.87
	Max	2.77	0.29	1.62	9.90	1.96	9.90
4	Items	61	61	61	61	61	61
	Mean	0.00	0.16	1.02	0.01	0.98	-0.06
	Standard Deviation	0.81	0.06	0.20	2.81	0.27	2.64
	Min	-1.72	0.05	0.69	-6.84	0.49	-5.76
	10th	-1.24	0.08	0.81	-3.57	0.65	-3.05
	25th	-0.56	0.09	0.88	-1.81	0.77	-1.74
	50th	0.02	0.17	0.98	-0.21	1.00	0.00
	75th	0.51	0.21	1.14	1.70	1.15	1.40
	90th	0.97	0.23	1.28	3.76	1.43	3.51
	Max	2.40	0.29	1.58	9.90	1.69	9.90

Exhibit 25. Reading and ELA Session 1 Item Measure Summary (cont.)

Grade	Statistic	Measure	SE	InMS	InZStd	OutMS	OutZStd
5	Items	60	60	60	60	60	60
	Mean	0.00	0.17	0.98	-0.19	0.93	-0.32
	Standard Deviation	0.81	0.07	0.13	2.27	0.21	2.29
	Min	-1.85	0.05	0.75	-4.15	0.62	-4.42
	10th	-1.3	0.08	0.84	-2.53	0.7	-2.57
	25th	-0.44	0.10	0.90	-1.51	0.76	-1.64
	50th	0.02	0.19	0.94	-0.72	0.89	-0.90
	75th	0.62	0.23	1.04	0.49	1.03	0.33
	90th	1.01	0.24	1.15	2.19	1.26	2.53
	Max	1.58	0.31	1.41	9.90	1.57	9.15
6	Items	59	59	59	59	59	59
	Mean	0.00	0.18	1.01	0.18	0.97	0.00
	Standard Deviation	0.99	0.06	0.14	1.83	0.27	2.05
	Min	-1.89	0.05	0.74	-3.68	0.53	-4.03
	10th	-1.47	0.08	0.85	-1.85	0.63	-2.29
	25th	-0.51	0.12	0.9	-1.1	0.75	-1.47
	50th	0.06	0.20	1.00	0.00	0.98	-0.26
	75th	0.69	0.23	1.10	1.21	1.13	1.43
	90th	1.11	0.25	1.20	2.56	1.30	2.13
	Max	2.04	0.27	1.38	5.89	1.66	5.79
7	Items	65	65	65	65	65	65
	Mean	0.00	0.17	1.00	0.03	0.94	-0.19
	Standard Deviation	0.89	0.06	0.16	2.27	0.25	2.35
	Min	-2.47	0.06	0.7	-5.67	0.55	-4.98
	10th	-1.18	0.08	0.81	-2.63	0.6	-2.83
	25th	-0.72	0.13	0.86	-1.65	0.71	-1.73
	50th	0.08	0.18	1.00	0.00	0.99	-0.14
	75th	0.64	0.20	1.08	1.04	1.07	0.82
	90th	0.96	0.22	1.24	3.19	1.28	2.67
	Max	2.24	0.31	1.44	7.65	1.52	8.45
8	Items	65	65	65	65	65	65
	Mean	0.00	0.17	1.00	0.20	0.98	0.03
	Standard Deviation	0.87	0.07	0.13	1.93	0.21	1.90
	Min	-2.56	0.05	0.66	-4.24	0.51	-4.20
	10th	-1.42	0.07	0.86	-1.70	0.68	-1.83
	25th	-0.47	0.13	0.91	-1.00	0.81	-1.21
	50th	0.13	0.18	0.99	-0.08	0.99	-0.11
	75th	0.64	0.22	1.07	1.31	1.09	1.22
	90th	1.03	0.24	1.22	2.79	1.31	2.28
	Max	1.55	0.33	1.32	8.84	1.52	6.96

Exhibit 25. Reading and ELA Session 1 Item Measure Summary (cont.)

	Items	57	57	57	57	57	57
	Mean	0.00	0.16	0.99	-0.09	0.93	-0.28
	Standard Deviation	0.89	0.07	0.16	2.56	0.27	2.54
	Min	-2.23	0.06	0.74	-5.31	0.51	-4.58
11	10th	-1.22	0.08	0.84	-2.23	0.62	-2.52
	25th	-0.45	0.10	0.89	-1.37	0.73	-1.56
	50th	0.00	0.16	0.96	-0.54	0.87	-0.83
	75th	0.67	0.22	1.08	0.80	1.07	0.49
	90th	1.05	0.26	1.21	2.03	1.31	2.29
	Max	1.86	0.30	1.57	9.90	1.72	9.90

