IOWA STATEWIDE ASSESSMENT OF STUDENT PROGRESS

Interpreting Mathematics Performance During the Pandemic: An Item-Level Perspective

Iowa Testing Programs
Prepared by Catherine Welch, Stephen Dunbar, Xi Wang and Jun Hee Park

January 2022

Interpreting Mathematics Performance During the Pandemic

Among the many challenges local and state education agencies (LEAs and SEAs) have faced during the nation's public health crisis, large-scale summative assessment for accountability might be ranked quite low from many social and political perspectives. Yet when schools reopened to operate a variety of models for teaching and learning, understanding how the pandemic affected student progress became a growing concern. Research on learning loss and recovery, although provocative, has been arguably inconclusive in the sense that findings have ranged widely (Camara, 2020). Some estimates of achievement have led researchers to infer that a full recovery could take years, whereas other analysts have presented evidence of progress similar to what would be expected during a normal school year (Center for Research on Education Outcomes, 2020; Lewis, Kuhfeld, Ruzek & McEachin, 2021; Renaissance Learning, 2021; Schwartz, 2021; Betebenner & Wenning, 2021; Lorié, 2020). A more consistent concern about learning loss and recovery has been evidence of the markedly greater impact of COVID-19 among communities of color and the widening of achievement gaps (Calfas, 2021). Researchers and policymakers are awaiting more comprehensive data to better understand the underlying realities of teaching and learning during a public health crisis.

The purpose of this paper is to summarize and interpret item-level census data from the state of Iowa on student progress in the context of a statewide, summative assessment program in mathematics. A central argument of the paper is that systematic administration of a peer-reviewed statewide assessment, aligned to state standards for purposes of accountability should provide results of interest to instruction and learning in a standards-based framework for interpretation and use.

With the abrupt interruption of instruction caused by the COVID-19 pandemic in March of 2020, local educators and policymakers quickly suspended regular school activities that were not considered essential in making the transition to remote learning, including large-scale, summative assessment. Accountability mandates were placed on hold, and waivers of federal and state testing requirements were granted to schools as a matter of course. As teachers and students adapted remote learning, LEA and SEA policymakers began to speculate on the degree of learning loss having limited information from local assessments. Academic planning for the 2020-21 school year was complicated by the uncertainties associated with the pandemic.

At the close of the 2020 legislative session in Iowa, lawmakers passed Senate File 2310, which stipulated there would be no waiver of the statewide summative assessment in the 2021-22 school year. All students in public and accredited private schools would take the Iowa Statewide Assessment of Student Progress (ISASP) as lawmakers prioritized large scale assessment as a means of understanding the possible achievement correlates of the public health and social upheaval experienced in local education communities. As schools grappled over the summer with transition plans for in-person, hybrid, and remote learning designs for instruction, it became clear that to satisfy the requirements of Senate File 2310, flexibility in the delivery of the ISASP was necessary, and a remote testing option was developed and offered. Three percent of the student population in Iowa completed the test remotely in the spring of 2021.

How worthwhile the results of the state assessments from 2021 are depends to a large extent on the approaches and strategies employed by states. The tests can provide a yardstick of student performance against a defined set of curricular standards and instructional emphasis during COVID. It was Iowa's approach to provide information to help evaluate the achievement and progress that has been made between the spring of 2019 and 2021, and to do so with an

assessment that had been validated by the federal peer review process. Users of the information were encouraged to view the results of the 2021 ISASP administration to help prepare students and teachers to move forward in ways that are most beneficial for the students.

With the input of stakeholders, the department of education and school districts, a framework for 2021 spring testing was developed and shared within the state. Critical to school districts seeing value in this testing, several key attributes of the test and administration were emphasized to help convey the integrity of the information that was produced. First, the test specifications were consistent with previous years. This information was available and known to all test takers and audiences. Extensive practice test items and materials were also available to all test takers. In addition, all items and related testing materials were reviewed for sensitivity due to COVID-related issues. For example, reading and science materials that may have referenced viruses or quarantines were eliminated from the available item banks. To aid in interpretation, proficiency levels remained consistent with previous years, including the thresholds for proficiency and definitions of performance levels. Parents were provided with proficiency indicators, growth metrics and domain-level information to better interpret student performance. As a result of the framework and the support of local school districts, participation rates for the state were above 95% for all grade levels and content areas, higher than anywhere else in the United States.

Spring 2021 statewide assessment data provided insights into state performance and helped districts move forward. Overall, state performance in mathematics at the test level did differ between 2019 and 2021 with fewer students being identified as proficient in all grades. Table 1 provides two indicators of this change. The percent of students that were identified as proficient decreased between 2019 and 2021 with differences that ranged from 2% to 8%. The

average scale score decreased between years, following a similar pattern to that of the proficiency indicators.

Table 1. Changes in Mathematics Performance by Grade between 2019 and 2021

	Percent Proficient		Average Scale Score	
Grade	**2019**	**2021**	**2019**	**2021**
5	69	63	454.1	448.6
6	69	67	476.0	473.2
7	70	64	499.9	492.6
8	72	69	526.5	521.0
9	69	61	544.8	536.6
10	66	64	567.9	562.9
11	66	65	594.1	593.3

Between March 2020 and spring 2021, districts in Iowa varied with respect to the opportunities to learn that were provided, the learning conditions available, and the level of pandemic-related disruptions experienced. Recognizing that there were overall changes in performance at the test level, the purpose of this report is to identify patterns in performance at the content (Iowa Core domain) and depth-of-knowledge (DOK) levels that may help users better understand the impact of these variations and associated curriculum modifications. The unit of analysis for this report is at the item level and the Iowa Core domains and DOK areas associated with each item. The design and collection of data allowed for direct comparisons between pre- and post-pandemic item-level statistics. Identifying the areas of greatest loss over the past two years can be used to inform instruction moving forward.

Data

The ISASP mathematics test was administered to students in the state of Iowa in the spring of 2019. Tests were aligned to the Iowa Core and assembled to predetermined test

specifications. The number of items per content domain and depth of knowledge (DOK) category are provided in Tables 2 and 3. Three levels of DOK are measured by ISASP. These include Essential Competencies (DOK 1) which involves recalling information such as facts, definitions, terms, or simple one-step procedures. Conceptual Understanding (DOK 2) which requires engaging in some cognitive processing beyond recalling or reproducing a response. A conceptual understanding item requires students to make some decisions as to how to approach the problem or activity and may require them to employ more than a single step. Extended Reasoning (DOK 3) requires problem solving, planning, and/or using evidence. These items require students to develop a strategy to connect and relate ideas in order to solve the problem, and the problem may require that the student use multiple steps and draw upon a variety of skills. Five content domains are measured at each grade level. Additional definitions for the Iowa Core domains and DOK can be found at: Iowa | Research (pearsonaccess.com).

Table 2. Number of Items by Iowa Core Domain by Grade in 2019 and 2021

				Grade			
Domain	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Number and Operations in Base Ten (NBT)	10						
Number and Operations – Fractions (NF)	10						
Operations and Algebraic Thinking (OA)	6						
Geometry (G)	6	6	8	9	5	10	8
Measurement and Data (MD)	8						
The Number System (NS)		10	10	5	7	7	7
Ratios and Proportional Relationships (RP)		8	9				
Expressions and Equations (EE)		13	11	15			
Statistics and Probability (SP)		5	7	9	5	6	5
Functions (F)				9	7	5	7
Algebra (A)					11	7	8

Table 3. Number of Items by Depth of Knowledge Classification by Grade in 2019 and 2021

				Grade			
Depth of Knowledge	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Essential competencies (DOK1)	12	15	8	15	8	10	9
Conceptual understanding (DOK2)	25	23	29	25	24	19	22
Extended reasoning (DOK3)	3	4	7	7	3	6	4

As part of the 2019 administration, embedded items were field tested in the operational forms. The item sets were spiraled among all test takers to create randomly equivalent groups of test takers. Using the results of the 2019 field test design, a parallel form of the test was assembled at each grade level for administration in 2021. The number of items per content domain and DOK category as reported in Tables 2 and 3 were matched. Approximately ten items per grade from the 2019 forms were identified as anchor items and included in the 2021 forms as a check on consistency and stability going forward. The anchor items were selected to be representative in terms of content coverage and DOK categories. All items appearing in the 2021 forms had item-level statistics from the 2019 administration, and those statistics are reported in the analyses that follow.

Students who completed the ISASP in both 2019 and 2021 were included in this analysis. Matching records for 2021 and 2019 ISASP was done so that only students who completed the tests in a proctored environment were included. All students in the 2021 and 2019 datasets had verified state identification numbers. Students were removed from the matched datasets if they did not complete the Mathematics test in both years. This process resulted in the final matched datasets including approximately 90% of all student records in grades 5 through 11 in 2021 matched to their records in grades 3 through 9 in 2019.

Results

Results for two item-level statistics are provided in this report. The first compares item-level p-values from 2019 to 2021. The p-value represents the proportion of test takers that answered the item correctly. The p-value has a floor of .00 and a ceiling of 1.00. Using the p-value, comparisons were made for both the depth-of-knowledge classifications as well as content domains. The second item statistic is the point-biserial correlation, which is an indication of how well an item differentiates among test takers. As each item on the test is aligned with the Iowa Core, item responses are expected to correlate with the total test score. The point-biserial is a product-moment correlation that can be positive or negative and ranges between .00 and 1.00. Items with higher coefficients are indicators of a stronger relationship with the overall test score. In addition to these two item statistics, a- and b-parameter estimates from the two-parameter logistic (2PL) model also were obtained for each item. The results for the a- and b-parameter estimates were very similar to those of the p-value and the point-biserial differences reported here.

Table 4 summarizes the distributions of p-values (PV 19 and PV 21) and point-biserial correlation (PBIS 19 and PBIS 21) for the 2019 and 2021 test administrations. In all grades, the p-values in 2019 were greater on average than the p-values in 2021, meaning that the items were more difficult in the assessment year following COVID-19. Differences ranged from about 1 to 5 percentage points depending on grade and level of difficulty, with the larger mean differences tending to be observed in grades 5 to 8. In terms of enrollment numbers in a typical grade in Iowa, anywhere from about 300 to 1800 fewer students answered the average math item correctly after the pandemic than before the pandemic. With respect to relations between items and total scores, math items in all grades were not as closely related to total scores after the pandemic. Possible implications of this finding are discussed later in the report.

Table 4. Mean and Standard Deviation (SD) of ISASP Item Statistics

Grade	Statistic	PV21	PBIS21	PV19	PBIS19
5	Mean	.54	.36	.58	.42
	SD	.16	.09	.14	.10
6	Mean	.56	.34	.58	.40
	SD	.20	.11	.19	.11
7	Mean	.53	.32	.56	.38
	SD	.17	.10	.15	.09
8	Mean	.51	.34	.55	.40
	SD	.17	.11	.16	.10
9	Mean	.49	.33	.52	.40
	SD	.16	.11	.15	.11
10	Mean	.49	.34	.50	.41
	SD	.19	.12	.17	.10
11	Mean	.52	.41	.53	.47
	SD	.15	.11	.13	.10

To determine whether p-value and point-biserial differences varied as a function of item difficulty, items were grouped into easy, medium and difficult categories. Easy items were those with p-values greater than .70, medium were items with p-values between .40 and .70, and difficult were items with p-values less than .40. Tables 5 and 6 summarize mean differences for the two item statistics of interest. In all but one grade, mean differences were larger for difficult items than easy items, which could suggest that instruction during the pandemic might have emphasized more straightforward aspects of mathematical thinking and problem-solving as reflected in the Iowa Core Standards. It should be noted, however, that the differences between p-values for difficult and easy items were no greater than 4 percent in any grade. No interaction between point-biserial differences in 2019 and 2021 and item difficulty was observed in these comparisons.

Table 5. Mean Differences between ISASP Item Difficulty Statistics (2021 minus 2019) for Easy, Medium and Difficult Items

Grade	Easy	Medium	Difficult
5	-.02	-.04	-.05
6	.00	-.03	-.03
7	-.01	-.03	-.05
8	-.03	-.04	-.02
9	-.04	-.02	-.05
10	.02	-.01	-.03
11	.01	-.01	-.01
All Grades	-.01	-.03	-.03

Note: Easy (p-value > .70), Medium (.40 < p-value < .70), Difficult (p-value < .40)

Table 6. Mean Differences between ISASP Item Discrimination Statistics (2021 minus 2019) for Easy, Medium and Difficult Items

Grade	Easy	Medium	Difficult
5	-.08	-.05	-.05
6	-.03	-.09	-.05
7	-.08	-.05	-.06
8	-.06	-.06	-.04
9	-.07	-.07	-.05
10	-.10	-.08	-.06
11	-.07	-.06	.03
All Grades	-.06	-.06	-.05

Note: Easy (p-value > .70), Medium (.40 < p-value < .70), Difficult (p-value < .40)

Differences between items statistics in 2019 and 2021 were further examined to determine whether there were any patterns that could be explained by depth-of knowledge (DOK) level or content domain in the Iowa Core. Tables 7 and 8 provide the mean p-values and point-biserial correlations in 2019 and 2021 in terms of DOK level and Iowa Core Domain for all grades, with results in grades 5 and 8 in bold to support the more detailed graphical display, analysis, and discussion that follows. These two grades were selected because the content domains represented are different and represent a transition point in the mathematics curriculum.

Similar results are displayed graphically for all additional grade levels in the Appendix to this document.

Table 7. Mean Item Statistics by DOK Level in Grades 5 through 11

Grade	DOK Level	PV 21	PBIS 21	PV 19	PBIS 19
	1	**.65**	**.32**	**.68**	**.38**
5	**2**	**.50**	**.38**	**.55**	**.43**
	3	**.33**	**.36**	**.37**	**.40**
	1	.66	.34	.67	.40
6	2	.52	.35	.54	.41
	3	.46	.31	.48	.38
	1	.64	.26	.66	.33
7	2	.55	.36	.57	.41
	3	.37	.27	.42	.34
	1	**.63**	**.34**	**.65**	**.40**
8	**2**	**.49**	**.34**	**.52**	**.40**
	3	**.37**	**.32**	**.43**	**.37**
	1	.57	.31	.61	.38
9	2	.47	.35	.50	.41
	3	.47	.26	.49	.31
	1	.54	.38	.54	.45
10	2	.46	.32	.48	.41
	3	.49	.30	.51	.35
	1	.59	.38	.60	.46
11	2	.51	.44	.51	.49
	3	.43	.34	.43	.41

Table 8. Mean Item Statistics by Iowa Core Domain in Grades 5 through 11

Grade	DOMAIN	PV 21	PBIS 21	PV 19	PBIS 19
	G	**.53**	**.32**	**.57**	**.38**
	MD	**.49**	**.38**	**.54**	**.42**
5	**NBT**	**.58**	**.36**	**.61**	**.41**
	NF	**.52**	**.38**	**.57**	**.47**
	OA	**.61**	**.35**	**.64**	**.38**
	EE	.54	.33	.57	.38
	G	.44	.29	.46	.35
6	NS	.64	.37	.64	.43
	RP	.67	.40	.64	.46
	SP	.47	.30	.52	.13
	EE	.50	.35	.54	.41
	G	.46	.29	.46	.33
7	NS	.55	.33	.58	.39
	RP	.57	.30	.62	.40
	SP	.58	.34	.58	.37
	EE	**.51**	**.37**	**.54**	**.41**
	F	**.49**	**.30**	**.53**	**.39**
8	**G**	**.51**	**.29**	**.53**	**.35**
	NS	**.45**	**.35**	**.52**	**.45**
	SP	**.58**	**.35**	**.60**	**.39**
	A	.51	.36	.54	.45
	F	.39	.32	.46	.41
9	G	.57	.33	.60	.36
	NS	.47	.33	.46	.33
	S	.56	.32	.60	.39
	A	.57	.40	.57	.49
	F	.44	.33	.46	.44
10	G	.53	.36	.52	.40
	N	.42	.31	.44	.40
	S	.44	.27	.48	.32
	A	.57	.45	.59	.48
	F	.53	.48	.53	.54
11	G	.52	.37	.51	.41
	NS	.46	.41	.47	.47
	S	.51	.35	.52	.47

Figure 1 shows the comparison of item difficulty between 2019 and 2021 for Grade 5 for the three DOK categories. The x-axis represents the p-value for each item based on the

administration in 2019 and the y-axis represents the p-value for the same items based on the administration in 2021. The numbers represent the item location in the 2021 test. Figure 1 shows a strong positive relationship between p-values for the 2019 and 2021 tests (r = .97), indicating that the rank ordering of p-value between the two years was very consistent. However, items below the diagonal line were easier when administered in 2019 and more difficult in 2021 as indicated previously. On the other hand, items above the line were easier in 2021 and more difficult in 2019. In grade 5, most items were the same difficulty or easier in the 2019 test. DOK1 items (items 1, 12 and 20) were an exception to this trend. Item 2 was an example of an item that showed no difference in item difficulty between the two years.

