Appendix 2-I: Pilot Development Partner Report: Pilot Phase 1 Test Results and Item Statistics

Test and Item Statistics

Item Statistics

The contractor completed analyses using Pilot Phase 1 results, critical questions about the developed test content and using the data to inform item data reviews (Pilot Phase 1 IDR, July 14 – 17, 2014). Pilot Phase 1 analyses evaluated the appropriateness and accessibility of the developed items to a sample of the student population, and it identified (flagged) weaknesses of items for their classical difficulty, discrimination, or time. The analyses also established evidence to understand and evaluate the performance of items across tiers.

The vendor completed classical item analysis (e.g., p-value, point-biserial correlation, omit rates, distractor analysis) and applied item flagging criteria to identify items that may be candidates for revision; with suppression criteria established jointly between the vendor and the NCSC consortium. It is important to note that because of low sample sizes, the interpretation of item statistics should be in the context of the small sample sizes. In addition, all of selected-response items were either two-option or three-option items.

Recommendations from the NCSC Technical Advisory Committee (TAC) and input from NCSC states partners informed the development of the final item flagging criteria as follows:

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

For Tier 1 items in both reading and mathematics, 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.

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

For Tiers 2--4 items in both reading and mathematics, 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.

(3) Easy item: High p-value > 0.90

If the p-value was greater than 0.90 for an item in reading or mathematics, the item was flagged and reviewed at IDR.

(4) 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 reviewed at IDR.

(5) Tier reversal: Tier 1 items harder than Tier 4

The vendor conducted an evaluation of item tiers for the Pilot Phase 1 item data reviews, 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 ELA, there were four variations of each passage, one for each tier (i.e., Tier 1 – Tier 4). Each tier of a passage has an associated set of items. The number of items associated with a passage varied by tier as well as by passage. Because there was not a direct one-to-one correspondence of items across the tiers, the vendor calculated mean p-values were 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. IDR included the review of items flagged for a tier reversal and 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 have meant that there was a miskey (i.e., correct response not correctly noted), a second possible correct response, or a distractor with elements of a correct response.

Reading Item Analyses

Exhibit 1 contains form summary statistics for reading. The first column of the table indicates the grade and the second column indicates the form. The next three columns contain the total number of items, total number of selected-response (SR) items and total number of open-response (OR) items. During Pilot Phase 1, each item was considered worth one score point, meaning the total number of points and number of items is equal. The last five columns of the table indicate the number of students taking each form, the mean raw score, standard deviation of the raw score, and the minimum and maximum raw score per form. Overall, the mean raw scores ranged from 47% to 67% of the total points/form.

Exhibit 1. Reading Form Summary Statistics

Grade	Form	Number of Items			Summary Statistics				
		Total (Points)	SR	OR	Number of Students	Mean	St. Dev.	Min	Max
3	1	32	27	5	109	18.50	5.94	6	31
	2	30	25	5	49	18.22	7.10	5	30
	3	32	27	5	69	17.65	5.73	8	31
	4	31	31	0	72	18.19	5.16	8	29
	5	29	29	0	46	17.22	4.25	9	27
	6	31	26	5	40	15.98	5.80	7	28
	7	30	25	5	64	16.52	6.19	3	29
	8	32	27	5	69	19.29	5.69	7	30
4	1	34	29	5	135	17.39	6.86	6	33
	2	33	28	5	66	18.52	6.40	6	30
	3	32	27	5	62	16.87	5.71	6	28
	4	31	31	0	53	20.55	4.66	8	28
	5	32	32	0	63	18.59	6.14	7	30
	6	35	30	5	67	19.09	5.35	8	30
	7	31	26	5	51	16.02	5.33	7	31
	8	34	29	5	79	17.81	6.50	7	33
5	1	25	25	0	109	13.61	4.26	3	24
	2	25	25	0	53	13.70	4.66	5	23
	3	24	24	0	54	13.80	5.04	4	22
	4	24	24	0	65	13.35	3.96	6	21
	5	25	25	0	51	14.82	4.65	5	24
	6	25	25	0	70	14.06	4.34	2	24
	7	23	23	0	55	12.31	3.85	4	20
	8	24	24	0	69	13.51	4.21	3	22

Exhibit 1. Reading Form Summary Statistics (cont.)

Grade	Form	Number of Items			Summary Statistics				
		Total (Points)	SR	CR	Number of Students	Mean	St. Dev.	Min	Max
6	1	23	23	0	52	14.23	4.01	4	22
	2	23	23	0	111	13.41	4.18	5	22
	3	22	22	0	54	14.07	4.20	5	21
	4	22	22	0	64	12.98	3.96	5	21
	5	24	24	0	88	13.27	4.62	4	24
	6	24	24	0	73	14.90	4.83	6	24
	7	23	23	0	88	12.83	4.38	5	23
	8	23	23	0	68	13.59	4.03	5	22
7	1	26	26	0	61	14.31	4.04	7	23
	2	26	26	0	73	14.58	4.56	5	24
	3	25	25	0	97	12.79	4.18	5	22
	4	25	25	0	83	13.49	4.61	4	23
	5	25	25	0	63	16.64	4.49	6	24
	6	25	25	0	63	13.35	4.99	2	23
	7	25	25	0	50	14.02	4.44	5	23
	8	25	25	0	43	12.81	3.81	6	20
8	1	26	26	0	76	15.37	4.86	7	26
	2	26	26	0	63	14.97	5.02	6	25
	3	25	25	0	66	11.97	3.73	3	20
	4	25	25	0	57	14.68	4.14	8	25
	5	25	25	0	83	14.53	4.41	4	23
	6	25	25	0	69	16.44	4.90	6	25
	7	25	25	0	68	14.52	5.56	2	25
	8	25	25	0	64	11.73	4.04	3	24
11	1	27	27	0	77	17.44	4.95	6	26
	2	26	26	0	64	16.59	3.18	10	22
	3	29	29	0	55	15.20	4.42	4	25
	4	27	27	0	74	15.78	5.04	6	27
	5	26	26	0	52	17.48	5.36	7	25
	6	23	23	0	78	13.42	4.33	3	23
	7	27	27	0	55	14.49	5.17	4	24
	8	24	24	0	80	12.74	4.72	4	22

Exhibit 2 contains the total number of items flagged overall as well as by content type in reading. Overall, 30.3% of all items were flagged for at least one reason. The majority of items were flagged for a low item point-biserial correlation or a high option point-biserial correlation. None of the items were flagged for a high omit rate. In evaluating the number of items flagged and adequacy of the flagging criteria, an important consideration was that all items were two-option or three-option selected-response or one-point constructed-response items.

Exhibit 2. Number of Items Flagged in Reading by Content Type

Grade	Content Type	Number of Items Flagged	Percent of Items Flagged
3	Foundational – Non-Verbal	5	13.9
	Foundational – Verbal	8	22.2
	Informational Passage	27	32.9
	Literary Passage	28	30.1
4	Foundational – Non-Verbal	4	11.1
	Foundational – Verbal	7	19.4
	Informational Passage	28	36.8
	Literary Passage	37	32.5
5	Informational Passage	36	43.4
	Literary Passage	28	25.0
6	Informational Passage	31	33.7
	Literary Passage	17	18.5
7	Informational Passage	52	45.6
	Literary Passage	17	19.3
8	Informational Passage	42	36.8
	Literary Passage	22	25.0
11	Informational Passage	55	40.4
	Literary Passage	11	15.1
Total		455	30.3

Exhibit 3 contains a summary of item p-values by grade and tier for reading. The first column of the table indicates the grade. The second column of the table indicates the tier. The next three columns contain the total number of items, total number of SR items and total number of OR 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 table indicate the mean p-value and standard deviation of p-value for each tier. Following the theoretical design of the items and tiers, the expectation was that the highest p-values would be at Tier 1 and the lowest p-values at Tier 4.

Exhibit 3. Summary of p-value by Tier in Reading

Grade	Tier	Number of Items			Summary Statistics	
		Total (Points)	SR	OR	Mean p-value	St. Dev. of p-value
3	1	49	49	0	0.79	0.11
	2	66	56	10	0.55	0.14
	3	66	56	10	0.54	0.14
	4	66	56	10	0.47	0.17
4	1	49	49	0	0.77	0.08
	2	72	62	10	0.53	0.14
	3	70	60	10	0.52	0.14
	4	71	61	10	0.46	0.14
5	1	40	40	0	0.78	0.07
	2	51	51	0	0.53	0.14
	3	52	52	0	0.51	0.13
	4	52	52	0	0.47	0.12
6	1	42	42	0	0.80	0.07
	2	46	46	0	0.60	0.12
	3	48	48	0	0.52	0.14
	4	48	48	0	0.48	0.11
7	1	40	40	0	0.76	0.09
	2	54	54	0	0.52	0.15
	3	54	54	0	0.51	0.13
	4	54	54	0	0.48	0.13
8	1	40	40	0	0.77	0.09
	2	54	54	0	0.55	0.13
	3	54	54	0	0.51	0.12
	4	54	54	0	0.48	0.13
11	1	42	42	0	0.81	0.09
	2	53	53	0	0.58	0.15
	3	57	57	0	0.54	0.14
	4	57	57	0	0.49	0.14

Exhibit 4 contains a summary of item p-values by grade and CCC for reading. The first column of the table indicates the grade. The second column of the table indicates the CCC. The next three columns contain the total number of items, total number of SR items and total number of OR items. The last two columns of the table indicate the mean p-value and standard deviation of p-value for each CCC.