Exhibit 26. Reading and ELA Session 1 Person Measure Summary

Grade	Statistic	Measure	SE	InMS	InZStd	OutMS	OutZStd
3	Persons	859	859	859	859	859	859
	Mean	0.79	0.52	1.00	0.10	0.93	-0.01
	Standard Deviation	1.25	0.17	0.27	1.04	0.39	0.99
	Min	-4.91	0.40	0.46	-2.83	0.14	-2.58
	10th	-0.78	0.41	0.67	-1.22	0.47	-1.16
	25th	-0.12	0.43	0.79	-0.61	0.62	-0.70
	50th	0.73	0.47	0.98	0.00	0.87	-0.16
	75th	1.54	0.53	1.18	0.75	1.17	0.67
	90th	2.26	0.67	1.36	1.43	1.45	1.30
	Max	5.22	1.86	2.14	4.16	2.82	4.12
4	Persons	964	964	964	964	964	964
	Mean	0.53	0.49	1.03	0.19	0.95	0.02
	Standard Deviation	1.25	0.16	0.23	0.92	0.34	0.92
	Min	-4.02	0.38	0.48	-2.54	0.17	-2.29
	10th	-0.94	0.39	0.73	-1.02	0.55	-1.12
	25th	-0.46	0.41	0.87	-0.37	0.73	-0.62
	50th	0.51	0.45	1.01	0.18	0.94	-0.05
	75th	1.40	0.50	1.18	0.79	1.14	0.62
	90th	2.12	0.63	1.33	1.34	1.33	1.25
	Max	5.03	1.90	1.89	3.27	5.75	3.43
5	Persons	765	765	765	765	765	765
	Mean	0.49	0.46	1.03	0.15	0.96	0.02
	Standard Deviation	1.04	0.13	0.27	0.98	0.31	1.04
	Min	-3.86	0.37	0.43	-2.38	0.25	-2.64
	10th	-0.71	0.38	0.71	-1.11	0.58	-1.26
	25th	-0.22	0.40	0.84	-0.52	0.73	-0.72
	50th	0.46	0.44	1.01	0.12	0.94	-0.12
	75th	1.28	0.47	1.19	0.78	1.17	0.75
	90th	1.87	0.56	1.38	1.41	1.37	1.48
	Max	3.90	2.18	2.51	4.49	2.30	3.12
6	Persons	777	777	777	777	777	777
	Mean	0.75	0.52	1.00	0.06	0.93	-0.02
	Standard Deviation	1.25	0.19	0.29	0.96	0.44	1.01
	Min	-3.08	0.41	0.41	-2.60	0.17	-2.45
	10th	-0.76	0.42	0.66	-1.10	0.51	-1.13
	25th	-0.15	0.43	0.81	-0.53	0.64	-0.65
	50th	0.60	0.47	0.98	0.02	0.84	-0.18
	75th	1.46	0.53	1.14	0.57	1.10	0.44
	90th	2.38	0.66	1.39	1.24	1.43	1.43
	Max	5.20	1.86	2.42	3.81	5.50	3.25

Exhibit 26. Reading and ELA Session 1 Person Measure Summary (cont.)

Grade	Statistic	Measure	SE	InMS	InZStd	OutMS	OutZStd
7	Persons	873	873	873	873	873	873
	Mean	0.62	0.51	1.02	0.12	0.93	-0.03
	Standard Deviation	1.19	0.18	0.30	1.10	0.39	1.15
	Min	-5.27	0.38	0.40	-3.58	0.24	-3.26
	10th	-0.81	0.39	0.67	-1.21	0.52	-1.21
	25th	-0.17	0.41	0.81	-0.52	0.67	-0.76
	50th	0.69	0.48	0.99	0.10	0.85	-0.27
	75th	1.33	0.54	1.19	0.74	1.14	0.54
	90th	2.08	0.64	1.44	1.52	1.46	1.58
	Max	4.77	1.90	2.03	3.39	2.83	4.01
8	Persons	936	936	936	936	936	936
	Mean	0.60	0.51	0.99	0.03	0.96	0.02
	Standard Deviation	1.09	0.20	0.23	0.87	0.33	0.87
	Min	-2.97	0.37	0.40	-2.23	0.21	-2.14
	10th	-0.59	0.38	0.73	-1.01	0.64	-0.95
	25th	-0.13	0.40	0.83	-0.50	0.75	-0.56
	50th	0.55	0.47	0.97	0.00	0.90	-0.12
	75th	1.18	0.52	1.10	0.47	1.11	0.49
	90th	1.93	0.64	1.30	1.15	1.33	1.25
	Max	4.80	2.06	2.03	4.10	5.20	4.30
11	Persons	879	879	879	879	879	879
	Mean	0.92	0.53	1.01	0.13	0.93	0.05
	Standard Deviation	1.30	0.22	0.26	0.97	0.40	1.01
	Min	-3.05	0.39	0.45	-2.41	0.10	-2.15
	10th	-0.56	0.40	0.69	-1.02	0.42	-1.07
	25th	-0.01	0.42	0.82	-0.47	0.66	-0.68
	50th	0.73	0.46	1.00	0.06	0.91	-0.06
	75th	1.78	0.56	1.15	0.60	1.15	0.63
	90th	2.69	0.68	1.35	1.30	1.43	1.39
	Max	5.55	2.03	2.05	3.62	3.34	3.92

Exhibit 27 contains a summary of the principal component analysis (PCA) of residual variation, as produced by WINSTEPS. The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the number of items, and the remaining three columns indicate the percent of variation associated with each dimension. As expected, the first dimension associated with the Rasch model accounts for the majority of score variation. It is important to note that since items were calibrated concurrently, the PCA is based on both reading forms and ELA Session 1.

Exhibit 27. Principal Components Analysis in Reading and ELA

Grade	Number of Items	Percent First Dimension	Percent Second Dimension	Percent Third Dimension
3	65	56.7	1.5	1.3
4	61	49.8	2.1	1.7
5	60	42.3	2.7	2.0
6	59	56.7	1.6	1.5
7	65	52.6	2.5	1.9
8	65	50.2	1.8	1.6
11	57	59.1	1.6	1.4

Exhibit 28 contains a summary of the writing selected-response items and Tier 1 prompt item sets flagged for misfit. The first column of the Exhibit indicates the grade. The second column of the Exhibit indicates the total number of items. The next three columns indicate the number of items flagged for infit, outfit, and the total number of items flagged by either infit or outfit. Overall, a small number of items were flagged for misfit, and no items were flagged for infit. Exhibit 29 provides the statistics for each flagged item.

