

# Interim Assessment Pattern-Based Item Report

---

Prepared for the Texas Education Agency

by Walter M. Stroup

with the Support of the GenEd Corporation

Submitted: December 12, 2019  
and approved with edits on June 12, 2020

About the author: Dr. Walter Stroup is the inventor of the pattern-based items, methods and technologies presented in this report. An education researcher, educator, and product developer for more than 30 years, he currently serves as Chair of the Department of STEM Education and Teacher Development at the University of Massachusetts, Dartmouth. Dr. Stroup is an experienced leader of research teams and a repeat recipient of funding from the National Science Foundation. Prior to accepting his current position, he was an Associate Professor in the Department of Curriculum and Instruction at The University of Texas, Austin. The teacher education program he helped found while at UT is used by more than 45 universities across the US. He has also been a consultant to state and federal ministries of education in Mexico. Dr. Stroup received his doctorate in education from Harvard University and is the Founder and the President of the GenEd Corp.

GenEd Corp. is a Delaware chartered company located in Austin, Texas established in 2019 to advance and support innovations in assessment and other educational services.

This report is presented to the Commissioner of the Texas Education Agency, Mike Morath, in partial fulfillment of terms of a Memorandum of Understanding (MOU) executed by the author of this report and the Texas Education Agency (TEA). The commissioned items developed for use in the 2018-2019 Interim Assessment Program, the analyses of results, and the preparation of this report were undertaken at no cost to the Texas Education Agency. The views and findings expressed in this report are solely those of the author and do not necessarily reflect the position of the TEA, the Commissioner of Education or of the GenEd Corp.

## Table of Contents

<b>Executive Summary.....</b>	<b>5</b>
<b>Introduction.....</b>	<b>7</b>
What is Distinctive about Pattern-Based Items? .....	7
What Are Some of the Advantages of Pattern-Based Items? .....	8
<b>Implementation .....</b>	<b>8</b>
<b>Illustrative Results .....</b>	<b>10</b>
Grade 3 Introductory Example – Three-Column Addition .....	10
Grade 5 Introductory Example – Language Arts .....	13
Introduction to the Significance of Moving to an Information-based Theoretical Model .....	14
Partial Credit and Full Credit Comparisons .....	15
Pilot Comparison of Released STAAR Item and a Corresponding PBI .....	16
Moving from Classroom Results to Statewide Results.....	20
A Related Example of Patterns Being Informative Across Scale .....	22
Practices to Improve Outcomes – Closing the Loop Back to instruction .....	25
Comparing Differences Across Scale .....	27
Differences in Reading Results Across Scale .....	27
Differences in Mathematics Results Across Scale .....	28
Comparing Differences Over Time - Assessing Growth.....	29
Online Reporting .....	30
<b>Summary and Conclusion.....</b>	<b>31</b>
<b>Appendices .....</b>	<b>33</b>
A: Items as They Appeared on 2018-2019 Interim Assessment (source: TEA) .....	33
Initial Interim Assessment Screen and Instructions to Students.....	33
Mathematics Grade 3 – Item 14 .....	34
Mathematics Grade 3 – Item 15 .....	35
Mathematics Grade 3 – Item 16 .....	36
Mathematics Grade 3 – Item 17 .....	37
Mathematics Grade 3 – Item 18 .....	38
Mathematics Grade 4 – Item 14 .....	39
Mathematics Grade 4 – Item 15 .....	40
Mathematics Grade 4 – Item 16 .....	41

Mathematics Grade 4 – Item 17 .....	42
Mathematics Grade 4 – Item 18 .....	43
Mathematics Grade 5 – Item 16 .....	44
Mathematics Grade 5 – Item 17 .....	45
Mathematics Grade 5 – Item 18 .....	46
Mathematics Grade 5 – Item 19 .....	47
Mathematics Grade 5 – Item 20 .....	48
Mathematics Grade 6 – Item 16 .....	49
Mathematics Grade 6 – Item 17 .....	50
Mathematics Grade 6 – Item 18 .....	51
Mathematics Grade 6 – Item 19 .....	52
Mathematics Grade 6 – Item 20 .....	53
Mathematics Grade 7 – Item 18 .....	54
Mathematics Grade 7 – Item 19 .....	55
Mathematics Grade 7 – Item 20 .....	56
Mathematics Grade 7 – Item 21 .....	57
Mathematics Grade 7 – Item 22 .....	58
Mathematics Grade 8 – Item 18 .....	59
Mathematics Grade 8 – Item 19 .....	60
Mathematics Grade 8 – Item 20 .....	61
Mathematics Grade 8 – Item 21 .....	62
Mathematics Grade 8 – Item 22 .....	63
Reading Grade 5 – “The Wind and the Moon” by George MacDonald.....	64
Reading Grade 5 – Item 15.....	65
Reading Grade 5 – Item 16.....	66
Reading Grade 5 – Item 17.....	66
Reading Grade 5 – Item 18.....	68
B: Partial and Full Credit Results for Pattern-Based Items and Released STAAR Items .....	69
<b>References .....</b>	<b>71</b>

## Executive Summary

During an assessment window from August 31, 2018 through March 31, 2019 a series of non-dichotomous, or “pattern-based,” items, focused primarily on mathematics but also including reading items, were implemented for grades three through eight on a statewide assessment program. The items were part of an opt-in Interim Assessment Program made available to districts by the Texas Education Agency. A central question to be addressed was whether the greater information density of pattern-based items would provide significant, actionable insight into student learning outcomes.

This report addresses the question by presenting results across scale – from individual classrooms and schools to statewide results – and across time with the use of repeat items to characterize student growth. Item by item alignment with the Texas Essential Knowledge and Skills (TEKS) and with the Blueprints for Assessment was ensured by developing each pattern-based question to parallel a released State of Texas Assessments of Academic Readiness (STAAR) item. The increased information density of PBIs also supports the assessment of the depth of student understanding, as well as additional related standards, within a single item instead of across a number of items. Consequently, assessments made up of PBIs can be both shorter *and* more informative.

Pattern-based Items are developed to provide a significant, fully scalable alternative to current dichotomously scored and analyzed items. While the PBIs analyzed in this report can appear similar to standard multiselect multiple-choice items, they should be seen as fundamentally distinct both in terms of how they are developed and in terms of how they are analyzed.

For an individual assessment item with four responses – A, B, C, or D – both multiselect and pattern-based items allow the students to select more than one response (e.g., A, C or B, C, D). For a standard multiselect item only one combination, the “exact match”, is scored as correct (“1”). All other combinations are scored as incorrect (“0”).

While legacy multiselect multiple-choice items can be viewed as an improvement over the more commonly used single-select multiple-choice items, the dichotomous scoring, and subsequent analyses, for *both single-select and multiselect items continues to be based on only two “states”, either 1 or 0*. In contrast, the pattern-based items appearing on the Interim Assessments were developed to *work with the full combinatoric space of student responses*. This means that for the PBIs analyzed in this report, there are sixteen possible selection combinations for the four available responses (fifteen if “no response” is not included). Rather than continue the current practice of reducing these states to only two states, this report illustrates how the greater information density of pattern-based items provides significant, actionable insights related to student learning outcomes. The results from PBIs can also be reporting in terms of partial credit, making the interpretation of item level results relatively transparent for students, parents/guardians, educators and the wider public.

Taken together, the greater information density and greater transparency of PBIs are intended to better enable large-scale assessment to support systematic improvement in outcomes for all students. For the purposes of this report, and to maintain the key distinctions between item types, both traditional single-select multiple-choice items (one right answer) and multiple-select multiple-choice items (one right combination of answers) will be referred to as “legacy” items. The non-dichotomous items will be referred to as either “pattern-based” or SmartQs™ (a commercial trade name).

Both legacy items (specifically, single-select multiple-choice) and pattern-based items appeared on the 2018-2019 Interim Assessments. Students were required to respond to the legacy items. They could then optionally complete the SmartQs™. Statewide, 418,147 students submitted responses to the SmartQs™. Figure 1 shows the breakdown by grade, subject area, and the specific item number on each of the Interim Assessments.

Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 5 Reading
14	60517	14	65263	16	65516	16
15	60212	15	64508	17	65163	17
16	59976	16	65044	18	64659	18
17	59884	17	64818	19	64352	19
18	59561	18	64750	20	64359	20
<b>Total</b>						
<b>418147</b>	<b>60517</b>	<b>65263</b>	<b>65516</b>	<b>60510</b>	<b>47774</b>	<b>48711</b>
						<b>69856</b>

**Figure 1. Distribution of responses to pattern-based items appearing on the 2018-2019 Interim Assessment**

The number of students responding to an individual item went from 46,144 for grade 8, item 20, mathematics up to 69,856 for grade 5, item 15, reading. To provide additional comparisons at the classroom and school levels, these results are supplemented by pilot implementations of the same items at two elementary schools in central Texas.

The findings from the implementation of pattern-based questions on the Interim Assessment as well as from the analyses that appear in this report are that:

- pattern-based items provide up to eight times as much information as legacy items, revealing meaningful patterns in student understanding
- because of the increase in analytical information, PBI-based tests can be much shorter than current legacy tests, which are designed to take from two to three hours to complete and where, beginning with the grade 3 STAAR, students are permitted to use up to four hours to complete an assessment
- shorter tests may mean less student fatigue, more student engagement, and more reliable item-level results
- pattern-based assessments have the potential to identify students who may have specific, as yet undiagnosed, learning challenges (e.g., dyslexia)
- educators see the patterns as having implications for improving instruction
- the reduced time lag in returning information from PBI-based assessments to teachers – as well as administrators, parents, and other stakeholders – increases the likelihood that the students who complete the assessments will also directly benefit from educational responses that are based on the results
- pattern-based items can be deployed at scale, even using existing vendor systems that were designed to implement legacy items
- comparisons of results within or across school years can address growth in understanding as well as identify areas of relative strength or ongoing difficulty
- partial credit scores are comparable to the percent correct on legacy items, and pattern-based items set a higher bar than legacy items in the percentage of students receiving full credit
- pattern-based items can be administered throughout the academic year, either as formative assessments to support ongoing instruction in classrooms or, as with the Interim Assessment Program, in a single deployment for benchmark or summative purposes.

## Introduction

In order for assessment based accountability to work as intended in improving education outcomes – “raising the bar and closing the gaps” – for all students, there is the requirement that assessment results can be used to characterize the effectiveness of instruction, these results can then be used to improve instruction, and this improvement will be reflected in improved outcomes. Rather than allowing “the soft bigotry of low expectations” to reinforce inequities in educational outcomes, and the long term consequences for underserved students’ lives, or allowing mediocrity to converge to a dispirited educational complacency that will undermine the long term prospects for our society as a whole, assessment-based accountability was intended to be a key driver of reform in the original No Child Left Behind legislation and its successors.

Framed by an interest in exploring a range of approaches to assessment that might serve the goal of improving student outcomes, an invitation was extended by the Commissioner of Education in Texas, Mike Morath, to the author of this report to make a presentation March of 2017 regarding the possible use of “non-dichotomous multiple-choice” (NDMC) items. After a detailed discussion of NDMC examples as well as aspects of their design and analysis, the meeting concluded with a commitment to pursue a statewide pilot of what are now called “pattern-based items” or SmartQs™ (a commercial name). The implementation of the items as part of the Interim Assessment Program (IAP) is the culmination of this initial piloting, and the results reported herein help frame a discussion of next steps in developing assessments meant to inform instruction, with the over-arching goal of improving outcomes for all students.

### What is Distinctive about Pattern-Based Items?

Pattern-based Items are developed to provide a significant, and fully scalable, alternative to current, dichotomously scored and analyzed, items. While the PBIs analyzed in this report can appear similar to standard multiselect multiple-choice items, they should be seen fundamentally distinct both in terms of how they are developed and in terms of how they are analyzed.

For an individual assessment question with four responses – A, B, C, or D – both multiselect and pattern-based items allow the students to select, or endorse, more than one response (e.g., A, C or B, C, D). With a standard multiselect item, however, only one combination, the “exact match”, is scored as correct (“1”). All other combinations are scored as incorrect (“0”).

While legacy multiselect multiple-choice items can be viewed as an improvement over the more commonly used single-select multiple-choice items, the dichotomous scoring, and subsequent analyses, for *both single-select and multiselect items continues to be based on only two “states”, either 1 or 0*. In contrast, the pattern-based items appearing on the Interim Assessments were developed to *work with the full combinatoric space of student responses*. This means for the PBIs analyzed in this report, there are sixteen possible selection combinations for the four available responses (fifteen if “no response” is not included). Rather than continue the current practice of reducing these states to only two states, this

report illustrates how the greater information density of pattern-based items (sixteen states vs. two states) provides significant, actionable insights related to student learning outcomes.

### What Are Some of the Advantages of Pattern-Based Items?

The increased information provided as well as the intentional design of pattern-based items allow the focus of assessment to be on evaluating depth of student understanding as well as providing all students with more ways to show what they do know. Elementary school teachers from the pilot sites reported that students find the questions engaging and, as will be reported below, this observation is consistent with how the completion rate for the pattern-based items appearing on the Interim Assessments remains constant across all the items for each grade level and subject assessed. As will also be illustrated in what follows, the ability to identifying patterns in student responses allows assessments to be a more effective diagnostic tool for characterizing student strengths and areas of concern. This allows teachers' instructional responses for all students in their classrooms to be more precise and the identification of students for additional interventions to be more accurate and informed than is possible with a similar number of dichotomous items. More information about their students supports teachers being more effective in improving student outcomes. Pattern-based assessments can be shorter and more informative in ways that can directly support better ongoing instruction. Pattern-based items can be administered regularly throughout the academic year as formative assessments meant to support ongoing instruction in classrooms or, as was the case with the Interim Assessment, in a single deployment for benchmark or summative purposes.

## Implementation

During an assessment window beginning on August 31, 2018 and extending through March 31, 2019 a series of non-dichotomous, or “pattern-based,” items focused primarily in mathematics but also including reading items were implemented for grades three through eight on a statewide assessment program. The items were deployed as part of an opt-in Interim Assessment Program made available to districts by the Texas Education Agency. A central question to be addressed in deploying these commissioned items at scale was whether the greater information density of pattern-based items would provide significant, actionable insight into student learning outcomes.

The students were required to complete a series of single-select multiple-choice STAAR items prior to then having the option to complete the pattern-based items. The directions stated: “The next set of questions is optional. These questions will not be counted as part of your score.” Moreover, due to terms agreed to by the Texas Education Agency and the schools participating in the Interim Assessment Program, no datasets containing student-, school-, or district-identifiable data would be provided by the vendor to the Agency.

Consistent with these terms, we received two large, fully anonymized, datasets containing only each student’s selections for the pattern-based items. As a result, the data cannot be used to provide an account of the makeup of the students who participated in the overall Interim Assessment Program or of those who then elected to complete the optional items. Comparisons between results for legacy items and results for PBIs on the Interim Assessments are also precluded. To be able to illustrate how PBIs support comparisons across scale – e.g., comparing individual classroom results or school-wide results

with statewide results – we have augmented the datasets from the Interim Formative Assessment Program with datasets from a December 2018 pilot implementation of the same items in two elementary schools in central Texas.

The overall number of students who elected to submit responses to the pattern-based items was 418,147. A table of question number, as well as the grades three through eight completion rates, for the pattern-based mathematics and reading items is shown in Figure 2.

	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 5 Reading			
	14	60517	14	65263	16	65516	16	60510	18	47774	18	48711	15	69856		
	15	60212	15	64508	17	65163	17	59130	19	46591	19	47676	16	69708		
	16	59976	16	65044	18	64659	18	58852	20	46144	20	47812	17	69634		
	17	59884	17	64818	19	64352	19	58902	21	46464	21	48167	18	69522		
	18	59561	18	64750	20	64359	20	58497	22	46483	22	47720				
<b>Total</b>		<b>418147</b>		<b>60517</b>		<b>65263</b>		<b>65516</b>		<b>60510</b>		<b>47774</b>		<b>48711</b>		<b>69856</b>

**Figure 2. Table of students submitting responses to the pattern-based item on the Interim Assessments**

Even with the items being optional and appearing after the required portion of the assessments, student engagement for each grade level in math and in reading remained consistent, with the average completion rate for the last item being more than 98% of the average completion rate for the first item. If there were significant difficulties with the implementation of items at scale, then one would expect much less consistency in the levels of participation.

A significant number of students in grade 3 through 8 were able to complete and submit responses to PBIs and no issues with the statewide implementation were reported to the author of this report. Given this robustness of the implementation at scale, the focus can now turn to directly addressing the question of whether the greater information density of pattern-based items provides significant, actionable insight into student learning outcomes.

## Illustrative Results

Using the results from the Interim Assessment Program, as augmented by results from pilot studies carried out in central Texas using the same items, this section of the report highlights some of the key findings related to the advantages of using pattern-based items as an alternative to dichotomously scored, legacy items. The presentation begins with results from introductory examples in mathematics and reading. After these introductory examples, partial credit and exact match comparisons of PBIs and 2018 STAAR items are made. We then use results from one third grade classroom to highlight a particular challenge related to learning fraction concepts as well as how a modification in the instructional approach to using standard manipulatives can address this challenge. Results from a 5<sup>th</sup> grade item, similar to the 3<sup>rd</sup> grade item discussed previously, are used to highlight how the analyses supported by the use of pattern-based items hold up across scale. This consistency across scale can help advance educational responses to student difficulties that are focused and coordinated across levels. In addition, the results from the use of PBIs illustrate how consistency in the patterns allows for detailed comparisons in outcomes across scale or across settings – to help identify practices that are effective – or over time – to serve as a measure of student growth. We begin our discussion of results with analyses of results from a Grade 3 mathematics item and a Grade 5 reading item.

### Grade 3 Introductory Example – Three-Column Addition

The following item (Figure 3) was developed for use on the Interim Assessment to address a third grade “number and operations” standard from the Texas Essential Knowledge and Skills (TEKS). The standard requires students to be able to, “solve with fluency one-step and two-step problems involving addition and subtraction within 1,000 using strategies based on place value, properties of operations, and the relationship between addition and subtraction.” For this question, students are asked to find two numbers that would make the addition correct. As was the case for each pattern-based item appearing on the Interim Assessment, the directions state that the question “may have more than one answer” and that the students are to select “all correct responses.”

This particular question has only one correct response, but as is also the case for all pattern-based items, *not* selecting certain responses contributes to an overall account of student learning outcomes. Consequently, not selecting an incorrect response contributes equally to the partial credit score on an item as does selecting the correct response(s). The utility of considering the patterns both in answers selected and ones not selected, is illustrated using the combinations of responses from the statewide results.

This addition problem is missing two numbers. Which numbers placed in the boxes below would make the addition correct?

$$\begin{array}{r} 7 \square 2 \\ + 1 \square 3 \\ \hline 9 4 5 \end{array}$$

***This question may have more than one correct answer. Select all correct responses.***

A 3 and 1

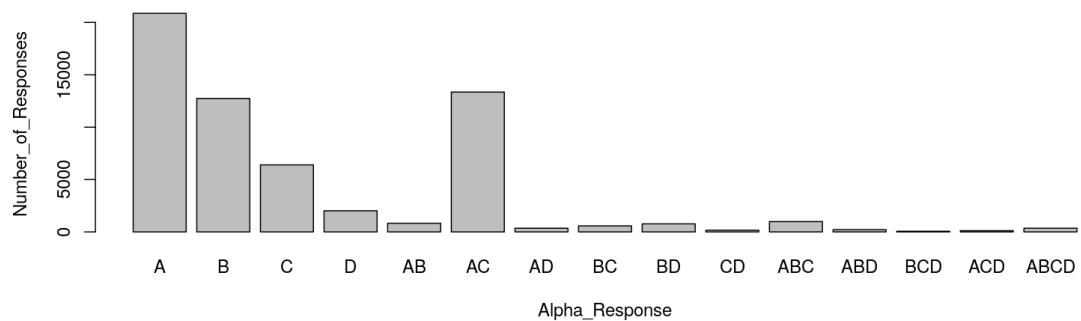
B 9 and 5

C 0 and 4

D 6 and 6

**Figure 3. Grade 3 Item 17 item assessing a number and operations curriculum standard**

59,884 students responded to this question on the Interim Assessment and the frequency of each of the letter responses is shown in both graphical form (Figure 4) and tabular form (Figure 5) below. The table converts the number of students into a percentage and also provides the partial credit score for each combination of letters.

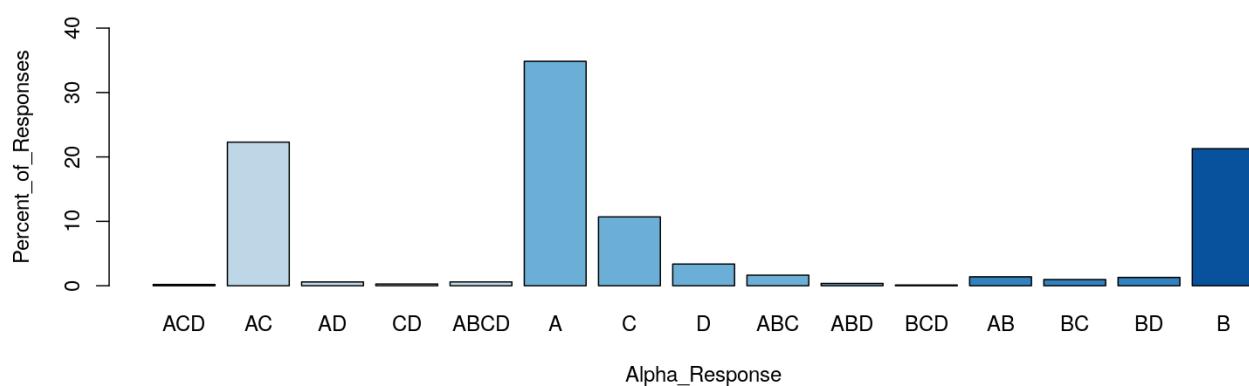


**Figure 4. Number of responses for each combination sorted by letter**

Alpha Responses	Number of Responses	Percentage	Scores
ACD	117	0.20	0.00
AC	13351	22.29	0.25
AD	354	0.59	0.25
CD	162	0.27	0.25
ABCD	357	0.60	0.25
A	20866	34.84	0.50
C	6401	10.69	0.50
D	2018	3.37	0.50
ABC	984	1.64	0.50
ABD	213	0.36	0.50
BCD	66	0.11	0.50
AB	829	1.38	0.75
BC	576	0.96	0.75
BD	771	1.29	0.75
B	12745	21.28	1.00

**Figure 5. Table of letter combinations, number and percentage of students selecting, and partial credit scores**

Teachers from central Texas involved in the pilot implementation of the items appearing on the IAP preferred that the letter combinations and partial credit score be integrated into one graphical display. Figure 6 shows this display. The letter combinations are shaded and sorted from lowest partial credit score (light shading) on the left up to the full credit score (dark shading) on the right (meaning, for this item, *selecting B and also not selecting A, C or D*).