Exhibit 4. Summary of p-value by CCC in ELA

Grade	CCC	Number of Items			Summary Statistics	
		Total (Points)	SR	CR	Mean p-value	St. Dev. of p-value
3	ELA_3_RI_H1	28	28	0	0.45	0.17
	ELA_3_RI_H4	16	16	0	0.69	0.16
	ELA_3_RI_I2	16	16	0	0.59	0.16
	ELA_3_RI_K5	14	14	0	0.57	0.14
	ELA_3_RL_H1	17	17	0	0.65	0.13
	ELA_3_RL_I2	32	32	0	0.70	0.11
	ELA_3_RL_K2	28	28	0	0.54	0.16
	ELA_3_RWL_H2	72	42	30	0.52	0.21
	ELA_3_RWL_I2	24	24	0	0.62	0.17
4	ELA_4_RI_H4	24	24	0	0.56	0.16
	ELA_4_RI_I3	16	16	0	0.60	0.16
	ELA_4_RI_L1	28	28	0	0.54	0.13
	ELA_4_RL_I1	28	28	0	0.60	0.15
	ELA_4_RL_K2	28	28	0	0.53	0.15
	ELA_4_RL_L1	42	42	0	0.57	0.15
	ELA_4_RWL_H2	72	42	30	0.53	0.21
	ELA_4_RWL_I2	16	16	0	0.58	0.14
	ELA_4_RWL_J1	8	8	0	0.47	0.19
5	ELA_5_RI_C4	28	28	0	0.51	0.15
	ELA_5_RI_D5	16	16	0	0.47	0.22
	ELA_5_RI_E2	31	31	0	0.54	0.18
	ELA_5_RL_B1	32	32	0	0.66	0.13
	ELA_5_RL_C2	40	40	0	0.55	0.13
	ELA_5_RL_D1	24	24	0	0.58	0.17
	ELA_5_RWL_A2	24	24	0	0.58	0.17

Exhibit 4. Summary of p-value by CCC in ELA (cont.)

Grade	CCC	Number of Items			Summary Statistics	
		Total (Points)	SR	CR	Mean p-value	St. Dev. of p-value
6	ELA_6_RI_B4	8	8	0	0.51	0.23
	ELA_6_RI_C2	12	12	0	0.56	0.20
	ELA_6_RI_G4	32	32	0	0.56	0.18
	ELA_6_RI_G6	24	24	0	0.55	0.17
	ELA_6_RL_B2	16	16	0	0.69	0.12
	ELA_6_RL_B3	16	16	0	0.63	0.17
	ELA_6_RL_C3	44	44	0	0.57	0.15
	ELA_6_RWL_A1	16	16	0	0.68	0.11
	ELA_6_RWL_C1	16	16	0	0.64	0.15
7	ELA_7_RI_J1	42	42	0	0.51	0.16
	ELA_7_RI_J5	24	24	0	0.56	0.18
	ELA_7_RI_K4	16	16	0	0.53	0.16
	ELA_7_RI_L1	16	16	0	0.52	0.18
	ELA_7_RL_I2	56	56	0	0.56	0.17
	ELA_7_RL_J1	16	16	0	0.66	0.14
	ELA_7_RWL_G1	32	32	0	0.58	0.16
8	ELA_8_RI_J1	42	42	0	0.53	0.15
	ELA_8_RI_K2	24	24	0	0.54	0.19
	ELA_8_RI_K4	24	24	0	0.56	0.20
	ELA_8_RI_L1	8	8	0	0.61	0.15
	ELA_8_RL_I2	56	56	0	0.57	0.16
	ELA_8_RL_J2	16	16	0	0.64	0.13
	ELA_8_RWL_G1	16	16	0	0.59	0.11
	ELA_8_RWL_I1	16	16	0	0.57	0.16
11	ELA_1112_RI_B1	42	42	0	0.50	0.18
	ELA_1112_RI_B5	46	46	0	0.55	0.18
	ELA_1112_RI_D1	24	24	0	0.58	0.20
	ELA_1112_RI_E1	8	8	0	0.64	0.21
	ELA_1112_RL_B1	28	28	0	0.61	0.15
	ELA_1112_RL_D1	29	29	0	0.69	0.13
	ELA_1112_RWL_B1	16	16	0	0.68	0.16
	ELA_1112_RWL_C3	16	16	0	0.62	0.15

Exhibit 5 contains a summary of median item response times by content type in reading. The first column of the table indicates the grade. The second column of the table indicates the content type. The third column indicates the median item response time in seconds for a particular content type. Median response times were calculated by first computing the median response time across students for each individual item and then taking the median response time of all items in a content type. Overall, the response time for non-verbal Foundational items was longer than verbal Foundational items and the response time for informational passage items was longer than literary passage items within each grade level.

Exhibit 5. Summary of Median Item Response Times by Content Type in ELA

Grade	Content Type	Median Response Time (Seconds)
3	Foundational – Non-Verbal	16.00
	Foundational – Verbal	12.50
	Informational Passage	24.25
	Literary Passage	21.00
4	Foundational – Non-Verbal	15.00
	Foundational – Verbal	11.75
	Informational Passage	23.75
	Literary Passage	23.00
5	Informational Passage	32.00
	Literary Passage	24.00
6	Informational Passage	28.50
	Literary Passage	23.00
7	Informational Passage	26.25
	Literary Passage	25.00
8	Informational Passage	26.00
	Literary Passage	24.75
11	Informational Passage	28.00
	Literary Passage	26.00

Exhibit 6 contains a summary of the number of flags for tier reversal according to CCC in reading. A tier reversal flag indicates when p-values are not in descending order from Tier 1 to Tier 4. In reading, item passages are written in families, where 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 varied by tier as well as by passage. Because there was 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.

First, the mean p-value for a tier was calculated as the mean of all items for a particular passage at a specific CCC and tier. For example, assume there is a passage, “Cats”, measuring two different CCCs, X and Y. At Tier 1, there are three items measuring X and two items measuring Y. The mean Tier 1 p-value for X was calculated as the average of the p-values for the three items measuring X. The same was done

at Tiers 2 through 4, resulting in four mean p-values for X in “Cats”. These four mean p-values constituted a single item family in reading for the purpose of flagging tier reversals.

Next, the four mean p-values were compared across tiers. It was expected that Tier 1 would have the highest p-values and Tier 4 would have the lowest p-values: Tier 1 items would be easiest and Tier 4 items would be most difficult. If the mean Tier 4 p-value for the item family and CCC (“Cats”, X in our example above) was greater than the mean Tier 3 p-value, meaning that Tier 4 was easier than Tier 3, then the item was flagged for a reversal between Tier 4 and Tier 3. This counts as one flag. The same process was then repeated for all passages with a given CCC. As in our example above, the vendor repeated this process for all passages with items measuring CCC, X. The values in Exhibit 6 represent the number of passages with items measuring a given CCC that contained a tier reversal flag.

In Exhibit 6, the first column indicates the grade, and the second column indicates the CCC. The third column is the number of item families, or passages. In reading, there were typically four different passages associated with a given CCC. Some exceptions to this included vocabulary or foundational items. The next three columns indicate the number of flags where the mean Tier 4 p-value was greater than the mean p-value at another tier. The column underneath Tier 4 labeled “> Tier 3” indicates the number of instances at a particular CCC where the mean Tier 4 p-value was greater than the mean Tier 3 p-value. For example, looking at the first row in Exhibit 6, for Grade 3, CCC 3.RI.H1, there are four passages containing this CCC. There was one flag for “>Tier 3” under the Tier 4 header. This means that there was one out of four passages where the mean Tier 4 p-value was greater than the mean Tier 3 p-value for that CCC. Similarly, for “>Tier 2” under the Tier 4 header, there were two passages where the mean Tier 4 p-value was greater than the mean Tier 2 p-value. It is important to note that these flags are not mutually exclusive. That is, a passage could have been flagged for Tier 4 greater than Tier 3 and Tier 2.

Similarly, the next two columns indicate the number of flags where the mean Tier 3 p-value was greater than the mean p-value at Tier 2 or Tier 1. The Tier 2 header indicates the number of flags where the mean Tier 2 p-value was greater than the mean p-value at Tier 1.

The last four columns contain the mean p-values for each CCC at Tier 1 through Tier 4. In the first step, p-values were averaged across items within a passage for a particular CCC. The mean p-values in the last four columns are these p-values averaged across passages.