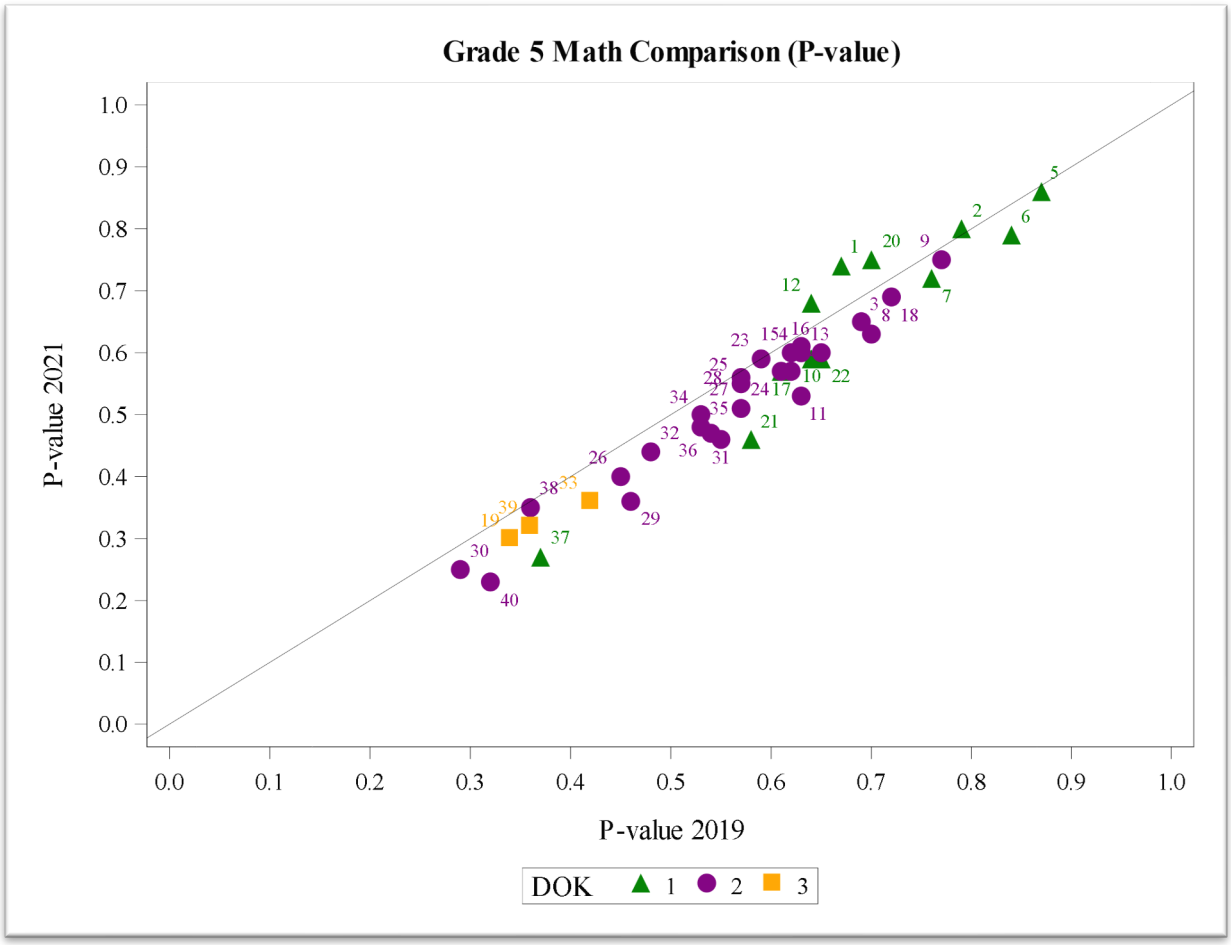


Figure 1. Scatterplot of Grade 5 Item Difficulty by DOK Level in 2019 and 2021

Figure 2 provides item difficulty values based on the Iowa Core domains, with same x- and y- axis values as in Figure 1. In grade 5, Iowa Core domains include Operations and Algebraic Thinking (OA), Number and Operations in Base Ten (NBT), Number and Operations – Fractions (NF), Measurement and Data (MD), and Geometry (G). All MD and OA items were below the diagonal line, indicating students in 2021 had more difficulty with MD and OA items than in 2019. Moreover, except for a few items, most of the G, NF, and NBT items were more difficult for 2021 test-takers than they were for 2019 test takers.

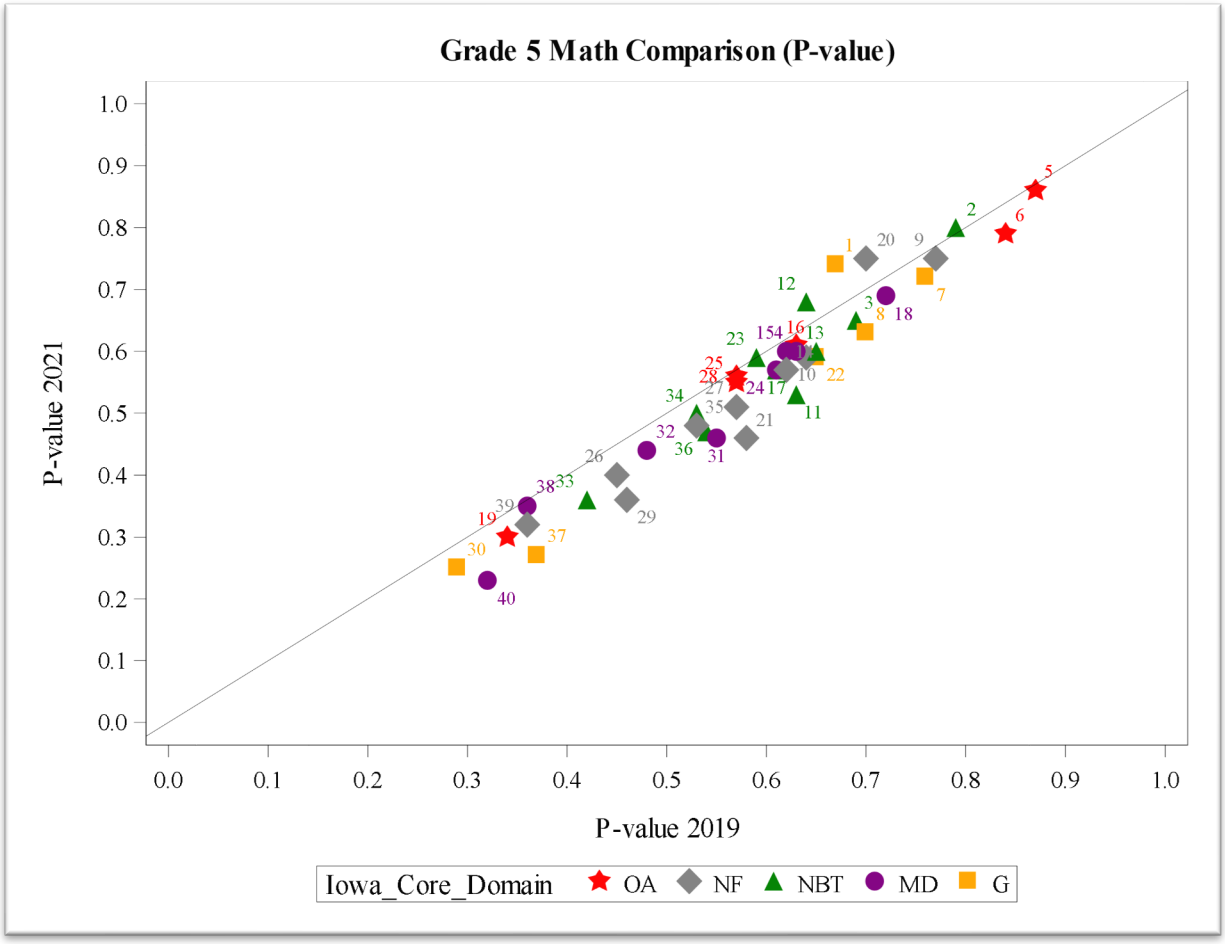


Figure 2. Scatterplot of Grade 5 Item Difficulty by Iowa Core Domain in 2019 and 2021

Figure 3 provides a comparison of point-biserial correlations between 2019 and 2021 for the DOK categories. The x-axis (ranging from .0 to 1.0) represents the point-biserial correlation

for each item in 2019, and the y-axis likewise represents the point-biserial correlation for each item in 2021. The correlation between the two sets of statistics is strong (r = .73) although there is more scatter, as expected, in these values. Items that fall below the diagonal line had higher correlations with total math score in 2019 than they did in 2021, which was the predominant finding with respect to the means presented previously. All items of DOK1 are below the line, indicating that these items were stronger differentiators of performance in 2019 than they were in 2021.

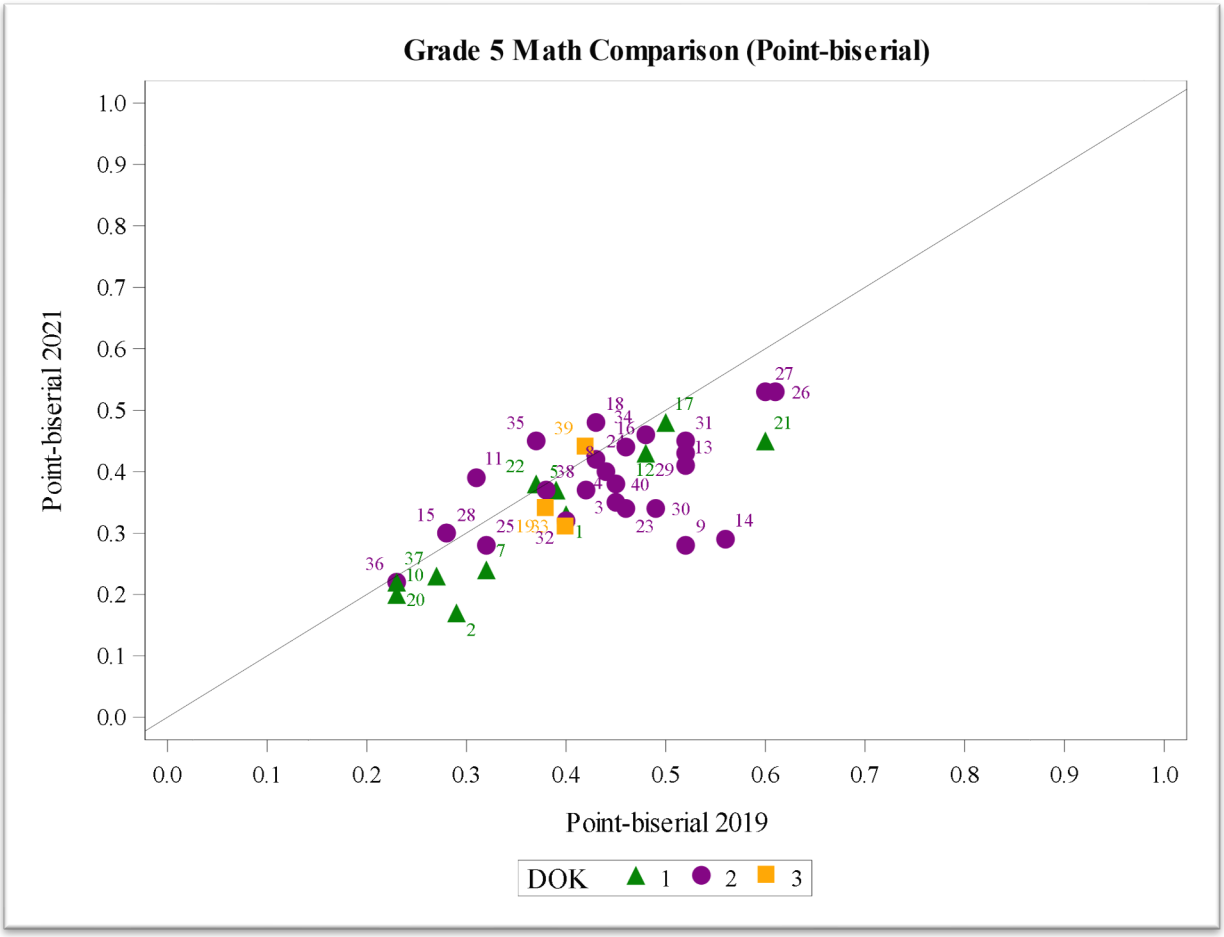


Figure 3. Scatterplot of Grade 5 Item-Total Correlations by DOK Level in 2019 and 2021

Figure 4 illustrates the same results for the Iowa Core domain categories in grade 5. Items measuring NBT, G, and OA discriminated better in 2019 than 2021, with the exception of one NBT item (# 11), and one OA item (# 28). Moreover, most MD and NF items were more highly correlated with total math score in 2019 rather than in 2021.