**Figure 6. Percentage of responses sorted and shaded by partial credit with full credit, or exact match, at far right**

Consistent with results for other PBIs appearing on the IAP, many students selected only one letter response (either A, B, C, or D). It is likely students' previous experience with single-select legacy items may have resulted in selecting what they see as the "one best answer" instead of selecting "all correct responses." Within a pattern-based analysis, however, the selection of one response, even when multiple correct responses are available, provides information about the response that the students may have the most confidence is correct. Then, when viewed in aggregate, the analyses of patterns in single selections by students (e.g., either A or C for this example) can be combined with the analyses of patterns in combinations of responses (e.g., AC for this example) to provide meaningful insight capable of informing instruction.

For this multi-column addition item, the patterns of selecting only A or only C or the combination of A and C are related by students successfully identifying combinations of values that sum to four in the tens place, but not recognizing that they *also* need values that will increase the sum in the hundreds place from eight to nine. By selecting A and/or C, students are able to show what they do know – and receive partial credit – even as the statewide results make it clear that more work needs to be done related to carrying values into the next column. A focused snapshot of where the students are in their development is provided in a way that can directly inform an instructional response. With just this one pattern-based item and an integrated display of the student results like that shown in Figure 6, both depth of student understanding and precise diagnostic information related to statewide curriculum standards are identified.

### Grade 5 Introductory Example – Language Arts

The ability of pattern-based items to assess depth of understanding within an item is also significant in the analyses of the IAP reading results. Figure 7 shows both item 16 as it appeared on the Interim Assessment and the responses from 69,708 fifth grade students. For the pattern-based language arts items, students were asked to respond to questions about a reading passage. These language arts items appeared on both the Interim Assessment and on pilot assessments implemented in central Texas.

For this pattern-based item, the reading passage was a poem, "The Wind and the Moon" by George MacDonald.

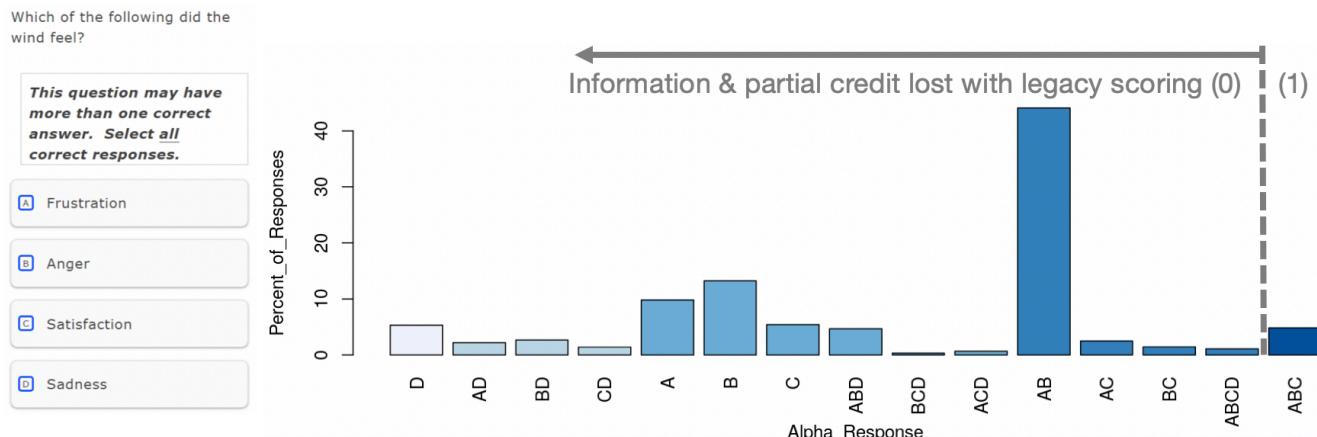


Figure 7. Grade 5 reading item 16 with indication of information and partial credit lost with dichotomous scoring

The students receive full credit for *selecting* A, B and C and *not selecting* D. For all of the pilot implementations, after completing the assessment the students received their results as well as

immediate item-by-item feedback. In the context of illustrating how the pattern-based items are able to assess depth of understanding for reading, it is useful to include the feedback provided in the pilot implementation.

[A] Line 17 states, “He turned in his bed; she was there again!” This suggests frustration that his previous effort to blow out the moon didn’t work. [B] Line 45 says, “He flew in a rage...” Rage is a kind of anger. [C] Phrases like “I’ve done for that Moon” (line 15) or “Sure and certain the moon was gone” (line 36) suggest that the wind was satisfied every time he thought he had covered the moon. [D] At no point in the poem is the wind sad.

Some of the correct responses (C in this case) reflect a deeper understanding than do others (A or B). When taken as whole, the item sets a high bar for receiving full credit. Only 5% of the nearly 70,000 students who responded to this question received full credit (selected A, B and C and not selecting D). Nonetheless, with a focus on patterns in student responses, over 48% the students would receive 0.75 partial credit for combinations with two of the three letters A, B or C selected and D not selected. Similar to the math example discussed previously, within *one* pattern-based reading item students are able to show what they do know and receive partial credit. Consequently, educators are provided with information about depth of student understanding and instruction can be focused on further improving students use of the text to clarify, or justify, their responses.

Much of the information about student understanding, as well as the associated partial credit, would be lost with dichotomous scoring. To illustrate the scope of the loss, if dichotomous scoring was applied to the results from this reading item, then all of the combinations to the left of the “exact match” shown in Figure 7 would be scored as a “0”. Any attempt to make up for the loss of information would require the use of additional items and each of these items would need to be focused the same curricular standard. This would have – and does have – the consequence that tests meant to assess specific learning standards using dichotomous items need to be longer.

Given the need to deploy multiple items to more fully assess student learning outcomes, there is an additional challenge related to how to represent the results across these multiple items. To attempt to match, through the continued use of IRT-based approaches, the kinds of integrated representations of results already supported by the use of PBIs (and illustrated in Figure 6), new forms of detailed, *cross-item*, analyses – well beyond that of a derived scale score – would need to be developed and deployed.

The complete set of the pattern-based items appearing on the 2018-2019 Interim Assessment, including the full reading passage for the fifth-grade language arts items, and the related results are provided in Appendix A. For each question, a sorted display of percent combinations, like that in Figure 7, is included.

### [Introduction to the Significance of Moving to an Information-based Theoretical Model](#)

With PBIs *not being limited to the use of item response theory* as the overarching psychometric model, pattern-based assessments, as illustrated in this report, are designed and developed with the intention of being able to assess the depth of student understanding within even a single item. An *information-based theoretical model* supports the development of each pattern-based item and this model can be readily extended to characterizing patterns across related items, or even across an assessment as a whole.

For dichotomous scoring there are only two states per item. This implies the theoretical upper limit for the information provided by a non-polytomous, IRT-based, assessment with  $N$  questions is  $2^N$  states. For the PBIs used on the Interim Assessment, the information provided by an assessment with the same number of items would be up to  $16^N$  states.

To illustrate the significance of this difference in the context of the implementation of the Interim Assessments, if the possible scores on the *thirteen dichotomous legacy questions* appearing on the Grade 3 mathematics 2018-2019 Interim Assessment are taken together, then the maximum number of states that can be represented across these items is  $2^{13}$ , or 8192, states. In contrast, with *only the five PBIs* appearing on the same Grade 3 IA, the maximum information that can be represented is  $16^5$  or 1,048,576 states. Not only is each PBI designed to provide up to eight times more information than a dichotomous legacy item but, by extending an information-based theoretical model across items, the five PBIs used in this comparison can, nonetheless, provide more than 100 times as much information related to student outcomes as the thirteen legacy items appearing before them on the same assessment.

Looking ahead to what deploying information-based, cross-item analyses at scale will be able to support in the near future, advanced computational techniques for characterizing patterns in information-rich datasets can be used to identify results associated with students who may have specific, as yet undiagnosed, learning challenges (e.g., dyslexia). Once the challenges have been identified using these computational techniques, focused follow-up assessments for the students can be carried out and, as needed, effective individualized instructional responses can be developed.

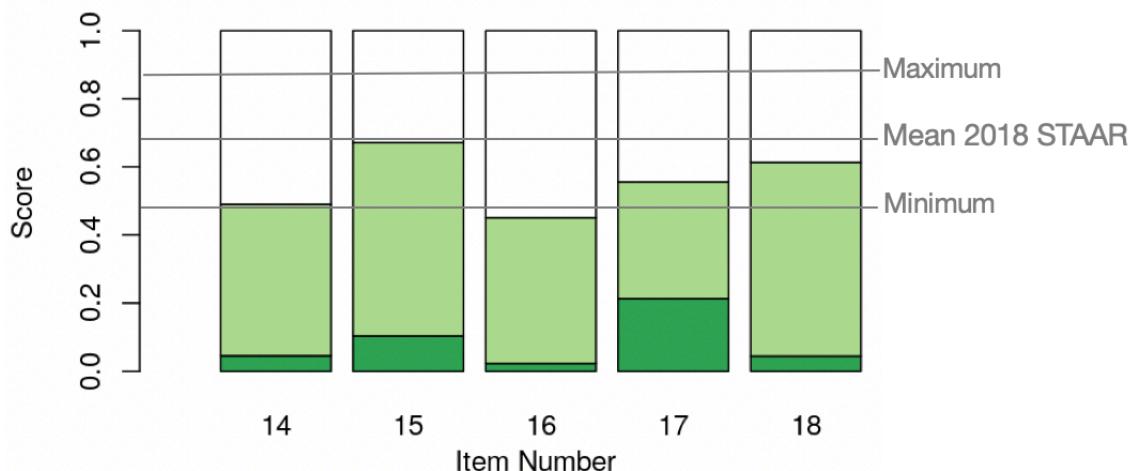
For the purpose of this report, however, the immediate focus remains on the analyses of the 15 states for the pattern-based items appearing on the Interim Assessments (students were not required to respond to the PBIs on the IA and, as a result of this aspect of the implementation, we were not able to make the 16<sup>th</sup>, no response, state a meaningful part of the analyses). Overall, the analyses can be seen to illustrates the utility of having items and assessment that are designed from within an information-based theoretical model.

### Partial Credit and Full Credit Comparisons

A number of the legacy items appearing on the non-optional portion of the Interim Assessments were drawn from items implemented on the Spring 2018 STAAR. For the purposes of making comparisons in results, the Texas Education Agency provided the respective *STAAR Item Analysis Summary Reports* for each of the grades and subject areas that overlapped with the implementation of the pattern-based items. These summary reports include the percent of students selecting the correct response for each single-select item. For comparison to results from the pattern-based items, the mean percent correct as well as the maximum and minimum percent correct for the STAAR items were computed.

Figures 8 and 9 present the overall partial credit results for each of the third grade, pattern-based items appearing on the Interim Assessments. The dark shading within each bar of the graph depicts the portion of students who received full credit. The values for the minimum percentage, maximum percentage and average percentage of students receiving full credit for the Grade 3 Spring STAAR results are represented as horizontal lines in the graph. The percent of students receiving full credit on the PBI's ranged between

2% and 21%. The partial credit scores, however, are generally within the range of the percent correct for the Spring 2018 STAAR results.



**Figure 8. Graph of full credit and partial credit results for Grade 3 Math Items as well as 2018 item-level STAAR results.**

Percent Exact Match	Average Score	Item Number	Total Responses
5	0.49	14	60517
10	0.67	15	60212
2	0.45	16	59976
21	0.56	17	59884
4	0.61	18	59561

**Figure 9. Table of full credit and partial credit results for Grade 3 Math PBIs**

As noted in the Implementation section of this report, only the student responses to the PBIs are in the IAP dataset. Absent results from the legacy portion of the assessments, or any demographic information about the students who participated, descriptive statistics cannot be provided for comparing the population of students participating in the IAP to the much larger population of students whose results are summarized in the *2018 STAAR Item Analysis Summary Reports*. Based on the *STAAR Item Analysis Summary Reports* the minimum, maximum and mean percent correct for the STAAR results tend to decrease in subsequent grades from what is shown in Figure 8. Over these same grades the corresponding PBI partial credit scores tend to align more centrally relative to the STAAR results.

The graphs and tables for partial credit and full credit match results for all subject areas are included in Appendix B. For comparison purposes, a table of results from the *STAAR Item Analysis Summary Reports* is also provided in Appendix B.

#### Pilot Comparison of Released STAAR Item and a Corresponding PBI

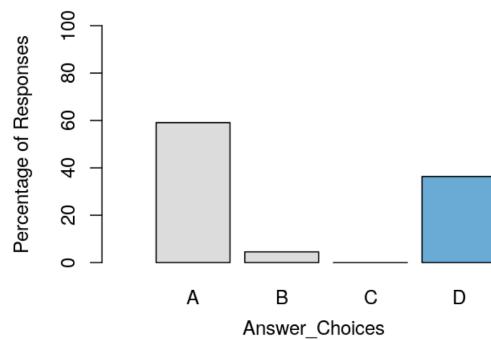
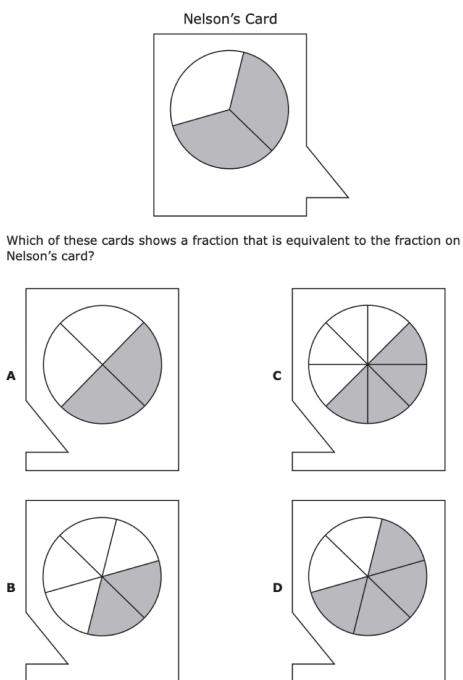
To help ensure the close alignment of the PBIs appearing on the Interim Assessments with the Texas Essential Knowledge and Skills (TEKS) and the Reporting Categories found in the respective STAAR

Blueprints, each PBI was developed to parallel a released STAAR item. Figure 10 shows the released Grade 3 2016 STAAR item that corresponds to the Grade 3 pattern-based item that appeared on the 2018-2019 Interim Assessment. The Grade 3 pattern-based item is shown in Figure 11. Both questions ask students to identify “equivalent” fractional areas.

For the released STAAR item there is only one equivalent fraction that appears in the responses. For the corresponding PBI there are three responses (A, B and C) representing fractional areas that are equivalent to the fractional area appearing in the prompt and one response (D) with a fractional area that is not equivalent. The three equivalent response were developed to assess increasing depth of understanding in moving from B (rearranging the three same-shaped vertical areas), to A (equivalently shading two horizontal shaped areas), to C (working with the shading of a triangular area).

### Item 25 on Grade 3 STAAR 2016

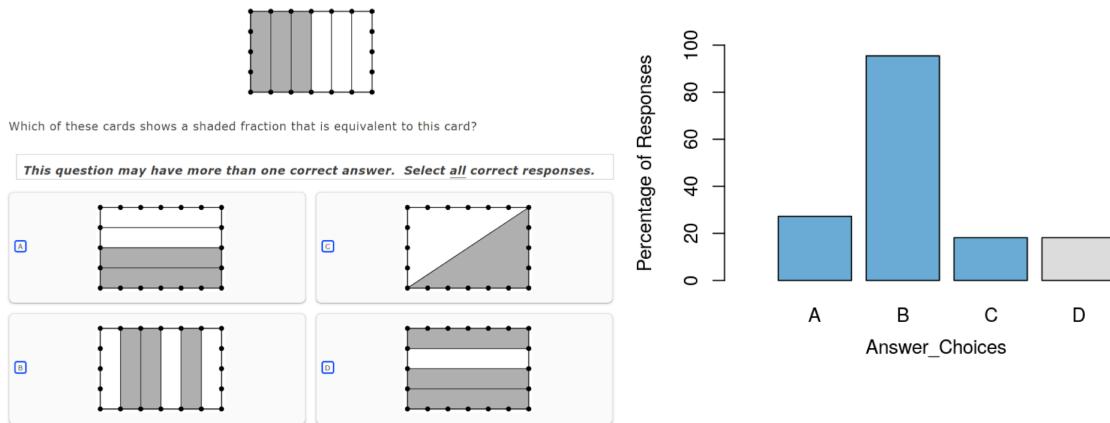
Nelson is playing a math game. He needs to match two cards that show equivalent shaded fractions.



**Figure 10. Released STAAR item and results from pilot implementation in one third grade classroom**

### Item 14 on Grade 3 IAP 2018-2019

A group of friends is making cards for a fraction game. The cards need to show equivalent shaded fractions. They start with this card.



**Figure 11. Corresponding pattern-based item and results from pilot implementation in one, third grade, classroom**

To begin to illustrate the explanatory capabilities of pattern based items, both within a given classroom and then across the statewide implementation, results are included in Figures 10 and 11 from a pilot implementation in one third grade classroom ( $N=22$ ) of an online assessment where both the specific STAAR items and the corresponding PBIs were included. This juxtaposition of legacy items and PBIs on one assessment is similar to the implementation of items in the Interim Assessment Program.

What is particularly noticeable about the results for the released STAAR item is how many more student selected the incorrect response A than the correct response D. Based on the dichotomous scoring method of the 2016 STAAR, also used for the legacy items appearing on the 2018-2019 Interim Assessments, these students would receive a score of “0” as an assessment of their ability to identify equivalent fractional areas. With the PBI, 21 of the 22 students (more than 95%) correctly identified B as an equivalent fractional area, 27% correctly selected A and 18% correctly selected C. Clearly the PBI allowed students to show what they do know in a way that the all-or-nothing results for the released STAAR item do not. Using partial credit scoring, all the students in this class would receive at least 0.25 credit, most would receive between 0.5 and 0.75 as credit. In terms of receiving full credit, however, this item is more demanding than the released STAAR item. Only 4 of the students would receive full credit on this item whereas 8 students would receive full credit on the release STAAR item. For this third-grade class, the PBI allowed all the students to receive some credit for what they do know while also setting a significantly higher bar for students to receive full credit (1) than the released STAAR item.

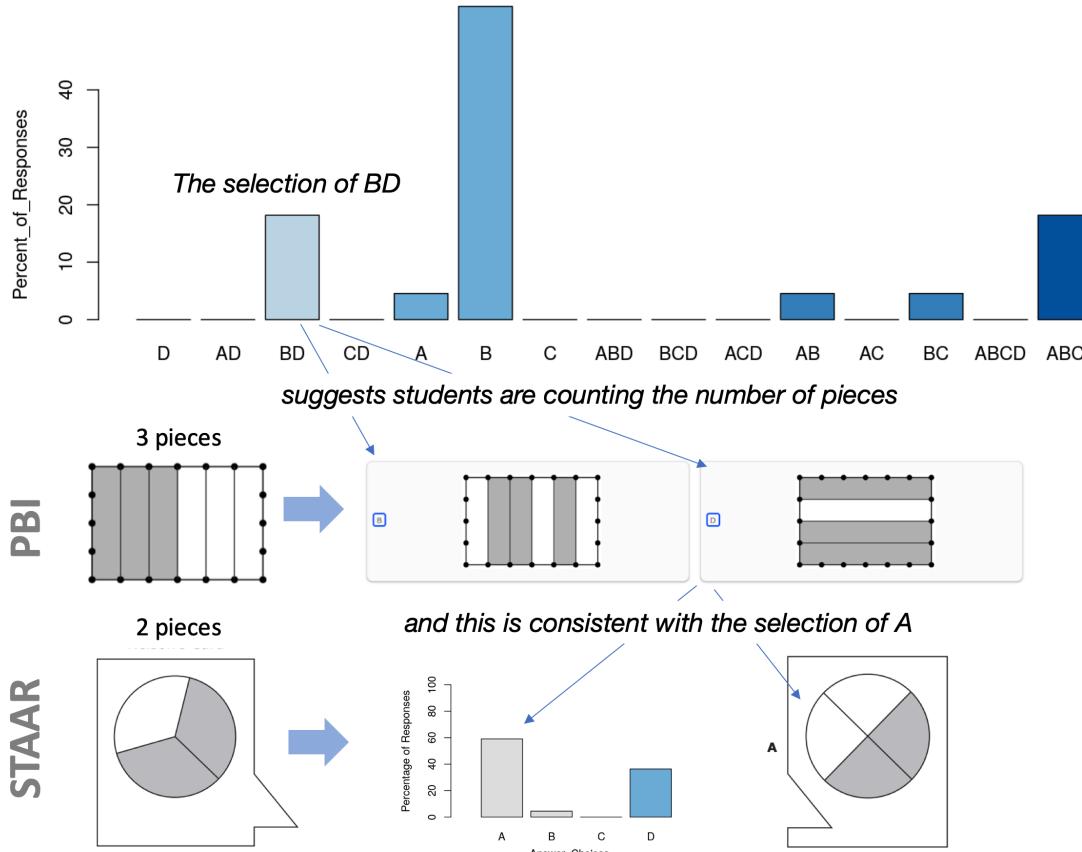
Beyond assessing what student do know and setting a higher bar for full credit than the released STAAR item, is the significant capabilities associated with being able to focus on patterns in student responses. As an example, we can use a frequently occurring pattern from the implementation with 22 students to identify a significant learning issue related to fractions. We can also use this result to explain why so many students selected the incorrect response A to the released STAAR item

Rather than limit the analysis of results to how many students selected each of the responses (see Figures 10 and 11), an essential feature of using a pattern-based approach in assessment is the ability to address the significance of *combinations* in the responses. For this fractional equivalence PBI, this means

the interpretation of a correct response like B can change in its significance based on whether it occurs alone or in combination with other responses. B, when selected with A but not C, would indicate students are able to identify two important equivalent representations. When B is combined with D, however, the combination points toward a distinct account of how student are reasoning about fractions. This account has explanatory significance both for the PBI results and for the results from the corresponding STAAR item.