Exhibit 28. Number of Items Flagged for Misfit in Writing

Grade	Total Number of Items	Infit	Outfit	Total
3	5	0	0	0
4	5	0	0	0
5	5	0	0	0
6	5	0	1	1
7	5	0	1	1
8	5	0	0	0
11	4	0	1	1

Exhibit 29. Item Statistics for Misfitting Items in Writing

Grade	Item	Measure	SE	InMS	InZstd	OutMS	OutZStd	Infit	Outfit
6	124263A	0.91	0.14	1.26	4.13	1.53	5.00		1
7	124271A	2.24	0.15	1.09	1.09	1.31	1.83		1
11	124319A	-0.23	0.12	1.26	4.47	1.52	4.16		1

Writing Constructed-Response Pilot Phase 2 Results

The test development partner scored the Tier 2 and 3 writing constructed-response items using trait scores for Organization, Idea Development and Conventions. Student could receive scores of Full Evidence (F), Partial Evidence (P), Limited Evidence (L) or Unrelated Evidence (U) on each of the three traits. The three trait scores were then resolved into a total score. Tier 1 writing prompt item sets consisted of a set of selected-response items, which were resolved into a total score based on the number of items a student answered correctly. Exhibit 30 contains a sample rubric for the explanatory writing constructed-response items. Students could also receive one of seven condition codes, as shown in Exhibit 31. The assignment of a condition code was applied across all trait scores.

Exhibit 30. Informative/Explanatory Writing Scoring Plan – Pilot Phase 2

	0	1	2	3	4	5	6	7
Tier 3	UUU	PUF, PUP PUL, PUU UPF, UPP, UPL, UPU LUF, LUP LUL, LUU	LFF, LFP, LFL, LFU LPF, LPP, LPL, LPU LLF, LLP LLL, LLU	FFU FPU, PFU PPU FLU PLU	PPF, PPP, PPL FLF, FLP, FLL PLF, PLP, PLL	FFL FPL PFL	FFP, FPF FPP PFF PFP	FFF
Tier 2	UUU	ULF, ULP ULL, ULU UUF, UUP UUL	PPU, PLU LFF, LFP, LFL, LFU, LPF,LPP, LPL, LPU, LLF, LLP, LLL, LLU	FFU, FPU, FLU, PLF, PLP, PLL, PFU PPF, PPP, PPL	FFL FPP, FPL, FLF, FLP, FLL PFF, PFP ,PFL	FFF, FFP, FPF		
Tier 1	No Items Correct	One or Two items correct	Three to Six items correct					

Exhibit 31. Condition Codes for Tier 2 and Tier 3 Writing Constructed-Response Items

Condition	Code
Graphic Organizer	G
No Annotation	N
Foreign Language	L
Corrupt Data	C
Direct Replication	D
Blank Document	B
Screen Capture Error	S

Exhibit 32 contains summary statistics for the total score for each of the writing prompts. The students included in these summary statistics include only those students not receiving a condition code. The first column of the exhibit indicates the grade, and the second column indicates the item ID. The next column contains the tier of the prompt. The next six columns of the Exhibit indicate the number of students, the mean total score and standard deviation, the median, minimum and maximum total score. The last column of the exhibit indicates the correlation between the writing prompt total score and the ELA Session 1 score, with the Tier 1 writing prompt excluded.

Exhibit 33 contains a summary of the number of students scoring at each total score point for each prompt. There is a maximum score of 2 for Tier 1 prompts, 5 for Tier 2 prompts, and 7 for Tier 3 prompts. Exhibit 54 contains the percentage of students at each total score point. Exhibit 55 contains a summary of the number of students score at each trait score point for each prompt. The Tier 1 item set does not have trait scores, since it is a compilation of selected-response items.

Exhibits 36 through 42 contain a summary of the number of students receiving each trait score pattern by grade and prompt. In these exhibits, 0 indicates unrelated evidence, 1 indicates limited evidence, 2 indicates partial evidence, and 3 indicates full evidence.

Exhibit 32. Writing Prompt Summary Statistics

Summary Statistics Grade	Item	Tier	N	Mean Total Score	SD Total Score	Median Total Score	Min Total Score	Max Total Score	Correlation
	300012W	TIER 1	350	1.79	0.46	2.0	0	2	0.43
	125981B	TIER 2	92	2.20	1.72	2.0	0	5	0.62
3	126013B	TIER 2	73	2.71	1.68	2.0	0	5	0.45
	125971B	TIER 3	126	1.98	1.57	2.0	0	7	0.34
	126006B	TIER 3	110	1.69	1.49	2.0	0	6	0.42
	400019W	TIER 1	405	1.56	0.54	2.0	0	2	0.52
	126171B	TIER 2	108	2.03	1.53	2.0	0	5	0.48
4	126203B	TIER 2	109	2.32	1.70	2.0	0	5	0.61
	126163B	TIER 3	136	1.49	1.50	1.0	0	6	0.50
	126195B	TIER 3	148	1.95	1.56	2.0	0	6	0.35
	500009W	TIER 1	293	1.57	0.52	2.0	0	2	0.50
	126982B	TIER 2	83	2.46	1.55	3.0	0	5	0.43
5	127091B	TIER 2	62	2.02	1.65	2.0	0	5	0.44
	126972B	TIER 3	87	2.00	1.64	2.0	0	6	0.54
	127081B	TIER 3	114	2.08	1.69	2.0	0	6	0.37
	600007W	TIER 1	281	1.71	0.51	2.0	0	2	0.39
	127291B	TIER 2	67	2.54	1.46	2.0	0	5	0.62
6	127369B	TIER 2	61	2.51	1.40	2.0	0	5	0.43
	127286B	TIER 3	95	1.86	1.39	2.0	0	7	0.35
	127365B	TIER 3	110	1.60	1.50	2.0	0	7	0.49
	700016W	TIER 1	319	1.81	0.42	2.0	0	2	0.34
	127642B	TIER 2	69	2.16	1.54	2.0	0	5	0.25
7	127663B	TIER 2	74	2.45	1.58	2.0	0	5	0.43
	127638B	TIER 3	128	1.63	1.37	2.0	0	7	0.44
	127658B	TIER 3	111	1.65	1.29	1.0	0	7	0.26
	800016W	TIER 1	422	1.88	0.35	2.0	0	2	0.36
	127798B	TIER 2	89	1.89	1.32	2.0	0	5	0.20
8	127819B	TIER 2	100	2.21	1.46	2.0	0	5	0.32
	127794B	TIER 3	141	1.84	1.69	1.0	0	7	0.39
	127815B	TIER 3	178	1.96	1.38	2.0	0	7	0.46
	110016W	TIER 1	378	1.88	0.35	2.0	0	2	0.37
	126802B	TIER 2	92	1.80	1.58	1.0	0	5	0.39
11	126865B	TIER 2	96	1.97	1.48	1.0	0	5	0.40
	126795B	TIER 3	132	1.72	1.70	1.0	0	7	0.32
	126858B	TIER 3	136	1.86	1.56	1.5	0	7	0.36