It is evident by looking at Exhibit 6 that very few items were flagged for having a Tier 2, Tier 3, or Tier 4 p-value that was greater than the Tier 1 p-value, meaning that Tier 1 was nearly always easier than Tiers 2, 3 or 4. The majority of the tier reversals occurred as a result of Tier 3 being easier than Tier 2, or as a result of Tier 4 being easier than either Tier 3 or Tier 2. This pattern is also evident by examining the mean p-values in the last four columns of the table. Mean Tier 1 p-values were nearly always higher than any other tier. Mean p-values for Tiers 2, 3 and 4 were often either very similar in magnitude or diverged from the expected pattern.

An evaluation of item tiers was conducted, including an ordering of items by difficulty and evaluating the order in light of items’ tier design. Exhibits 6 and 12 contain a summary of tier flags by grade and CCC.

Exhibit 6. Summary of Tier Flags by CCC in Reading

Grade	CCC	Number of Passages	Tier 4			Tier 3		Tier 2	Tier 1 Mean p-value	Tier 2 Mean p-value	Tier 3 Mean p-value	Tier 4 Mean p-value
			> Tier 3	> Tier 2	> Tier 1	> Tier 2	> Tier 1	> Tier 1				
3	3.RI.H1	4	1	2	0	4	0	0	0.65	0.38	0.47	0.41
	3.RI.H4	4	2	0	0	0	0	0	0.87	0.72	0.61	0.55
	3.RI.I2	4	1	1	0	3	0	0	0.83	0.55	0.49	0.50
	3.RI.K5	2	0	0	0	1	1	0	0.76	0.57	0.60	0.43
	3.RL.H1	4	1	2	0	3	0	0	0.79	0.56	0.63	0.57
	3.RL.I2	4	1	2	0	4	1	0	0.78	0.65	0.72	0.63
	3.RL.K2	4	2	1	0	0	0	0	0.77	0.56	0.49	0.45
	3.RWL.H2	4	1	1	0	1	0	0	0.79	0.51	0.48	0.41
	3.RWL.I2	6	1	1	0	2	0	0	0.85	0.58	0.55	0.50
4	4.RI.H4	4	1	0	0	2	0	1	0.73	0.60	0.51	0.42
	4.RI.I3	4	1	2	0	1	0	0	0.83	0.55	0.52	0.49
	4.RI.L1	4	2	3	1	3	0	0	0.71	0.47	0.55	0.53
	4.RL.I1	4	0	0	0	1	0	0	0.75	0.57	0.54	0.47
	4.RL.K2	4	0	1	0	1	0	0	0.78	0.52	0.53	0.42
	4.RL.L1	4	3	2	0	1	0	1	0.75	0.61	0.54	0.56
	4.RWL.H2	4	0	0	0	2	0	0	0.78	0.49	0.53	0.39
	4.RWL.I2	4	1	2	0	1	0	1	0.77	0.54	0.52	0.50
	4.RWL.J1	2	1	1	0	0	0	0	0.76	0.42	0.34	0.36
5	5.RI.C4	4	1	1	0	1	0	0	0.79	0.51	0.46	0.43
	5.RI.D5	2	2	1	0	1	0	0	0.78	0.37	0.35	0.38
	5.RI.E2	4	2	1	0	3	0	0	0.76	0.47	0.50	0.41
	5.RL.B1	4	0	1	0	1	0	0	0.81	0.62	0.62	0.58
	5.RL.C2	4	1	0	0	0	0	0	0.82	0.57	0.51	0.48
	5.RL.D1	4	0	1	0	2	0	0	0.79	0.52	0.52	0.41
	5.RWL.A2	6	2	1	0	3	1	1	0.75	0.55	0.53	0.48
6	6.RI.B4	2	1	1	0	1	0	0	0.85	0.41	0.41	0.36
	6.RI.C2	2	1	0	0	0	0	0	0.87	0.74	0.42	0.46
	6.RI.G4	4	2	0	0	1	0	0	0.80	0.58	0.44	0.42
	6.RI.G6	4	1	0	0	0	0	0	0.78	0.62	0.44	0.41
	6.RL.B2	4	1	1	0	2	0	0	0.79	0.71	0.68	0.58
	6.RL.B3	4	1	1	0	2	0	0	0.83	0.61	0.61	0.45
	6.RL.C3	4	0	1	0	1	0	0	0.81	0.57	0.51	0.46
	6.RWL.A1	4	1	1	0	2	2	0	0.76	0.69	0.67	0.61
	6.RWL.C1	4	1	3	0	4	0	0	0.79	0.54	0.65	0.57

Exhibit 6. Summary of Tier Flags by CCC in Reading

Grade	CCC	Number of Passages	Tier 4			Tier 3		Tier 2	Tier 1 Mean p-value	Tier 2 Mean p-value	Tier 3 Mean p-value	Tier 4 Mean p-value
			> Tier 3	> Tier 2	> Tier 1	> Tier 2	> Tier 1	> Tier 1				
7	7.RI.J1	4	2	1	0	1	0	0	0.80	0.48	0.48	0.46
	7.RI.J5	4	3	3	0	3	0	0	0.82	0.45	0.51	0.49
	7.RI.K4	4	2	2	0	1	0	0	0.76	0.48	0.45	0.44
	7.RI.L1	2	1	2	0	1	0	0	0.75	0.40	0.46	0.47
	7.RL.I2	4	1	2	0	1	0	0	0.76	0.56	0.53	0.49
	7.RL.J1	4	2	0	1	1	1	2	0.72	0.72	0.65	0.53
	7.RWL.G1	8	3	2	0	3	2	1	0.72	0.57	0.51	0.52
8	8.RI.J1	4	2	2	0	2	0	0	0.77	0.53	0.51	0.45
	8.RI.K2	4	3	3	0	3	0	0	0.80	0.39	0.46	0.46
	8.RI.K4	4	1	2	0	1	0	0	0.81	0.56	0.54	0.41
	8.RI.L1	2	0	1	0	1	0	0	0.81	0.57	0.56	0.49
	8.RL.I2	4	2	0	0	0	0	1	0.78	0.61	0.50	0.48
	8.RL.J2	4	2	3	0	3	0	0	0.79	0.56	0.59	0.62
	8.RWL.G1	4	2	1	1	1	1	3	0.63	0.66	0.52	0.57
	8.RWL.I1	4	1	0	0	1	1	0	0.76	0.54	0.51	0.45
11	1112.RI.B1	4	2	0	0	0	0	0	0.84	0.51	0.42	0.39
	1112.RI.B5	4	1	1	0	1	0	0	0.82	0.55	0.49	0.42
	1112.RI.D1	4	0	2	0	3	1	0	0.82	0.55	0.60	0.47
	1112.RI.E1	2	0	1	0	2	1	0	0.85	0.50	0.71	0.52
	1112.RL.B1	4	2	2	0	1	0	0	0.81	0.64	0.52	0.55
	1112.RL.D1	4	1	1	0	3	0	0	0.81	0.67	0.65	0.62
	1112.RWL.B1	4	2	2	0	1	0	1	0.87	0.64	0.61	0.61
	1112.RWL.C3	4	1	1	0	0	1	1	0.75	0.64	0.59	0.49

Note: 59 out of 64 passages or foundational item families (92.2%) were flagged for tier reversal.

Mathematics Item Analyses

Exhibit 7 contains form summary statistics for Mathematics. The first column of the table indicates the grade and the second column indicates the form. The next three columns contain the total number of items, total number of SR items and total number of constructed-response (CR) items. Each item was worth one score point, meaning the total number of points and number of items was equal. The last five columns of the table indicate the number of students taking each form, the mean raw score, standard deviation of the raw score, and the minimum and maximum raw score per form. Overall, the mean raw scores ranged from 38% to 56% of the total points/form.

Exhibit 7. Mathematics Form Summary Statistics

		Number of Items			Summary Statistics				
Grade	Form	Total (Points)	SR	CR	Number of Students	Mean	St. Dev.	Min	Max
3	1	25	24	1	102	12.05	4.36	2	23
	2	25	24	1	63	13.21	4.31	6	23
	3	25	23	2	73	12.12	3.89	4	20
	4	25	23	2	56	12.36	4.91	2	23
	5	25	23	2	61	11.98	3.96	5	21
	6	25	24	1	54	13.89	5.03	4	24
	7	25	23	2	58	12.64	5.36	4	24
	8	25	24	1	66	12.29	5.04	5	25
4	1	25	20	5	126	9.71	3.43	3	20
	2	25	20	5	45	10.87	3.51	4	20
	3	25	21	4	67	11.05	4.22	3	24
	4	25	21	4	50	11.50	3.34	6	20
	5	25	21	4	59	11.25	3.90	4	22
	6	25	20	5	62	9.90	2.67	5	18
	7	25	20	5	46	11.15	3.85	5	22
	8	25	21	4	59	11.25	3.86	5	22
5	1	25	24	1	125	10.17	3.09	4	21
	2	25	23	2	66	9.38	3.45	1	20
	3	25	23	2	50	11.88	3.54	4	19
	4	25	24	1	63	11.75	3.70	3	21
	5	25	23	2	44	10.27	3.10	5	19
	6	25	23	2	93	9.98	2.81	4	17
	7	25	24	1	44	11.59	3.62	4	23
	8	25	24	1	65	11.48	3.51	6	24

Exhibit 7. Mathematics Form Summary Statistics (cont.)