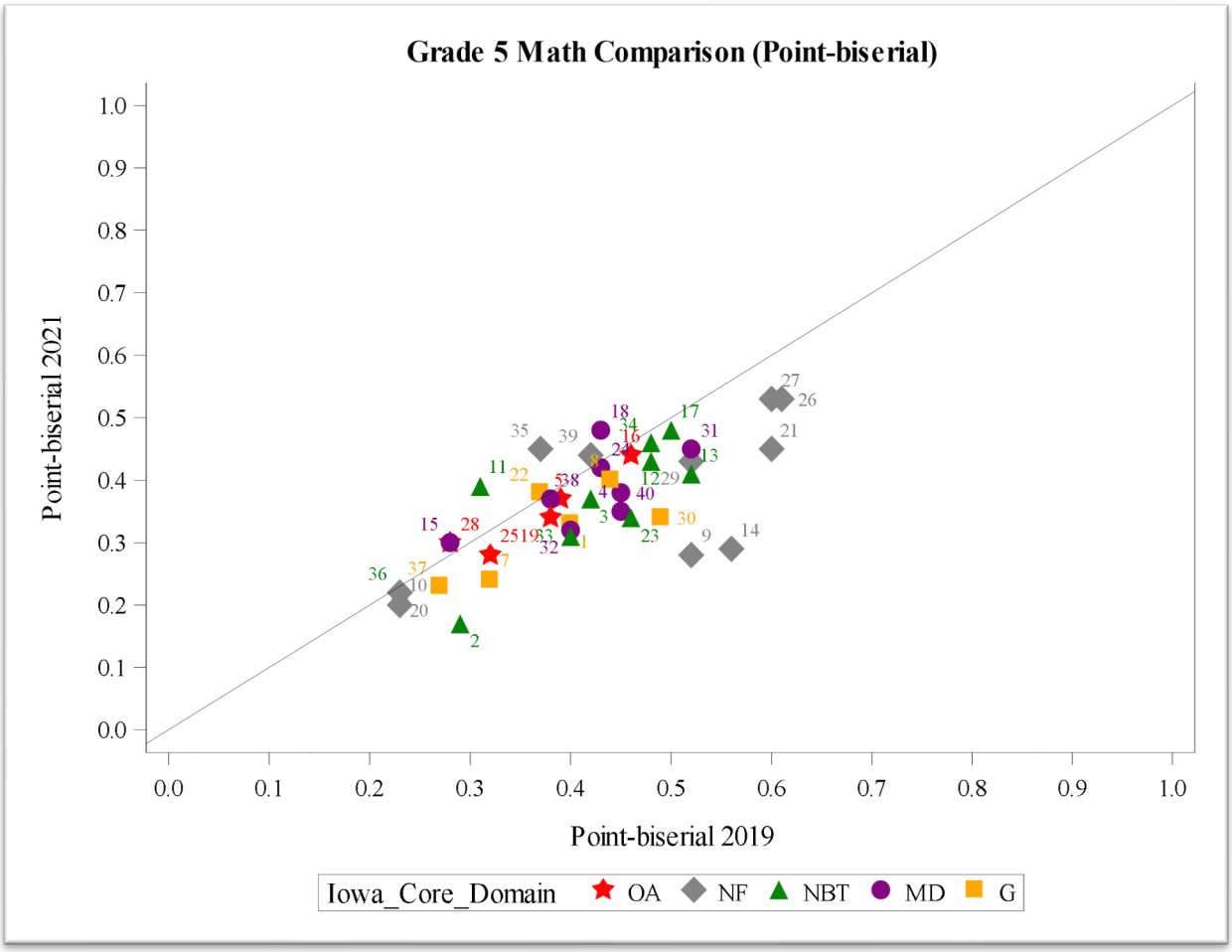


Figure 4. Scatterplot of Grade 5 Item-Total Correlations by Iowa Core Domain in 2019 and 2021

Figures 5 through 8 provide for grade 8 parallel information to that in Figures 1 to 4 for grade 5. The definitions of the DOK levels in Figures 5 and 7 for grade 8 are consistent with the grade 5 definitions. The correlation of the p-values between 2019 and 2021 is .95, again indicating a strong linear relationship. Consistent with the results for grade 5, most items were easier in 2019 than in 2021. With respect to DOK, most DOK3 items were more difficult than

DOK1 and DOK2 items, as perhaps expected, but changed in difficulty between 2019 and 2021 appeared to be idiosyncratic with respect to DOK. Despite some outliers such as item 36, no discernable patterns were observed to indicate, for example, that student responses to more cognitively complex items were adversely affected by the pandemic.

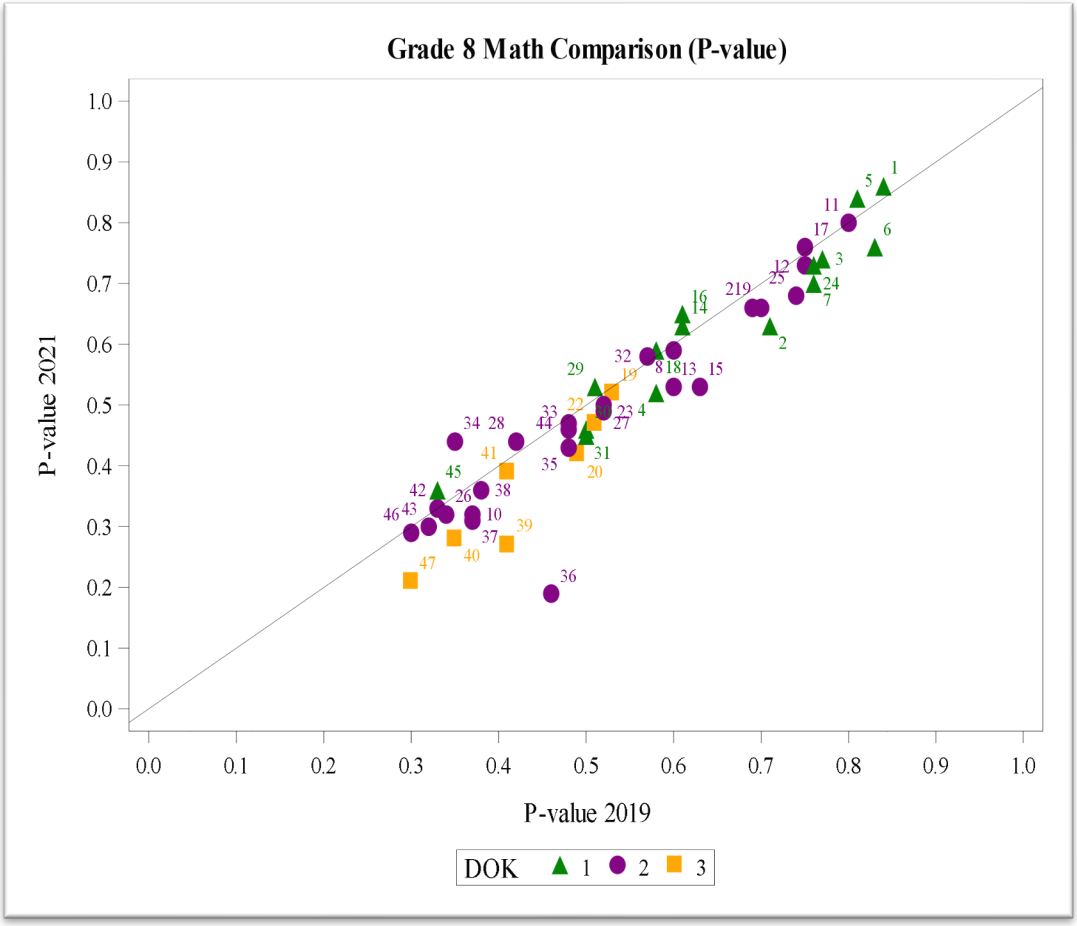


Figure 5. Scatterplot of Grade 8 Item Difficulty by DOK Level in 2019 and 2021

Figure 6 provides a comparison of the p-values between 2019 and 2021 for grade 8 in terms of Iowa Core domains. In grade 8, there are five core domains: Geometry (G), Expressions and Equations (EE), Functions (F), Number System (NS) and Statistics and Probability (SP). Each domain requires different mathematical skills to solve problems. With respect to Iowa Core domain, despite some outliers such as item 36, no discernable patterns were found to indicate

that student responses to items in some Iowa Core domains were impacted to a greater degree by the pandemic than those in other domains.

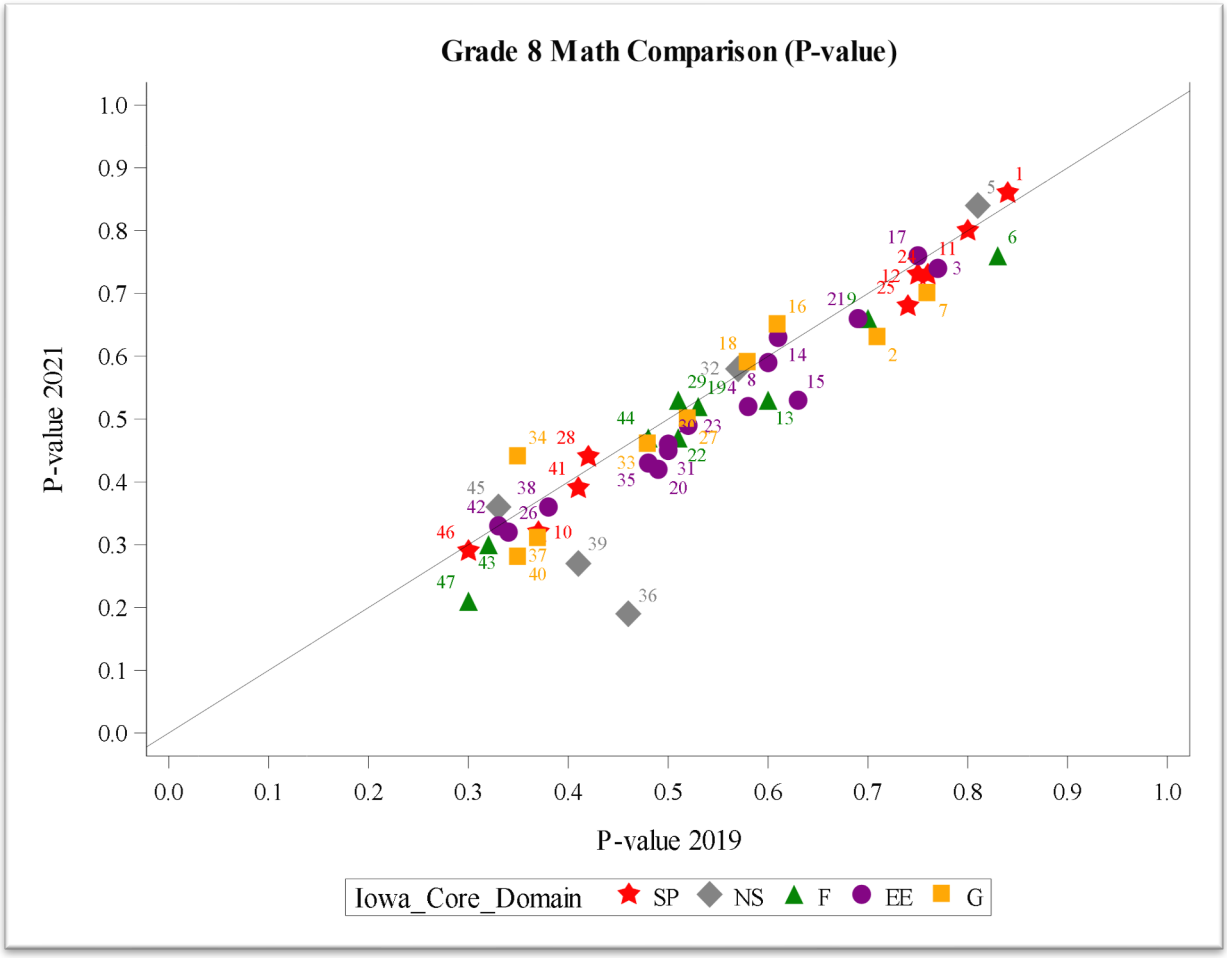


Figure 6. Scatterplot of Grade 8 Item Difficulty by Iowa Core Domain in 2019 and 2021

Figures 7 and 8 display a comparison of the point-biserial correlations from 2019 and 2021 by DOK and Iowa Core domain, respectively. With respect to both DOK and Iowa Core domain, no discernable patterns were found as items from each DOK level (Figure 7) and domain (Figure 8) are more or less scattered in the same direction as the diagonal line, though clearly below it. The most noteworthy characteristic of the scatterplots of point-biserial correlations is the systematic tendency for items in 2021 to not be as highly related to total math

score as they were in the 2019 test administration. Detailed item-level statistics used to create the scatterplots in Figures 1 through 8 are included in the Appendix.

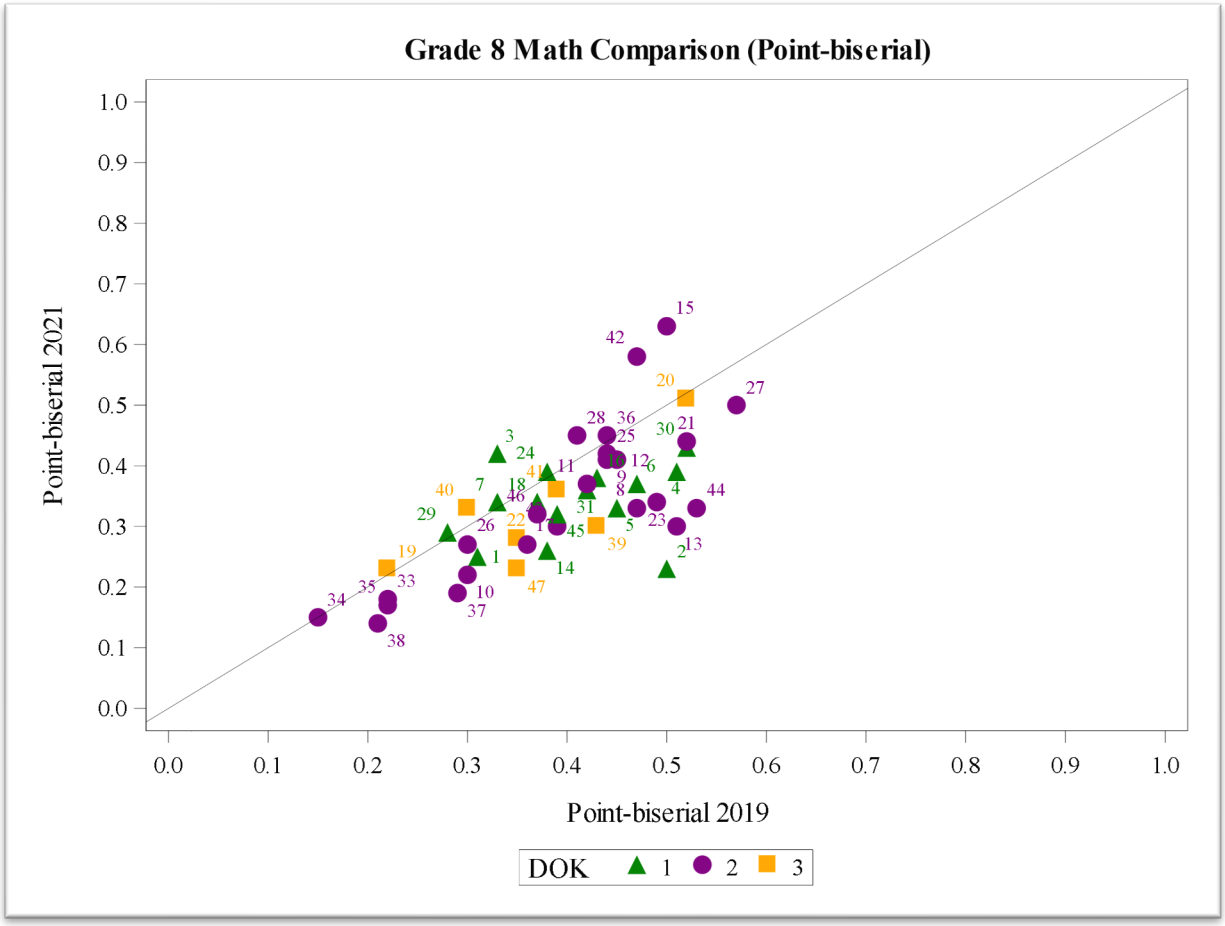


Figure 7. Scatterplot of Grade 8 Item-Total Correlations by DOK Level in 2019 and 2021