Exhibit 33. Writing Prompt Total Score Frequencies

Grade	Item	Tier	0	1	2	3	4	5	6	7	B	C	D	G	N
	300012W	TIER 1	8	59	283						5				
	125981B	TIER 2	21	11	26	13	5	16			24	53	0	0	1
3	126013B	TIER 2	9	9	20	5	16	14			35	71	2	1	2
	125971B	TIER 3	10	41	59	0	4	0	10	2	48	0	4	0	6
	126006B	TIER 3	25	23	49	0	6	1	6	0	42	0	5	0	13
	400019W	TIER 1	8	163	234						6				
	126171B	TIER 2	19	25	30	12	12	10			32	77	0	0	6
4	126203B	TIER 2	19	20	29	5	20	16			9	66	1	0	3
	126163B	TIER 3	37	43	39	0	8	4	5	0	36	0	3	0	13
	126195B	TIER 3	22	37	63	3	9	4	10	0	9	0	5	0	8
	500009W	TIER 1	4	117	172						4				
	126982B	TIER 2	12	11	16	27	5	12			19	45	1	0	2
5	127091B	TIER 2	14	13	13	10	4	8			27	53	1	0	4
	126972B	TIER 3	17	17	33	0	14	1	5	0	52	0	3	0	5
	127081B	TIER 3	18	34	30	0	24	1	7	0	32	0	2	0	2
	600007W	TIER 1	7	67	207										
	127291B	TIER 2	3	13	26	5	10	10			18	55	6	0	6
6	127369B	TIER 2	1	18	14	12	9	7			19	47	1	0	2
	127286B	TIER 3	10	20	59	0	0	0	3	3	29	0	3	0	3
	127365B	TIER 3	21	33	49	0	1	0	1	5	30	0	4	1	5
	700016W	TIER 1	4	52	263						4				
	127642B	TIER 2	8	22	13	10	9	7			18	66	1	0	3
7	127663B	TIER 2	6	20	18	4	17	9			14	67	5	0	6
	127638B	TIER 3	18	43	59	0	2	0	2	4	3	0	6	0	7
	127658B	TIER 3	9	52	40	0	5	0	4	1	2	0	3	0	7
	800016W	TIER 1	3	46	373						7				
	127798B	TIER 2	10	31	24	11	9	4			40	111	1	1	3
8	127819B	TIER 2	15	16	31	16	15	7			17	62	0	0	3
	127794B	TIER 3	17	59	44	0	9	1	5	6	9	0	1	0	5
	127815B	TIER 3	17	49	82	0	22	1	6	1	4	0	3	1	4
	110016W	TIER 1	4	36	338						6				
	126802B	TIER 2	15	42	8	10	7	10			29	54	4	1	1
11	126865B	TIER 2	9	43	12	15	8	9			23	77	2	1	2
	126795B	TIER 3	22	61	27	0	11	1	7	3	13	0	2	0	6
	126858B	TIER 3	13	55	48	1	9	0	7	3	5	0	2	0	2

Exhibit 34. Writing Prompt Total Score Percentages

Grade	Item	Tier	0	1	2	3	4	5	6	7	B	C	D	G	N
	300012W	TIER 1	2.3	16.6	79.7						1.4				
	125981B	TIER 2	12.4	6.5	15.3	7.6	2.9	9.4			14.1	31.2	0	0	0.6
3	126013B	TIER 2	4.9	4.9	10.9	2.7	8.7	7.6			19.0	38.6	1.1	0.5	1.1
	125971B	TIER 3	5.4	22.3	32.1	0	2.2	0	5.4	1.1	26.1	0	2.2	0	3.3
	126006B	TIER 3	14.7	13.5	28.8	0	3.5	0.6	3.5	0	24.7	0	2.9	0	7.6
	400019W	TIER 1	1.9	39.7	56.9						1.5				
	126171B	TIER 2	8.5	11.2	13.5	5.4	5.4	4.5			14.3	34.5	0	0	2.7
4	126203B	TIER 2	10.1	10.6	15.4	2.7	10.6	8.5			4.8	35.1	0.5	0	1.6
	126163B	TIER 3	19.7	22.9	20.7	0	4.3	2.1	2.7	0	19.1	0	1.6	0	6.9
	126195B	TIER 3	12.9	21.8	37.1	1.8	5.3	2.4	5.9	0	5.3	0	2.9	0	4.7
	500009W	TIER 1	1.3	39.4	57.9						1.3				
	126982B	TIER 2	8.0	7.3	10.7	18.0	3.3	8.0			12.7	30.0	0.7	0	1.3
5	127091B	TIER 2	9.5	8.8	8.8	6.8	2.7	5.4			18.4	36.1	0.7	0	2.7
	126972B	TIER 3	11.6	11.6	22.4	0	9.5	0.7	3.4	0	35.4	0	2.0	0	3.4
	127081B	TIER 3	12.0	22.7	20.0	0	16.0	0.7	4.7	0	21.3	0	1.3	0	1.3
	600007W	TIER 1	2.5	23.8	73.7						0.0				
	127291B	TIER 2	2.0	8.7	17.3	3.3	6.7	6.7			12.0	36.7	4.0	0	2.7
6	127369B	TIER 2	0.8	13.8	10.8	9.2	6.9	5.4			14.6	36.2	0.8	0	1.5
	127286B	TIER 3	7.7	15.4	45.4	0	0	0	2.3	2.3	22.3	0	2.3	0	2.3
	127365B	TIER 3	14.0	22.0	32.7	0	0.7	0	0.7	3.3	20.0	0	2.7	0.7	3.3
	700016W	TIER 1	1.2	16.1	81.4						1.2				
	127642B	TIER 2	5.1	14.0	8.3	6.4	5.7	4.5			11.5	42.0	0.6	0	1.9
7	127663B	TIER 2	3.6	12.0	10.8	2.4	10.2	5.4			8.4	40.4	3.0	0	3.6
	127638B	TIER 3	12.5	29.9	41.0	0	1.4	0	1.4	2.8	2.1	0	4.2	0	4.9
	127658B	TIER 3	7.3	42.3	32.5	0	4.1	0	3.3	0.8	1.6	0	2.4	0	5.7
	800016W	TIER 1	0.7	10.7	86.9						1.6				
	127798B	TIER 2	4.1	12.7	9.8	4.5	3.7	1.6			16.3	45.3	0.4	0.4	1.2
8	127819B	TIER 2	8.2	8.8	17.0	8.8	8.2	3.8			9.3	34.1	0	0	1.6
	127794B	TIER 3	10.9	37.8	28.2	0	5.8	0.6	3.2	3.8	5.8	0	0.6	0	3.2
	127815B	TIER 3	8.9	25.8	43.2	0	11.6	0.5	3.2	0.5	2.1	0	1.6	0.5	2.1
	110016W	TIER 1	1.0	9.4	88.0										
	126802B	TIER 2	8.3	23.2	4.4	5.5	3.9	5.5			16.0	29.8	2.2	0.6	0.6
11	126865B	TIER 2	4.5	21.4	6.0	7.5	4.0	4.5			11.4	38.3	1.0	0.5	1.0
	126795B	TIER 3	14.4	39.9	17.6	0	7.2	0.7	4.6	2.0	8.5	0	1.3	0	3.9
	126858B	TIER 3	9.0	37.9	33.1	0.7	6.2	0	4.8	2.1	3.4	0	1.4	0	1.4