Figure 12 shows the display of combinations selected by students for the PBI. When elementary teachers were asked during a workshop why students might choose D, an incorrect response, they suggested the students may be thinking of equivalence in terms of “counting pieces” instead of reasoning in terms of equivalent fractional areas of the whole rectangle. For the students selecting D, the given area has three pieces and response D also has three pieces. It would be consistent with this explanation if these students also selected the correct response B, which also has three pieces. In this context, B in combination with D would indicate the students were not reasoning about fractions in relation to a unit whole (the rectangle), as would be indicated by B being selected on its own or in combination with A and/or C.

The display of the combinations of responses shown in Figure 12 allows the frequency of B occurring with D to be identified separate from B occurring on its own or in combination with A and/or C. The suggestion by teachers that students selecting D are counting pieces, and not thinking in terms of fraction of a whole, is supported by these students also selecting B. As is summarized in Figure 12, the selection of the combination BD suggests students are counting the number of pieces in the responses and, further, that this way of thinking about equivalence is consistent with students selecting A for their (one) response to the released STAAR item. For the PBI, the prompt has three pieces and as a result these students chose responses with three pieces (B and D). For the STAAR item, the prompt has two pieces and the students chose the (incorrect) response also having two pieces (A).



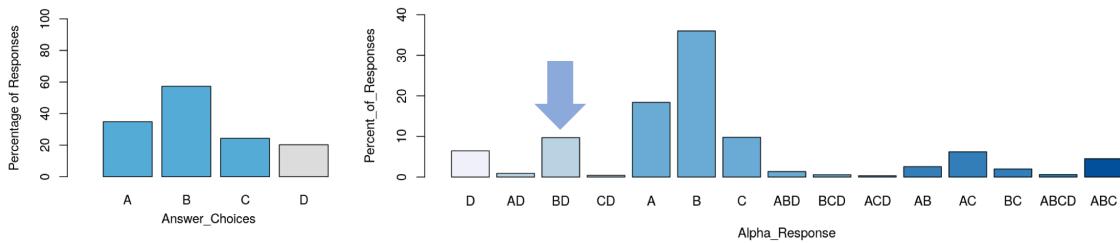
**Figure 12. Illustration of the explanatory capabilities of PBIs**

As with the three-column addition example discussed earlier, the combination of selected responses (e.g., for the earlier example identifying values that add to four in the tens place but that do not increase the value in the 100's place or, in this case, D occurring with B) provides significant potential insight into students' understanding and, as will be discussed in a later section, has implications for how to improve classroom instruction. The next section illustrates how the patterns hold up across scale in moving from one classroom's results to statewide results.

### Moving from Classroom Results to Statewide Results

Using the same pattern-based item discussed in the previous section, the effectiveness of pattern-based assessment in providing consistent meaningful information *across scale* can be illustrated. The analysis moves from one third grade classroom,  $N = 22$  students, to results from the statewide Interim Assessment,  $N = 60,517$  students. By design, the available responses for pattern-based items are used to assess depth of understanding. For this item, the relative percentages of students selecting B, A and C for both the individual classroom (Figure 12) and statewide (Figure 13 below) are consistent with the increasing depth of understanding students would need in moving from B (the same number of bars of the same shape), to A (horizontal bars), to C (dividing the area diagonally).

As important for comparison purposes is how the *patterns in combinations of responses* continue to be informative across scale.

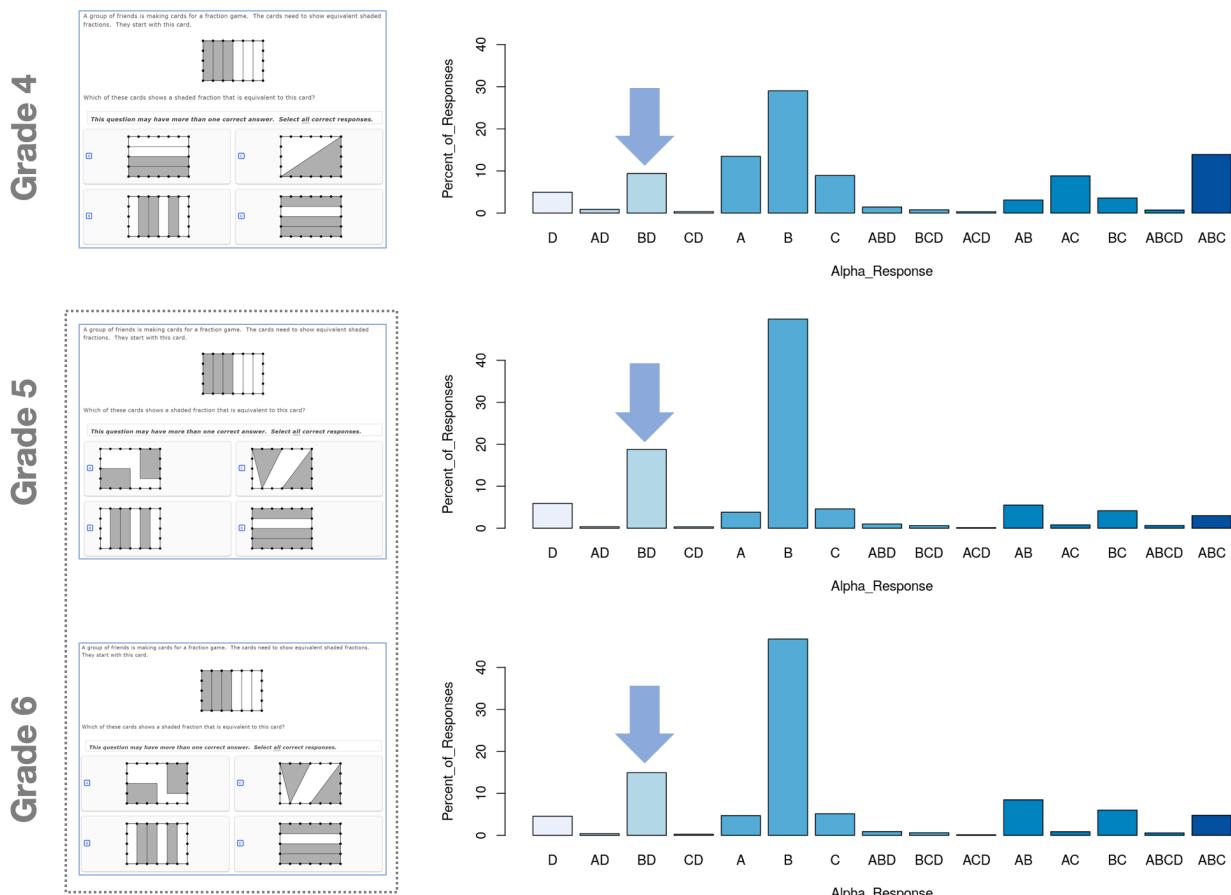


**Figure 13. Moving to full display allows BD combination to be identified**

The BD combination discussed earlier also occurs more frequently than selecting D alone. This suggests there is a particular significance to this combination (see arrow in Figure 13) *that would not be visible without the full display of states shown*. As was discussed earlier, this pattern is consistent with students “counting pieces” (further discussed later in the report), and not thinking of fractions in relation to a specific “whole” (in this case the area of the given rectangle).

Not only does this specific pattern appear across scale but by repeating the use of the same item on the grade 4 Interim Assessment, and then preserving the B and D responses in the corresponding grade 5 and grade 6 items, it is possible to track the relative presence of this persistent pattern of student reasoning across multiple years.

## Identification of Meaningful Patterns Across Years



**Figure 14. The same BD combination appears across grades**

Before going on to discuss possible instructional responses, given the importance of fractions in the early mathematics curriculum – and then discussing a subsequent set of related issues for ratio and proportion, which are highlighted in middle-school curricula (and beyond) – it is worth illustrating how the identification of patterns with a second item can add significantly to framing our understanding of student learning outcomes relative to multi-year curricular standards (e.g., comparing fractions).

### A Related Example of Patterns Being Informative Across Scale

The PBI shown in Figure 15 asks students which statements “must be true” about two friends eating lunch together. One friend, Felicia, “ate  $\frac{1}{3}$  of her sandwich” and the other friend, Daria, “ate  $\frac{1}{4}$  of her pie.” The responses require students to attend to the fractions being expressed relative to either the sandwich as a whole or the pie as a whole. Response A is correct because  $\frac{1}{3}$  is a “greater part” of Felicia’s sandwich than  $\frac{1}{4}$  is a part of Daria’s pie. As is then shown in Figures 15, more than 50% of the respondents were able to identify the comparison of fractions,  $\frac{1}{3} > \frac{1}{4}$ , correctly.

Once again, a display of all the combinations provides significant additional insights related to student understanding. As with the combination of B (a correct response) with D (an incorrect response) in the previous example, the meaning for this item of a correct response when combined with an incorrect response is distinct in its interpretation from either response being on its own.

Of particular relevance to fractions being expressed relative to a specific whole, response B is incorrect because it may not be true that Felicia “ate more food than Darian ate.” While  $\frac{1}{3}$  is a greater fraction of the sandwich than  $\frac{1}{4}$  is a fraction of the pie, there is no sufficient way, based on the information provided, to compare how much food each of these fractions represents. Without more information, the “whole” of the sandwich cannot be related to the whole of the pie. As a result, B being correct does *not* follow from A being correct. As presented, it is not the case that B “must be true.”

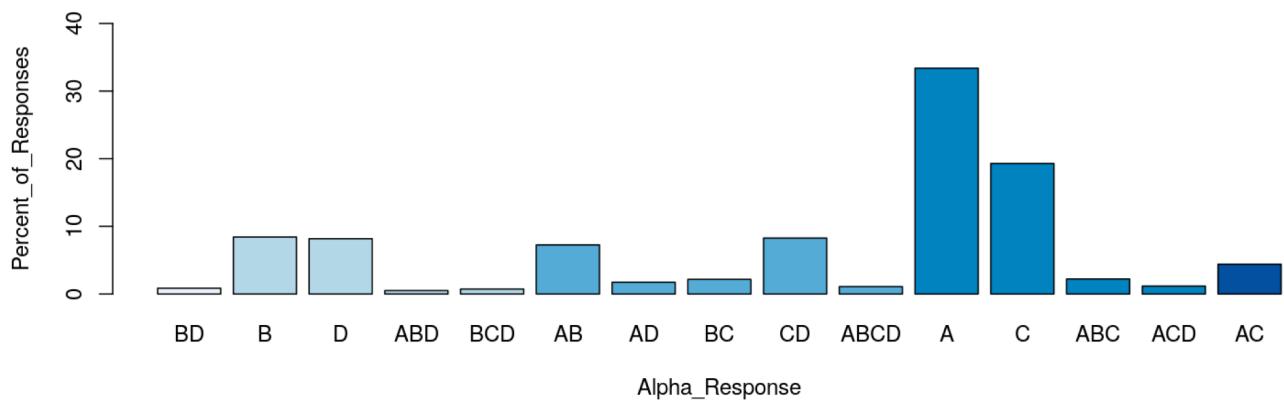
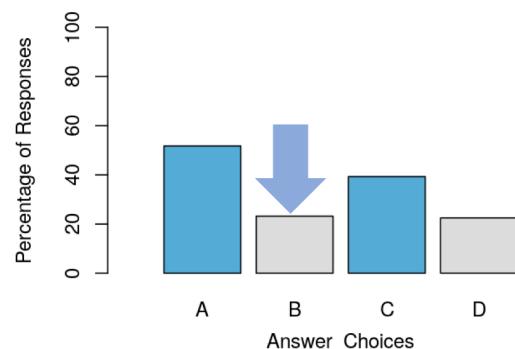
Two friends ate lunch together.

Felicia ate  $\frac{1}{3}$  of her sandwich. Daria ate  $\frac{1}{4}$  of her pie.

Which of these statements must be true?

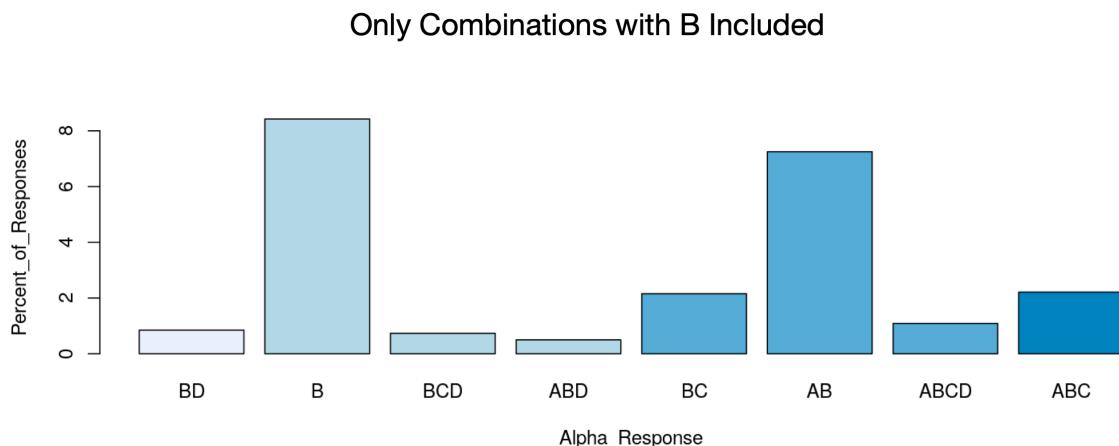
*This question may have more than one correct answer. Select all correct responses.*

A Felicia ate a greater part of her sandwich than Daria ate of her pie because  $\frac{1}{3} > \frac{1}{4}$   
 B Felicia ate more food than Daria ate  
 C  $\frac{2}{3}$  of the sandwich was not eaten by Felicia  
 D  $\frac{3}{3}$  of the pie was not eaten by Daria



**Figure 15. Result from a second fraction-related item.**

To further illustrate the importance of the additional information provided by this PBI, Figure 16 and Figure 17 show the relative percentages only for combinations that include B. The AB combination accounts for 31% of these responses, and nearly half (48%) of the responses that included B also included A (e.g., ABC). As is the case in general with PBIs, students are able to show what they do know and receive partial credit. The combination ABC, even though it includes B, receives partial credit of 0.75 because this response involves *selecting* both correct responses and *not selecting* one of the two incorrect responses. It is also important that the results from this item make visible the students who do not realize that comparisons of different wholes, or units, cannot be made without additional information.



**Figure 16. Display of only the combinations including B**

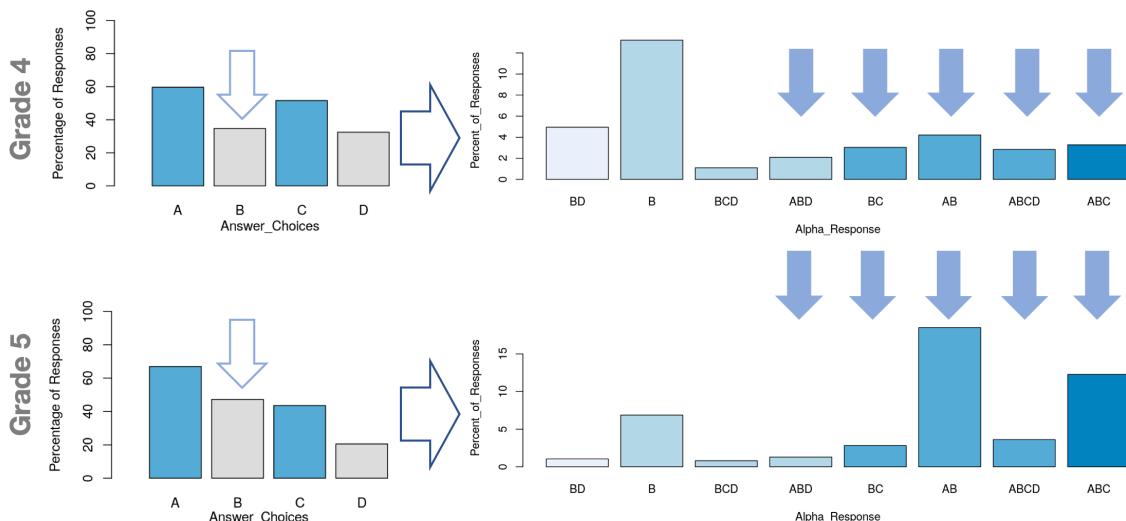
Percent of Responses	Alpha Responses
3.7	BD
36.3	B
3.2	BCD
2.1	<u>ABD</u>
9.3	BC
31.2	<u>AB</u>
4.7	<u>ABCD</u>
9.5	<u>ABC</u>

**Figure 17. Table of percents for combinations including B**

What these results suggest for instruction – in a way that goes beyond what would be the case if A and B were treated independently – is that while many students can correctly compare the fractions  $\frac{1}{3}$  and  $\frac{1}{4}$ , they also need to understand better the importance of fractions being expressed relative to a specific whole or unit. The AB combination, like with the BD combination discussed with the previous example, provides additional information about student understanding to educators. Greater depth of understanding is required to be able to both compare fractions (A) and also realize that B may not be the case.

Similar to the example discussed previously, this item was repeated on the Grade 4 and Grade 5 Interim Assessments. A more informative account is provided by moving from the aggregate percentages of students selecting each response to the combinations of responses. For this item we can focus on all the responses that include B.

## Increased Information Provided from Answers Including B



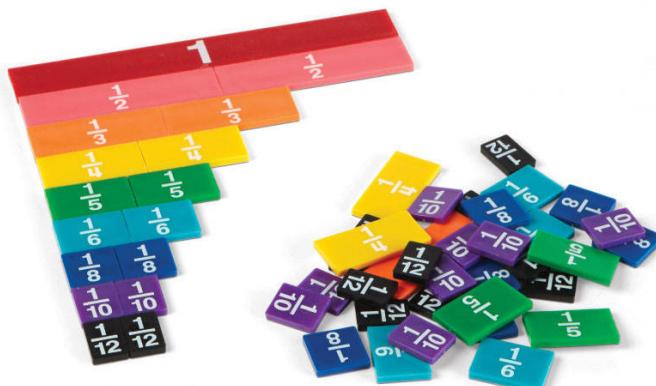
**Figure 18. Analysis over grade 4 and grade 5 results of combinations involving both B and A**

In Figure 18, the arrows on the right point to all the combinations where the correct response A (comparing fractions) is combined with the incorrect response B (not realizing that the fraction comparisons are relative to two different wholes).

Teachers at PBI-related professional development workshops discussed how current classroom practices could result in students “counting pieces” (first item discussed earlier) or in students not realizing that comparisons of fractions have to be relative to a given unit or whole (the second item). They also discussed ways to improve students’ understanding. For the purposes of this report, their insights illustrate how the increased information provided by PBIs can be used to directly inform professional development aimed at improving classroom instruction and, ultimately, student outcomes.

### Practices to Improve Outcomes – Closing the Loop Back to instruction

In reflecting on how current classroom practices might result in students “counting pieces” and not referencing respective unit wholes, teachers noted that many versions of fraction manipulatives come with the blocks already labelled relative to a given unit.



**Figure 19. An example of a commonly used type of fraction manipulative**

Although the intention is for students to always realize the respective pieces, and corresponding fraction labels, are relative to the unit whole shown in the upper left of the image in Figure 19, the teachers at the workshops suggested that students might readily dissociate the pieces from any reference to the given unit whole.

Then, as might be suggested by the pile of pieces on the right, students might simply count the number of pieces labelled “1/6.”  $3/6$  would then just be a way of counting how many of these pieces you have (that is, three pieces that happen to be labelled  $1/6$ ) without any connection to the unit whole. In addition, students would be unlikely to realize that the unit whole can change, and that this would have the effect of changing how the pieces would need to be labelled (e.g., a block labelled a sixth could change to become a third, if the unit whole was changed).

To address the confusion that can be created, an experienced teacher at one of the workshops described how the first thing she has students do is turn the tiles over, so that the labels are no longer visible. During her lessons she will change which piece is to represent the unit whole, and then ask students to show her what  $1/3$  would now look like (students hold up a piece) or what  $3/4$  would look like (students point to, or hold up, multiple pieces). Then to further deepen students’ understanding, she would hold up a piece and ask the students to come up with, and be able to explain, “What other fractions could this piece be?” Once the students realize they can now “create their own wholes,” she reported students “become very creative” and will combine pieces to make increasingly complex wholes and then work hard to figure out what fraction the given piece would be of their “more interesting” unit whole.