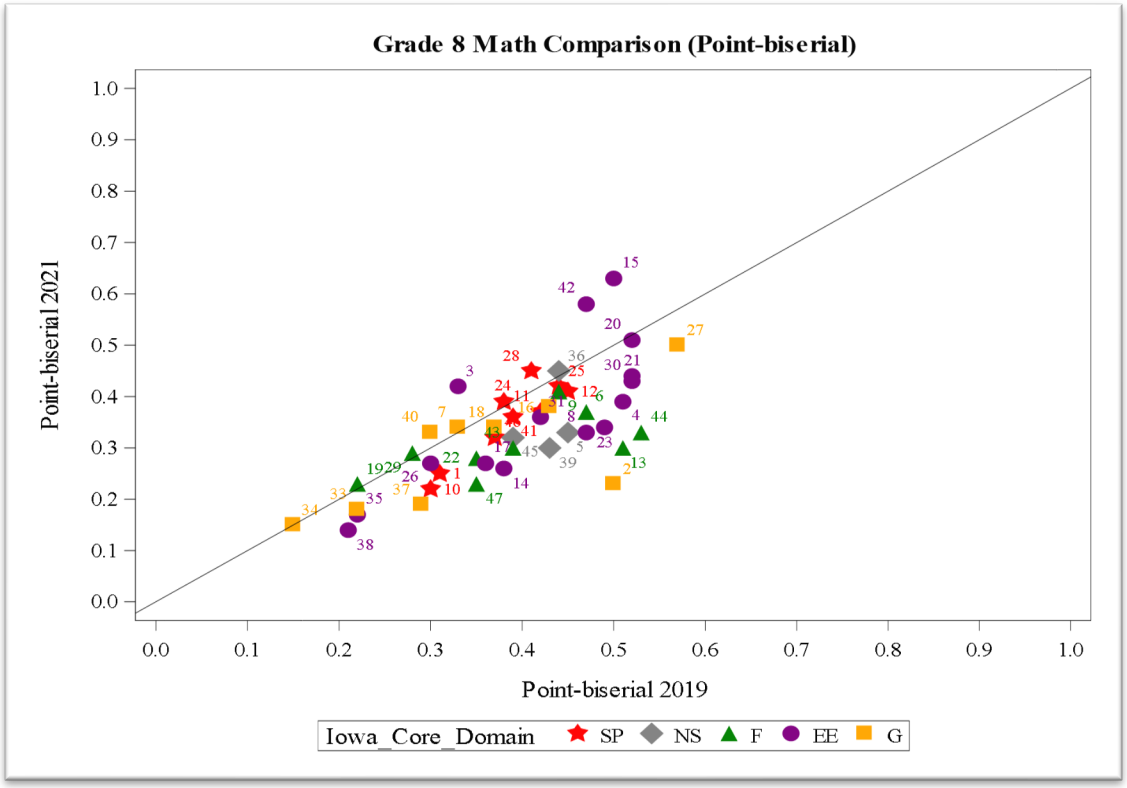


Figure 8. Scatterplot of Grade 8 Item-Total Correlations by Iowa Core Domain in 2019 and 2021

Discussion

Summative assessments such as ISASP can provide proficiency status and trend information, but they can also help users analyze and interpret changes in performance at the item level to consider patterns of change with respect to specific attributes of items such as cognitive complexity and content specifications. Under current conditions of disruptions to instructions due to implementation of district and state policy with respect to the coronavirus pandemic, it is important to examine every source of information available to better understand changes, gaps, and areas most in need of improvement. The results of the current analyses identified an overall downward shift in performance by content areas of mathematics between the 2019 and 2021 ISASP administrations, although there were some items that became easier

across the two years of study. However, the relative standing of individual items with respect to difficulty remained stable across years, indicating that there was not a single content area in which students were advantaged or disadvantaged because of potentially disproportionate instructional time across the mathematics domains of the Iowa Core.

A review of the items that demonstrated the largest differences in performance between the years of interest did not yield any obvious patterns with respect to item format (multiple-choice or constructed-response, for example) or the reading demands of the items. With respect to DOK level, in some grades, the DOK3 items showed a small tendency toward greater difficulty in 2021 relative to the change in difficulty observed for the less cognitively complex items of DOK levels 1 and 2.

As indicated in the analysis of items deemed Easy, Medium and Difficult, there was some evidence to support a hypothesis that more difficult items, regardless of math domain, were even more difficult after the pandemic. The mean p-values of Difficult items dropped by a greater amount than did the p-values of Easy items, and this finding occurred in all but one grade in the current analysis. This result is consistent with the finding with respect to cognitively complex DOK level 3 items.

Perhaps of greatest interest considering the potentially pervasive yet complex effects of disruptions to instruction during the period of time between the 2019 and 2021 ISASP administrations is the clear evidence at the item level that the 2021 student responses in mathematics reflected a less cohesive achievement construct than was evidenced after the initial ISASP administration in 2019. Nearly all items in all content areas, at all DOK levels and at all grades were *less highly correlated* with total math score in 2021 than in 2019.

An assessment that is fully aligned with content standards adopted by the state and by all schools, public and private, is by definition assembled from items designed and developed to be instructionally sensitive, and instructional sensitivity has been described as a psychometric property of such assessments (Polikoff, 2010). This attribute of items from tests used for accountability is required by the federal assessment peer review process (US Department of Education, 2018)The items in such assessments will possess cohesive item-level characteristics not because they all measure the same thing but rather because the instruction students receive is presumed to be uniform and balanced across all the content standards and cognitive levels represented in the assessment. Under uniform instructional practices, point-biserial item-total correlations, as well as IRT discrimination parameters, will tend to be high and somewhat uniform in magnitude. The systematic decline in item-total correlations in the 2021 item-response data suggests less cohesion in the achievement construct as it relates to instruction. Given that the test items themselves are the same in the 2019 and 2021 data, it seems reasonable to consider instructional effects as a potential cause of the decline in item-total correlations.

If one were to accept the argument that, whatever was the specific impact of the pandemic in particular schools, the overall effect statewide was unpredictable or at least uneven coverage of the broadly defined mathematics standards of the Iowa Core. Many math teachers would agree with the idea that under normal circumstances, coverage of all math standards documented in the Iowa Core requires careful attention to the scope and sequence of instruction and adherence to a schedule of lesson plans that flesh out the breadth and depth of the standards. Intermittent partial or complete quarantines, unscheduled needs to move back and forth between remote and in-person learning, and other factors that result in disruption of instructional planning and delivery all represent potential influences on the psychometric characteristics of

instructionally sensitive items. Evidence from this analysis demonstrates the subtle but pervasive effects of the pandemic on student performance in mathematics.

References

Betebenner, D. W. & Wenning, R. J. (2021). Understanding Pandemic Learning Loss and Learning Recovery: The Role of Student Growth & Statewide Testing. *National Center for the Improvement of Educational Assessment*.

Calfas, J. (2021, April). New test scores show students rebounding after pandemic-related learning loss. *Wall Street Journal*. Retrieved from https://webreprints.djreprints.com/5056661172457.html

Camara, W. (2020). Never let a crisis go to waste: Large-scale assessment and the response to COVID-19. *Educational Measurement: Issues and Practice*, *39* (3), 10-18. doi.org/10.1111/emip.12358

Center for Research on Education Outcomes (2020, October). *Estimates of learning loss in the 2019-2020 school year*. Palo Alto, CA: Stanford University.

Lewis, K., Kuhfeld, M., Ruzek, E. & McEachin, A. (2021). Learning during COVID-19: Reading and math achievement in the 2020-2021 school year. NWEA.

Lorié, W. (2020, June). *Contextualizing COVID-19 "Learning Loss" and "Learning Recovery."* Dover, NH: Center for Assessment. Retrieved from https://www.nciea.org/blog/school-disruption/contextualizing-covid-19-learning-loss-and-learning-recovery

Polikoff, M. (2010). Instructional sensitivity as a psychometric property of assessments. *Educational Measurement: Issues and Practice*, *29* (4), 3-14. doi.org/10.1111/j.1745-3992.2010.00189.x

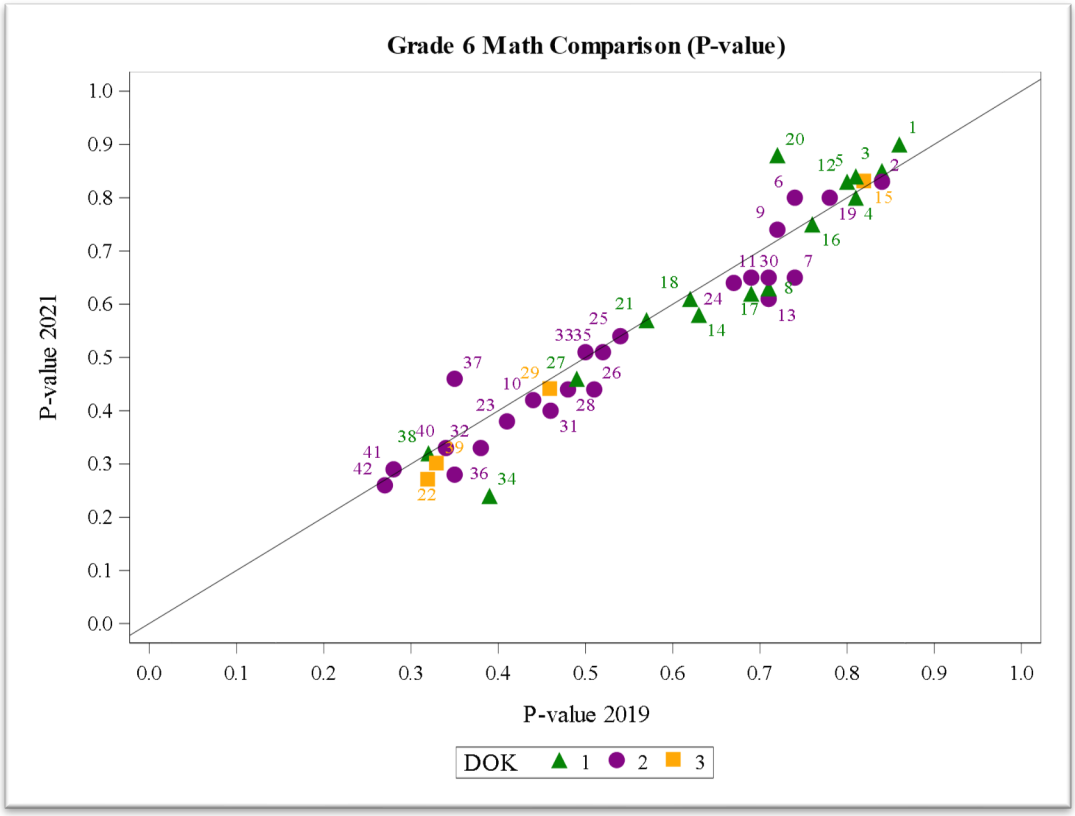
Renaissance Learning (2021). *How kids are performing: Tracking the midyear impact of COVID-19 on reading and mathematics achievement*. Wisconsin Rapids, WI: Author. Retrieved from https://research.renaissance.com/research/646.asp

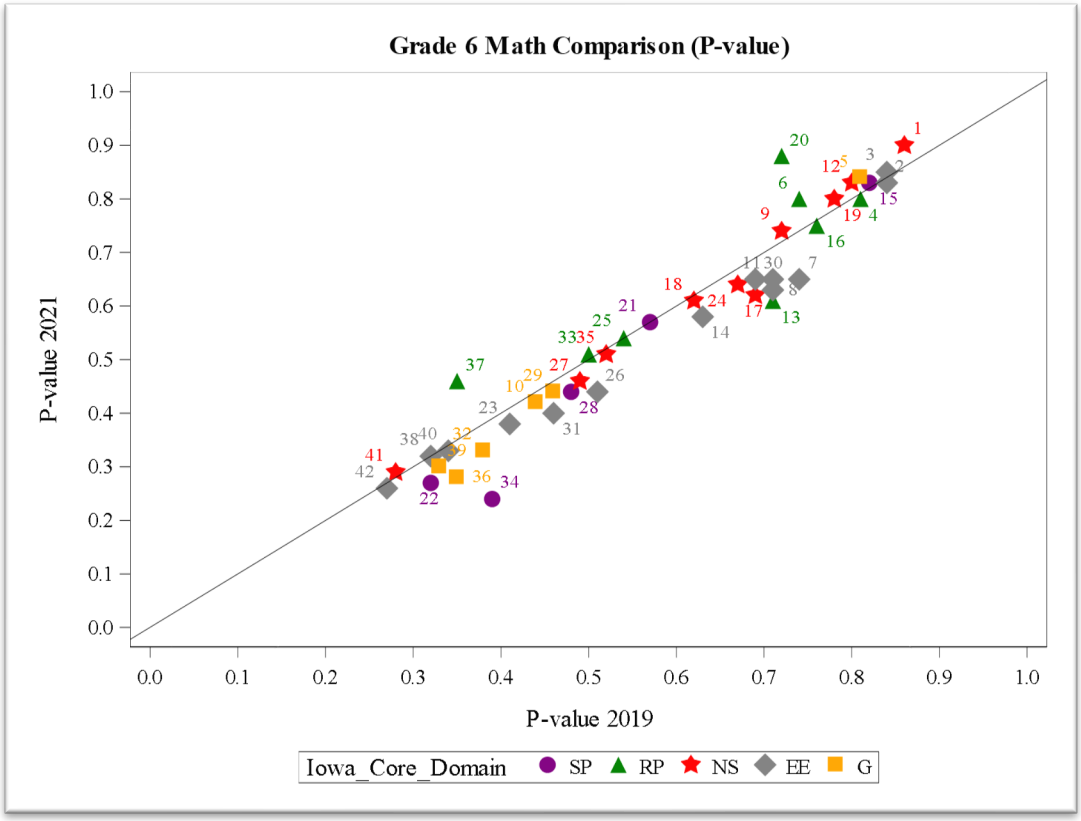
Schwartz, S. (2021, April). "Learning loss, in general, is a misnomer': Study shows kids made progress during COVID-19. *Education Week*, Retrieved from https://www.edweek.org/leadership/learning-loss-in-general-is-a-misnomer-study-shows-kids-made-progress-during-covid-19/2021/04