Exhibit 35. Writing Prompt Trait Score Frequencies

Grade	Item	Tier	Organization				Idea Development				Conventions				Condition Codes				
			U	L	P	F	U	L	P	F	U	L	P	F	B	C	D	G	N
3	125981B	TIER 2	21	35	15	21	32	23	15	22	33	13	9	37	24	53	0	0	1
	126013B	TIER 2	9	28	6	30	18	28	16	11	12	19	9	33	35	71	2	1	2
	125971B	TIER 3	21	89	8	8	51	38	20	17	13	27	54	32	48	0	4	0	6
			32	64	8	6	48	29	21	12	26	16	42	26	42	0	5	0	13
4	126171B	TIER 2	20	53	13	22	43	26	16	23	33	33	7	35	32	77	0	0	6
	126203B	TIER 2	19	45	6	39	39	27	18	25	30	24	5	50	9	66	1	0	3
	126163B	TIER 3	64	53	16	3	80	17	23	16	42	35	22	37	36	0	3	0	13
	126195B	TIER 3	29	91	23	5	59	31	21	37	42	51	20	35	9	0	5	0	8
5	126982B	TIER 2	13	25	28	17	23	28	17	15	18	21	8	36	19	45	1	0	2
	127091B	TIER 2	15	24	11	12	27	17	4	14	22	10	4	26	27	53	1	0	4
	126972B	TIER 3	20	46	20	1	34	38	9	6	21	31	15	20	52	0	3	0	5
	127081B	TIER 3	32	47	29	6	52	34	21	7	28	33	23	30	32	0	2	0	2
6	127291B	TIER 2	7	32	14	14	16	4	24	23	7	21	0	39	18	55	6	0	4
	127369B	TIER 2	2	28	22	9	19	3	22	17	11	19	9	22	19	47	1	0	2
	127286B	TIER 3	26	63	0	6	22	37	22	14	15	35	13	32	29	0	3	0	3
	127365B	TIER 3	47	56	1	6	38	32	30	10	28	39	11	32	30	0	4	1	5
7	127642B	TIER 2	16	26	17	10	30	15	7	17	15	27	1	26	18	66	1	0	3
	127663B	TIER 2	9	32	15	18	26	12	5	31	14	25	7	28	14	67	5	0	6
	127638B	TIER 3	59	61	3	5	31	61	15	21	20	50	16	42	3	0	6	0	7
	127658B	TIER 3	59	42	8	2	21	62	17	11	14	52	18	27	2	0	3	0	7
8	127798B	TIER 2	17	46	22	4	41	1	29	18	24	29	4	32	40	111	1	1	3
	127819B	TIER 2	19	41	32	8	29	2	38	31	30	25	4	41	17	62	0	0	3
	127794B	TIER 3	22	95	10	14	76	44	9	12	35	44	21	41	9	0	1	0	5
	127815B	TIER 3	20	126	23	9	65	93	12	8	41	58	30	49	4	0	3	1	4
11	126802B	TIER 2	41	22	15	14	57	19	2	14	19	30	2	41	29	54	4	1	1
	126865B	TIER 2	15	37	28	16	52	28	4	12	21	25	1	49	23	77	2	1	2
	126795B	TIER 3	45	65	10	12	83	27	9	13	34	40	22	36	13	0	2	0	6
	126858B	TIER 3	18	98	9	11	68	42	18	8	27	44	25	40	5	0	2	0	2