Grade	Form	Number of Items			Summary Statistics				
		Total (Points)	SR	CR	Number of Students	Mean	St. Dev.	Min	Max
6	1	25	25	0	55	12.76	5.24	0	25
	2	25	25	0	46	13.98	4.54	3	25
	3	25	25	0	62	13.71	4.40	6	23
	4	25	25	0	66	13.35	4.72	6	24
	5	25	25	0	118	12.80	3.94	4	25
	6	25	25	0	68	13.22	4.16	2	22
	7	25	25	0	66	12.44	3.90	4	25
	8	25	25	0	63	13.57	4.00	7	23
7	1	25	25	0	70	13.16	3.51	6	20
	2	25	25	0	79	13.65	4.15	6	23
	3	25	25	0	81	11.37	3.51	5	20
	4	25	25	0	72	11.72	3.13	6	22
	5	25	25	0	53	12.85	3.89	7	23
	6	25	25	0	64	11.14	3.78	4	22
	7	25	25	0	63	12.29	4.23	5	24
	8	25	25	0	58	11.78	4.13	0	21
8	1	25	25	0	92	13.44	4.18	3	23
	2	25	25	0	46	12.07	5.11	4	25
	3	25	24	1	44	11.59	3.82	4	22
	4	25	24	1	67	11.88	4.13	5	22
	5	25	25	0	73	11.56	3.11	5	20
	6	25	25	0	103	12.91	3.53	4	22
	7	25	24	1	68	12.90	4.18	4	23
	8	25	24	1	62	11.84	3.96	5	22
11	1	25	24	1	86	12.17	4.41	4	21
	2	25	24	1	30	12.53	3.03	8	19
	3	25	24	1	43	12.19	4.47	6	24
	4	25	24	1	31	12.74	4.88	6	24
	5	25	24	1	70	11.06	3.94	4	24
	6	25	24	1	74	11.69	3.41	5	20
	7	25	24	1	62	10.71	4.14	2	25
	8	25	24	1	71	12.63	3.48	7	22

Exhibit 8 contains the total number of items flagged overall as well as by content type in Mathematics. The flagging criteria applied to all items are listed below.

- Low p-value, <0.20
- High p-value, >0.90
- Low point-biserial correlation, < 0.20
- High option point-biserial correlation, >0.05
- Omit rate, >5%

Overall, 50.6% of all items were flagged for at least one reason. The majority of items were flagged for a low item point-biserial correlation or a high option point-biserial correlation. None of the items were flagged for a high omit rate. In evaluating the number of items flagged and adequacy of the flagging criteria, an important consideration was that all items were two-option or three-option selected-response or one-point constructed-response items.

Exhibit 8. Number of Items Flagged in Mathematics by Content Type

Grade	Domain	Number of Items Flagged	Percent of Items Flagged
3	Geometry	3	18.8
	Measurement and Data	7	21.9
	Number and Operations Base Ten	13	36.1
	Number and Operations Fractions	8	25.0
	Operations and Algebraic Thinking	18	34.6
4	Geometry	4	25.0
	Measurement and Data	20	55.6
	Number and Operations Base Ten	11	68.8
	Number and Operations Fractions	38	79.2
	Operations and Algebraic Thinking	32	61.5
5	Geometry	14	70.0
	Measurement and Data	22	61.1
	Number and Operations Base Ten	45	70.3
	Number and Operations Fractions	23	71.9
	Operations and Algebraic Thinking	9	56.3
6	Expressions and Equations	17	47.2
	Geometry	3	18.8
	Ratio and Proportions	23	44.2
	Statistics and Probability	6	37.5
	The Number System	12	25.0
7	Expressions and Equations	10	62.5
	Geometry	22	61.1
	Ratio and Proportions	33	51.6
	Statistics and Probability	12	75.0
	The Number System	17	47.2

Exhibit 8. Number of Items Flagged in Mathematics by Content Type (cont.)

Grade	Domain	Number of Items Flagged	Percent of Items Flagged
8	Expressions and Equations	20	55.6
	Functions	22	68.8
	Geometry	22	45.8
	Statistics and Probability	15	46.9
	The Number System	10	50.0
11	Algebra and Functions	37	46.3
	Geometry	12	75.0
	Number and Quantity	18	50.0
	Statistics and Probability	17	47.2
Total		**595**	**50.6**

Exhibit 9 contains a summary of item p-values by grade and tier for Mathematics. The first column of the table indicates the grade. The second column of the table indicates the tier. The next three columns contain the total number of items, total number of SR items and total number of CR items. Each item was worth one score point, meaning the total number of score points was the same as the number of items. The last two columns of the table indicate the mean p-value and standard deviation of p-value for each tier. Following the theoretical design of the items and tiers, the expectation was that the highest p-values would be at Tier 1 and the lowest p-values at Tier 4.

Exhibit 9. Summary of p-value by Tier in Mathematics

Grade	Tier	Number of Items			Summary Statistics	
		Total (Points)	SR	CR	Mean p-value	St. Dev. of p-value
3	1	42	42	0	0.67	0.10
	2	42	38	4	0.50	0.12
	3	42	38	4	0.42	0.11
	4	42	38	4	0.42	0.11
4	1	42	42	0	0.66	0.14
	2	42	33	9	0.40	0.14
	3	42	33	9	0.37	0.12
	4	42	33	9	0.30	0.11
5	1	42	42	0	0.67	0.11
	2	42	42	0	0.40	0.11
	3	42	37	5	0.34	0.11
	4	42	37	5	0.33	0.10
6	1	42	42	0	0.70	0.10
	2	42	42	0	0.54	0.13
	3	42	42	0	0.48	0.10
	4	42	42	0	0.42	0.13
7	1	42	42	0	0.69	0.10
	2	42	42	0	0.47	0.11
	3	42	42	0	0.43	0.08
	4	42	42	0	0.38	0.12
8	1	42	42	0	0.65	0.13
	2	42	42	0	0.45	0.12
	3	42	40	2	0.44	0.11
	4	42	40	2	0.42	0.10
11	1	42	42	0	0.65	0.15
	2	42	42	0	0.47	0.10
	3	42	38	4	0.41	0.11
	4	42	38	4	0.38	0.12

Exhibit 10 contains a summary of item p-values by grade and CCC for Mathematics. The first column of the table indicates the grade. The second column of the table indicates the CCC. The next three columns contain the total number of items, total number of SR items and total number of CR items. The last two columns of the table indicate the mean p-value and standard deviation of p-value for each CCC.

Exhibit 10. Summary of p-value by CCC in Mathematics

Grade	CCC	Number of Items			Summary Statistics	
		Total (Points)	SR	CR	Mean p-value	SD of p-value
3	MA_3_DPS_1G1	16	4	12	0.52	0.13
	MA_3_GM_1I1	16	16	0	0.62	0.08
	MA_3_ME_1D2	16	16	0	0.47	0.16
	MA_3_NO_1J3	16	16	0	0.47	0.14
	MA_3_NO_1L3	16	16	0	0.50	0.16
	MA_3_NO_2C1	20	20	0	0.48	0.19
	MA_3_NO_2D3	16	16	0	0.48	0.14
	MA_3_NO_2E1	16	16	0	0.53	0.18
	MA_3_PRF_2D1	20	20	0	0.48	0.14
	MA_3_SE_1G1	16	16	0	0.49	0.13
4	MA_4_DPS_1G3	20	5	15	0.33	0.28
	MA_4_GM_1H2	16	4	12	0.55	0.20
	MA_4_ME_1G2	16	16	0	0.49	0.17
	MA_4_NO_1J5	16	16	0	0.45	0.12
	MA_4_NO_1M1	16	16	0	0.39	0.20
	MA_4_NO_1N2	16	16	0	0.40	0.14
	MA_4_NO_2D7	16	16	0	0.50	0.15
	MA_4_NO_2E2	16	16	0	0.41	0.16
	MA_4_PRF_1E3	20	20	0	0.40	0.13
	MA_4_SE_1G2	16	16	0	0.43	0.19

Exhibit 10. Summary of p-value by CCC in Mathematics (cont.)