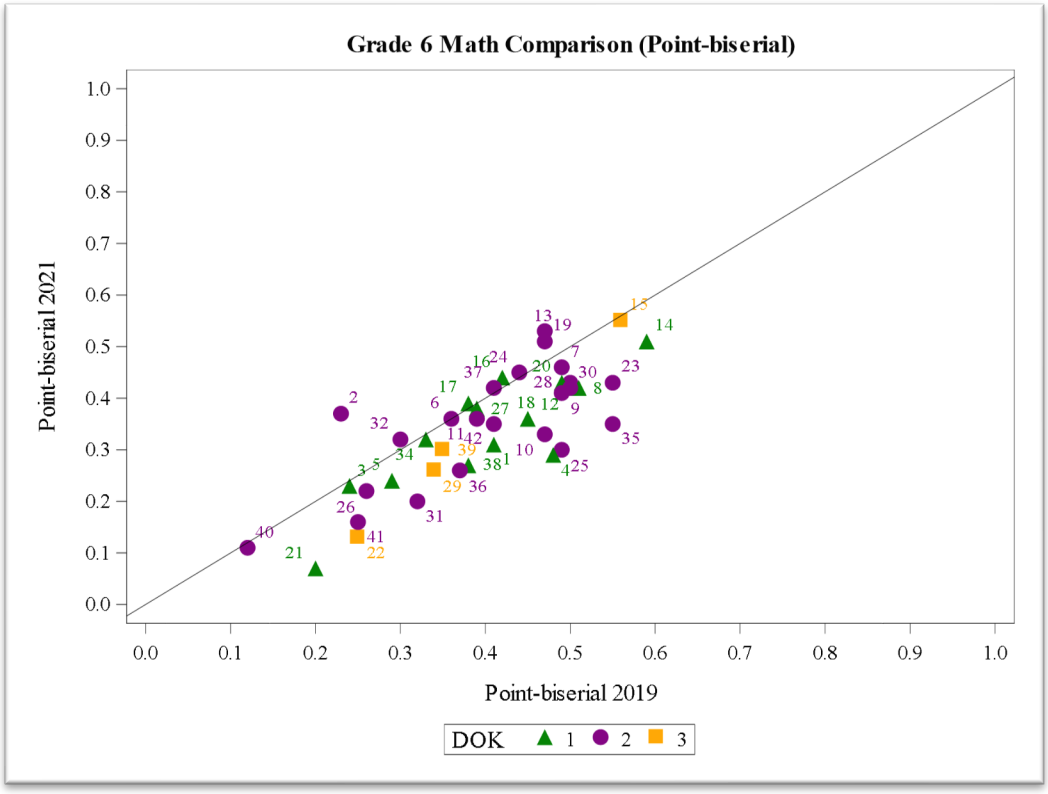
US Department of Education (2018). A state's guide to the US Department of Education's assessment peer review process. Retrieved May 25, 2021 from assessmentpeerreviewguidepubliccommentrevisionsjuly2018(OGC)(8.28)(PDF (ed.gov)

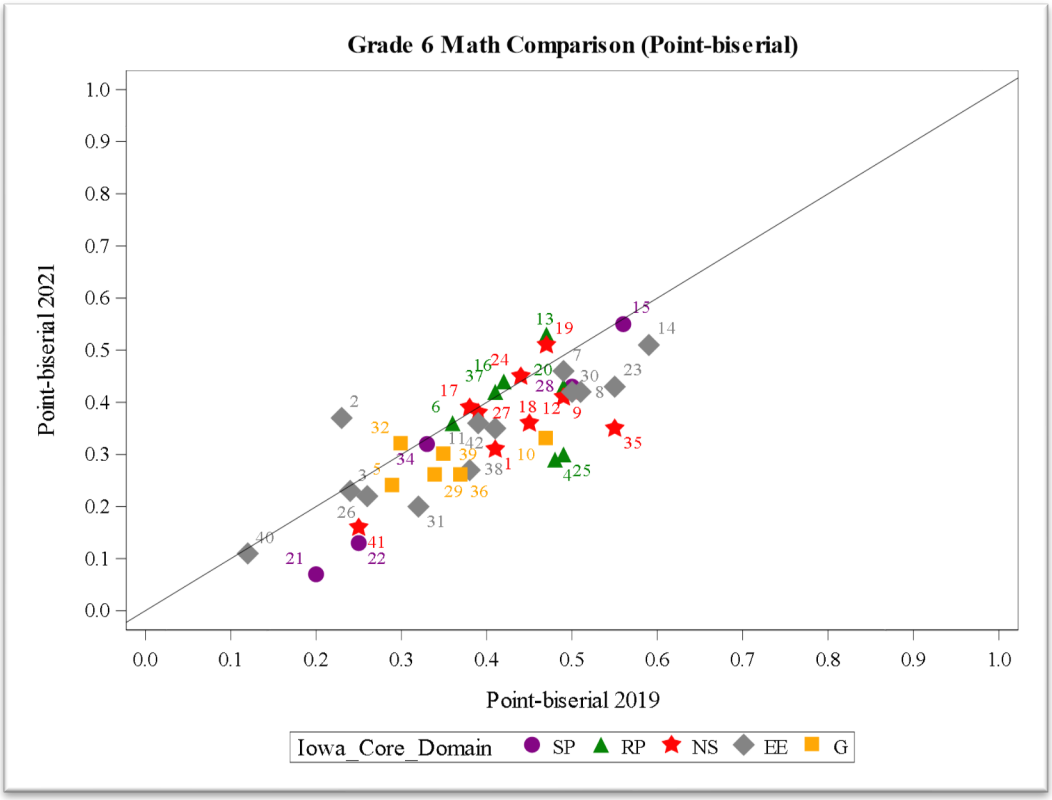
Appendix

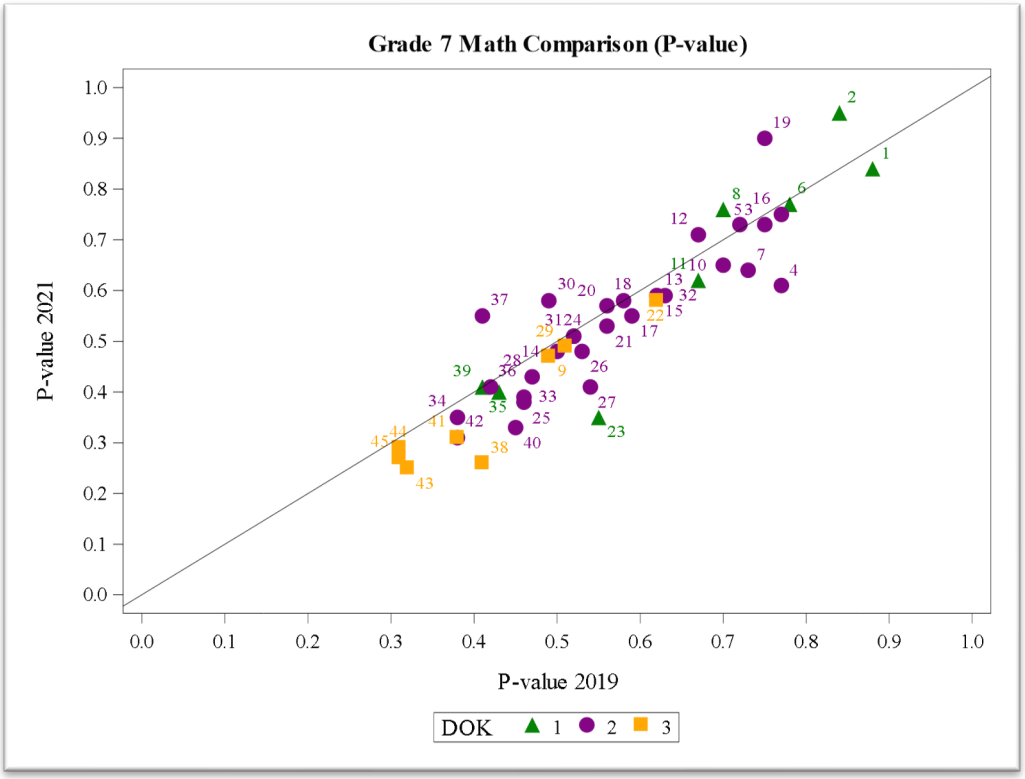
1. Figures for grades 6, 7, 9, 10 and 11.
2. Tables A.1 through A.7 of item-level statistics for grades 5-11.
    a. Item = item location in the 2021 test
    b. PV21 = p-value of the item in 2021
    c. PBIS21 = point-biserial correlation of the item in 2021
    d. PV19 = p-value of the item in 2019
    e. PBIS21 = point-biserial correlation of the item in 2021
    f. Domain = Iowa Core content domain designation
    g. DOK = Depth of Knowledge category

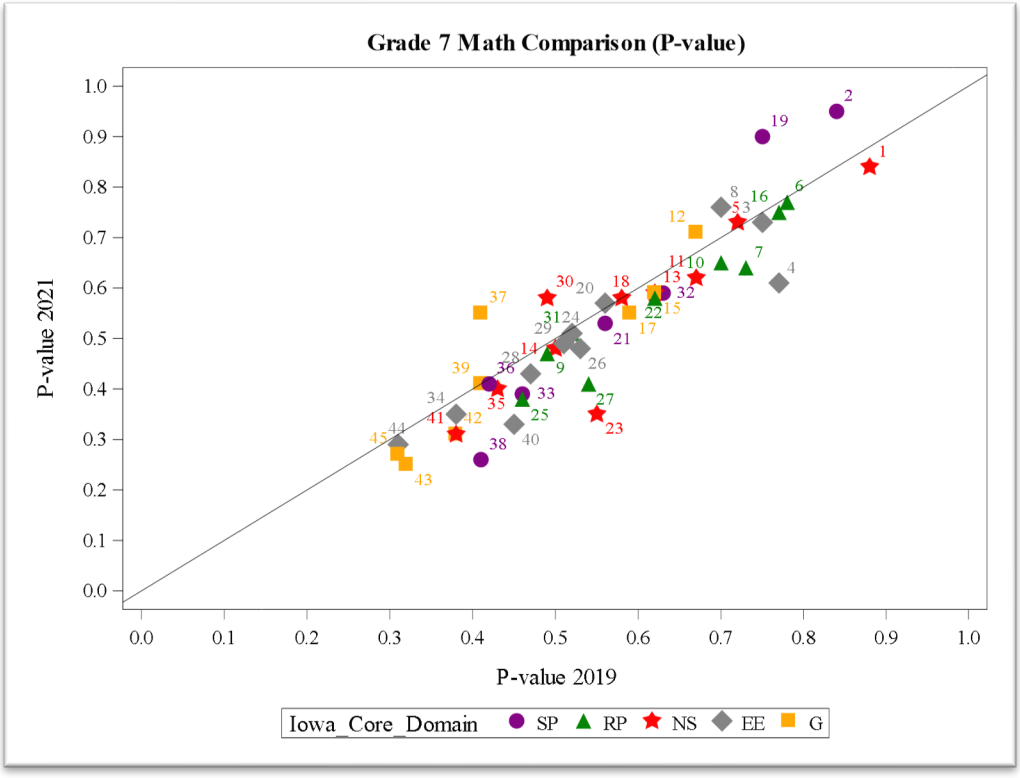

Grade 6 Math Comparison (P-value)
P-value 2021
P-value 2019
DOK 1 2 3

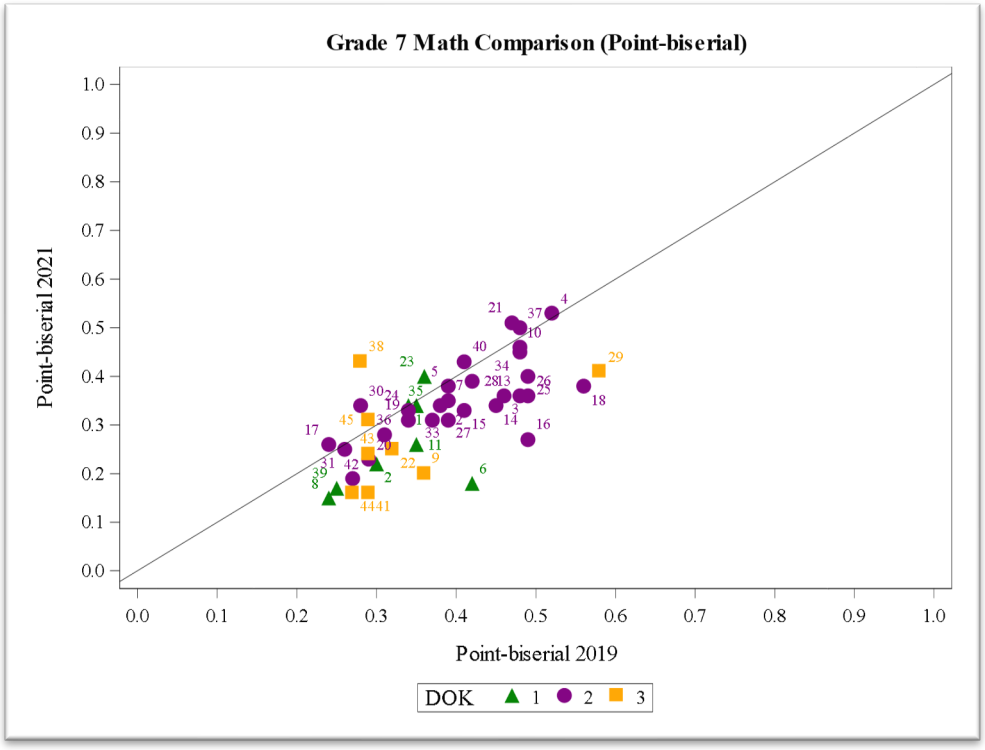

Grade 6 Math Comparison (P-value)
P-value 2021
P-value 2019
Iowa_Core_Domain SP RP NS EE G

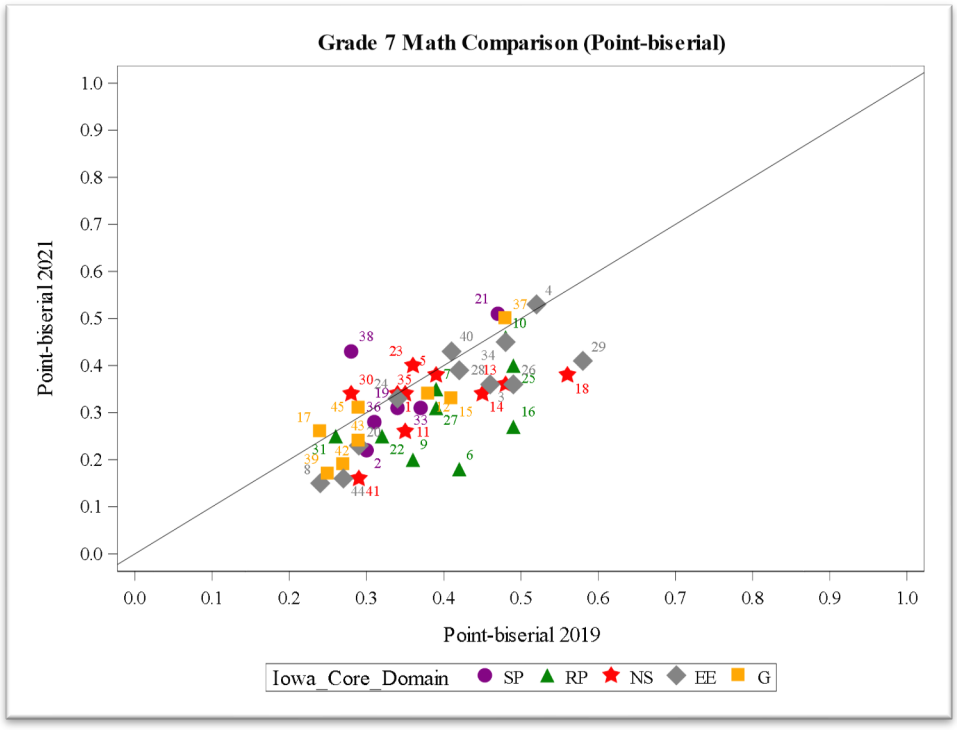

Grade 6 Math Comparison (Point-biserial)
Point-biserial 2021
Point-biserial 2019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
DOK 1 2 3