Exhibit 36. Grade 3 Trait Score Pattern Frequency Distributions

125981B			126013B			125971B			126006B		
Pattern	**N**	**%**	**Pattern**	**N**	**%**	**Pattern**	**N**	**%**	**Pattern**	**N**	**%**
000	21	12.4	000	9	4.9	000	10	5.4	000	25	14.7
100	6	3.5	100	3	1.6	001	7	3.8	001	5	2.9
101	2	1.2	101	5	2.7	002	3	1.6	002	1	0.6
103	2	1.2	111	6	3.3	003	1	0.5	003	1	0.6
110	4	2.4	112	5	2.7	100	2	1.1	101	5	2.9
111	5	2.9	113	1	0.5	101	9	4.9	102	8	4.7
112	4	2.4	121	2	1.1	102	15	8.2	103	2	1.2
113	2	1.2	122	1	0.5	103	4	2.2	110	1	0.6
121	2	1.2	123	4	2.2	110	1	0.5	111	2	1.2
123	2	1.2	133	1	0.5	111	5	2.7	112	15	8.8
131	1	0.6	203	1	0.5	112	19	10.3	113	8	4.7
132	1	0.6	212	1	0.5	113	10	5.4	121	2	1.2
133	4	2.4	213	2	1.1	121	4	2.2	122	8	4.7
200	1	0.6	223	1	0.5	122	3	1.6	123	5	2.9
210	1	0.6	233	1	0.5	123	8	4.3	131	1	0.6
211	1	0.6	311	4	2.2	131	1	0.5	132	2	1.2
212	2	1.2	312	1	0.5	132	4	2.2	133	5	2.9
222	2	1.2	313	8	4.3	133	4	2.2	203	1	0.6
223	2	1.2	321	1	0.5	212	2	1.1	212	1	0.6
233	6	3.5	322	1	0.5	221	1	0.5	213	1	0.6
311	1	0.6	323	6	3.3	232	2	1.1	222	3	1.8
313	3	1.8	331	1	0.5	233	3	1.6	232	2	1.2
321	1	0.6	333	8	4.3	312	1	0.5	312	1	0.6
323	6	3.5	BBB	35	19.0	322	4	2.2	323	3	1.8
333	10	5.9	CCC	71	38.6	332	1	0.5	331	1	0.6
BBB	24	14.1	DDD	2	1.1	333	2	1.1	332	1	0.6
CCC	53	31.2	GGG	1	0.5	BBB	48	26.1	BBB	42	24.7
NNN	1	0.6	NNN	2	1.1	DDD	4	2.2	DDD	5	2.9

Exhibit 37. Grade 4 Trait Score Pattern Frequency Distributions

126171B			126203B			126163B			126195B		
Pattern	N	%	Pattern	N	%	Pattern	N	%	Pattern	N	%
000	19	8.5	000	19	10.1	000	37	19.7	000	22	12.9
033	1	0.4	100	8	4.3	001	12	6.4	001	5	2.9
100	11	4.9	101	3	1.6	002	5	2.7	002	2	1.2
101	8	3.6	102	2	1.1	003	10	5.3	100	8	4.7
103	4	1.8	103	4	2.1	100	2	1.1	101	13	7.6
110	3	1.3	110	1	0.5	101	7	3.7	102	2	1.2
111	9	4.0	111	5	2.7	102	4	2.1	103	5	2.9
112	1	0.4	113	7	3.7	103	1	0.5	110	4	2.4
113	4	1.8	121	1	0.5	110	3	1.6	111	13	7.6
121	1	0.4	123	4	2.1	111	6	3.2	112	5	2.9
123	6	2.7	131	4	2.1	112	1	0.5	113	5	2.9
132	1	0.4	133	6	3.2	113	5	2.7	120	2	1.2
133	5	2.2	201	1	0.5	121	2	1.1	121	5	2.9
201	1	0.4	211	1	0.5	122	5	2.7	122	3	1.8
211	2	0.9	220	1	0.5	123	8	4.3	123	6	3.5
212	1	0.4	223	3	1.6	131	2	1.1	130	3	1.8
213	2	0.9	303	2	1.1	132	2	1.1	131	7	4.1
221	2	0.9	311	4	2.1	133	5	2.7	132	2	1.2
222	1	0.4	312	1	0.5	203	1	0.5	133	8	4.7
231	2	0.9	313	8	4.3	213	2	1.1	203	2	1.2
232	1	0.4	322	2	1.1	221	2	1.1	211	2	1.2
233	1	0.4	323	7	3.7	222	1	0.5	213	1	0.6
311	1	0.4	330	1	0.5	223	3	1.6	221	2	1.2
313	3	1.3	331	5	2.7	231	3	1.6	222	3	1.8
321	4	1.8	333	9	4.8	232	2	1.1	230	1	0.6
322	1	0.4	BBB	9	4.8	233	2	1.1	231	3	1.8
323	1	0.4	CCC	66	35.1	302	1	0.5	232	2	1.2
331	3	1.3	DDD	1	0.5	321	1	0.5	233	7	4.1
332	1	0.4	NNN	3	1.6	322	1	0.5	313	1	0.6
333	8	3.6				BBB	36	19.1	330	2	1.2
BBB	32	14.3				DDD	3	1.6	331	1	0.6

Exhibit 38. Grade 5 Trait Score Pattern Frequency Distributions

126982B			127091B			126972B			127081B		
Pattern	N	%	Pattern	N	%	Pattern	N	%	Pattern	N	%