Grade	CCC	Number of Items			Summary Statistics	
		Total (Points)	SR	CR	Mean p-value	SD of p-value
5	MA_5_GM_1C3	20	10	10	0.37	0.19
	MA_5_ME_1B2	20	20	0	0.38	0.13
	MA_5_ME_2A1	16	16	0	0.43	0.24
	MA_5_NO_1B1	16	16	0	0.48	0.19
	MA_5_NO_1B4	16	16	0	0.50	0.15
	MA_5_NO_2A5	16	16	0	0.46	0.19
	MA_5_NO_2C1	16	16	0	0.46	0.18
	MA_5_NO_2C2	16	16	0	0.42	0.14
	MA_5_PRF_1A1	16	16	0	0.47	0.11
	MA_5_PRF_2B1	16	16	0	0.43	0.18
6	MA_6_DPS_1D3	16	16	0	0.50	0.15
	MA_6_GM_1D1	16	16	0	0.54	0.19
	MA_6_ME_2A2	16	16	0	0.51	0.19
	MA_6_NO_1D2	16	16	0	0.65	0.08
	MA_6_NO_1D4	16	16	0	0.46	0.15
	MA_6_NO_1F1	20	20	0	0.57	0.16
	MA_6_NO_2A6	20	20	0	0.43	0.19
	MA_6_NO_2C3	16	16	0	0.56	0.07
	MA_6_PRF_1C1	16	16	0	0.59	0.12
	MA_6_PRF_1D1	16	16	0	0.54	0.15
7	MA_7_DPS_1K1	16	16	0	0.54	0.19
	MA_7_GM_1H2	20	20	0	0.51	0.17
	MA_7_ME_2D1	16	16	0	0.45	0.17
	MA_7_NO_2F1	16	16	0	0.52	0.19
	MA_7_NO_2F2	16	16	0	0.49	0.21
	MA_7_NO_2F6	16	16	0	0.51	0.18
	MA_7_NO_2I1	20	20	0	0.49	0.12
	MA_7_NO_2I2	16	16	0	0.44	0.12
	MA_7_PRF_1F1	16	16	0	0.46	0.12
	MA_7_PRF_1G2	16	16	0	0.50	0.11

Exhibit 10. Summary of p-value by CCC in Mathematics (cont.)

Grade	CCC	Number of Items			Summary Statistics	
		Total (Points)	SR	CR	Mean p-value	SD of p-value
8	MA_8_DPS_1H1	16	12	4	0.57	0.14
	MA_8_DPS_1K2	16	16	0	0.48	0.15
	MA_8_GM_1G1	16	16	0	0.52	0.20
	MA_8_ME_1E1	16	16	0	0.49	0.16
	MA_8_ME_2D2	16	16	0	0.51	0.13
	MA_8_NO_1K3	20	20	0	0.51	0.17
	MA_8_PRF_1E2	20	20	0	0.44	0.12
	MA_8_PRF_1F2	16	16	0	0.50	0.13
	MA_8_PRF_1G3	16	16	0	0.45	0.10
	MA_8_PRF_2E2	16	16	0	0.45	0.17
11	MA_H_DPS_1B1	16	8	8	0.43	0.10
	MA_H_DPS_1C1	20	20	0	0.53	0.17
	MA_H_GM_1B1	16	16	0	0.40	0.17
	MA_H_ME_1A2	16	16	0	0.45	0.19
	MA_H_ME_1B2	16	16	0	0.49	0.16
	MA_H_NO_1A1	20	20	0	0.43	0.13
	MA_H_PRF_1C1	16	16	0	0.46	0.14
	MA_H_PRF_2B1	16	16	0	0.48	0.18
	MA_H_PRF_2B2	16	16	0	0.58	0.11
	MA_H_PRF_2C1	16	16	0	0.52	0.15

Exhibit 11 contains a summary of median item response times by content type in Mathematics. The first column of the table indicates the grade. The second column of the table indicates the content type. The third column indicates the median item response time in seconds for a particular content type. Median response times were calculated by first computing the median response time across students for each individual item and then taking the median response time of all items in a content type.

Exhibit 11. Summary of Median Item Response Times by Content Type in Mathematics

Grade	Domain	Median Response Time (Sec.)
3	Geometry	29.50
	Measurement and Data	47.50
	Number and Operations Base Ten	37.00
	Number and Operations Fractions	43.00
	Operations and Algebraic Thinking	38.00
4	Geometry	64.75
	Measurement and Data	70.25
	Number and Operations Base Ten	43.00
	Number and Operations Fractions	40.50
	Operations and Algebraic Thinking	46.25
5	Geometry	72.00
	Measurement and Data	35.50
	Number and Operations Base Ten	35.25
	Number and Operations Fractions	41.00
	Operations and Algebraic Thinking	68.50
6	Expressions and Equations	46.00
	Geometry	46.50
	Ratio and Proportions	40.00
	Statistics and Probability	58.25
	The Number System	49.00
7	Expressions and Equations	55.50
	Geometry	65.00
	Ratio and Proportions	48.50
	Statistics and Probability	66.25
	The Number System	49.75
8	Expressions and Equations	61.50
	Functions	62.50
	Geometry	65.75
	Statistics and Probability	72.00
	The Number System	63.00
11	Algebra and Functions	77.25
	Geometry	61.00
	Number and Quantity	54.50
	Statistics and Probability	72.00

Exhibit 12 contains a summary of the tier reversal flags according to CCC in mathematics. A tier reversal flag indicates when p-values are not in descending order from Tier 1 to Tier 4. An item family in mathematics was a set of four items, one item each at Tier 1 through Tier 4. Each CCC in mathematics had between four and five item families, for a total of 16 to 20 items. To calculate tier reversal flags in mathematics, the p-value was first calculated for each item tier in an item family. The p-values were then compared across tiers within an item family. 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 4 tier reversal if the Tier 4 p-value was greater than any of the other tiers. Similarly, items were flagged for a Tier 3 reversal if the Tier 3 p-value was greater than the p-value for Tier 2 or Tier 1. Finally, items were flagged for a Tier 2

reversal if the Tier 2 p-value was greater than the p-value for Tier 1. In mathematics, the values in Table C-6 represent the number of item families measuring a given CCC that contained a tier reversal flag.

In Exhibit 12, the first column indicates the grade, and the second column indicates the CCC. The third column is the number of item families. In mathematics, there are always four or five item families. The next three columns indicate the number of flags where the Tier 4 p-value was greater than the p-value at another tier. The column underneath Tier 4 labeled "> Tier 3" indicates the number of instances at a particular CCC where the Tier 4 p-value was greater than the Tier 3 p-value. For example, looking at the first row in Table 12, we can see that for Grade 3 CCC 3.DPS.1G1, there are three flags for ">Tier 3" under the Tier 4 header. This means that there were three item families where the Tier 4 p-value was greater than the Tier 3 p-value for that CCC. Similarly, for ">Tier 2" under the Tier 4 header, there was one item where the Tier 4 p-value was greater than the Tier 2 p-value. It is important to note that these flags were not mutually exclusive. That is, an item could have been flagged for Tier 4 greater than Tier 3 and Tier 2.

Similarly, the next two columns indicate the number of flags where the Tier 3 p-value was greater than the p-value at Tier 2 or Tier 1. The Tier 2 header indicates the number of flags where the Tier 2 p-value was greater than the p-value at Tier 1. It is immediately evident by looking at Table 12 that very few items were flagged for having a Tier 2, Tier 3, or Tier 4 p-value that was greater than the Tier 1 p-value. The last four columns contain the mean p-values across item families for each CCC at Tiers 1 through Tiers 4.

Similar to Reading, very few items were flagged for having a Tier 2, Tier 3, or Tier 4 p-value that was greater than the Tier 1 p-value, meaning that Tier 1 was nearly always easier than Tiers 2, 3 or 4. The majority of the tier reversals occurred as a result of Tier 3 being easier than Tier 2, or as a result of Tier 4 being easier than either Tier 3 or Tier 2. This pattern was also evident by examining the mean p-values in the last four columns of the table. Mean Tier 1 p-values are nearly always higher than any other tier. Mean p-values for Tiers 2, 3 and 4 are often either very similar in magnitude or diverge from the expected pattern.