Another teacher, after reviewing the results from the first item, suggested the task described in the item “could make a good classroom activity.” Students would start with a card shaded as shown in the item, and then be asked to come up with their own unique shadings of additional, same-sized but blank, cards. The “rule of the game” would be that the new “shaded fraction” of the area had to be “equivalent” to the shaded fraction of the original card. Students could then share and discuss their cards with the class.

The fraction manipulatives activity would likely help students realize why they cannot compare the amount of food Felicia and Daria ate without knowing more about the units or wholes (the sandwich or the pie in the second item discussed earlier). One teacher, in the midst of the discussion of this activity, suggested illustrating the issue by asking students, “If you were really hungry, which would you want, one-third of a small donut or one-quarter of a really big pizza?” It is also much less likely students would simply “count pieces” to decide what fractions were the same (i.e., the BD combination from the first item discussed). The second activity might also support students in developing a deeper understanding of related fraction concepts. The students would be likely to realize that there are many ways of shading an equivalent area and, further, they might be much less likely to resort to simply dividing the area into an arbitrary number of pieces and then shade in (count out) three of these pieces (D in the first item).

Whether or not these practices actually result in better student outcomes would need to be assessed empirically. Nonetheless, these suggestions by teachers illustrate how assessment and improvement in instruction can be better integrated by the use of items that provide much more detailed and precise information than is possible with existing, dichotomously scored items. More information makes it easier to “close the loop” back to improving instruction in classrooms.

## Comparing Differences Across Scale

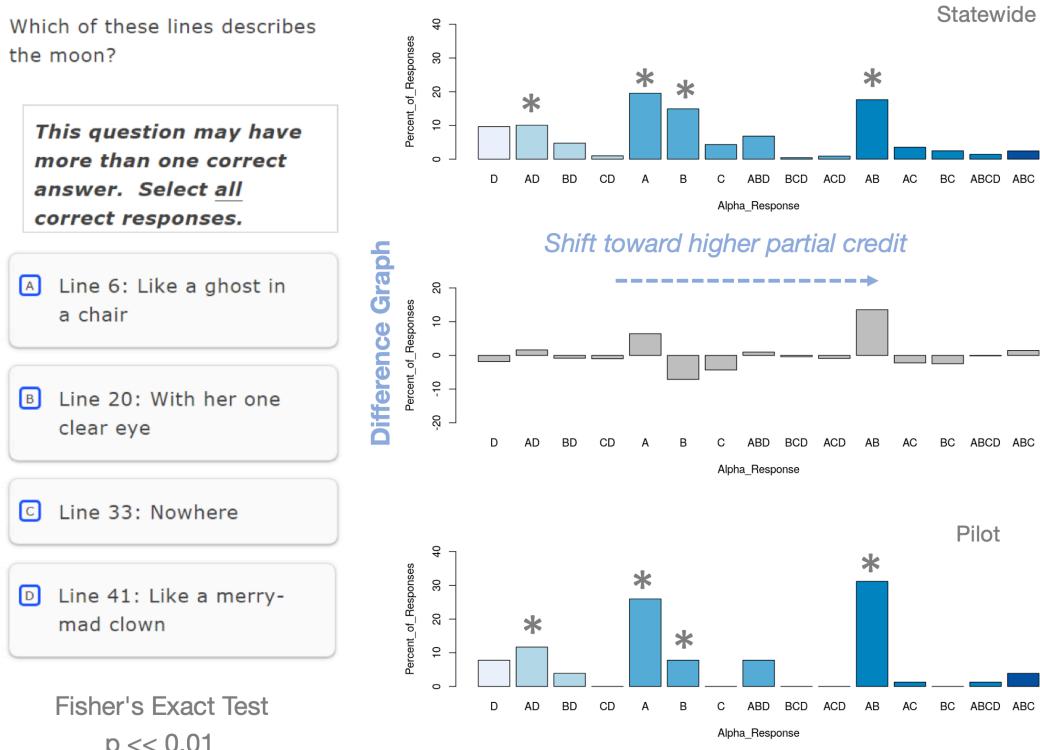
In addition to results from pattern-based items providing meaningfully information about learning outcomes that can inform instruction, the results from PBIs can also support comparisons of outcomes between groups of students. In this section this capability is illustrated by comparing differences in results *across scale* for both reading and mathematics. Then, in a subsequent section, comparisons will be made be made *over time* between grade levels.

### Differences in Reading Results Across Scale

Figure 20 shows a comparison of combinations of reading responses in moving from the statewide results (top) to results for the same item from a pilot implementation at an elementary school ( $N=77$ ). The item asks which of the lines from the reading passage describe the moon – one of the two main characters in the passage. For the pilot implementation student received immediate feedback on their responses.

[A] This line is the wind talking about the moon. [B] This line is also about the moon, shaped like “one clear eye.” [C] “Nowhere” is followed by “Was a moonbeam bare,” making it clear this is about the moon. [D] The wind was “in town” followed by “Like a merry-mad clown.” This is about the wind, not the moon.

As is borne out in the results, the responses required different levels of attention to the text, with some references to the moon being easier to discern (A) than others (B and then C).



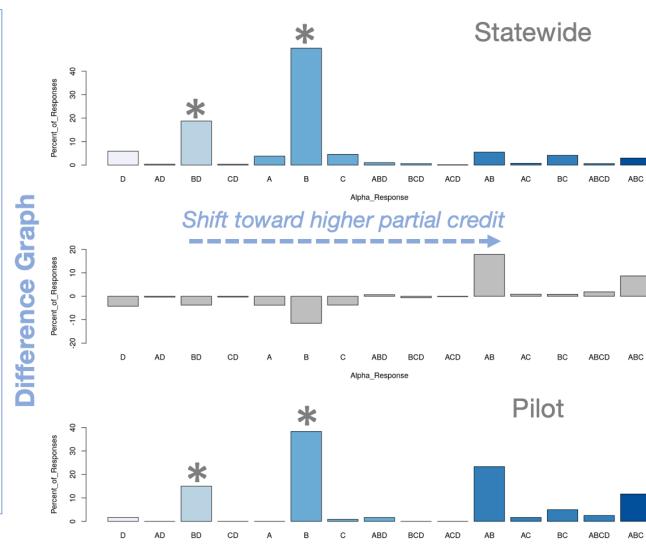
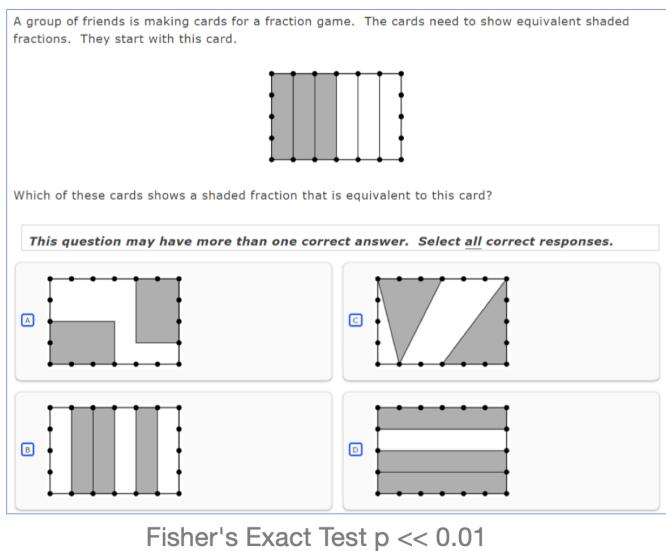
**Figure 20. Analysis of differences in reading results across scale (statewide to pilot)**

Similar to what was reported previously, the patterns in responses remain relatively consistent across scale. This is illustrated in Figure 20 with asterisks placed over the four most frequently occurring responses for the statewide results at the top and for the pilot implementation at the bottom.

Even with this consistency, however, an evaluation of differences in comparative outcomes also can be made. The graph in the middle of Figure 20 plots the changes (differences) in the percentages for each of the combinations of responses. Generally, there is movement from lower partial credit responses to higher partial credit responses. The respective percentage receiving full credit are 2% and 4%, and the respective partial credit scores are 49% and 55%. With the p-value for a non-parametric Fisher's Exact Test being much less than 0.01, these results are both highly statistically significant and consistent with the positive differences in results at the pilot site representing a better overall student outcome for this item.

### Differences in Mathematics Results Across Scale

A comparison of combinations of responses to a pattern-based mathematics item is shown in Figure 21. As was the case with the reading item, the statewide grade 5 results are shown in the top graph and the results from a grade 5 pilot implementation ( $N= 121$ ) are shown in the bottom graph. This item is the grade 5 version of the grade 3 fractional area questions discussed earlier in this report. Asterisks are placed over the two most frequently occurring responses for the statewide results at the top and for the pilot implementation at the bottom. They are indicative of the consistency of the patterns across scale. Similar to the grade 3 version, the responses are designed to assess a range of depth of understanding from B, to A and then to C. The A and C responses for the grade 5 version require a more in depth understanding than the corresponding responses for grade 3. The significance of selecting B versus BD was addressed earlier and remains the same across the grade 3 and the grade 5 versions.



**Figure 21. Analysis of differences in mathematics results across scale (statewide to pilot)**

Changes in the percentages for each of the combinations of responses support an evaluation of the comparative differences in outcomes across scale for the two groups of students. The overall shift from lower partial credit responses to higher partial credit responses depicted by the difference graph suggests better learning outcomes for the students in the pilot implementation. The respective percentage receiving full credit being 3% and 12% and the respective partial credit scores being 46% and 59% also suggest better outcomes. Using a non-parametric Fisher's Exact Test, the p-value is much less than 0.01. Like with the reading results, the high statistical significance is consistent with the claim that

the learning outcomes for the pilot students were better than for the students in the statewide implementation.

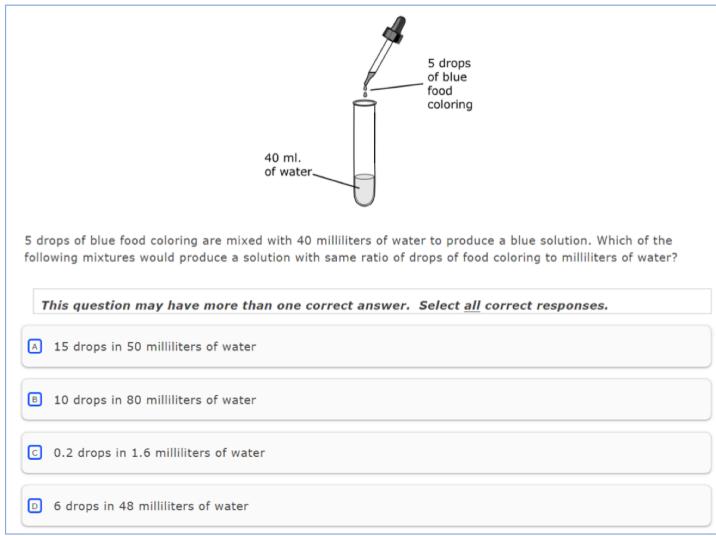
### Comparing Differences Over Time - Assessing Growth

Even absent information about student, school, district or temporal implementation in the Interim Assessment Program dataset, the sensitivity of pattern-based items to student growth can be assessed by comparing results for items deployed across grade levels. The item shown Figure 22 assesses student understanding of equivalent ratios in relation to adding drops of blue food coloring to water to create a blue solution. A key curricular transition in moving from late elementary grades to middle school grades, is extending emergent multiplicative forms of reasoning about fractions to comparisons of ratios and proportions. To be able to assess the development of students' ratio and proportional reasoning in middle school, this pattern-based item was included on both the grade 6 ( $N=58,947$ ) and grade 7 ( $N=46,483$ ) Interim Assessments.

The Grade 6 mathematics TEKS require students to be able to, "apply qualitative and quantitative reasoning to solve prediction and comparison of real-world problems involving ratios and rates" (6.b.4.B) and "give examples of ratios as multiplicative comparisons of two quantities describing the same attribute" (6.b.4.C). To satisfy the Grade 7 TEKS, students must be able to "solve problems involving ratios" framed more explicitly in terms of proportional reasoning (7.b.4.D). With the item shown in Figure 22, students are assessed on their ability to use "multiplicative comparisons of two quantities" (drops of food coloring and amount of water) in a real-world context to describe the "same attribute" of what would be the blue-ness of the resultant solutions.

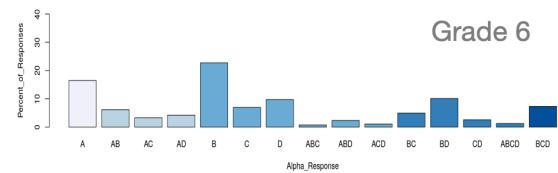
The item closely parallels an activity done in classrooms where students add drops of food coloring to water in clear bottles to create solutions with "the same blue-ness." Often students attempt to extend additive forms of reasoning to a task that requires multiplicative comparisons. Response A is intended to assess whether students are attempting to reason additively – adding 10 to both the number of drops and to the amount of water – about a task requiring multiplicative reasoning. The difference graph shown in Figure 22 reflects a 6.8% decrease in moving from 16.4% of the grade 6 students selecting only A to 9.6% of the grade 7 students selecting only A.

In the overall context of the design of pattern-based items, this decrease illustrates the significance of students *not selecting a response* in contributing to an overall assessment of depth of understanding. In contrast to this additive, incorrect, answer, response B correctly multiplies the 5 drops of food coloring by 2 and the 40 milliliters of water by 2. Response D requires students to simplify the given ratio of 5 drops of food coloring to 40 milliliters of water to the equivalent ratio of 1 drop of food coloring to 8 milliliters of water, and then correctly multiply each quantity by 6. Response C also requires this simplification of the original ratio and then correctly multiplying each quantity by 0.2. Increasing depth of understanding is assessed in this pattern-based item in moving from not selecting A, to selecting B, to selecting D and then, at the highest level of understanding, selecting C.

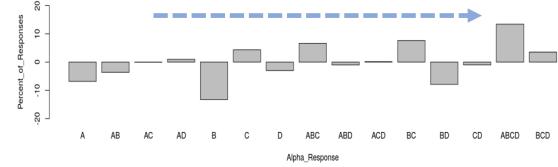


Pearson's Chi-squared Test  $p << 0.01$

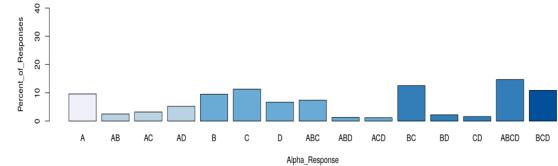
### Difference Graph



*Shift toward higher partial credit*



### Grade 7



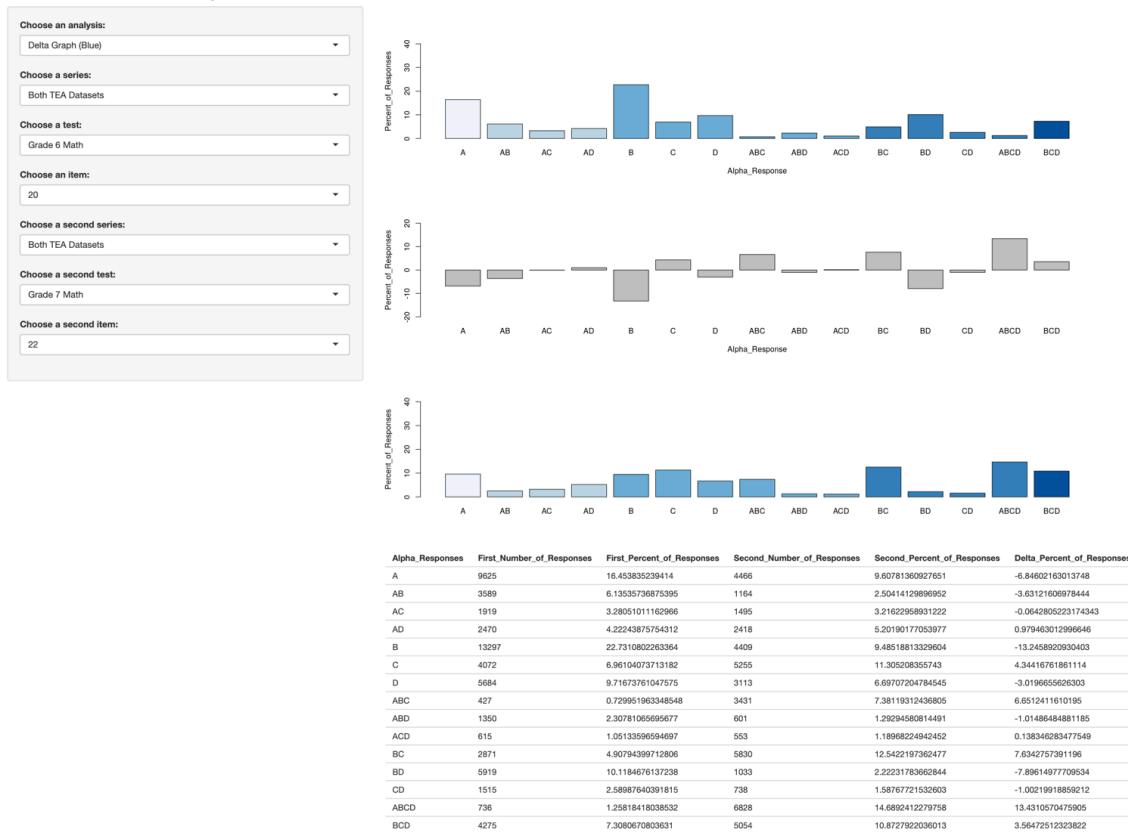
**Figure 22. Analysis of differences in mathematics results over time to assess growth**

As with other comparative results reported in this section, the graph in the middle of Figure 22 shows all the relative changes in the percentages for each of the combinations of responses. Although the shifts are more complex across the various combinations than for the results reported previously, there remains general movement from lower partial credit responses to higher partial credit responses. The respective percentage receiving full credit are 7% for grade 6 and 11% for grade 7 and the respective average partial credit scores are 47% and 56%. For this comparison, the substantial number of students at both grade levels supports the use of a non-parametric Pearson's Chi-squared Test to evaluate statistical significance. With the p-value being much less than 0.01, these results are highly statistically significant. The very low p-value is consistent with the positive differences in results representing an improvement in student outcomes on this item. This improvement over time from grade 6 to grade 7 illustrates the ability of pattern-based items to assess student growth.

### Online Reporting

All the graphs, tables and comparisons used in the preparation of this report were created using a secure, menu-driven, online reporting and analysis system with access to the (fully anonymized) datasets. Figure 23 shows an image of this system as it appears in a browser window.

## GenEd Assessment Report

**Figure 23.** Interactive, online system used to prepare report

Many more analyses and comparisons than can be included in this report are able to be carried out using this system. To support possible additional analyses, as part of the deliverables to the Texas Education Agency, we have provided logins to this automated reporting system. A robust system for integrating these kinds of automated reporting capabilities at-scale with an online implementation of assessments also exists and can be reported on in the future.

## Summary and Conclusion

The central question in using pattern-based items at scale as part of the Interim Assessment Program was whether, with their richer information about student content knowledge, they would generate significant, actionable insight into student learning outcomes. As this report clearly shows, the design of pattern-based items allows assessment to focus on evaluating depth of understanding, giving all students more ways to show what they know. More information about their students supports teachers in improving student outcomes. Pattern-based assessments are shorter and more informative than legacy assessments in ways that can directly support better instruction. Differences in student outcomes can be analyzed across scale and over time, allowing for more detailed longitudinal evaluation of teaching and learning at the class, school, district, and state levels. Pattern-based items can be administered relatively effortlessly throughout the academic year, either as formative assessments to support ongoing instruction in classrooms or, as in the Interim Assessment, in a single deployment for benchmark or summative purposes. In short, the increased information provided by pattern-based items

and assessments can better support, at scale, accountability working as intended in improving educational and workforce outcomes. Pattern-based items are a tangible step toward “raising the bar and closing the gaps” for all students.

By transitioning from existing vendor platforms to a robust, state-of-the-art system capable of fully implementing both pattern-based items and legacy items, we can provide instantaneous feedback to students and provide the kinds of analysis presented in this report for individual classrooms as well as for schools, districts, regions, and the entire state. A fully integrated system for implementing pattern-based assessments can improve day-to-day instruction for all students. In addition, the advanced computational analyses supported by moving to an information-based model can be used to more precisely and efficiently identify students who may need additional, individualized, interventions. Using a pattern-based approach with state-of-the-art implementation, analysis, and reporting will help ensure that assessment and instruction work together for students’ and schools’ benefit. Realistic, responsive assessment, at scale, can now advance meaningful educational reform.

## Appendices

A: Items as They Appeared on 2018-2019 Interim Assessment (source: TEA)

Initial Interim Assessment Screen and Instructions to Students

The screenshot shows the STAAR Interim Grade 5 Reading assessment interface. At the top, there is a header bar with the STAAR logo, the text "Interim Grade 5 Reading", and session information "s21 No PNP". On the right side of the header are links for "Zoom", "Color", "Help", and "Logout". Below the header is the STAAR logo, which features a star on the left and the text "STAAR® State of Texas Assessments of Academic Readiness" on the right. The main content area displays the title "Interim Assessment" in bold, followed by "Grade 5 Reading" and "Fall 2018". A copyright notice at the bottom states: "Copyright © 2018, Texas Education Agency. All rights reserved. Reproduction of all or portions of this work is prohibited without express written permission from the Texas Education Agency." At the bottom of the screen are navigation buttons for "Back" and "Next".

**STAAR** Interim Grade 5 Reading s21 No PNP Zoom Color Help Logout

**STAAR**  
State of Texas  
Assessments of  
Academic Readiness

**Interim Assessment**  
**Grade 5 Reading**  
**Fall 2018**

Copyright © 2018, Texas Education Agency. All rights reserved. Reproduction of all or portions of this work is prohibited without express written permission from the Texas Education Agency.

