PIE

Pathways for Instructionally Embedded Assessment

Evaluating Pathways for Instructionally Embedded Assessment (PIE) Assessments: Evidence From Cognitive Labs

Technical Report

September 2025

Accessible Teaching, Learning, and Assessment Systems. (2025). *Evaluating Pathways for Instructionally Embedded Assessment (PIE) assessments: Evidence from cognitive labs*.

Acknowledgments

This technical report represents the work carried out in evaluating item prototypes for the PIE project. We acknowledge the contributions of many people. Namely, Ashley Hirt and Gail Tiemann made significant writing contributions to this report. The individuals who supported development of PIE item prototypes and assisted with cognitive lab sessions are listed below. We express our gratitude to them. We additionally thank Gail Tiemann for contributing to the design and coding schematic for this study, Lisa Denney for supporting the facilitation of cognitive lab sessions, and Delaney Wilson for assisting in processing and analyzing codes.

Project Director: **Shaun Bates**, Standards, Curriculum, and Assessment Coordinator, Missouri Department of Elementary and Secondary Education (DESE)
Principal Investigator: **Brooke Nash**, ATLAS, University of Kansas
Mathematics Pathways Director: **Mary Majerus, Ed.D.**, Missouri Department of Elementary and Secondary Education (DESE)

Project Advisory Committee (PAC)
Brian Gong, **Ph.D.**, *Center for Assessment*
Bob Henson, **Ph.D.**, *University of North Carolina at Greensboro*
Meagan Karvonen, **Ph.D.**, *Accessible Teaching, Learning, and Assessment Systems, University of Kansas*
Leanne Ketterlin-Geller, **Ph.D.**, *Southern Methodist University*
Ed Roeber, **Ph.D.**, *Michigan Assessment Consortium*
Lisa Sireno, *Missouri Department of Elementary and Secondary Education (DESE)*
David Williamson, **Ph.D.**, *Independent consultant*
Phoebe Winter, **Ph.D.**, *Independent consultant*

Accessible Teaching, Learning, and Assessment Systems (ATLAS)

Allison Barkley	Kristine David	David Whitcomb
Alexus Cafferty	Sarah Hardman	Marjorie Wine
Amanda Caffman	Sarah Randol	Breonna Yungeberg

The project reported here is supported by the U.S. Department of Education (USED) through a Competitive Grants for State Assessments (CGSA) program under Grant No. S368A220019. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the USED and CGSA.

Table of Contents

Introduction

This report describes a cognitive lab study aimed at gathering evidence of student response processes and interactions with prototype assessment items from the Pathways for Instructionally Embedded Assessment (PIE) project.

Funded through a Competitive Grant for State Assessments (CGSA) and led by the Missouri Department of Elementary and Secondary Education (MO-DESE) and Accessible Teaching, Learning, and Assessment Systems (ATLAS) at the University of Kansas, the PIE project constructed and evaluated a prototype assessment system built on cognitive models of student learning, or *learning pathways*, aligned with fifth-grade-level academic expectations.

A project goal for PIE was to design, develop, administer, and evaluate the PIE assessment system, a prototype instructionally embedded assessment built on cognitive models of student learning (learning pathways) aligned to grade-level content standards. A related objective of this goal was to iteratively develop and test prototype items. To this end, ATLAS staff conducted cognitive (cog) lab sessions in Missouri in December 2023 and January 2024. The primary purpose of the cognitive labs was to gather evidence of student response process and interactions with PIE prototype items. Findings from the cog labs may be used as evidence within the assessment validity argument for the prototype (American Educational Research Association et al., 2014) PIE assessment system. The cog labs explored several research questions (RQs).

1. How much time is required to complete the prototype items?
2. Do students interpret items as intended?
3. How do students interact with the features of the prototype items?
4. Do students' responses represent the constructs the item prototypes were designed to measure?
5. What are students' perceptions of the item prototypes?

Prototype Items

PIE staff developed item prototypes for three learning pathways (LPs): Data and Statistics (DS), Mathematics Geometry and Measurement (GM), and Mathematics Relationships and Algebraic Thinking (RA). Table 1 displays the domains and related Missouri standards selected for item prototypes. These content standards were selected for the cog lab study because they provided opportunities to use a variety of item types. The goal of this approach was to explore whether different item types might increase student engagement while also maximizing access and minimizing barriers for students.

Table 1. Missouri Grade 5 Mathematics Content Standards Selected for the PIE Item Prototypes

Domain	Priority Code	Missouri Priority Standard
Relationships and Algebraic Thinking (RA)	**5.RA.A**	**Represent and analyze patterns and relationships.**
	5.RA.A.1a	Generate two numeric patterns given two rules.
Geometry and Measurement (GM)	**5.GM.A**	**Classify two- and three-dimensional geometric shapes.**
	5.GM.A.2	Classify figures in a hierarchy based on properties.
Data and Statistics (DS)	**5.DS.A**	**Represent and analyze data.**
	5.DS.A.2	Create a line plot to represent a given or generated data set, and analyze the data to answer questions and solve problems, recognizing the outliers and generating the median.