Exhibit 12. Summary of Tier Flags by CCC in Mathematics

Grade	CCC	Number of Item Families	Tier 4			Tier 3		Tier 2	Tier 1 Mean p-value	Tier 2 Mean p-value	Tier 3 Mean p-value	Tier 4 Mean p-value
			> Tier 3	> Tier 2	> Tier 1	> Tier 2	> Tier 1	> Tier 1				
3	3.DPS.1G1	4	3	1	0	0	0	0	0.71	0.48	0.43	0.48
	3.GM.1I1	4	2	2	0	2	0	0	0.72	0.57	0.60	0.57
	3.ME.1D2	4	1	0	0	1	0	0	0.66	0.52	0.39	0.30
	3.NO.1J3	4	2	3	1	3	0	0	0.62	0.40	0.42	0.44
	3.NO.1L3	4	4	4	1	2	0	0	0.70	0.35	0.40	0.55
	3.NO.2C1	5	3	3	0	2	1	0	0.73	0.38	0.42	0.38
	3.NO.2D3	4	3	1	0	0	0	0	0.65	0.48	0.34	0.43
	3.NO.2E1	4	1	0	0	1	0	1	0.70	0.65	0.45	0.32
	3.PRF.2D1	5	1	0	0	0	1	4	0.55	0.63	0.39	0.35
	3.SE.1G1	4	2	0	0	0	0	0	0.68	0.53	0.38	0.38
4	4.DPS.1G3	5	1	0	0	2	0	0	0.78	0.24	0.18	0.12
	4.GM.1H2	4	0	0	0	0	0	0	0.76	0.62	0.48	0.32
	4.ME.1G2	4	3	2	0	1	0	0	0.73	0.49	0.34	0.40
	4.NO.1J5	4	1	1	0	1	0	1	0.61	0.44	0.39	0.35
	4.NO.1M1	4	1	3	0	3	0	0	0.69	0.26	0.34	0.28
	4.NO.1N2	4	1	0	1	2	4	3	0.39	0.47	0.47	0.27
	4.NO.2D7	4	1	1	0	2	0	0	0.71	0.44	0.45	0.39
	4.NO.2E2	4	0	0	0	3	0	0	0.65	0.36	0.34	0.29
	4.PRF.1E3	5	3	0	0	2	1	1	0.54	0.41	0.34	0.32
	4.SE.1G2	4	1	1	0	2	0	0	0.70	0.32	0.38	0.30
5	5.GM.1C3	5	2	1	0	1	0	0	0.63	0.40	0.22	0.23
	5.ME.1B2	5	2	1	0	2	0	1	0.53	0.38	0.35	0.25
	5.ME.2A1	4	4	2	0	1	0	0	0.81	0.30	0.25	0.34
	5.NO.1B1	4	2	2	0	3	0	0	0.75	0.37	0.39	0.41
	5.NO.1B4	4	1	0	0	3	0	1	0.69	0.52	0.43	0.36
	5.NO.2A5	4	2	2	0	2	0	0	0.74	0.33	0.34	0.41
	5.NO.2C1	4	3	1	0	0	0	0	0.72	0.43	0.30	0.37
	5.NO.2C2	4	3	1	0	0	0	0	0.62	0.44	0.28	0.34
	5.PRF.1A1	4	0	1	0	2	1	0	0.58	0.46	0.46	0.36
	5.PRF.2B1	4	2	1	0	1	0	0	0.67	0.37	0.37	0.32

Exhibit 12. Summary of Tier Flags by CCC in Mathematics (cont.)

Grade	CCC	Number of Item Families	Tier 4			Tier 3		Tier 2	Tier 1 Mean p-value	Tier 2 Mean p-value	Tier 3 Mean p-value	Tier 4 Mean p-value
			> Tier 3	> Tier 2	> Tier 1	> Tier 2	> Tier 1	> Tier 1				
6	6.DPS.1D3	4	1	1	0	1	0	0	0.71	0.44	0.45	0.41
	6.GM.1D1	4	3	0	0	1	0	0	0.82	0.49	0.42	0.41
	6.ME.2A2	4	0	0	0	0	0	1	0.68	0.63	0.43	0.30
	6.NO.1D2	4	2	1	1	1	1	1	0.71	0.67	0.59	0.64
	6.NO.1D4	4	0	0	0	2	0	0	0.68	0.44	0.40	0.34
	6.NO.1F1	5	1	0	0	0	0	0	0.77	0.60	0.54	0.37
	6.NO.2A6	5	1	1	0	4	0	0	0.70	0.33	0.43	0.27
	6.NO.2C3	4	2	2	3	2	3	2	0.55	0.57	0.56	0.56
	6.PRF.1C1	4	3	2	0	2	0	1	0.72	0.57	0.52	0.56
	6.PRF.1D1	4	0	0	0	0	0	1	0.69	0.66	0.45	0.38
7	7.DPS.1K1	4	0	0	0	0	0	1	0.78	0.58	0.51	0.31
	7.GM.1H2	5	1	1	1	0	0	1	0.66	0.57	0.46	0.35
	7.ME.2D1	4	0	0	0	3	0	0	0.68	0.37	0.45	0.28
	7.NO.2F1	4	0	1	0	1	0	1	0.77	0.55	0.43	0.34
	7.NO.2F2	4	1	0	0	0	0	0	0.76	0.58	0.35	0.26
	7.NO.2F6	4	1	2	0	2	0	0	0.80	0.42	0.40	0.43
	7.NO.2I1	5	5	5	1	4	0	0	0.64	0.39	0.40	0.53
	7.NO.2I2	4	3	3	0	3	0	0	0.60	0.37	0.41	0.40
	7.PRF.1F1	4	3	2	1	1	0	0	0.59	0.46	0.34	0.44
	7.PRF.1G2	4	0	1	0	3	1	0	0.64	0.44	0.50	0.40
8	8.DPS.1H1	4	4	2	0	0	0	0	0.77	0.54	0.44	0.52
	8.DPS.1K2	4	2	2	1	2	0	1	0.62	0.44	0.43	0.44
	8.GM.1G1	4	2	3	2	4	1	0	0.58	0.40	0.57	0.52
	8.ME.1E1	4	1	0	0	0	0	0	0.71	0.48	0.41	0.37
	8.ME.2D2	4	2	0	0	1	0	0	0.67	0.51	0.46	0.39
	8.NO.1K3	5	2	3	0	4	0	0	0.74	0.38	0.49	0.45
	8.PRF.1E2	5	3	3	1	2	0	0	0.56	0.44	0.37	0.39
	8.PRF.1F2	4	0	0	0	0	0	0	0.69	0.51	0.43	0.38
	8.PRF.1G3	4	1	3	1	3	1	0	0.51	0.38	0.45	0.44
	8.PRF.2E2	4	3	1	0	1	0	0	0.68	0.47	0.32	0.32

Exhibit 12. Summary of Tier Flags by CCC in Mathematics (cont.)

Grade	CCC	Number of Item Families	Tier 4			Tier 3		Tier 2	Tier 1 Mean p-value	Tier 2 Mean p-value	Tier 3 Mean p-value	Tier 4 Mean p-value
			> Tier 3	> Tier 2	> Tier 1	> Tier 2	> Tier 1	> Tier 1				
11	H.DPS.1B1	4	2	3	0	1	1	1	0.50	0.39	0.41	0.41
	H.DPS.1C1	5	1	2	0	2	0	0	0.78	0.47	0.47	0.42
	H.GM.1B1	4	1	2	0	2	0	0	0.66	0.32	0.33	0.27
	H.ME.1A2	4	0	0	0	2	0	0	0.71	0.46	0.43	0.22
	H.ME.1B2	4	1	0	0	2	0	0	0.70	0.48	0.44	0.33
	H.NO.1A1	5	3	1	1	1	1	2	0.52	0.50	0.33	0.37
	H.PRF.1C1	4	0	1	3	1	3	3	0.45	0.52	0.48	0.39
	H.PRF.2B1	4	2	3	0	1	0	1	0.72	0.44	0.36	0.41
	H.PRF.2B2	4	2	2	1	2	0	0	0.68	0.59	0.50	0.55
	H.PRF.2C1	4	3	1	0	1	0	0	0.73	0.48	0.41	0.46

Note: 218 out of 294 Mathematics item families (74.1%) were flagged for at least one tier reversal.

Common Items

Common items sets were not included to allow for common-item linking across forms in English language arts. Minimal numbers of common items were included in mathematics, though the number of common items varied by form and were not representative of the content blueprint. Exhibits 13 through 19 contain a comparison of p-values in Mathematics for items in common between forms. In each grade, some items were administered on more than one test form. Each form contained only a small number of common items, and not all forms may have contained items in common. Further, the common items were not representative of the content specifications of the test forms. However, by comparing the p-values of these items across forms, it was possible to gain some sense of the comparability of student groups' performance taking common items on different Mathematics forms. Each table contains the individual item ID, the unique form on which the item appeared, the number of students taking that item, and the p-value for that item on a given form. In addition, the "Combined p-value" indicates the p-value when including all students from both forms in the p-value calculation.