000	12	8.0	000	14	9.5	000	17	11.6	000	18	12.0
001	1	0.7	001	1	0.7	001	2	1.4	001	9	6.0
100	3	2.0	100	6	4.1	003	1	0.7	002	2	1.3
101	5	3.3	101	3	2.0	100	1	0.7	003	3	2.0
102	1	0.7	103	3	2.0	101	9	6.1	100	4	2.7
110	2	1.3	111	3	2.0	102	2	1.4	101	9	6.0
111	4	2.7	112	3	2.0	103	1	0.7	102	2	1.3
112	2	1.3	113	2	1.4	110	3	2.0	103	2	1.3
113	3	2.0	123	1	0.7	111	12	8.2	110	4	2.7
121	2	1.3	131	1	0.7	112	7	4.8	111	9	6.0
122	1	0.7	133	2	1.4	113	6	4.1	112	1	0.7
123	1	0.7	210	1	0.7	121	2	1.4	113	7	4.7
133	1	0.7	213	4	2.7	122	3	2.0	122	2	1.3
201	1	0.7	223	2	1.4	201	1	0.7	123	5	3.3
211	2	1.3	230	1	0.7	211	2	1.4	130	1	0.7
212	1	0.7	231	2	1.4	212	2	1.4	133	1	0.7
213	10	6.7	233	1	0.7	213	6	4.1	200	1	0.7
221	3	2.0	313	4	2.7	221	2	1.4	201	1	0.7
222	1	0.7	323	1	0.7	223	2	1.4	202	1	0.7
223	5	3.3	332	1	0.7	232	1	0.7	211	1	0.7
230	1	0.7	333	6	4.1	233	4	2.7	212	7	4.7
232	1	0.7	BBB	27	18.4	331	1	0.7	213	5	3.3
233	3	2.0	CCC	53	36.1	BBB	52	35.4	221	3	2.0
311	2	1.3	DDD	1	0.7	DDD	3	2.0	222	3	2.0
312	1	0.7	NNN	4	2.7	NNN	5	3.4	223	5	3.3
313	1	0.7							231	1	0.7
323	4	2.7							233	1	0.7
331	1	0.7							322	2	1.3
333	8	5.3							323	1	0.7
BBB	19	12.7							332	3	2.0
CCC	45	30.0							BBB	32	21.3

Exhibit 39. Grade 6 Trait Score Pattern Frequency Distributions

127291B	127369B	127286B	127365B

Pattern	N	%	Pattern	N	%	Pattern	N	%	Pattern	N	%
000	3	2.0	000	1	0.8	000	10	7.7	000	21	14.0
001	3	2.0	001	1	0.8	001	6	4.6	001	10	6.7
003	1	0.7	100	6	4.6	003	2	1.5	010	4	2.7
100	3	2.0	101	7	5.4	010	2	1.5	011	8	5.3
101	1	0.7	102	3	2.3	011	3	2.3	012	1	0.7
103	2	1.3	110	1	0.8	012	3	2.3	013	2	1.3
110	1	0.7	113	1	0.8	101	2	1.5	021	1	0.7
111	2	1.3	120	1	0.8	103	2	1.5	100	2	1.3
121	7	4.7	121	3	2.3	110	3	2.3	101	1	0.7
123	11	7.3	122	2	1.5	111	15	11.5	103	4	2.7
131	1	0.7	131	2	1.5	112	4	3.1	110	1	0.7
133	4	2.7	133	2	1.5	113	7	5.4	111	9	6.0
201	2	1.3	201	1	0.8	121	8	6.2	112	2	1.3
203	1	0.7	210	1	0.8	122	4	3.1	113	5	3.3
221	1	0.7	220	1	0.8	123	9	6.9	121	10	6.7
223	4	2.7	221	2	1.5	131	1	0.8	122	5	3.3
233	6	4.0	222	3	2.3	133	8	6.2	123	12	8.0
311	1	0.7	223	7	5.4	323	1	0.8	132	2	1.3
321	1	0.7	231	1	0.8	332	2	1.5	133	3	2.0
331	2	1.3	233	6	4.6	333	3	2.3	223	1	0.7
333	10	6.7	321	1	0.8	BBB	29	22.3	322	1	0.7
BBB	18	12.0	323	2	1.5	DDD	3	2.3	333	5	3.3
CCC	55	36.7	331	1	0.8	NNN	3	2.3	BBB	30	20.0
DDD	6	4.0	332	1	0.8				DDD	4	2.7
NNN	4	2.7	333	4	3.1				GGG	1	0.7
			BBB	19	14.6				NNN	5	3.3
			CCC	47	36.2						
			DDD	1	0.8						

Exhibit 40. Grade 7 Trait Score Pattern Frequency Distributions

127642B	127663B	127638B	127658B

Pattern	N	%	Pattern	N	%	Pattern	N	%	Pattern	N	%
000	8	5.1	000	6	3.6	000	18	12.5	000	9	7.3
001	6	3.8	001	2	1.2	001	9	6.3	001	10	8.1
003	2	1.3	002	1	0.6	002	1	0.7	010	4	3.3
100	4	2.5	100	7	4.2	003	1	0.7	011	26	21.1
101	5	3.2	101	2	1.2	010	1	0.7	012	4	3.3
103	4	2.5	102	2	1.2	011	19	13.2	013	6	4.9
110	1	0.6	103	3	1.8	012	5	3.5	100	1	0.8
111	4	2.5	110	1	0.6	013	5	3.5	102	1	0.8
113	1	0.6	111	5	3.0	103	2	1.4	111	11	8.9
121	3	1.9	113	1	0.6	110	1	0.7	112	5	4.1
123	1	0.6	123	1	0.6	111	16	11.1	113	6	4.9
130	1	0.6	131	2	1.2	112	5	3.5	121	1	0.8
131	1	0.6	132	2	1.2	113	9	6.3	122	4	3.3
133	1	0.6	133	6	3.6	121	3	2.1	123	7	5.7
200	1	0.6	201	1	0.6	122	2	1.4	131	1	0.8
211	4	2.5	211	2	1.2	123	8	5.6	132	1	0.8
212	1	0.6	221	1	0.6	131	3	2.1	133	4	3.3
213	2	1.3	223	1	0.6	132	1	0.7	221	3	2.4
221	1	0.6	231	3	1.8	133	11	7.6	223	2	1.6
223	2	1.3	233	7	4.2	222	1	0.7	232	2	1.6
231	2	1.3	301	2	1.2	223	1	0.7	233	1	0.8
233	4	2.5	311	2	1.2	233	1	0.7	332	1	0.8
313	2	1.3	313	1	0.6	332	1	0.7	333	1	0.8
331	1	0.6	321	1	0.6	333	4	2.8	BBB	2	1.6
333	7	4.5	322	1	0.6	BBB	3	2.1	DDD	3	2.4
BBB	18	11.5	331	2	1.2	DDD	6	4.2	NNN	7	5.7
CCC	66	42.0	332	1	0.6	NNN	7	4.9			
DDD	1	0.6	333	8	4.8						
NNN	3	1.9	BBB	14	8.4						
			CCC	67	40.4						