Back Next

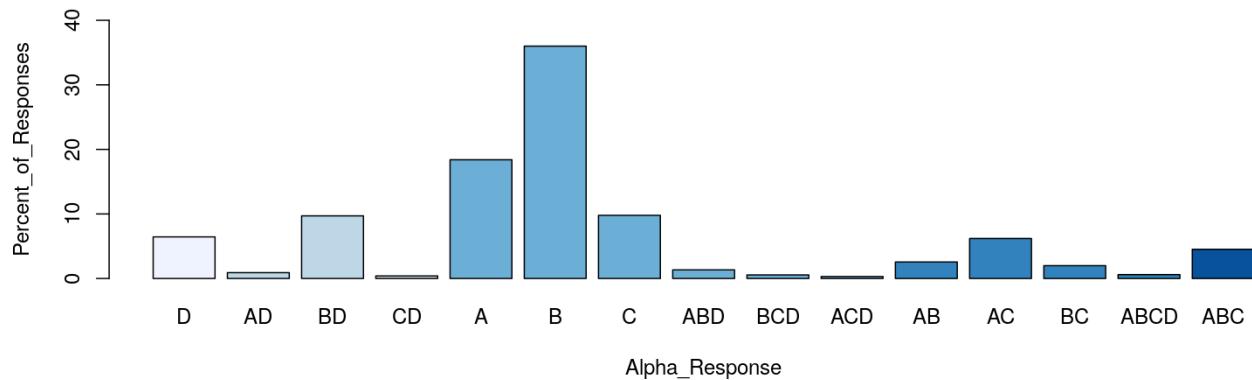
**STAAR** Interim Grade 5 Reading s21 No PNP Zoom Color Help Logout

Tools Sticky Notes

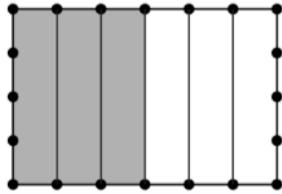
The next set of questions is optional. These questions will not be counted as part of your score. To continue, click the **Next** button.

If you do not want to complete these questions, click the **Back** button at the bottom of this screen, then click the **Review/End** button to review your answers and end the test.

Back Next



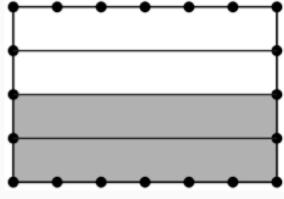
A group of friends is making cards for a fraction game. The cards need to show equivalent shaded fractions. They start with this card.



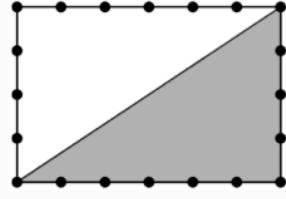
Which of these cards shows a shaded fraction that is equivalent to this card?

***This question may have more than one correct answer. Select all correct responses.***

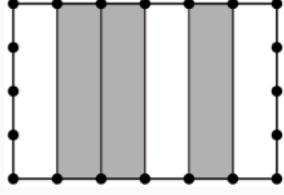
A



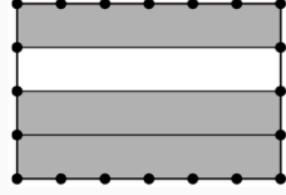
C

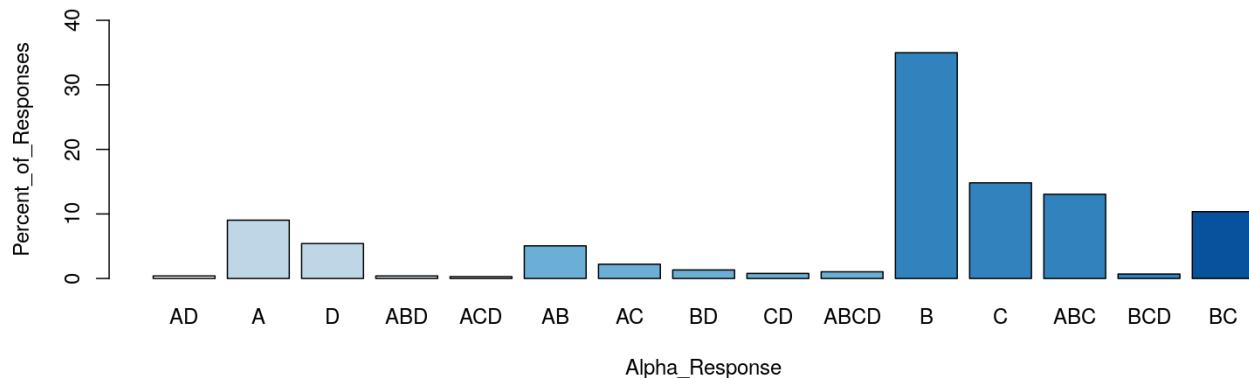


B



D





There are a total of 36 bicycles in rows. There are the same number of bicycles in each row. Which equation can be used to find the number of bicycles in each row?

***This question may have more than one correct answer. Select all correct responses.***

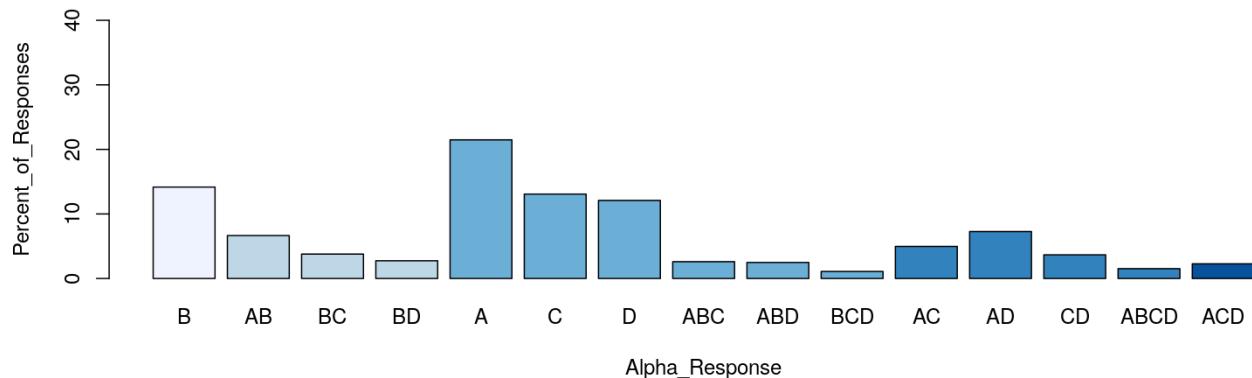
A  $42 - 6 = 36$

B  $6 + 6 + 6 + 6 + 6 + 6 = 36$

C  $9 \times 4 = 36$

D  $3 \times 6 = 18$

Mathematics Grade 3 – Item 16

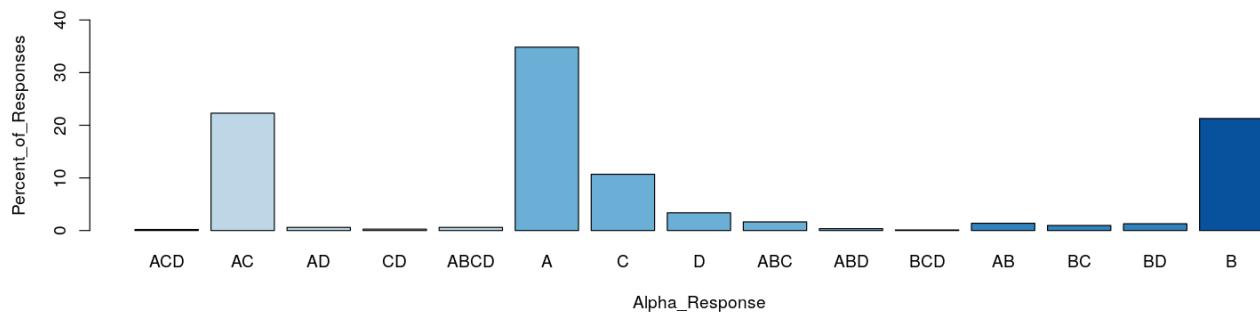


David's class is working on a project comparing amounts of food. These comparisons are made. Which of these could be correct?

***This question may have more than one correct answer. Select all correct responses.***

- A Gallons of milk are compared to quarts of milk
- B Pounds of flour are compared to fluid ounces of milk
- C Pounds of flour are compared to pounds of cheese
- D Fluid ounces of juice are compared to fluid ounces of juice

Mathematics Grade 3 – Item 17



This addition problem is missing two numbers. Which numbers placed in the boxes below would make the addition correct?

$$\begin{array}{r} 7 \square 2 \\ + 1 \square 3 \\ \hline 9 4 5 \end{array}$$

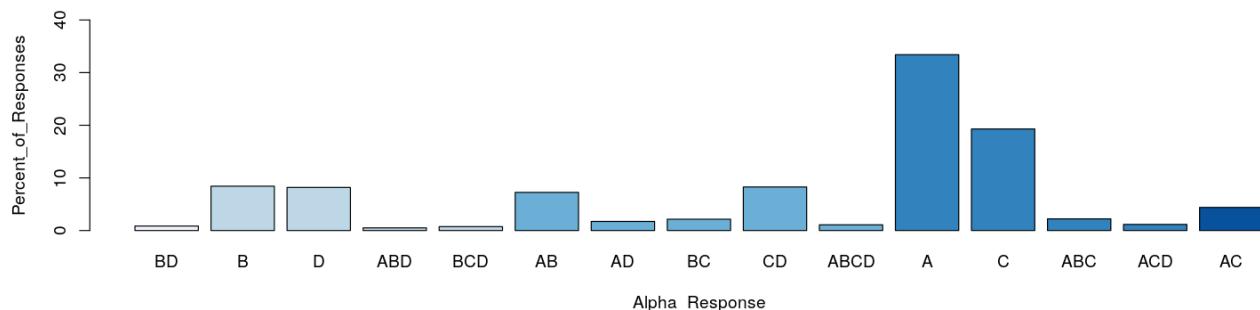
***This question may have more than one correct answer. Select all correct responses.***

A 3 and 1

B 9 and 5

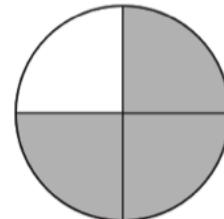
C 0 and 4

D 6 and 6



Two friends ate lunch together.

Felicia ate  $\frac{1}{3}$  of her sandwich. Daria ate  $\frac{1}{4}$  of her pie.

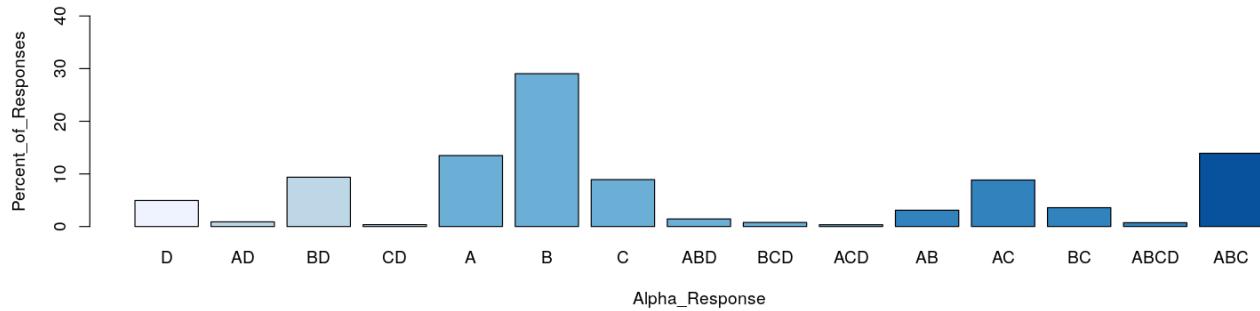


Which of these statements must be true?

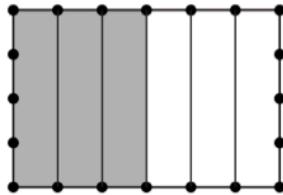
***This question may have more than one correct answer. Select all correct responses.***

- A Felicia ate a greater part of her sandwich than Daria ate of her pie because  $\frac{1}{3} > \frac{1}{4}$
- B Felicia ate more food than Daria ate
- C  $\frac{2}{3}$  of the sandwich was not eaten by Felicia
- D  $\frac{3}{3}$  of the pie was not eaten by Daria

Mathematics Grade 4 – Item 14



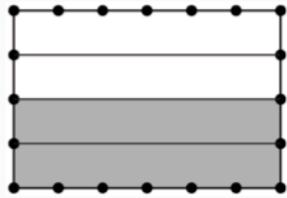
A group of friends is making cards for a fraction game. The cards need to show equivalent shaded fractions. They start with this card.



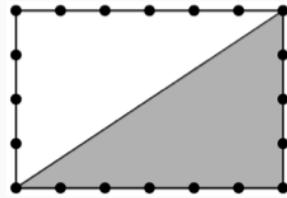
Which of these cards shows a shaded fraction that is equivalent to this card?

***This question may have more than one correct answer. Select all correct responses.***

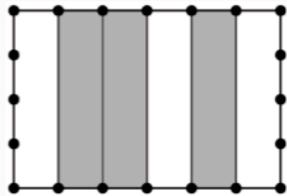
A



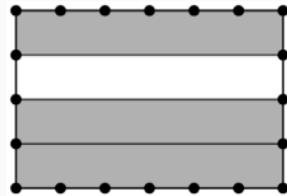
C



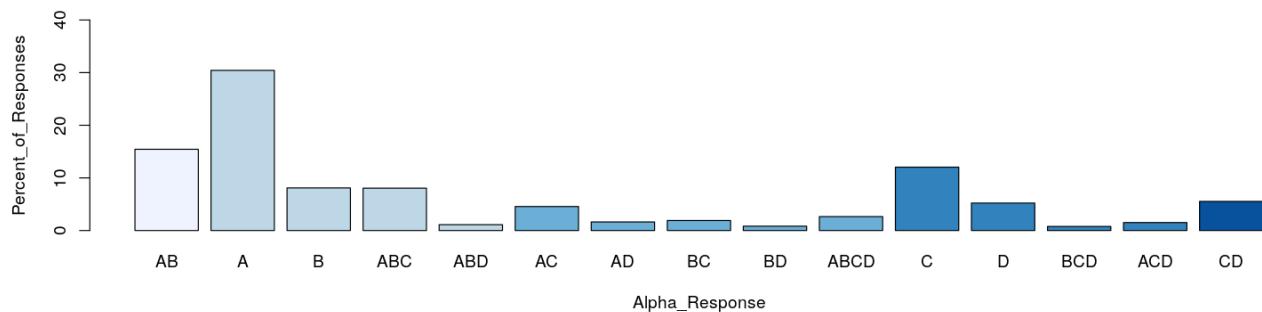
B



D



Mathematics Grade 4 – Item 15



Which of these comparisons is true?

***This question may have more than one correct answer. Select all correct responses.***

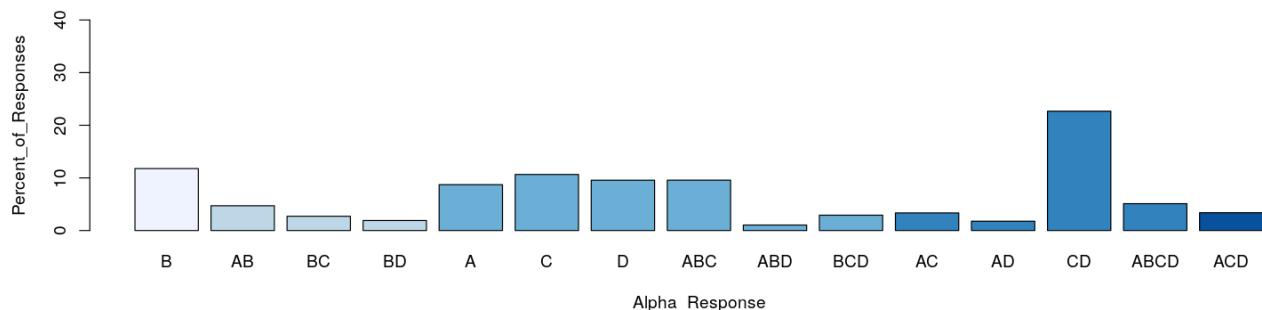
A  $\frac{3}{6} < \frac{3}{7}$

B  $\frac{1}{5} < \frac{2}{10}$

C  $\frac{2}{5} < \frac{3}{5}$

D  $\frac{3}{7} < \frac{4}{5}$

Mathematics Grade 4 – Item 16



Mike has 42 baseball cards. Maria starts with twice as many cards as Mike. Maria then gives away 6 cards. Which equation can be used to find  $y$ , the number of baseball cards Maria has now?

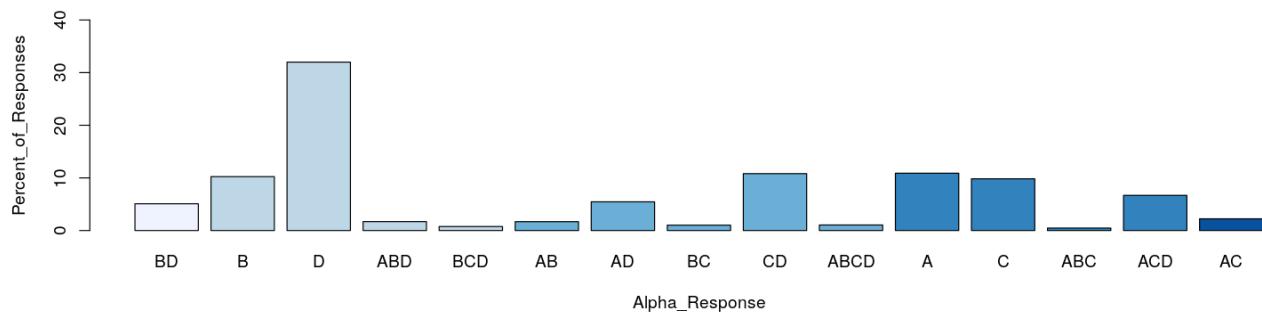
***This question may have more than one correct answer. Select all correct responses.***

A  $42 - 6 + 42 = y$

B  $2 \times 42 + 6 = y$

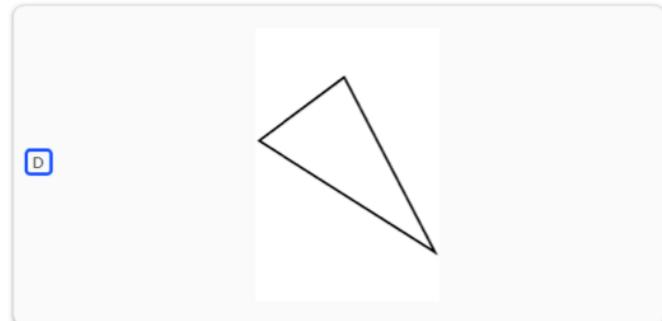
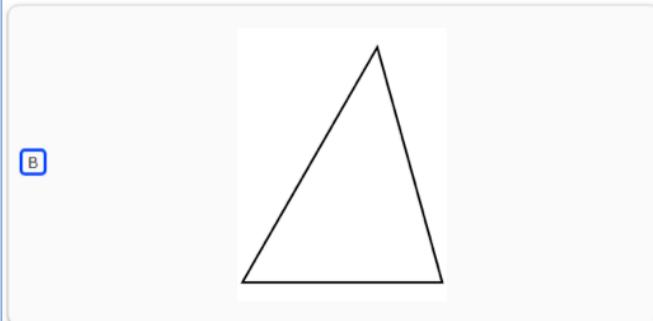
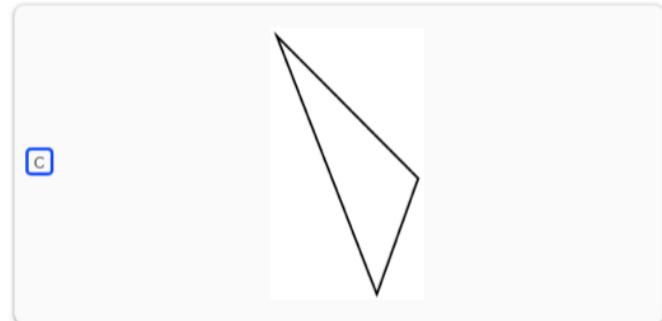
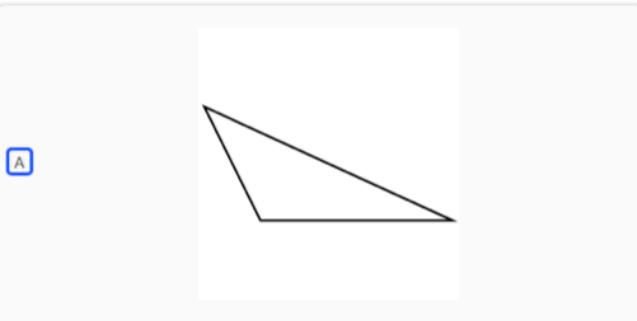
C  $84 - 6 = y$

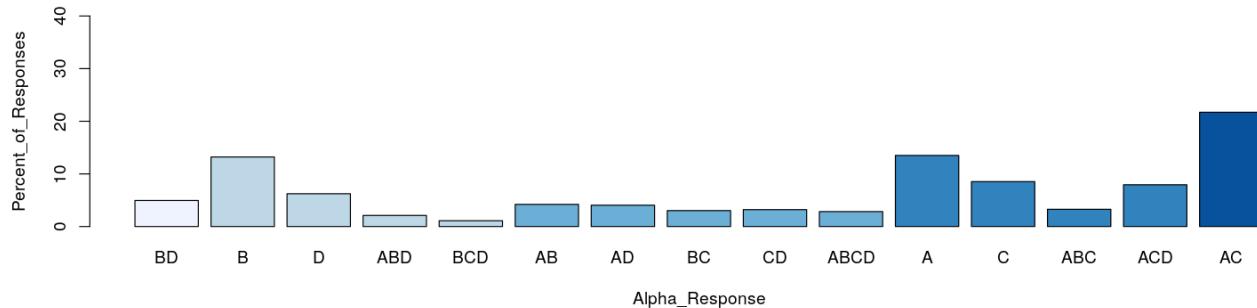
D  $42 \times 2 - 6 = y$



Which of these triangles appears to be obtuse?