Exhibit 13. Comparison of p-values Between Forms for Common Items in Mathematics Grade 3

Item	Form	N	*P*- Value	Item	Form	N	*P*- Value	Item	Form	N	*P*- Value
	4	56	0.71		1	102	0.32		6	54	0.76
110828A	7	58	0.69	111387A	8	66	0.38	112555A	8	66	0.74
	Combined	114	0.70		Combined	168	0.35		Combined	120	0.75
	2	63	0.71		1	102	0.55		3	73	0.58
110850A	8	66	0.65	111390A	4	56	0.61	112557A	5	61	0.51
	Combined	129	0.68		Combined	158	0.57		Combined	134	0.54
	3	73	0.41		6	54	0.41		3	73	0.42
110855A	6	54	0.44	111392A	7	58	0.29	112565A	8	66	0.38
	Combined	127	0.43		Combined	112	0.35		Combined	139	0.40
	4	56	0.64		1	102	0.36		4	56	0.39
110869A	5	61	0.69	111407A	3	73	0.37	112566A	7	58	0.34
	Combined	117	0.67		Combined	175	0.37		Combined	114	0.37
	2	63	0.41		4	56	0.55		5	61	0.44
110876A	8	66	0.44	111408A	7	58	0.67	112570A	8	66	0.33
	Combined	129	0.43		Combined	114	0.61		Combined	127	0.39
	2	63	0.71		1	102	0.59		3	73	0.56
110912A	5	61	0.64	111409A	6	54	0.57	112575A	8	66	0.52
	Combined	124	0.68		Combined	156	0.58		Combined	139	0.54
	6	54	0.56		2	63	0.32		3	73	0.36
110919A	7	58	0.48	111410A	7	58	0.50	112580A	6	54	0.33
	Combined	112	0.52		Combined	121	0.40		Combined	127	0.35
	3	73	0.59		5	61	0.46		1	102	0.32
110964A	5	61	0.51	111417A	8	66	0.42	112581A	8	66	0.33
	Combined	134	0.55		Combined	127	0.44		Combined	168	0.33
	4	56	0.52		1	102	0.36		4	56	0.61
111381A	5	61	0.48	111422A	5	61	0.33	120548A	7	58	0.53
	Combined	117	0.50		Combined	163	0.35		Combined	114	0.57
	7	58	0.62		6	54	0.57		5	61	0.52
111382A	8	66	0.62	112552A	7	58	0.62	120932A	6	54	0.57
	Combined	124	0.62		Combined	112	0.60		Combined	115	0.55
	5	61	0.64		3	73	0.32				
111385A	7	58	0.62	112554A	8	66	0.32				
	Combined	119	0.63		Combined	139	0.32				

Exhibit 14. Comparison of p-values Between Forms for Common Items in Mathematics Grade 4

Item	Form	N	*P*- Value	Item	Form	N	*P*- Value	Item	Form	N	*P*- Value
	3	67	0.57		6	62	0.63		3	67	0.25
111135A	5	59	0.47	111728A	7	46	0.76	112837A	6	62	0.18
	Combined	126	0.52		Combined	108	0.69		Combined	129	0.22
	2	45	0.38		3	67	0.42		7	46	0.70
111138A	6	61	0.31	111730A	5	59	0.42	120697A	8	59	0.71
	Combined	106	0.34		Combined	126	0.42		Combined	105	0.70
	5	59	0.25		3	67	0.66		1	126	0.60
111161A	8	59	0.32	112778A	8	59	0.71	121740A	6	62	0.56
	Combined	118	0.29		Combined	126	0.68		Combined	188	0.59
	5	59	0.37		2	45	0.73		4	50	0.38
111179A	8	59	0.41	112786A	6	62	0.63	122153A	7	46	0.50
	Combined	118	0.39		Combined	107	0.67		Combined	96	0.44
	3	67	0.19		4	50	0.38		5	59	0.51
111185A	7	46	0.35	112787A	5	59	0.41	122157A	7	46	0.57
	Combined	113	0.26		Combined	109	0.39		Combined	105	0.53
	3	67	0.69		5	59	0.59		1	126	0.13
111663A	8	59	0.81	112811A	6	62	0.60	122259A	4	50	0.16
	Combined	126	0.75		Combined	121	0.60		Combined	176	0.14
	5	59	0.81		7	46	0.33		3	67	0.51
111672A	6	62	0.77	112813A	8	59	0.39	122267A	5	59	0.56
	Combined	121	0.79		Combined	105	0.36		Combined	126	0.53
	5	59	0.39		5	58	0.43		3	67	0.31
111687A	6	62	0.47	112819A	8	59	0.25	122370A	7	46	0.30
	Combined	121	0.43		Combined	117	0.34		Combined	113	0.31
	4	50	0.46		5	59	0.36		1	126	0.15
111688A	8	59	0.42	112833A	6	62	0.35	122393A	8	59	0.17
	Combined	109	0.44		Combined	121	0.36		Combined	185	0.16
	1	126	0.30		4	50	0.36		3	67	0.09
111691A	8	59	0.51	112834A	8	59	0.22	122424A	7	46	0.13
	Combined	185	0.37		Combined	109	0.28		Combined	113	0.11
	2	45	0.27		4	50	0.74				
111692A	4	50	0.36	112836A	8	59	0.76				
	Combined	95	0.32		Combined	109	0.75				

Exhibit 15. Comparison of p-values Between Forms for Common Items in Mathematics Grade 5

Item	Form	N	*P*- Value	Item	Form	N	*P*- Value	Item	Form	N	*P*- Value
	1	125	0.35		3	50	0.32		1	125	0.32
111228A	7	44	0.36	112391A	8	65	0.48	113876A	6	93	0.25
	Combined	169	0.36		Combined	115	0.41		Combined	218	0.29
	2	66	0.17		1	125	0.26		2	66	0.27
111238A	5	44	0.20	112392A	7	44	0.11	113877A	7	44	0.48
	Combined	110	0.18		Combined	169	0.22		Combined	110	0.35
	5	44	0.66		1	125	0.60		2	66	0.30
111262A	7	44	0.57	112401A	8	65	0.57	113881A	7	44	0.39
	Combined	88	0.61		Combined	190	0.59		Combined	110	0.34
	3	50	0.38		2	66	0.39		3	50	0.30
111263A	4	63	0.44	112402A	6	93	0.44	113882A	4	63	0.37
	Combined	113	0.42		Combined	159	0.42		Combined	113	0.34
	3	49	0.73		3	50	0.42		1	125	0.40
111268A	5	44	0.77	112415A	5	44	0.41	113898A	7	44	0.48
	Combined	93	0.75		Combined	94	0.41		Combined	169	0.42
	2	66	0.24		1	125	0.53		4	63	0.75
111300A	8	65	0.51	113834A	7	44	0.70	113899A	6	93	0.75
	Combined	131	0.37		Combined	169	0.57		Combined	156	0.75
	2	66	0.27		2	66	0.73		3	49	0.29
111304A	8	65	0.45	113844A	4	63	0.83	113901A	7	44	0.25
	Combined	131	0.36		Combined	129	0.78		Combined	93	0.27
	3	50	0.40		1	125	0.44		4	63	0.38
111305A	6	93	0.29	113856A	7	44	0.43	113902A	6	93	0.24
	Combined	143	0.33		Combined	169	0.44		Combined	156	0.29
	1	125	0.33		2	66	0.73		2	66	0.21
112358A	6	93	0.25	113859A	3	50	0.84	120736A	8	65	0.18
	Combined	218	0.29		Combined	116	0.78		Combined	131	0.20
	4	63	0.21		1	125	0.39		1	125	0.22
112360A	5	44	0.32	113863A	6	93	0.54	120739A	2	66	0.14
	Combined	107	0.25		Combined	218	0.45		Combined	191	0.19
	6	93	0.19		5	44	0.73				
112364A	8	65	0.29	113864A	8	65	0.72				
	Combined	158	0.23		Combined	109	0.72				

Exhibit 16. Comparison of p-values Between Forms for Common Items in Mathematics Grade 6

Item	Form	N	*P*- Value	Item	Form	N	*P*- Value	Item	Form	N	*P*- Value
	1	54	0.65		2	46	0.65		2	46	0.37
110878A	7	66	0.70	111025A	6	68	0.63	111503A	6	68	0.50
	Combined	120	0.68		Combined	114	0.64		Combined	114	0.45
	2	46	0.30		2	46	0.57		4	66	0.65
110887A	8	63	0.41	111440A	5	118	0.66	111506A	7	66	0.53
	Combined	109	0.37		Combined	164	0.63		Combined	132	0.59
	1	54	0.83		3	62	0.42		3	62	0.58
110889A	8	63	0.86	111443A	7	66	0.39	111508A	6	68	0.65
	Combined	117	0.85		Combined	128	0.41		Combined	130	0.62
	4	66	0.73		1	54	0.39		2	46	0.43
110893A	7	66	0.61	111445A	5	118	0.43	112633A	8	63	0.49
	Combined	132	0.67		Combined	172	0.42		Combined	109	0.47
	3	62	0.60		3	62	0.40		5	118	0.39
110905A	7	66	0.58	111446A	7	66	0.29	112662A	7	66	0.36
	Combined	128	0.59		Combined	128	0.34		Combined	184	0.38
	2	46	0.52		7	66	0.27		1	54	0.37
110907A	5	118	0.42	111452A	8	63	0.37	112669A	7	66	0.26
	Combined	164	0.45		Combined	129	0.32		Combined	120	0.31
	3	62	0.58		1	54	0.48		2	46	0.83
110938A	7	66	0.53	111453A	4	66	0.35	112673A	7	66	0.52
	Combined	128	0.55		Combined	120	0.41		Combined	112	0.64
	4	66	0.35		4	66	0.79		3	62	0.48
110942A	7	66	0.48	111454A	6	68	0.78	112674A	6	68	0.60
	Combined	132	0.42		Combined	134	0.78		Combined	130	0.55
	7	66	0.52		2	46	0.37		3	62	0.69
110986A	8	63	0.71	111457A	7	66	0.38	112677A	4	66	0.61
	Combined	129	0.61		Combined	112	0.38		Combined	128	0.65
	1	54	0.46		4	66	0.30		2	46	0.65
111022A	5	118	0.31	111463A	5	118	0.20	112679A	5	118	0.53
	Combined	172	0.36		Combined	184	0.24		Combined	164	0.57
	2	46	0.33		2	46	0.57				
111023A	8	63	0.38	111466A	5	118	0.50				
	Combined	109	0.36		Combined	164	0.52				