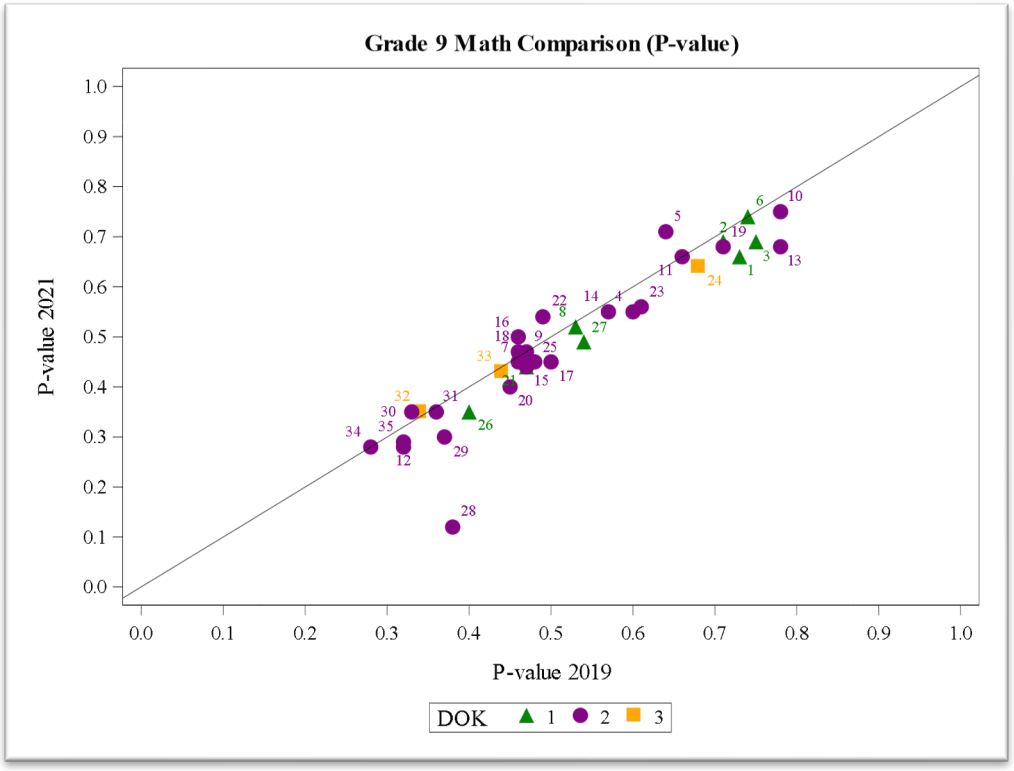

Grade 6 Math Comparison (Point-biserial)
Point-biserial 2021
Point-biserial 2019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Iowa_Core_Domain SP RP NS EE G

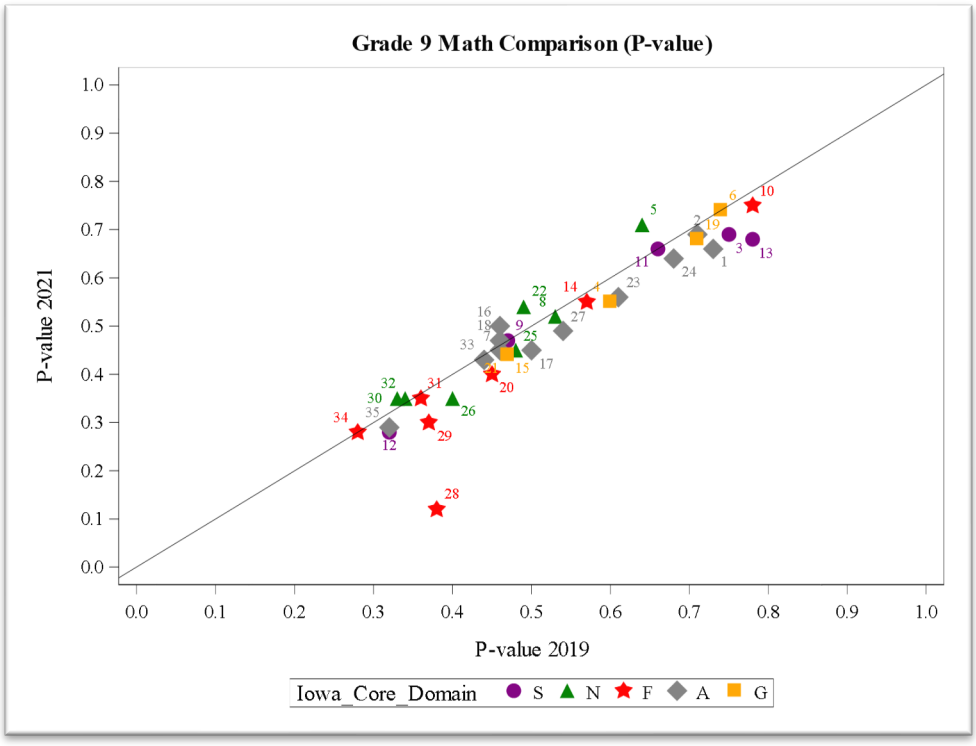

Grade 7 Math Comparison (P-value)
P-value 2021
P-value 2019
DOK 1 2 3

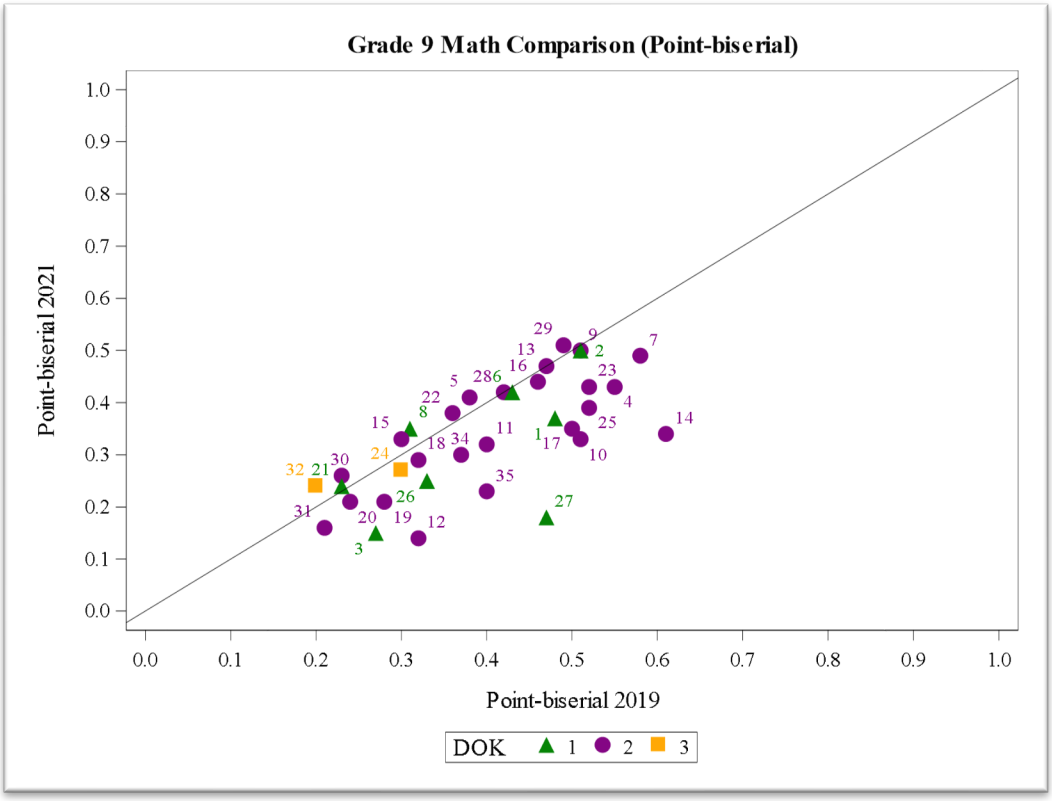

Grade 7 Math Comparison (P-value)
P-value 2021
P-value 2019
Iowa_Core_Domain SP RP NS EE G

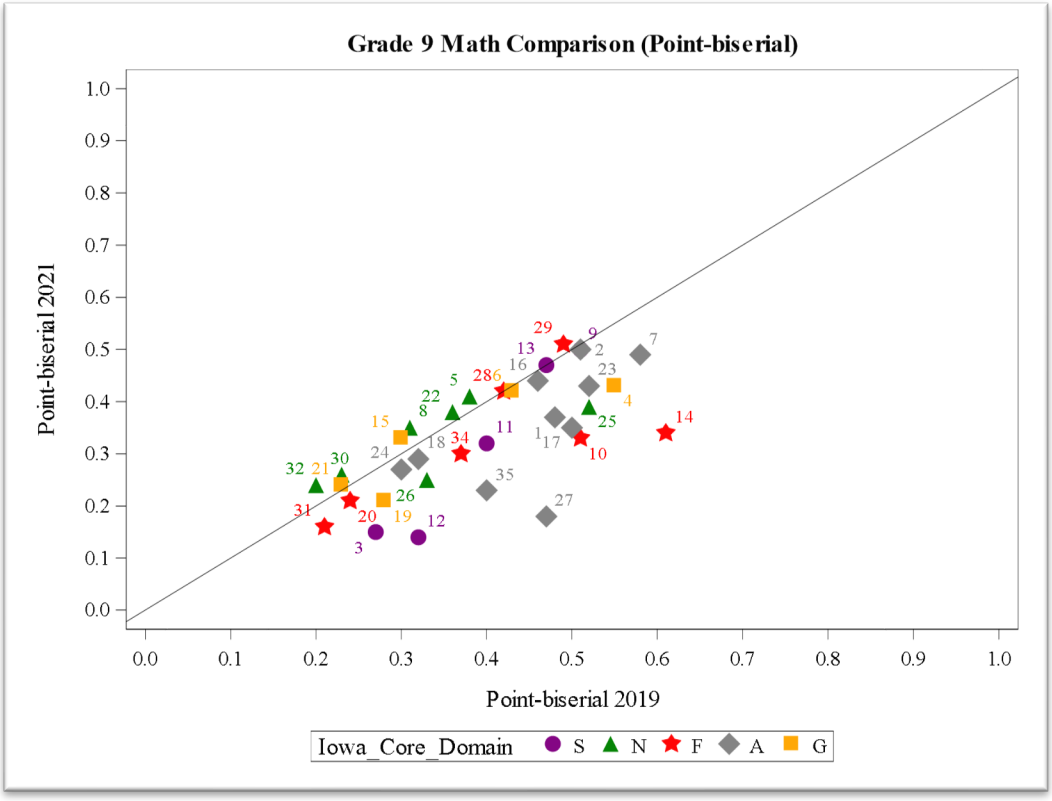

Grade 7 Math Comparison (Point-biserial)
Point-biserial 2021
Point-biserial 2019
DOK 1 2 3

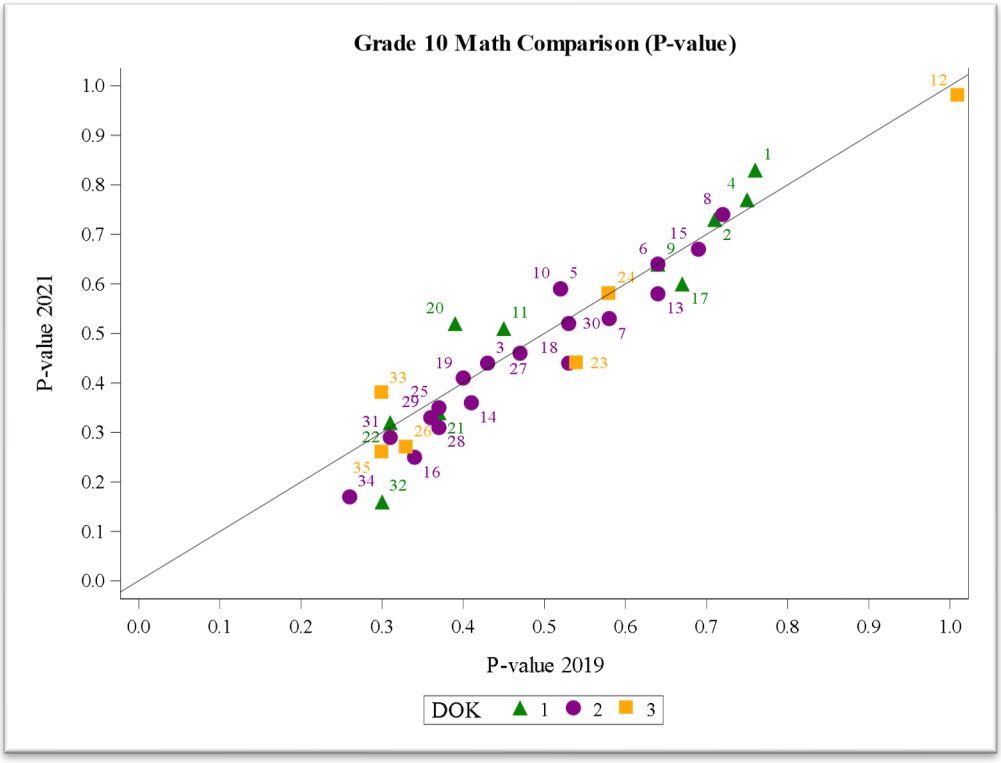

Grade 7 Math Comparison (Point-biserial)
Point-biserial 2021
Point-biserial 2019
Iowa_Core_Domain SP RP NS EE G