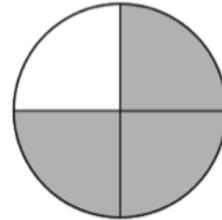
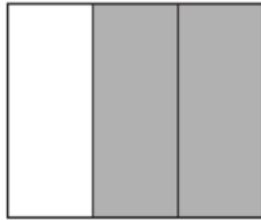
***This question may have more than one correct answer. Select all correct responses.***





Two friends ate lunch together.

Felicia ate  $\frac{1}{3}$  of her sandwich. Daria ate  $\frac{1}{4}$  of her pie.

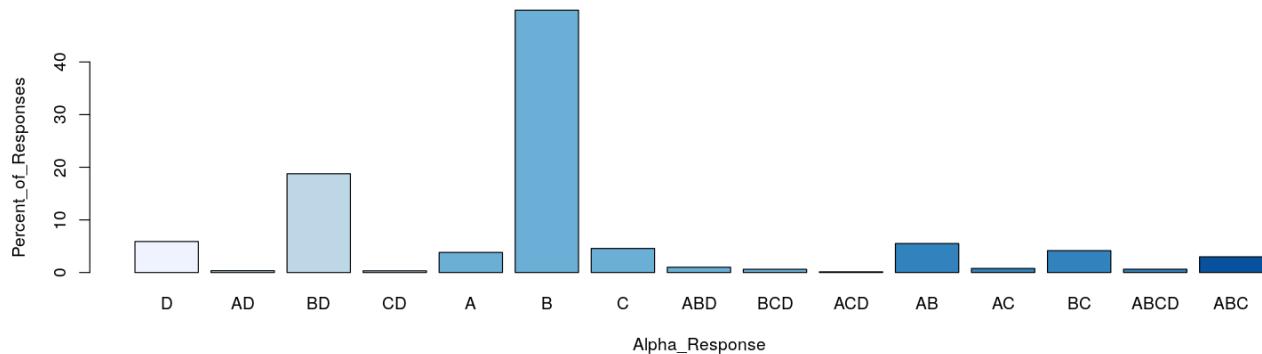


Which of these statements must be true?

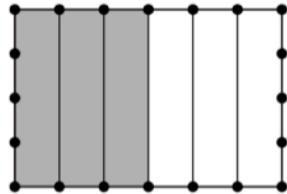
***This question may have more than one correct answer. Select all correct responses.***

- A Felicia ate a greater part of her sandwich than Daria ate of her pie because  $\frac{1}{3} > \frac{1}{4}$
- B Felicia ate more food than Daria ate
- C  $\frac{2}{3}$  of the sandwich was not eaten by Felicia
- D  $\frac{3}{4}$  of the pie was not eaten by Daria

Mathematics Grade 5 – Item 16

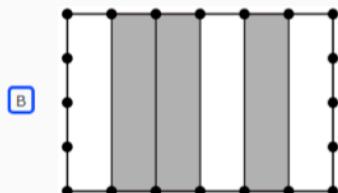
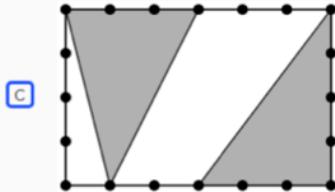
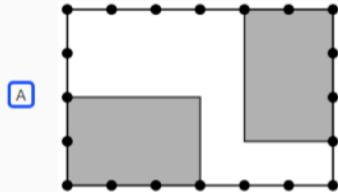


A group of friends is making cards for a fraction game. The cards need to show equivalent shaded fractions. They start with this card.

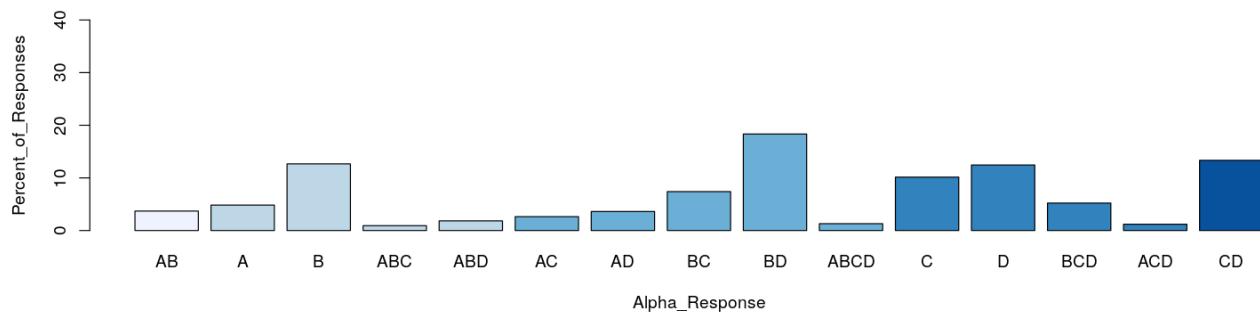


Which of these cards shows a shaded fraction that is equivalent to this card?

***This question may have more than one correct answer. Select all correct responses.***



Mathematics Grade 5 – Item 17

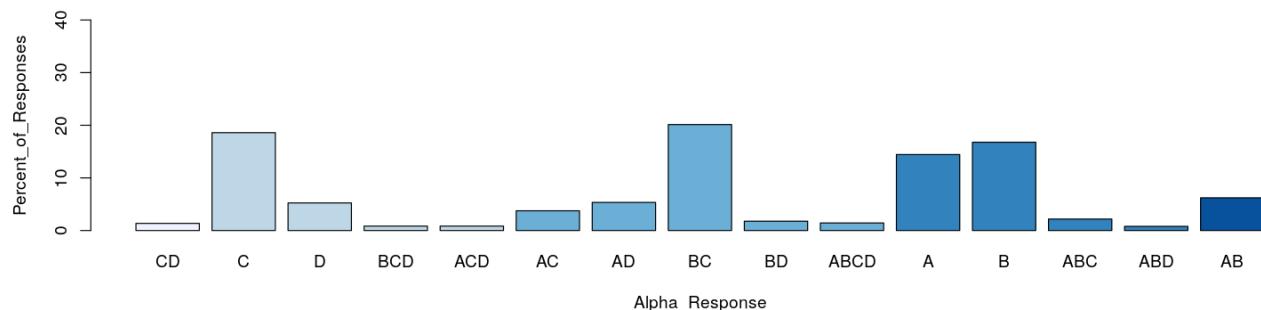


Which of these is possible?

***This question may have more than one correct answer. Select all correct responses.***

- A rhombus with five sides
- B A triangle with two obtuse angles
- C A polygon with fewer than four sides
- D A polygon with only obtuse angles

Mathematics Grade 5 – Item 18



A scatter plot with a Cartesian coordinate system. The x-axis is labeled "x" and ranges from 0 to 5. The y-axis is labeled "y" and ranges from 0 to 5. A shaded gray rectangle is plotted on the grid, bounded by vertices at approximately (1.5, 1.2), (3.5, 1.2), (3.5, 2.5), and (1.5, 2.5).

Which points are within the rectangle shown?

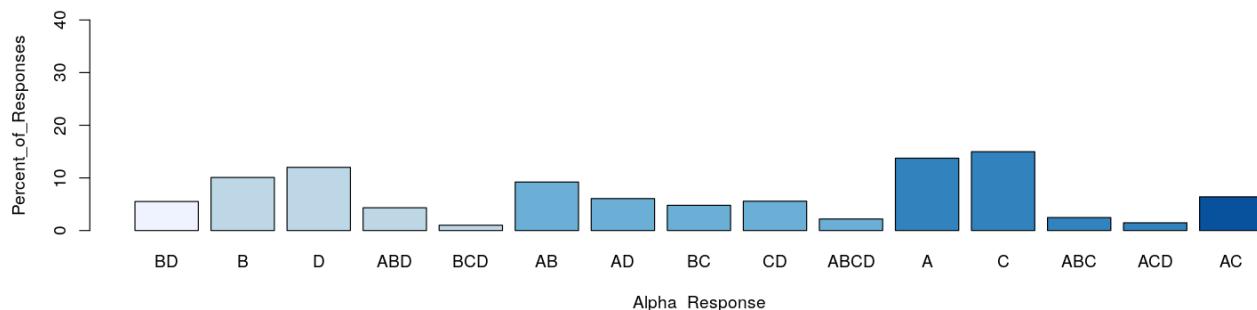
*This question may have more than one correct answer. Select all correct responses.*

A (2.711, 1.989)

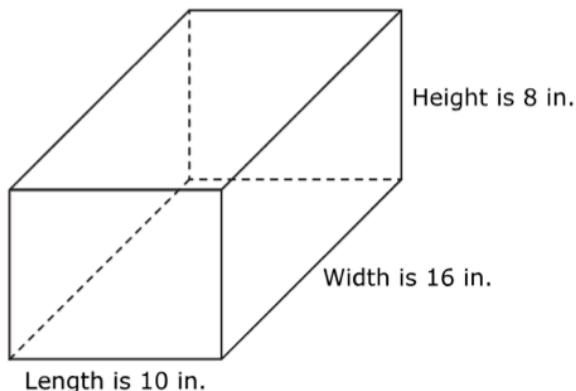
B (2, 2)

C (2, 3)

D (3.181, 3.181)



Duane wants to make a box shaped like a rectangular prism with the same volume as the box shown below.



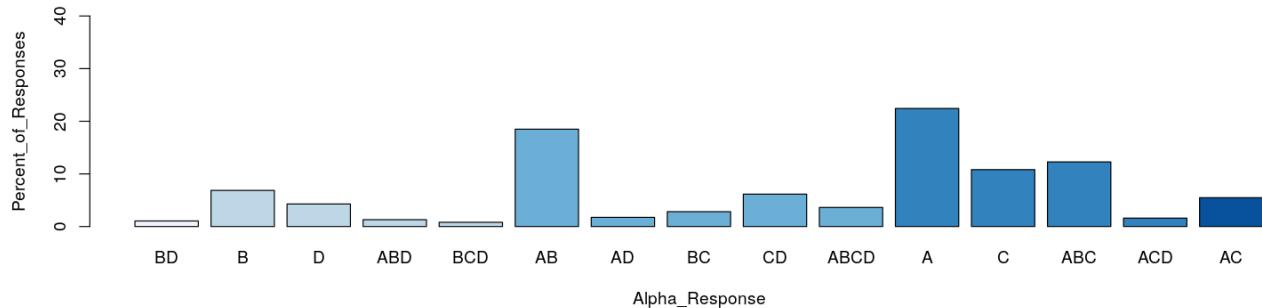
He wants the new box to have a height of 12 inches.

Which of these could be true of the new box?

***This question may have more than one correct answer. Select all correct responses.***

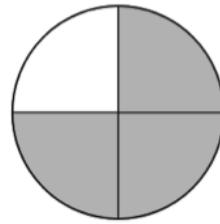
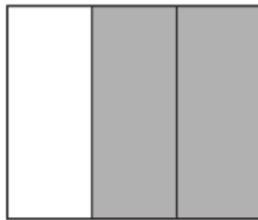
- A The new box could have a length greater than 10 inches
- B The new box could have a length greater than 10 inches and a width greater than 16 inches
- C The width of the new box could be less than 16 inches and the length could be less than 10 inches
- D The area of the base for the new box could be greater than the area of the base for the box shown

Mathematics Grade 5 – Item 20



Two friends ate lunch together.

Felicia ate  $\frac{1}{3}$  of her sandwich. Daria ate  $\frac{1}{4}$  of her pie.

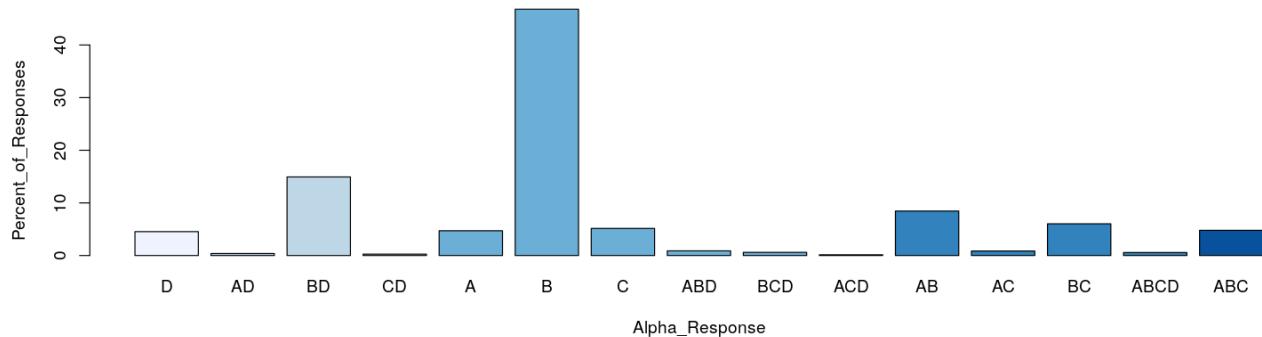


Which of these statements must be true?

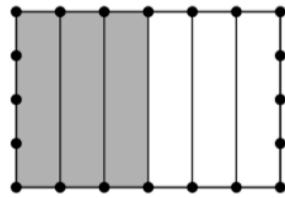
***This question may have more than one correct answer. Select all correct responses.***

- A Felicia ate a greater part of her sandwich than Daria ate of her pie because  $\frac{1}{3} > \frac{1}{4}$
- B Felicia ate more food than Daria ate
- C  $\frac{2}{3}$  of the sandwich was not eaten by Felicia
- D  $\frac{3}{4}$  of the pie was not eaten by Daria

Mathematics Grade 6 – Item 16

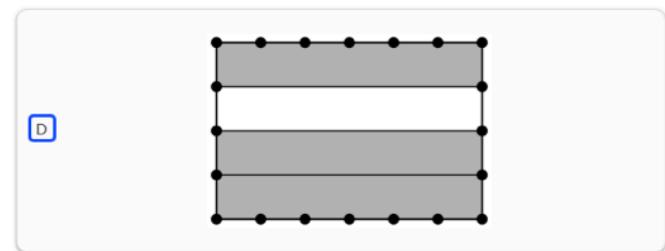
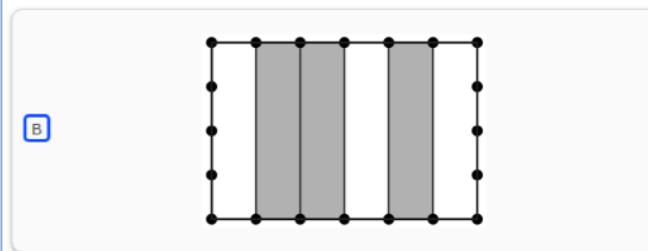
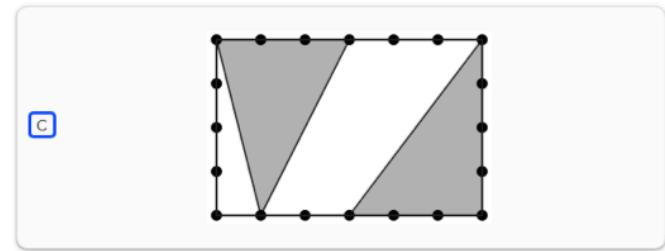
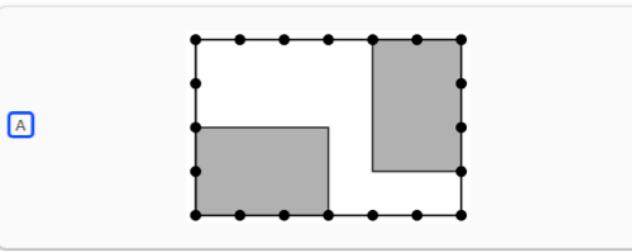


A group of friends is making cards for a fraction game. The cards need to show equivalent shaded fractions. They start with this card.

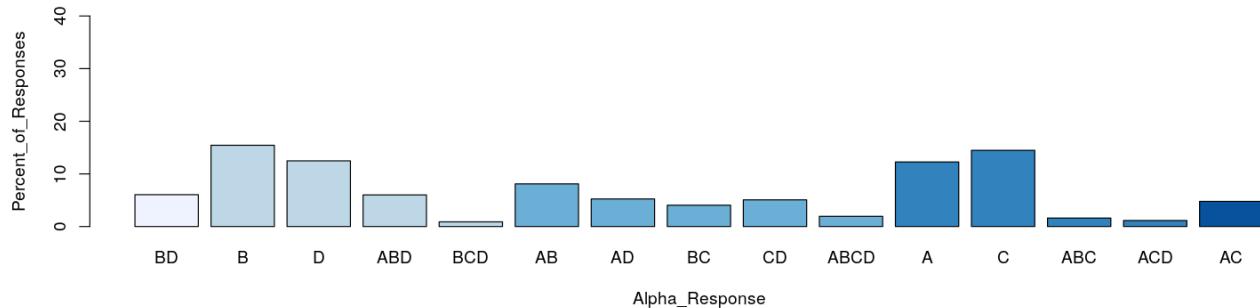


Which of these cards shows a shaded fraction that is equivalent to this card?

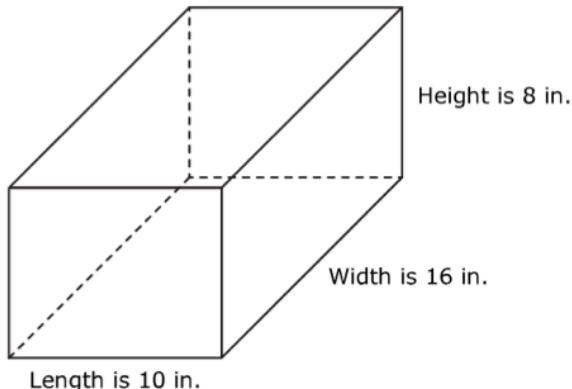
*This question may have more than one correct answer. Select all correct responses.*



Mathematics Grade 6 – Item 17



Duane wants to make a box shaped like a rectangular prism with the same volume as the box shown below.



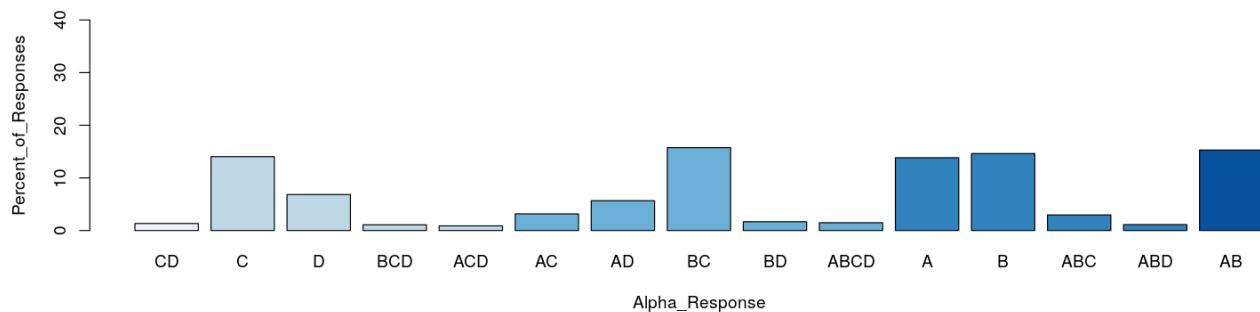
He wants the new box  
to have a height of  
12 inches.

Which of these could be  
true of the new box?

***This question may have more than one correct answer. Select all correct responses.***

- A The new box could have a length greater than 10 inches
- B The new box could have a length greater than 10 inches and a width greater than 16 inches
- C The width of the new box could be less than 16 inches and the length could be less than 10 inches
- D The area of the base for the new box could be greater than the area of the base for the box shown

Mathematics Grade 6 – Item 18



Which points are within the rectangle shown?

*This question may have more than one correct answer. Select all correct responses.*

A coordinate plane with x and y axes ranging from -5 to 2. A rectangle is shaded with a grid pattern, representing the region where  $-2 \leq x < -1$  and  $-2 \leq y < -1$ .

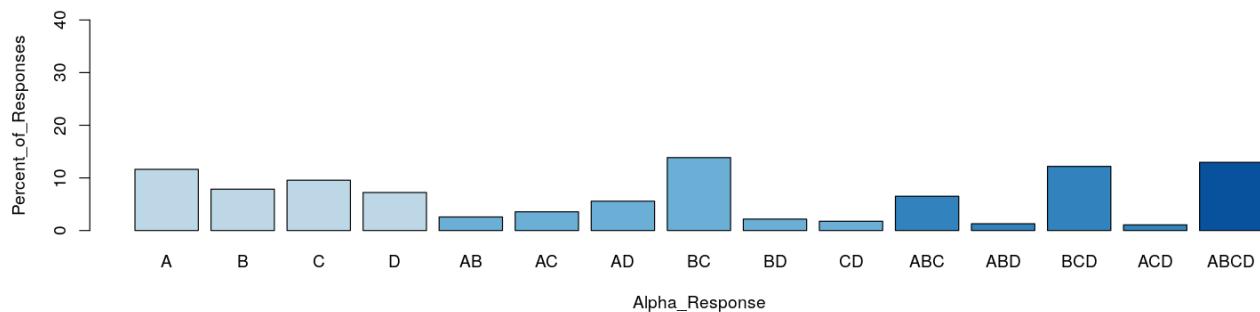
A  $(-2.711, -1.989)$

B  $(-2, -2)$

C  $(-2, -3)$

D  $(-3.181, -3.181)$

Mathematics Grade 6 – Item 19



Which of these is a rational number?