Exhibit 17. Comparison of p-values Between Forms for Common Items in Mathematics Grade 7

Item	Form	N	*P*- Value	Item	Form	N	*P*- Value	Item	Form	N	*P*- Value
	1	70	0.87		2	79	0.84		1	70	0.30
111046A	3	81	0.78	111733A	8	57	0.79	111796A	8	57	0.49
	Combined	151	0.82		Combined	136	0.82		Combined	127	0.39
	2	79	0.56		1	70	0.53		1	70	0.57
111048A	6	64	0.42	111734A	5	53	0.57	111805A	5	53	0.62
	Combined	143	0.50		Combined	123	0.54		Combined	123	0.59
	1	70	0.56		7	63	0.52		3	81	0.70
111055A	7	63	0.56	111739A	8	57	0.54	112847A	5	53	0.66
	Combined	133	0.56		Combined	120	0.53		Combined	134	0.69
	5	53	0.72		1	70	0.29		3	81	0.75
111074A	6	64	0.69	111741A	4	72	0.28	112864A	7	63	0.78
	Combined	117	0.70		Combined	142	0.28		Combined	144	0.76
	2	79	0.56		2	79	0.62		2	79	0.41
111075A	3	81	0.62	111744A	6	64	0.47	112866A	4	72	0.35
	Combined	160	0.59		Combined	143	0.55		Combined	151	0.38
	2	79	0.51		4	72	0.19		3	81	0.41
111099A	6	64	0.50	111751A	6	64	0.22	112891A	8	57	0.39
	Combined	143	0.50		Combined	136	0.21		Combined	138	0.40
	4	72	0.43		1	70	0.40		6	64	0.42
111101A	8	57	0.42	111775A	6	64	0.39	112894A	8	57	0.53
	Combined	129	0.43		Combined	134	0.40		Combined	121	0.47
	3	81	0.22		4	72	0.28		1	70	0.43
111106A	8	57	0.28	111776A	5	53	0.30	112895A	5	53	0.34
	Combined	138	0.25		Combined	125	0.29		Combined	123	0.39
	4	72	0.64		1	70	0.51		5	53	0.38
111117A	6	64	0.50	111779A	7	63	0.63	112910A	7	63	0.41
	Combined	136	0.57		Combined	133	0.57		Combined	116	0.40
	2	79	0.43		3	81	0.30		2	79	0.49
111120A	5	53	0.51	111781A	6	64	0.31	112911A	6	64	0.41
	Combined	132	0.46		Combined	145	0.30		Combined	143	0.45
	2	79	0.47		2	79	0.56				
111127A	6	64	0.42	111793A	7	63	0.44				
	Combined	143	0.45		Combined	142	0.51				

Exhibit 18. Comparison of p-values Between Forms for Common Items in Mathematics Grade 8

Item	Form	N	*P*- Value	Item	Form	N	*P*- Value	Item	Form	N	*P*- Value

	3	44	0.50		1	92	0.48		6	103	0.73
111280A	5	73	0.38	112462A	5	73	0.33	113921A	7	68	0.66
	Combined	117	0.43		Combined	165	0.41		Combined	171	0.70
	1	92	0.52		3	44	0.30		3	44	0.43
111281A	7	68	0.60	112465A	6	103	0.35	113922A	8	62	0.50
	Combined	160	0.56		Combined	147	0.33		Combined	106	0.47
	6	103	0.39		1	92	0.50		4	67	0.40
111331A	7	68	0.37	112466A	8	62	0.47	113933A	5	73	0.45
	Combined	171	0.38		Combined	154	0.49		Combined	140	0.43
	3	44	0.50		2	46	0.37		3	44	0.41
111332A	8	62	0.61	112481A	7	68	0.34	113937A	5	73	0.30
	Combined	106	0.57		Combined	114	0.35		Combined	117	0.34
	2	46	0.72		5	73	0.30		4	67	0.40
111352A	5	73	0.78	112482A	8	62	0.34	113941A	7	68	0.53
	Combined	119	0.76		Combined	135	0.32		Combined	135	0.47
	6	103	0.50		5	73	0.27		1	92	0.52
111353A	7	68	0.68	112486A	8	62	0.37	113942A	3	44	0.41
	Combined	171	0.57		Combined	135	0.32		Combined	136	0.49
	3	44	0.30		3	44	0.45		1	92	0.43
111583A	6	103	0.39	112487A	6	103	0.42	113952A	4	67	0.49
	Combined	147	0.36		Combined	147	0.43		Combined	159	0.46
	4	67	0.75		5	73	0.62		3	44	0.48
111591A	5	73	0.77	112502A	6	103	0.61	113959A	8	62	0.37
	Combined	140	0.76		Combined	176	0.61		Combined	106	0.42
	2	46	0.41		1	92	0.66		3	44	0.45
111594A	8	62	0.27	112504A	6	103	0.75	117072A	8	62	0.40
	Combined	108	0.33		Combined	195	0.71		Combined	106	0.42
	4	67	0.57		2	46	0.43		4	67	0.39
111596A	7	68	0.63	112505A	7	68	0.56	117075A	7	68	0.34
	Combined	135	0.60		Combined	114	0.51		Combined	135	0.36
	6	103	0.78		5	73	0.64				
112459A	8	62	0.85	112509A	7	68	0.57				
	Combined	165	0.81		Combined	141	0.61				

Exhibit 19. Comparison of p-values Between Forms for Common Items in Mathematics Grade 11

Item	Form	N	*P*- Value	Item	Form	N	*P*- Value	Item	Form	N	*P*- Value
110841A	5	69	0.43	111544A	2	30	0.23	112724A	6	74	0.50

	8	71	0.46		6	74	0.32		8	71	0.66
	Combined	140	0.45		Combined	104	0.30		Combined	145	0.58
	2	30	0.43		5	70	0.59		4	31	0.77
110847A	8	71	0.48	111546A	6	74	0.77	112726A	6	74	0.70
	Combined	101	0.47		Combined	144	0.68		Combined	105	0.72
	1	86	0.42		3	43	0.30		2	30	0.67
110852A	6	74	0.32	111547A	8	71	0.28	112727A	3	43	0.56
	Combined	160	0.38		Combined	114	0.29		Combined	73	0.60
	3	43	0.53		6	74	0.28		4	31	0.48
110882A	5	69	0.42	111819A	8	71	0.37	112732A	6	74	0.49
	Combined	112	0.46		Combined	145	0.32		Combined	105	0.49
	4	31	0.42		1	86	0.60		1	86	0.57
111000A	5	69	0.46	111823A	5	68	0.50	112733A	5	69	0.32
	Combined	100	0.45		Combined	154	0.56		Combined	155	0.46
	2	30	0.53		4	31	0.52		2	30	0.40
111002A	6	74	0.55	111824A	8	71	0.59	112742A	4	31	0.42
	Combined	104	0.55		Combined	102	0.57		Combined	61	0.41
	1	86	0.21		4	31	0.90		1	86	0.62
111107A	4	31	0.26	111832A	8	71	0.80	112744A	7	62	0.48
	Combined	117	0.22		Combined	102	0.83		Combined	148	0.56
	5	68	0.35		5	68	0.38		2	30	0.40
111109A	7	62	0.42	111835A	7	62	0.45	112927A	7	62	0.27
	Combined	130	0.38		Combined	130	0.42		Combined	92	0.32
	4	31	0.39		5	70	0.29		1	86	0.37
111534A	5	68	0.24	112712A	6	74	0.46	112930A	8	71	0.30
	Combined	99	0.28		Combined	144	0.38		Combined	157	0.34
	1	86	0.78		7	62	0.50		4	31	0.52
111536A	6	74	0.74	112716A	8	71	0.75	112931A	5	69	0.48
	Combined	160	0.76		Combined	133	0.63		Combined	100	0.49
	1	86	0.66		2	30	0.67				
111542A	3	43	0.53	112721A	5	70	0.73				
	Combined	129	0.62		Combined	100	0.71				