Exhibit 41. Grade 8 Trait Score Pattern Frequency Distributions

127798B	127819B	127794B	127815B

Pattern	N	%	Pattern	N	%	Pattern	N	%	Pattern	N	%
000	10	4.1	000	15	8.2	000	17	10.9	000	17	8.9
001	3	1.2	003	2	1.1	001	2	1.3	001	2	1.1
002	1	0.4	020	2	1.1	002	1	0.6	011	1	0.5
003	3	1.2	100	5	2.7	003	2	1.3	100	14	7.4
100	6	2.4	101	4	2.2	100	14	9.0	101	15	7.9
101	9	3.7	103	1	0.5	101	21	13.5	102	8	4.2
102	2	0.8	111	2	1.1	102	11	7.1	103	7	3.7
103	5	2.0	120	5	2.7	103	5	3.2	110	8	4.2
110	1	0.4	121	9	4.9	110	4	2.6	111	31	16.3
120	5	2.0	122	1	0.5	111	17	10.9	112	12	6.3
121	6	2.4	123	5	2.7	112	7	4.5	113	21	11.1
123	6	2.4	130	2	1.1	113	11	7.1	120	2	1.1
130	1	0.4	131	1	0.5	123	3	1.9	121	3	1.6
131	3	1.2	133	6	3.3	131	2	1.3	122	2	1.1
132	1	0.4	203	2	1.1	213	4	2.6	123	1	0.5
133	1	0.4	221	5	2.7	223	4	2.6	131	1	0.5
201	2	0.8	222	2	1.1	232	1	0.6	133	1	0.5
221	5	2.0	223	8	4.4	233	1	0.6	202	1	0.5
223	5	2.0	230	1	0.5	301	1	0.6	203	1	0.5
230	1	0.4	231	3	1.6	303	2	1.3	211	4	2.1
231	1	0.4	232	1	0.5	313	1	0.6	212	4	2.1
233	8	3.3	233	10	5.5	323	2	1.3	213	7	3.7
323	2	0.8	323	1	0.5	331	1	0.6	223	2	1.1
333	2	0.8	331	1	0.5	332	1	0.6	231	1	0.5
BBB	40	16.3	333	6	3.3	333	6	3.8	233	3	1.6
CCC	111	45.3	BBB	17	9.3	BBB	9	5.8	312	2	1.1
DDD	1	0.4	CCC	62	34.1	DDD	1	0.6	313	3	1.6
GGG	1	0.4	NNN	3	1.6	NNN	5	3.2	323	2	1.1
NNN	3	1.2							332	1	0.5
									333	1	0.5
									BBB	4	2.1
									DDD	3	1.6
									GGG	1	0.5

Exhibit 42. Grade 11 Trait Score Pattern Frequency Distributions

126802B			126865B			126795B			126858B		
Pattern	N	%	Pattern	N	%	Pattern	N	%	Pattern	N	%
000	15	8.3	000	9	4.5	000	22	14.4	000	13	9.0
001	17	9.4	001	2	1.0	001	14	9.2	001	3	2.1
002	1	0.6	003	4	2.0	002	1	0.7	002	2	1.4
003	8	4.4	100	9	4.5	003	8	5.2	100	10	6.9
100	3	1.7	101	8	4.0	100	11	7.2	101	28	19.3
101	5	2.8	102	1	0.5	101	13	8.5	102	7	4.8
103	6	3.3	103	8	4.0	102	9	5.9	103	5	3.4
110	1	0.6	111	6	3.0	103	5	3.3	110	3	2.1
111	3	1.7	113	4	2.0	111	7	4.6	111	10	6.9
112	1	0.6	131	1	0.5	112	4	2.6	112	5	3.4
113	2	1.1	200	1	0.5	113	8	5.2	113	16	11.0
133	1	0.6	201	3	1.5	120	1	0.7	121	3	2.1
203	1	0.6	203	4	2.0	122	1	0.7	122	6	4.1
211	4	2.2	210	1	0.5	123	2	1.3	123	4	2.8
213	5	2.8	211	4	2.0	131	1	0.7	132	1	0.7
223	1	0.6	213	10	5.0	132	1	0.7	212	1	0.7
233	4	2.2	231	1	0.5	133	2	1.3	213	3	2.1
303	1	0.6	233	4	2.0	211	3	2.0	222	1	0.7
311	1	0.6	303	3	1.5	212	1	0.7	223	1	0.7
313	2	1.1	313	3	1.5	223	3	2.0	233	3	2.1
323	1	0.6	320	1	0.5	233	3	2.0	310	1	0.7
333	9	5.0	323	3	1.5	311	1	0.7	312	1	0.7
BBB	29	16.0	333	6	3.0	312	3	2.0	313	2	1.4
CCC	54	29.8	BBB	23	11.4	323	2	1.3	323	3	2.1
DDD	4	2.2	CCC	77	38.3	331	1	0.7	332	1	0.7
GGG	1	0.6	DDD	2	1.0	332	2	1.3	333	3	2.1
NNN	1	0.6	GGG	1	0.5	333	3	2.0	BBB	5	3.4
			NNN	2	1.0	BBB	13	8.5	DDD	2	1.4
						DDD	2	1.3	NNN	2	1.4

There are no “flags” for the Tier 2 and Tier 3 writing constructed-response items as no flagging criteria was applied (Exhibits 33, 34, 35). Overall, the Tier 2 and Tier 3 writing constructed-response items at grades 3, 4, 5 and 6 reflected a high correlation between students who performed well on these writing constructed-response items and students who performed better overall on the ELA items (see Exhibit 32). The Tiers 2 and 3 writing constructed-response items at grades 7, 8, 11 showed a slightly lower correlation between overall performance on the ELA items and performance on these writing constructed-response items.