***This question may have more than one correct answer. Select all correct responses.***

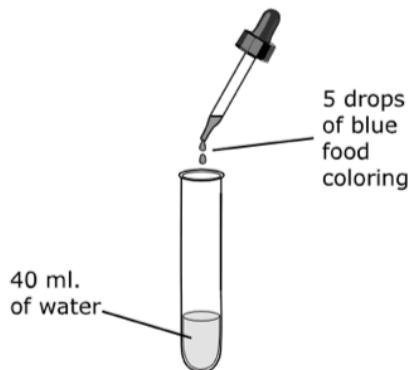
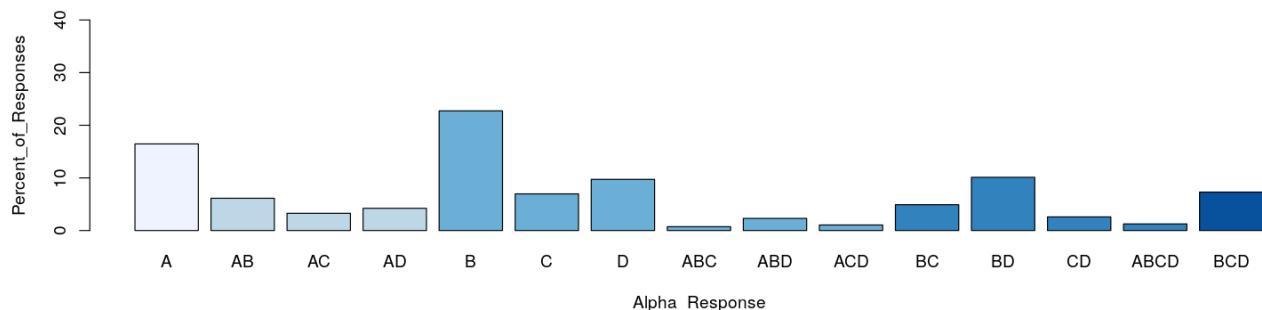
A  $-2$

B  $\frac{1}{3}$

C  $1\frac{1}{2}$

D  $0.6789992$

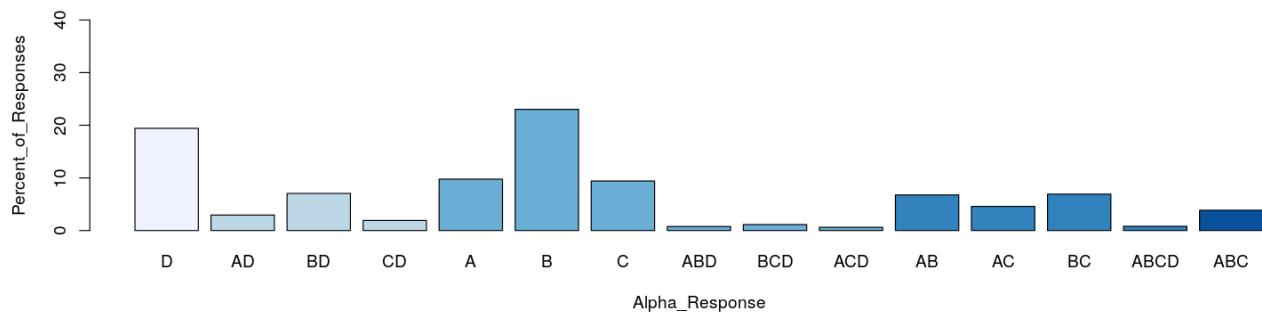
Mathematics Grade 6 – Item 20



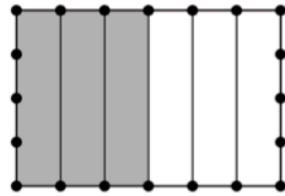
5 drops of blue food coloring are mixed with 40 milliliters of water to produce a blue solution. Which of the following mixtures would produce a solution with same ratio of drops of food coloring to milliliters of water?

***This question may have more than one correct answer. Select all correct responses.***

- A 15 drops in 50 milliliters of water
- B 10 drops in 80 milliliters of water
- C 0.2 drops in 1.6 milliliters of water
- D 6 drops in 48 milliliters of water

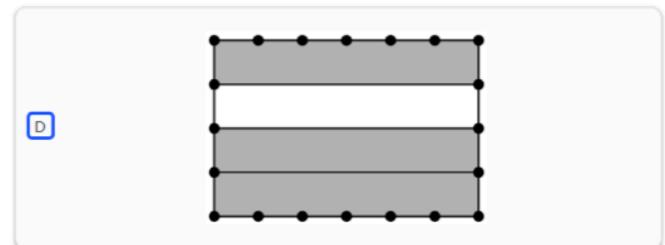
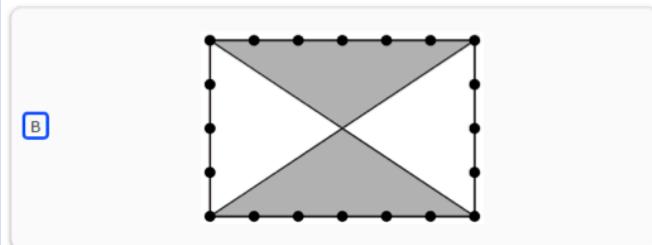
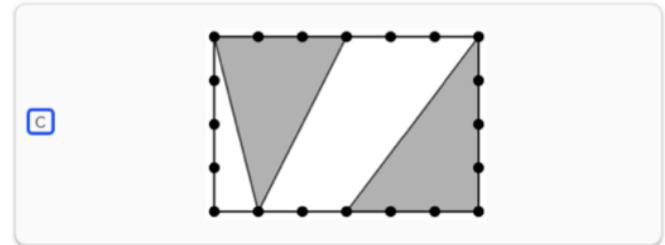
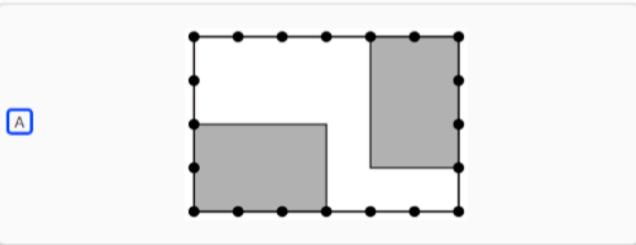


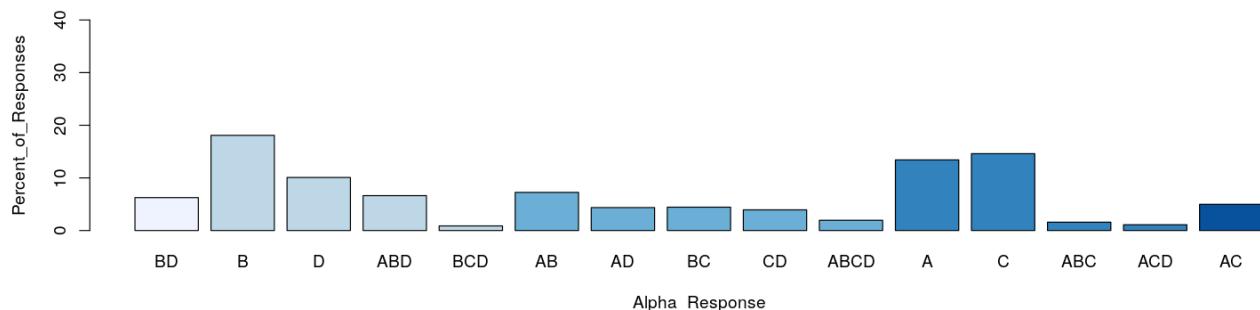
A group of friends is making cards for a fraction game. The cards need to show equivalent shaded fractions. They start with this card.



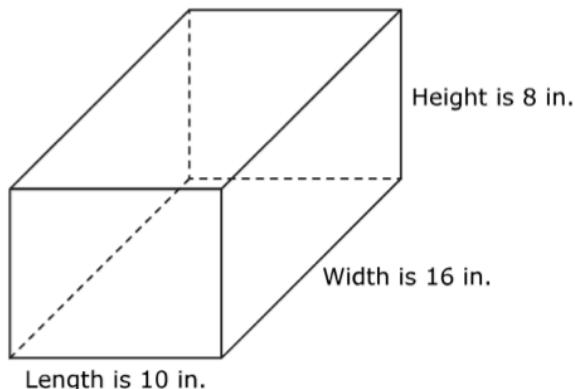
Which of these cards shows a shaded fraction that is equivalent to this card?

*This question may have more than one correct answer. Select all correct responses.*





Duane wants to make a box shaped like a rectangular prism with the same volume as the box shown below.



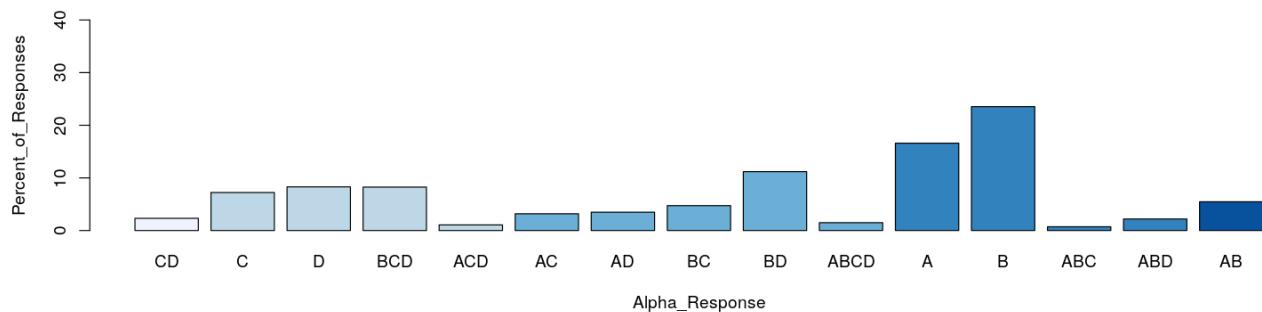
He wants the new box to have a height of 12 inches.

Which of these could be true of the new box?

***This question may have more than one correct answer. Select all correct responses.***

- A The new box could have a length greater than 10 inches
- B The new box could have a length greater than 10 inches and a width greater than 16 inches
- C The width of the new box could be less than 16 inches and the length could be less than 10 inches
- D The area of the base for the new box could be greater than the area of the base for the box shown

Mathematics Grade 7 – Item 20



A coordinate plane with x and y axes ranging from -5 to 2. A rectangle is shaded gray and outlined in black. Its vertices are at (-3, 1), (-3, 2), (-2, 1), and (-2, 2).

Which points are within the rectangle shown?

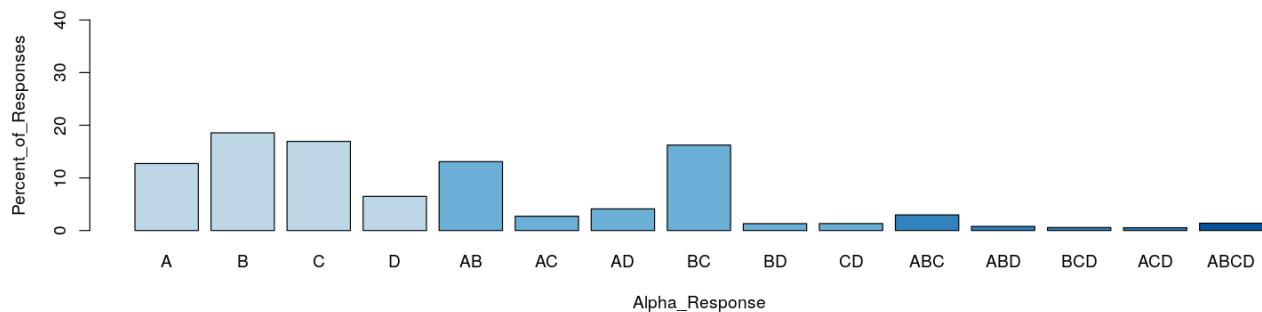
*This question may have more than one correct answer. Select all correct responses.*

A  $(-2.711997, 1.98979)$

B  $(-2, 2)$

C  $(2, -3)$

D  $(3.181995, -3.181995)$



Which of these is a rational number?

***This question may have more than one correct answer. Select all correct responses.***

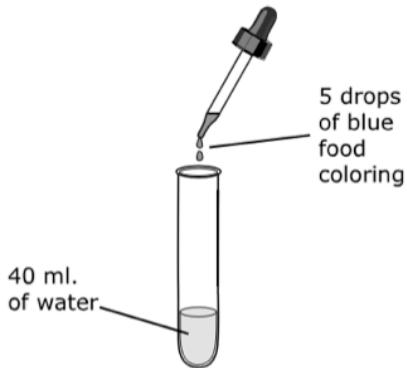
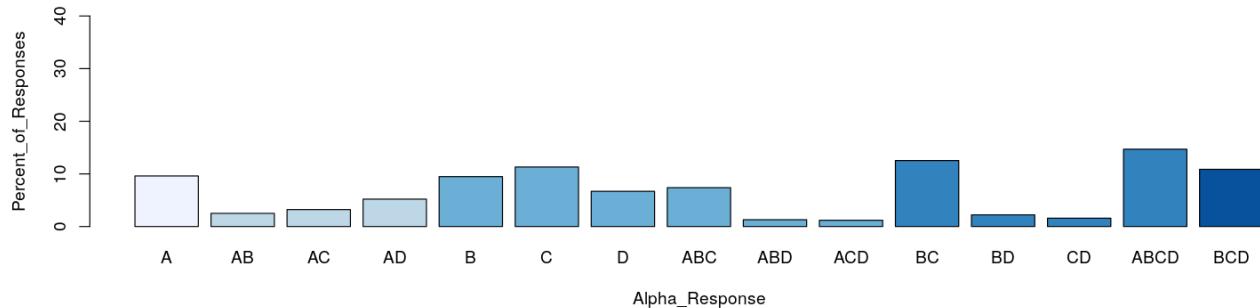
A -2

B  $\frac{1}{3}$

C  $1\frac{1}{2}$

D 0.6789992

Mathematics Grade 7 – Item 22

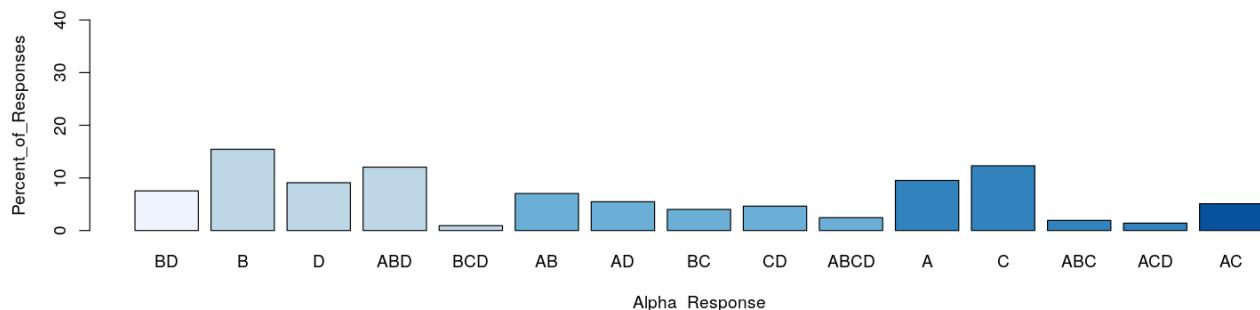


5 drops of blue food coloring are mixed with 40 milliliters of water to produce a blue solution. Which of the following mixtures would produce a solution with same ratio of drops of food coloring to milliliters of water?

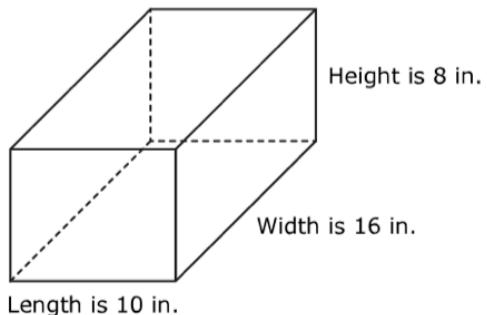
***This question may have more than one correct answer. Select all correct responses.***

- A 15 drops in 50 milliliters of water
- B 10 drops in 80 milliliters of water
- C 0.2 drops in 1.6 milliliters of water
- D 6 drops in 48 milliliters of water

Mathematics Grade 8 – Item 18



Duane wants to make a box shaped like a rectangular prism with the same volume as the box shown below.



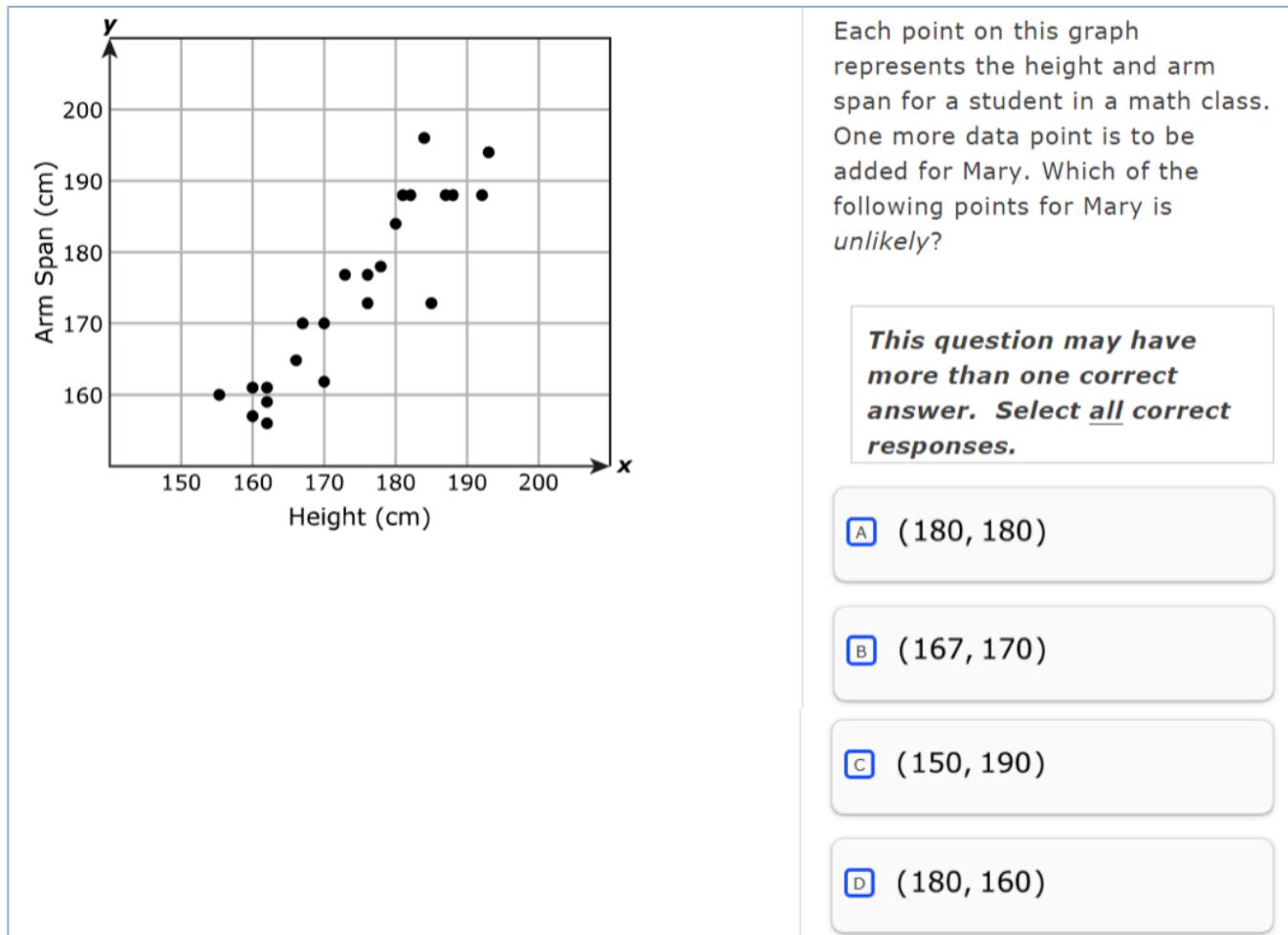
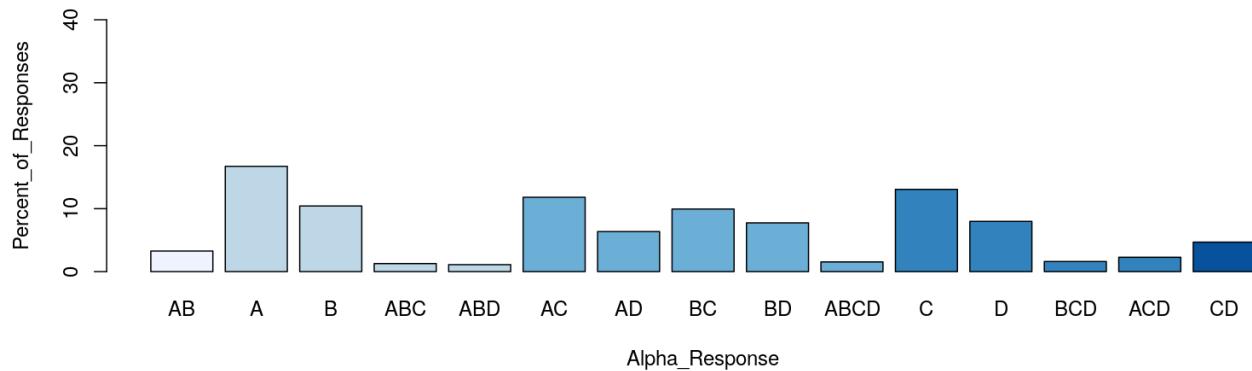
He wants the new box  
to have a height of  
12 inches.