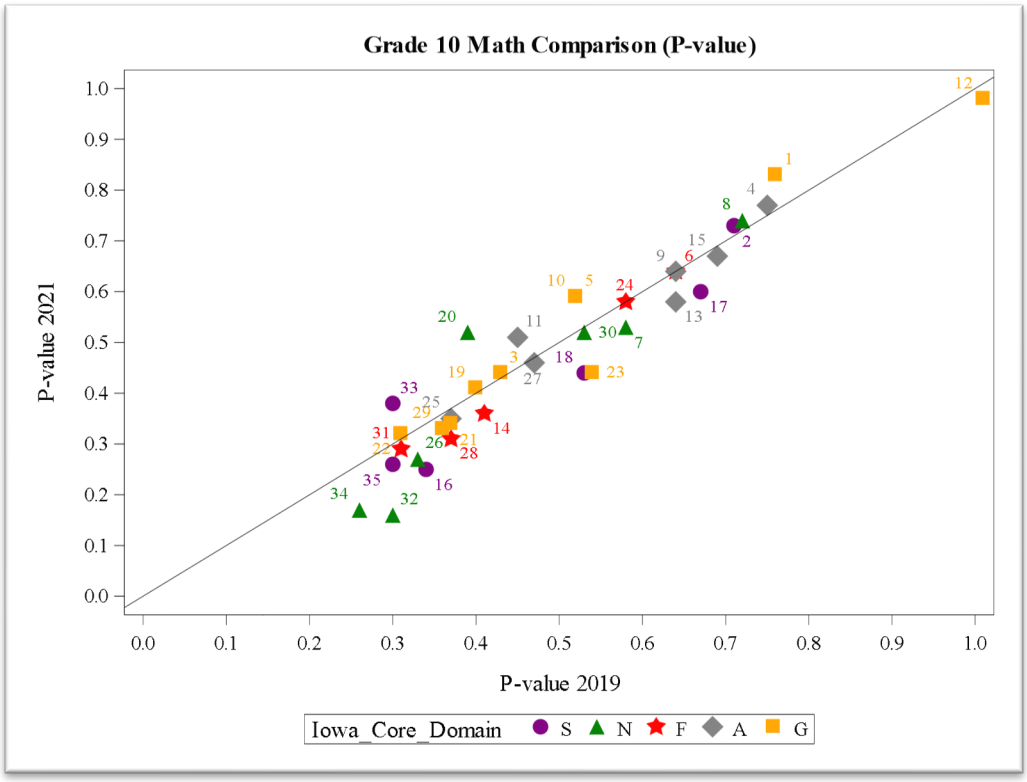

Grade 9 Math Comparison (P-value)
P-value 2021
P-value 2019
DOK 1 2 3

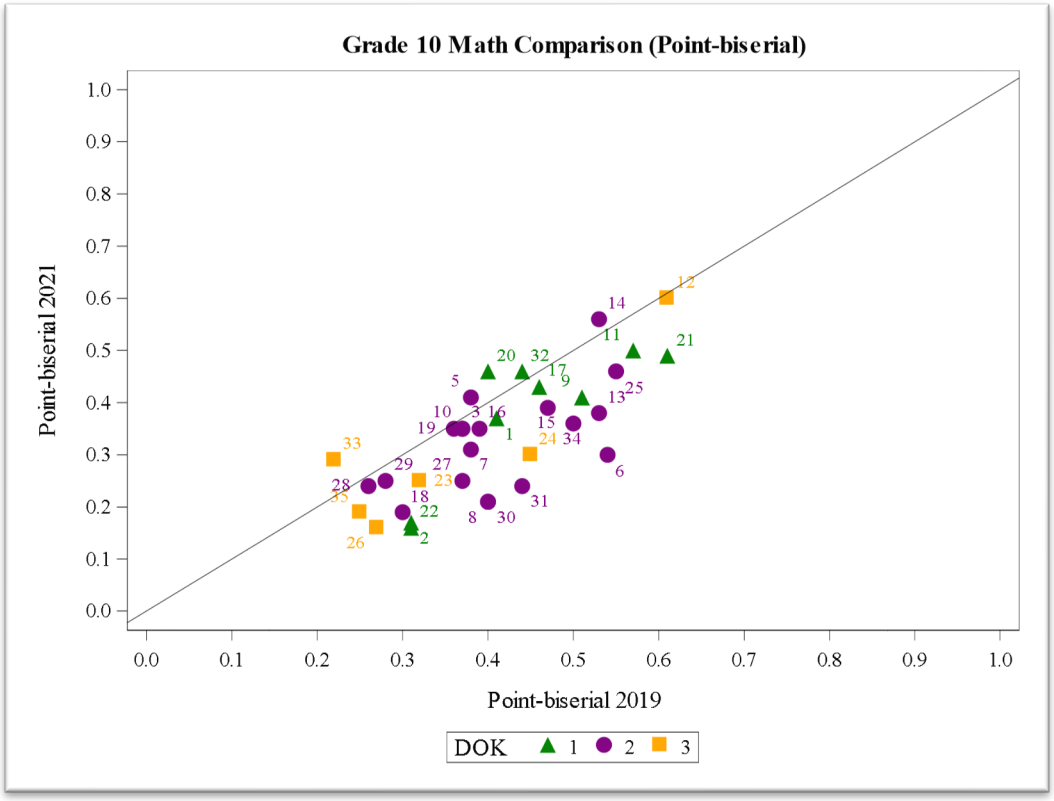

Grade 9 Math Comparison (P-value)
P-value 2021
P-value 2019
Iowa_Core_Domain S N F A G

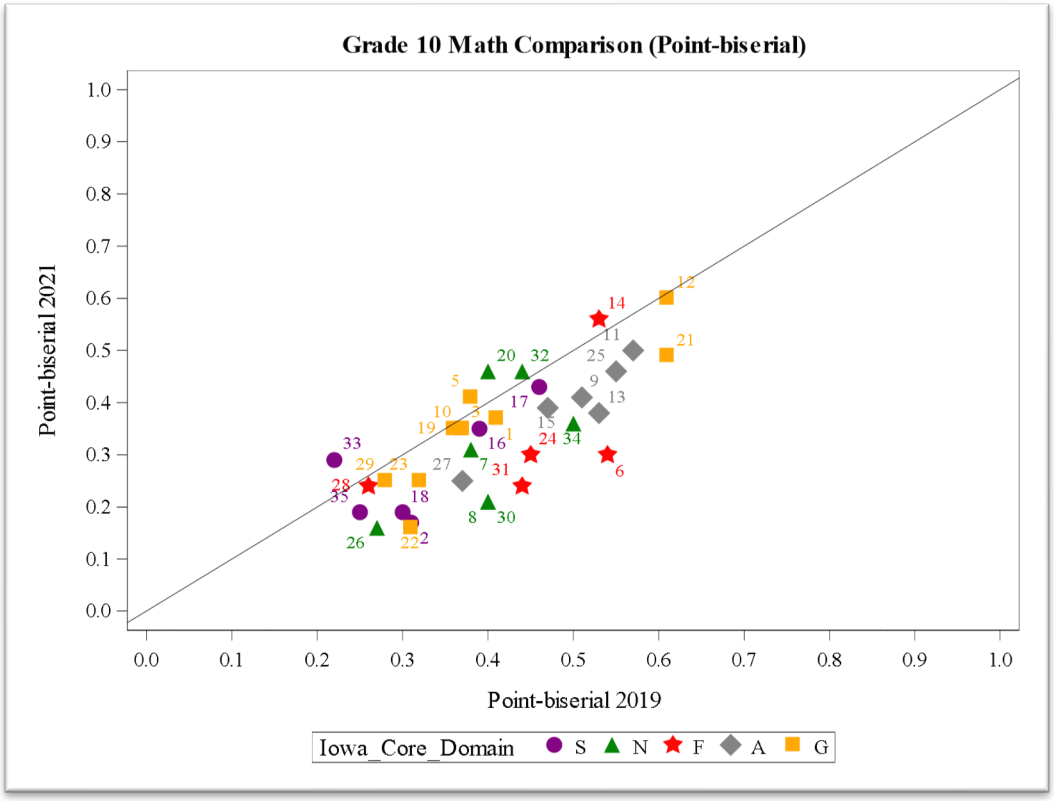

Grade 9 Math Comparison (Point-biserial)
Point-biserial 2021
Point-biserial 2019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
DOK 1 2 3

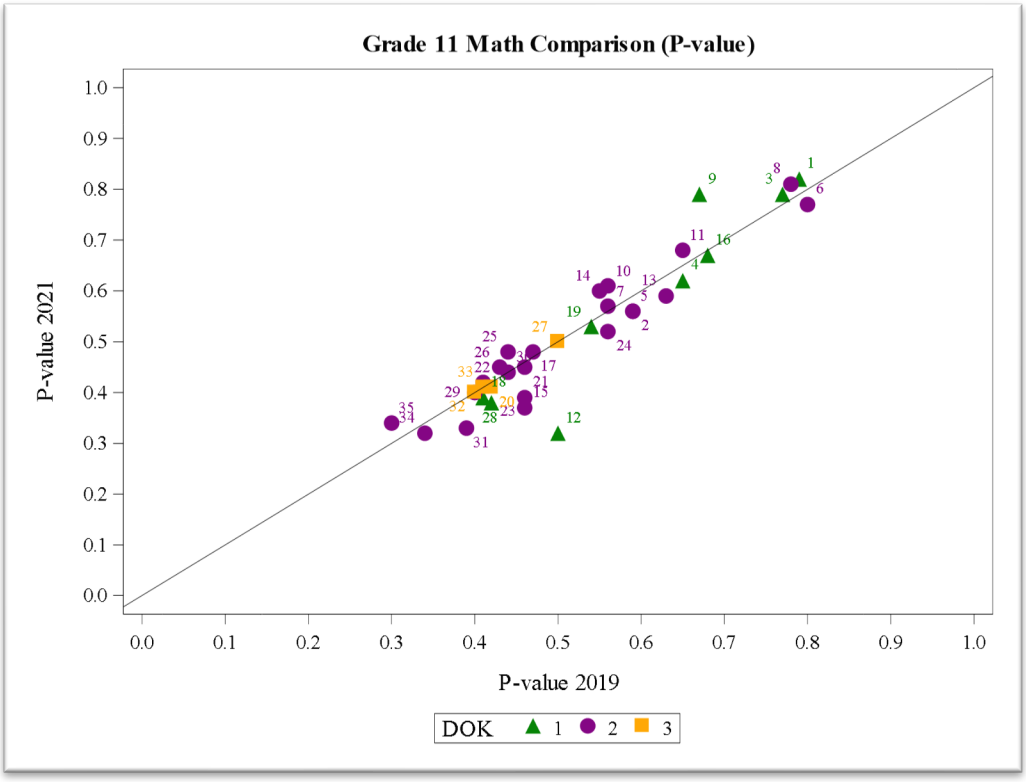

Grade 9 Math Comparison (Point-biserial)
Point-biserial 2021
Point-biserial 2019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Iowa_Core_Domain S N F A G

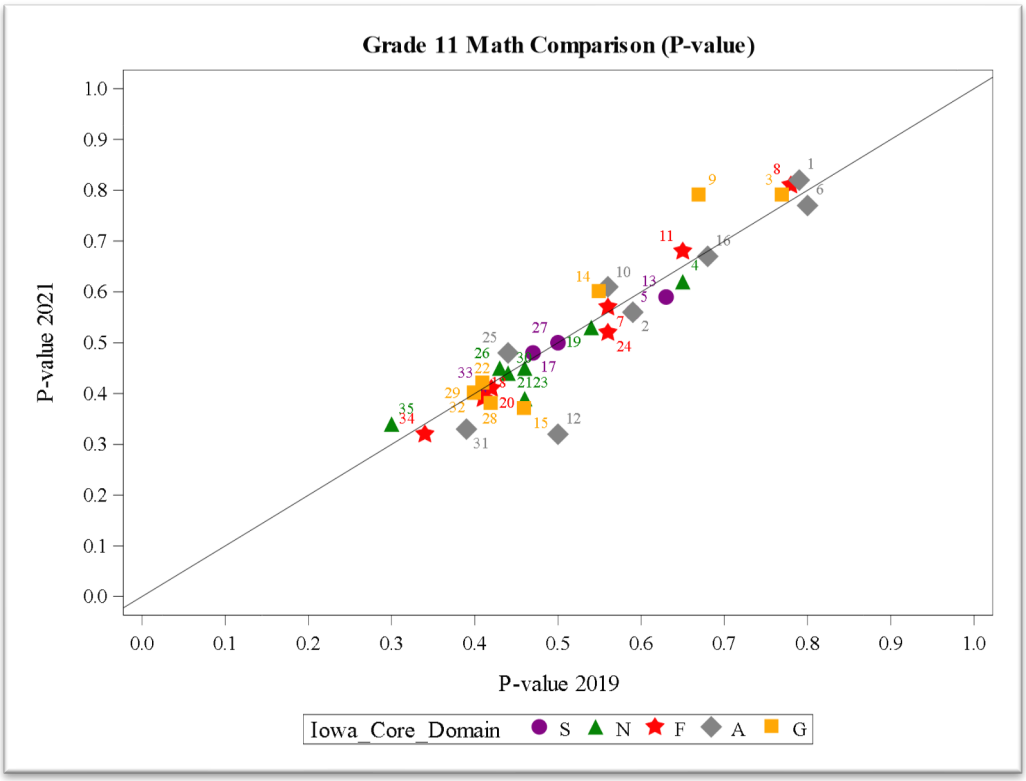

Grade 10 Math Comparison (P-value)
P-value 2021
P-value 2019
DOK 1 2 3

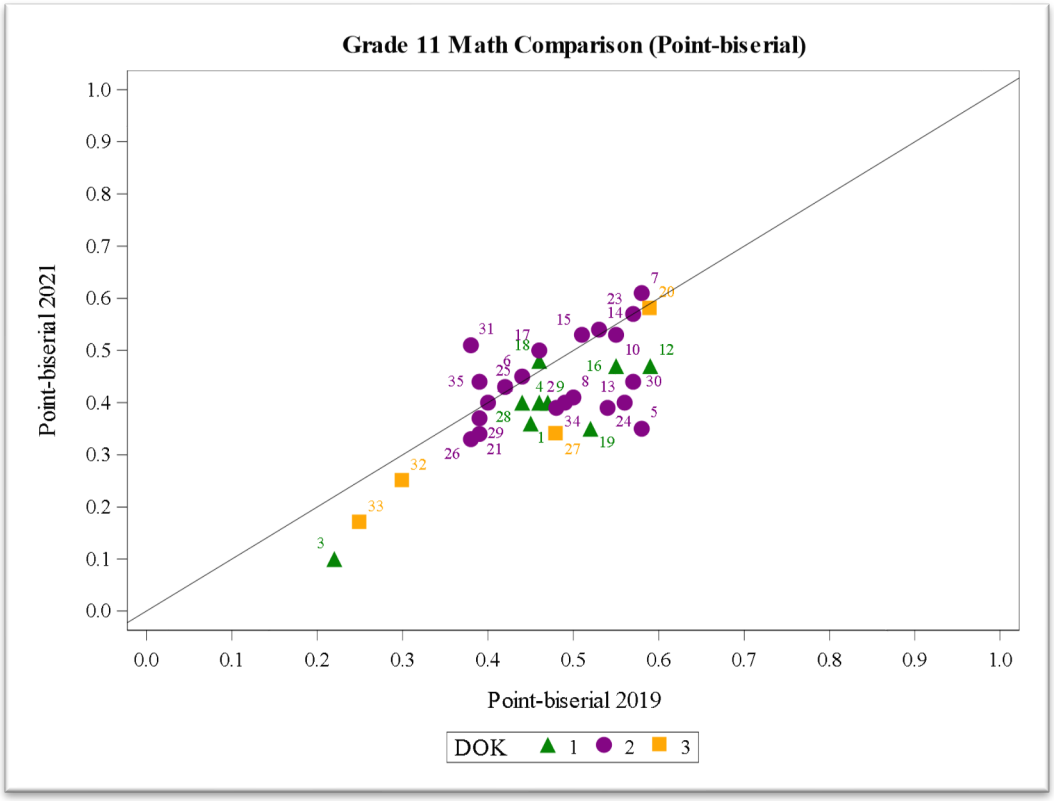

Grade 10 Math Comparison (P-value)
P-value 2021
P-value 2019
Iowa_Core_Domain S N F A G


Grade 10 Math Comparison (Point-biserial)
Point-biserial 2021
Point-biserial 2019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
DOK 1 2 3


Grade 10 Math Comparison (Point-biserial)
Point-biserial 2021
Point-biserial 2019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Iowa_Core_Domain S N F A G


Grade 11 Math Comparison (P-value)
P-value 2021
P-value 2019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
DOK 1 2 3


Grade 11 Math Comparison (P-value)
P-value 2021
P-value 2019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Iowa_Core_Domain S N F A G


Grade 11 Math Comparison (Point-biserial)
Point-biserial 2021
Point-biserial 2019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
DOK 1 2 3