Which of these could be  
true of the new box?

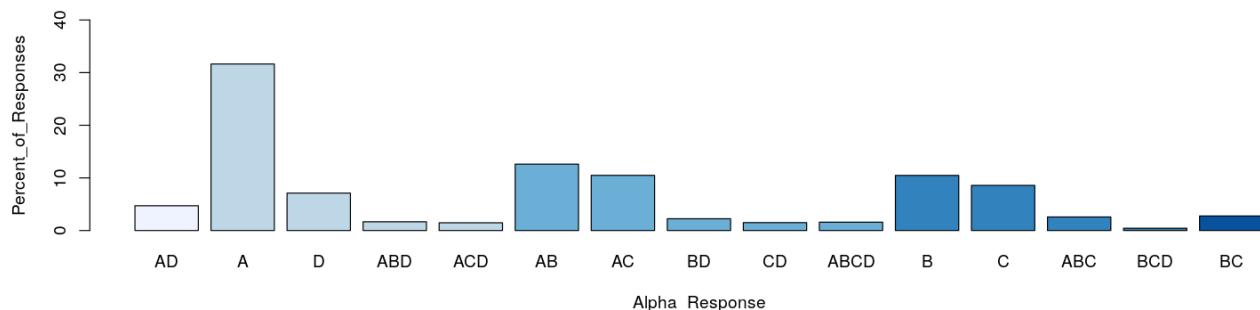
***This question may have more than one correct answer. Select all correct responses.***

- A The new box could have a length greater than 10 inches
- B The new box could have a length greater than 10 inches and a width greater than 16 inches
- C The width of the new box could be less than 16 inches and the length could be less than 10 inches
- D The area of the base for the new box could be greater than the area of the base for the box shown

Mathematics Grade 8 – Item 19



Mathematics Grade 8 – Item 20



A scatter plot with "Height (cm)" on the x-axis (150 to 200) and "Arm Span (cm)" on the y-axis (160 to 200). A line of best fit passes through the origin. Most data points are below the line, indicating arm spans are generally less than heights.

A bar graph titled "Count of Points Below, On and Above the Line". The x-axis categories are "Below", "On", and "Above". The y-axis ranges from 0 to 12. The bars show approximately 9 points below, 5 points on, and 11 points above the line.

A student sees a **pattern** in the graph of height and arm span data for her class. She says, "**Our arm spans are the same as our heights,**" and she draws the line for arm span being equal to height on the graph.

To investigate this idea, the class counts the number of students below, on, and above this line and represents this data with a bar graph. Which of the following can be concluded?

**This question may have more than one correct answer. Select all correct responses.**

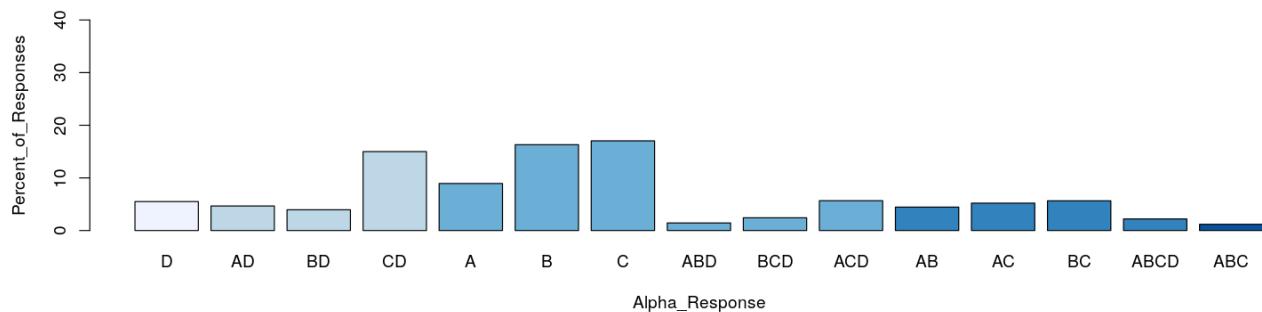
A The bar graph shows the pattern is unlikely to be valid

B The line graph shows the pattern is likely to be valid

C In this class there are more students who have arm spans greater than their heights

D In this class there are more students who have heights greater than their arm spans

Mathematics Grade 8 – Item 21



Given only the following, which of these could be used in the transformation of **A** to **A'**?

*This question may have more than one correct answer. Select all correct responses.*

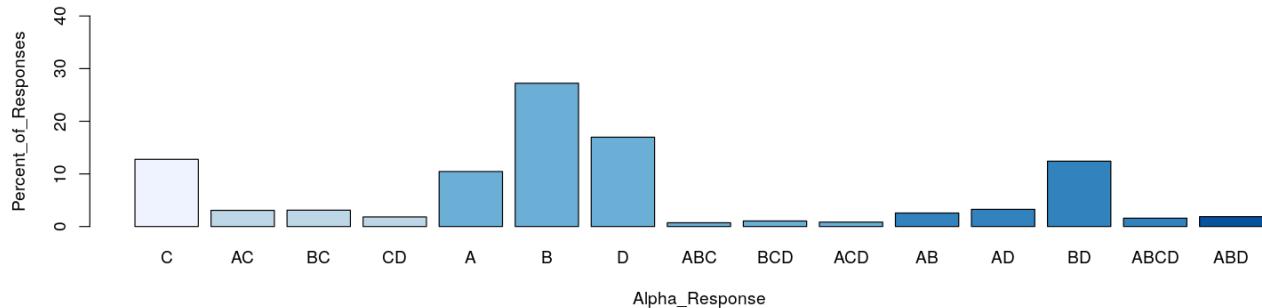
A Rotation

B Two Horizontal Reflections

C Translation

D 1:2 Dilation

Mathematics Grade 8 – Item 22



A student is given the equation

$$2x + 4 = 6x + 2$$

and is asked to solve for x. Which of the following could be used in solving for x?

***This question may have more than one correct answer. Select all correct responses.***

- A As a first step, divide both sides of the equation by 2
- B As a first step, subtract 2x from both sides of the equation
- C As a first step, subtract 2 from *only* the right side of the equation.
- D As a first step, subtract 2 from both sides of the equation

## The Wind and the Moon

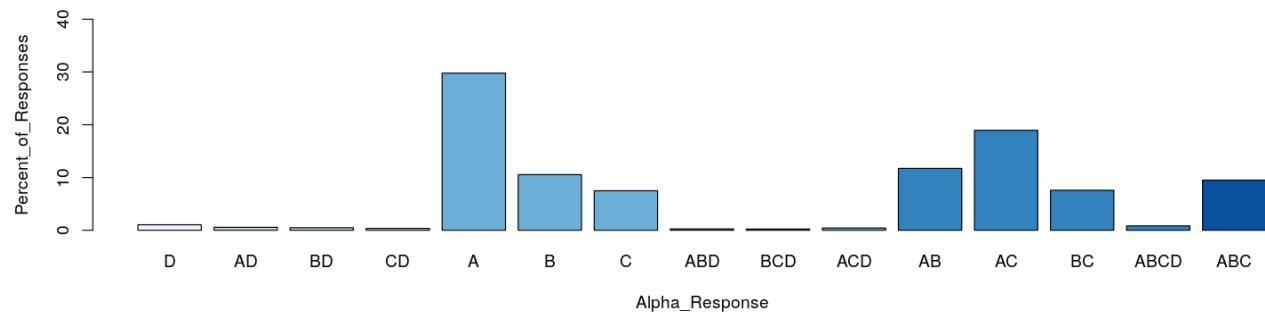
by George MacDonald

### 01 THE WIND AND THE MOON

02  
03 Said the Wind to the Moon, "I will blow you out.  
04        You stare  
05        In the air  
06        Like a ghost in a chair,  
07        Always looking what I am about;  
08        I hate to be watched – I will blow you out."  
09  
10 The Wind blew hard, and out went the Moon.  
11        So deep,  
12        On a heap  
13        Of clouds, to sleep  
14 Down lay the Wind, and slumbered soon –  
15 Muttering low, "I've done for that Moon."  
16  
17 He turned in his bed; she was there again!  
18        On high,  
19        In the sky,  
20        With her one clear eye,  
21 The Moon shone white and alive and plain.  
22 Said the Wind – "I will blow you out again."  
23  
24 The Wind blew hard, and the Moon grew dim.  
25        "With my sledge  
26        And my wedge  
27        I have knocked off her edge!  
28 If only I blow right fierce and grim,  
29 The creature will soon be dimmer than dim."  
30  
31 He blew a great blast, and the thread was gone;  
32        In the air  
33        Nowhere  
34        Was a moonbeam bare;  
35 Far off and harmless the shy stars shone;  
36 Sure and certain the Moon was gone.  
37  
38 The Wind, he took to his revels once more;  
39        On down  
40        In town,  
41        Like a merry-mad clown,  
42 He leaped and hallooed with whistle and roar,

43 "What's that?" The glimering thread once more!  
44  
45 He flew in a rage – he danced and blew;  
46        But in vain  
47        Was the pain  
48        Of his bursting brain;  
49 For still the broader the Moon-scrap grew,  
50 The broader he swelled his big cheeks and blew.  
51  
52 Slowly she grew – till she filled the night,  
53        And shone  
54        On her throne

Public Domain

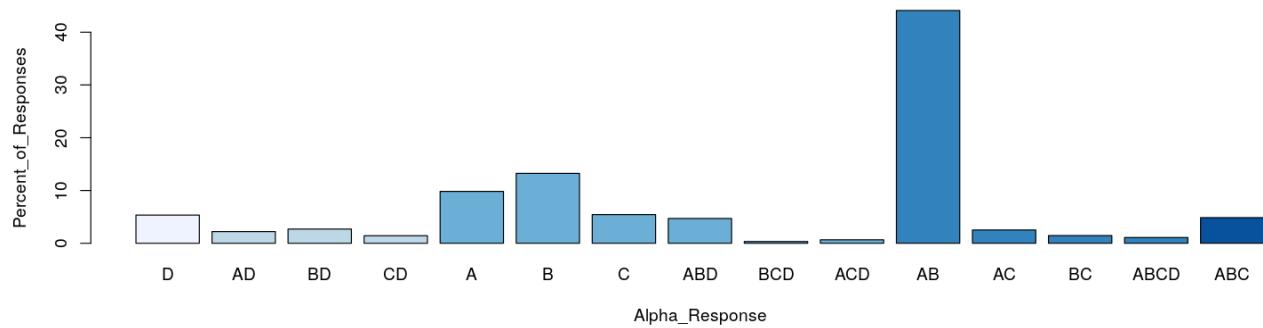


Why did the wind want to hide the moon?

*This question may have more than one correct answer. Select all correct responses.*

- A The wind did not want to be watched
- B The light from the moon was too bright
- C The wind wanted to sleep
- D The wind was hungry

Reading Grade 5 – Item 16



Which of the following did the wind feel?

***This question may have more than one correct answer. Select all correct responses.***

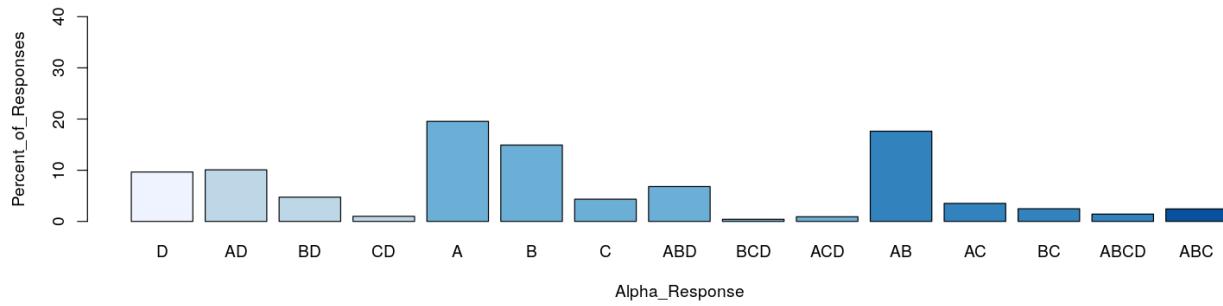
A Frustration

B Anger

C Satisfaction

D Sadness

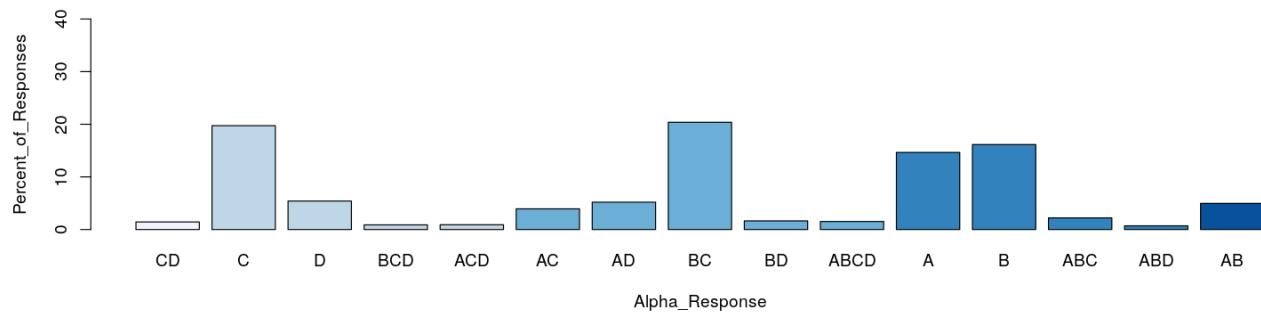
Reading Grade 5 – Item 17



Which of these lines describes the moon?

***This question may have more than one correct answer. Select all correct responses.***

- A Line 6: Like a ghost in a chair
- B Line 20: With her one clear eye
- C Line 33: Nowhere
- D Line 41: Like a merry-mad clown



Read line 42 in the poem.

He leaped and hallooed  
with whistle and roar,

Why did the author include sensory language in this line from the poem?

***This question may have more than one correct answer. Select all correct responses.***

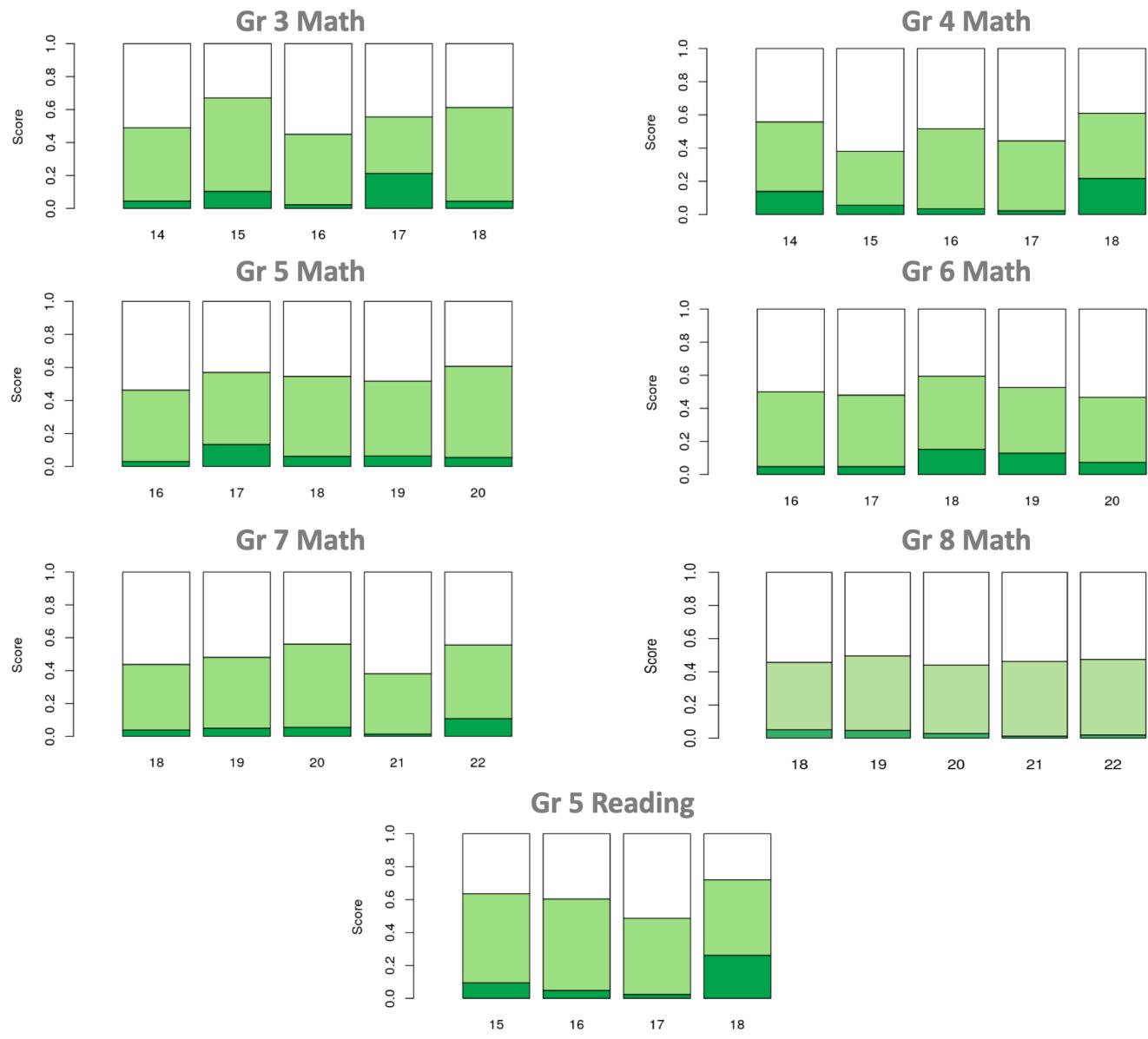
A To describe the movement and sound of the wind

B To describe how the wind felt pleased

C To describe how the wind felt upset

D To describe the shape of the wind

B: Partial and Full Credit Results for Pattern-Based Items and Released STAAR Items



**Figure 24. Partial and full credit results for pattern-based items**

	Percent Exact Match	Average Partial Credit	Item_Number	Responses
<b>Grade 3 Math</b>	5%	0.49	14	60,517
	10%	0.67	15	60,212
	2%	0.45	16	59,976
	21%	0.56	17	59,884
	4%	0.61	18	59,561
	<b>Averages</b>	<b>8%</b>	<b>0.56</b>	<b>60,030</b>
<b>Grade 4 Math</b>	14%	0.56	14	65,263
	6%	0.38	15	64,508
	3%	0.52	16	65,044
	2%	0.44	17	64,818
	22%	0.61	18	64,750
	<b>Averages</b>	<b>9%</b>	<b>0.50</b>	<b>64,877</b>
<b>Grade 5 Math</b>	3%	0.46	16	65,516
	13%	0.57	17	65,163
	6%	0.55	18	64,659
	6%	0.52	19	64,352
	5%	0.61	20	64,359
	<b>Averages</b>	<b>7%</b>	<b>0.54</b>	<b>64,810</b>
<b>Grade 6 Math</b>	5%	0.5	16	60,510
	5%	0.48	17	59,130
	15%	0.59	18	58,852
	13%	0.53	19	58,902
	7%	0.47	20	58,497
	<b>Averages</b>	<b>9%</b>	<b>0.51</b>	<b>59,178</b>
<b>Grade 7 Math</b>	4%	0.44	18	47,774
	5%	0.48	19	46,591
	5%	0.56	20	46,144
	1%	0.38	21	46,464
	11%	0.56	22	46,483
	<b>Averages</b>	<b>5%</b>	<b>0.48</b>	<b>46,691</b>
<b>Grade 8 Math</b>	5%	0.46	18	48,711
	5%	0.5	19	47,676
	3%	0.44	20	47,812
	1%	0.46	21	48,167
	2%	0.47	22	47,720
	<b>Averages</b>	<b>3%</b>	<b>0.47</b>	<b>48,017</b>
<b>Grade 5 Reading</b>	10%	0.64	15	69,856
	5%	0.6	16	69,708
	2%	0.49	17	69,634
	26%	0.72	18	69,522
	<b>Averages</b>	<b>11%</b>	<b>0.61</b>	<b>69,680</b>

Figure 25. Partial, full credit, responses and averages for pattern-based items

<b>Spring 2018 STAAR</b>	<b>Gr 3 Math</b>	<b>Gr 4 Math</b>	<b>Gr 5 Math</b>	<b>Gr 6 Math</b>	<b>Gr 7 Math</b>	<b>Gr 8 Math</b>	<b>Gr 5 Reading</b>
average	66.7	66.8	68.1	55.1	54.2	54.2	69.6
median	69	67.5	72	54.5	50	50.5	70
min	47	43	44	37	37	37	49
max	88	87	86	77	75	75	88

Figure 26. Average, median, minimum and maximum percent correct on Spring 2018 STAAR

## References

- American Statistical Association (2014), “ASA Statement on Using Value-Added Models for Educational Assessment,” available at [http://www.amstat.org/policy/pdfs/ASA\\_VAM\\_Statement.pdf](http://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf).
- Linn, Robert. (1990). Has Item Response Theory Increased the Validity of Achievement Test Scores?. *Applied Measurement in Education - APPL MEAS EDUC.* 3. 115-141.
- Pivovarova, M., Amrein-Beardsley & Broatch. A. (2016). Value-Added Models (VAMs): Caveat Emptor, *Statistics and Public Policy*, 3:1, 1-9, DOI: [10.1080/2330443X.2016.1164641](https://doi.org/10.1080/2330443X.2016.1164641)
- Polikoff, M. S., & Porter, A. C. (2014). Instructional Alignment as a Measure of Teaching Quality. *Educational Evaluation and Policy Analysis*, 36(4), 399–416.