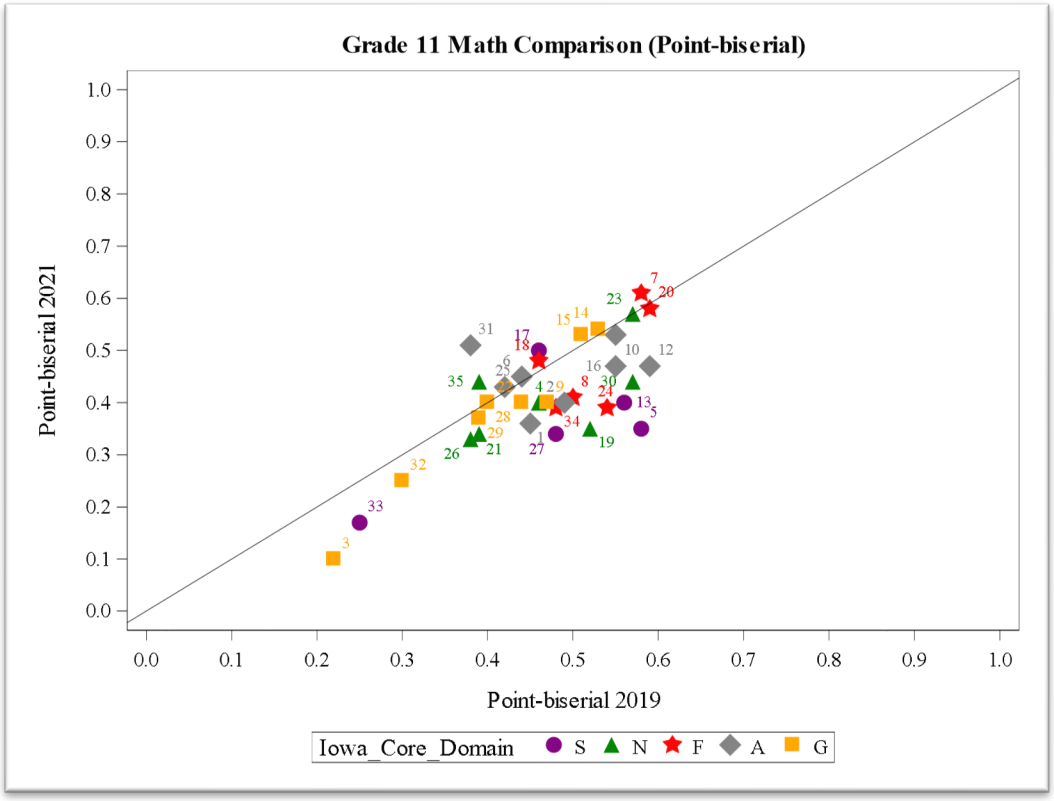

Grade 11 Math Comparison (Point-biserial)
Point-biserial 2021
Point-biserial 2019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Iowa_Core_Domain S N F A G

Table A.1. Item-Level Statistics for Grade 5

Item	PV 21	PBIS 21	PV 19	PBIS 19	Domain	DOK
1	74	.33	.67	.40	G	1
2	.80	.17	.79	.29	NBT	1
3	.65	.37	.69	.42	NBT	2
4	.6	.35	.63	.45	MD	2
5	.86	.37	.87	.39	OA	1
6	.79	NA	.84	.43	OA	1
7	.72	.24	.76	.32	G	1
8	.63	.40	.70	.44	G	2
9	.75	.28	.77	.52	NF	2
10	.59	.20	.64	.23	NF	1
11	.53	.39	.63	.31	NBT	2
12	.68	.43	.64	.48	NBT	1
13	.60	.41	.65	.52	NBT	2
14	.57	.29	.62	.56	NF	2
15	.60	.30	.62	.28	MD	2
16	.61	.44	.63	.46	OA	2
17	.57	.48	.61	.50	NBT	1
18	.69	.48	.72	.43	MD	2
19	.30	.34	.34	.38	OA	3
20	.75	.22	.70	.23	NF	1
21	.46	.45	.58	.60	NF	1
22	.59	.38	.65	.37	G	1
23	.59	.34	.59	.46	NBT	2
24	.57	.42	.61	.43	MD	2
25	.55	.28	.57	.32	OA	2
26	.40	.53	.45	.61	NF	2
27	.51	.53	.57	.60	NF	2
28	.56	.30	.57	.28	OA	2
29	.36	.43	.46	.52	NF	2
30	.25	.34	.29	.49	G	2
31	.46	.45	.55	.52	MD	2
32	.44	.32	.48	.40	MD	2
33	.36	.31	.42	.40	NBT	3
34	.50	.46	.53	.48	NBT	2
35	.48	.45	.53	.37	NF	2
36	.47	.22	.54	.23	NBT	2
37	.27	.23	.37	.27	G	1
38	.35	.37	.36	.38	MD	2
39	.32	.44	.36	.42	NF	3
40	.23	.38	.32	.45	MD	2

Table A.2. Item-Level Statistics for Grade 6

Item	PV 21	PBIS 21	PV 19	PBIS 19	Domain	DOK
1	.90	.31	.86	.41	NS	1
2	.83	.37	.84	.23	EE	2
3	.85	.23	.84	.24	EE	1
4	.8	.29	.81	.48	RP	1
5	.84	.24	.81	.29	G	1
6	.80	.36	.74	.36	RP	2
7	.65	.46	.74	.49	EE	2
8	.63	.42	.71	.51	EE	1
9	.74	.41	.72	.49	NS	2
10	.42	.33	.44	.47	G	2
11	.65	.36	.71	.39	EE	2
12	.83	.42	.80	.50	NS	1
13	.61	.53	.71	.47	RP	2
14	.58	.51	.63	.59	EE	1
15	.83	.55	.82	.56	SP	3
16	.75	.44	.76	.42	RP	1
17	.62	.39	.69	.38	NS	1
18	.61	.36	.62	.45	NS	1
19	.80	.51	.78	.47	NS	2
20	.88	.43	.72	.49	RP	1
21	.57	.07	.57	.20	SP	1
22	.27	.13	.32	.25	SP	3
23	.38	.43	.41	.55	EE	2
24	.64	.45	.67	.44	NS	2
25	.54	.30	.54	.49	RP	2
26	.44	.22	.51	.26	EE	2
27	.46	.38	.49	.39	NS	1
28	.44	.43	.48	.50	SP	2
29	.44	.26	.46	.34	G	3
30	.65	.42	.69	.50	EE	2
31	.40	.20	.46	.32	EE	2
32	.33	.32	.38	.30	G	2
33	.51	NA	.50	.52	RP	2
34	.24	.32	.39	.33	SP	1
35	.51	.35	.52	.55	NS	2
36	.28	.26	.35	.37	G	2
37	.46	.42	.35	.41	RP	2
38	.32	.27	.32	.38	EE	1
39	.30	.30	.33	.35	G	3
40	.33	.11	.34	.12	EE	2
41	.29	.16	.28	.25	NS	2
42	.26	.35	.27	.41	EE	2

Table A.3. Item-Level Statistics for Grade 7

Item	PV 21	PBIS 21	PV 19	PBIS 19	Domain	DOK
1	.84	.34	.88	.34	NS	1
2	.95	.22	.84	.30	SP	1
3	.73	.36	.75	.46	EE	2
4	.61	.53	.77	.52	EE	2
5	.73	.38	.72	.39	NS	2
6	.77	.18	.78	.42	RP	1
7	.64	.35	.73	.39	RP	2
8	.76	.15	.70	.24	EE	1
9	.47	.20	.49	.36	RP	3
10	.65	.46	.70	.48	RP	2
11	.62	.26	.67	.35	NS	1
12	.71	.34	.67	.38	G	2
13	.59	.36	.62	.48	NS	2
14	.48	.34	.50	.45	NS	2
15	.59	.33	.62	.41	G	2
16	.75	.27	.77	.49	RP	2
17	.55	.26	.59	.24	G	2
18	.58	.38	.58	.56	NS	2
19	.90	.31	.75	.34	SP	2
20	.57	.23	.56	.29	EE	2
21	.53	.51	.56	.47	SP	2
22	.58	.25	.62	.32	RP	3
23	.35	.40	.55	.36	NS	1
24	.51	.33	.52	.34	EE	2
25	.38	.40	.46	.49	RP	2
26	.48	.36	.53	.49	EE	2
27	.41	.31	.54	.39	RP	2
28	.43	.39	.47	.42	EE	2
29	.49	.41	.51	.58	EE	3
30	.58	.34	.49	.28	NS	2
31	.51	.25	.52	.26	RP	2
32	.59	NA	.63	.50	SP	2
33	.39	.31	.46	.37	SP	2
34	.35	.45	.38	.48	EE	2
35	.40	.34	.43	.35	NS	1
36	.41	.28	.42	.31	SP	2
37	.55	.50	.41	.48	G	2
38	.26	.43	.41	.28	SP	3
39	.41	.17	.41	.25	G	1
40	.33	.43	.45	.41	EE	2
41	.31	.16	.38	.29	NS	3
42	.31	.19	.38	.27	G	2
43	.25	.24	.32	.29	G	3
44	.29	.16	.31	.27	EE	3
45	.27	.31	.31	.29	G	3

Table A.4. Item-Level Statistics for Grade 8

Item	PV 21	PBIS 21	PV 19	PBIS 19	Domain	DOK
1	.86	.25	.84	.31	SP	1
2	.63	.23	.71	.50	G	1
3	.74	.42	.77	.33	EE	1
4	.52	.39	.58	.51	EE	1
5	.84	.33	.81	.45	NS	1
6	.76	.37	.83	.47	F	1
7	.70	.34	.76	.33	G	1
8	.59	.33	.60	.47	EE	2
9	.66	.41	.70	.44	F	2
10	.32	.22	.37	.30	SP	2
11	.80	.37	.80	.42	SP	2
12	.73	.41	.75	.45	SP	2
13	.53	.30	.60	.51	F	2
14	.63	.26	.61	.38	EE	1
15	.53	.63	.63	.50	EE	2
16	.65	.38	.61	.43	G	1
17	.76	.27	.75	.36	EE	2
18	.59	.34	.58	.37	G	1
19	.52	.23	.53	.22	F	3
20	.42	.51	.49	.52	EE	3
21	.66	.44	.69	.52	EE	2
22	.47	.28	.51	.35	F	3
23	.49	.34	.52	.49	EE	2
24	.73	.39	.76	.38	SP	1
25	.68	.42	.74	.44	SP	2
26	.32	.27	.34	.30	EE	2
27	.50	.50	.52	.57	G	2
28	.44	.45	.42	.41	SP	2
29	.53	.29	.51	.28	F	1
30	.46	.43	.50	.52	EE	1
31	.45	.36	.50	.42	EE	1
32	.58	NA	.57	.52	NS	2
33	.46	.18	.48	.22	G	2
34	.44	.15	.35	.15	G	2
35	.43	.17	.48	.22	EE	2
36	.19	.45	.46	.44	NS	2
37	.31	.19	.37	.29	G	2
38	.36	.14	.38	.21	EE	2
39	.27	.30	.41	.43	NS	3
40	.28	.33	.35	.30	G	3
41	.39	.36	.41	.39	SP	3
42	.33	.58	.33	.47	EE	2
43	.30	.30	.32	.39	F	2
44	.47	.33	.48	.53	F	2
45	.36	.32	.33	.39	NS	1

46	.29	.32	.30	.37	SP	2
47	.21	.23	.30	.35	F	3

Table A.5. Item-Level Statistics for Grade 9

Item	PV 21	PBIS 21	PV 19	PBIS 19	Domain	DOK
1	.66	.37	.73	.48	A	1
2	.69	.50	.71	.51	A	1
3	.69	.15	.75	.27	S	1
4	.55	.43	.60	.55	G	2
5	.71	.41	.64	.38	N	2
6	.74	.42	.74	.43	G	1
7	.45	.49	.46	.58	A	2
8	.52	.35	.53	.31	N	1
9	.47	.50	.47	.51	S	2
10	.75	.33	.78	.51	F	2
11	.66	.32	.66	.40	S	2
12	.28	.14	.32	.32	S	2
13	.68	.47	.78	.47	S	2
14	.55	.34	.57	.61	F	2
15	.44	.33	.47	.30	G	2
16	.50	.44	.46	.46	A	2
17	.45	.35	.50	.50	A	2
18	.47	.29	.46	.32	A	2
19	.68	.21	.71	.28	G	2
20	.40	.21	.45	.24	F	2
21	.44	.24	.47	.23	G	1
22	.54	.38	.49	.36	N	2
23	.56	.43	.61	.52	A	2
24	.64	.27	.68	.30	A	3
25	.45	.39	.48	.52	N	2
26	.35	.25	.40	.33	N	1
27	.49	.18	.54	.47	A	1
28	.12	.42	.38	.42	F	2
29	.30	.51	.37	.49	F	2
30	.35	.26	.33	.23	N	2
31	.35	.16	.36	.21	F	2
32	.35	.24	.34	.20	N	3
33	.43	NA	.44	.42	A	3
34	.28	.30	.28	.37	F	2
35	.29	.23	.32	.40	A	2

Table A.6. Item-Level Statistics for Grade 10

Item	PV 21	PBIS 21	PV 19	PBIS 19	Domain	DOK
1	.83	.37	.76	.41	G	1
2	.73	.17	.71	.31	S	1
3	.44	.35	.43	.37	G	2
4	.77	NA	.75	.45	A	1
5	.59	.41	.52	.38	G	2
6	.64	.30	.64	.54	F	2
7	.53	.31	.58	.38	N	2
8	.74	.21	.72	.40	N	2
9	.64	.41	.64	.51	A	1
10	.59	.35	.52	.37	G	2
11	.51	.50	.45	.57	A	1
12	.98	.60	.99	.61	G	3
13	.58	.38	.64	.53	A	2
14	.36	.56	.41	.53	F	2
15	.67	.39	.69	.47	A	2
16	.25	.35	.34	.39	S	2
17	.60	.43	.67	.46	S	1
18	.44	.19	.53	.30	S	2
19	.41	.35	.40	.36	G	2
20	.52	.46	.39	.40	N	1
21	.34	.49	.37	.61	G	1
22	.32	.16	.31	.31	G	1
23	.44	.25	.54	.32	G	3
24	.58	.30	.58	.45	F	3
25	.35	.46	.37	.55	A	2
26	.27	.16	.33	.27	N	3
27	.46	.25	.47	.37	A	2
28	.31	.24	.37	.26	F	2
29	.33	.25	.36	.28	G	2
30	.52	.21	.53	.40	N	2
31	.29	.24	.31	.44	F	2
32	.16	.46	.30	.44	N	1
33	.38	.29	.30	.22	S	3
34	.17	.36	.26	.50	N	2
35	.26	.19	.30	.25	S	3

Table A.7. Item-Level Statistics for Grade 11

Item	PV 21	PBIS 21	PV 19	PBIS 19	Domain	DOK
1	.82	.36	.79	.45	A	1
2	.56	.40	.59	.49	A	2
3	.79	.10	.77	.22	G	1
4	.62	.40	.65	.46	N	1
5	.56	.35	.59	.58	S	2
6	.77	.45	.80	.44	A	2
7	.57	.61	.56	.58	F	2
8	.81	.41	.78	.50	F	2
9	.79	.40	.67	.47	G	1
10	.61	.53	.56	.55	A	2
11	.68	NA	.65	.63	F	2
12	.32	.47	.50	.59	A	1
13	.59	.40	.63	.56	S	2
14	.60	.54	.55	.53	G	2
15	.37	.53	.46	.51	G	2
16	.67	.47	.68	.55	A	1
17	.48	.50	.47	.46	S	2
18	.39	.48	.41	.46	F	1
19	.53	.35	.54	.52	N	1
20	.41	.58	.42	.59	F	3
21	.45	.34	.46	.39	N	2
22	.42	.40	.41	.40	G	2
23	.39	.57	.46	.57	N	2
24	.52	.39	.56	.54	F	2
25	.48	.43	.44	.42	A	2
26	.45	.33	.43	.38	N	2
27	.50	.34	.50	.48	S	3
28	.38	.40	.42	.44	G	1
29	.40	.37	.40	.39	G	2
30	.44	.44	.44	.57	N	2
31	.33	.51	.39	.38	A	2
32	.40	.25	.40	.30	G	3
33	.41	.17	.41	.25	S	3
34	.32	.39	.34	.48	F	2
35	.34	.44	.30	.39	N	2