Note. Missouri Priority Standards are from Missouri DESE (Missouri Department of Elementary and Secondary Education, 2021). Boldface indicates cluster (i.e., priority code and description) as an anchor linking the standards to their domain.

Each LP had three vertical levels. Level 1 represented emerging concepts and skills related to the learning target. Level 2 represented concepts and skills approaching the learning target. Level 3 represented the learning target and aligned with the grade-level content standard.

For the cog lab sessions, staff developed three test forms, each aligned to one of the three selected standards. Each form offered three or four items at each of the three levels in their respective LPs for a total of nine or 10 items per form. Table 2 gives the number of items per form, by learning pathway levels.

Table 2. Number and Domain of Items, by Learning Pathway Levels

Form and Domain	Level 1 Items	Level 2 Items	Level 3 Items	Total
Form 1: PIE.5.DS.A.2 Data and Statistics (DS)	3	3	3	9
Form 2**:** PIE.5.GM.A.2 Mathematics Geometry and Measurement (GM)	4	3	3	10
Form 3**:** PIE.5.RA.A.1a Mathematics Relationships and Algebraic Thinking (RA)	3	3	3	9

For cog lab sessions, PIE staff developed items from three item categories: selected response, constructed response, and technology enhanced. Table 3 describes the item categories and item types used in the cog labs.

Table 3. Number of Prototype Item Types, by Category and Test Form

Item Categories	Prototype item types	Form 5.DS.A.2	Form 5.RA.A.1a	Form 5.GM.A.2
Selected Response	Multiple choice (single select)	1	0	0
Constructed Response	Short answer (quantitative)	2	5	1
Technology Enhanced	Multiple drop bucket	1	0	2
	Drop Down	0	1	0
	Gap Match	0	1	0
	Hot Spot	2	0	4
	Matching Lines	0	1	0
	Matrix Interaction (including multiple select)	1	1	3
	Background graphic	2	0	0
Total		9	9	10

Please see *PIE Assessment Design and Development Technical Report* (Accessible Teaching, Learning, and Assessment Systems [ATLAS], 2025) for a full description of the PIE item features and development processes and procedures.

Sample

Eligible Students

Missouri fifth grade students participating in general education programs of study were eligible to participate. Because of resource limitations for cog lab sessions, braille forms were not available. Students with hearing impairments were able to use a sign-language interpreter.

Other Inclusion Criteria

To maximize the heterogeneity of the sample, PIE staff requested that districts recruit students representing different genders, races, and ethnicities, as well as students with disabilities and English learners. The recruitment goal was to balance representation, but the sampling plan did not have a specific quota for group representation.

Sample Size

PIE staff initially sought to recruit six students for each of the three assessment forms, for a total of 18 students. From the sites that volunteered to participate, PIE staff overrecruited by one or two students per form to guard against attrition between the time of parent or guardian consent and data collection.

Recruitment

PIE and MO-DESE staff worked together to recruit districts and schools to participate. Staff contacted potential districts by email and then followed up via virtual meetings to share information about the purpose of the sessions, session procedures, eligibility criteria, informed consent procedures, scheduling, and parent/guardian and teacher information letters. Once the district agreed to participate, staff arranged sessions by working directly with the building principals. Teachers consented to participate electronically via DocuSign. Parents and guardians signed consent forms that teachers sent home with students. Each student received a $20 gift card for participating in a cog lab session.

Procedures

Facilitator Training

PIE staff trained cog lab session facilitators on the semi-structured interview protocol, facilitation techniques, data collection processes, and student assent procedures. Facilitators were staff who had prior experience as math educators and had worked with elementary students.

General Procedures

PIE staff worked with teachers before the sessions to determine which accessibility features students might need. Although text-to-speech was not offered, facilitators provided human read-aloud support according to student need.

Two staff members conducted the sessions with students during school hours. One staff member facilitated the session while the other recorded notes in an observation form. Staff conducted the sessions individually in a quiet location. Students completed the items on a laptop provided by PIE staff, with access to an external mouse and the test administration system. Students used anonymous, generic logins to access the items. For ease of data collection, facilitators assigned students to test forms in repeating, sequential order (e.g., Student 1 received Form 1, Student 2 received Form 2).

Staff asked students to think aloud while interacting with the items. At the beginning of the session, staff modeled thinking aloud with a simple math problem, which students then practiced. Facilitators probed for additional information about the students' thought processes and reasoning as needed.

Data Analysis

Data Sources

With parent/guardian consent and student assent, staff used Zoom video-conferencing software to record video, audio, and students' screens (via screen sharing) as students interacted with the items. Completed observation forms served as back-up for screen recordings and captured observer notes to assist with data interpretation.

Finally, staff sent teachers an online Qualtrics survey to collect information on (a) whether students had received instruction on the LPs reflected in the item content, (b) whether students had previously mastered those LPs, (c) students' typical testing accommodations, and (d) whether students were receiving special education or English learner services. Results were intended to provide additional context to interpretations of all other data sources.

Analysis Approach

PIE staff applied a top-down cognitive coding approach (Willis, 2015) to the sources of data. The team coded for (a) students' response processes for each item (intended or alternate representations of the constructs), (b) item-completion time, (c) difficulties encountered including usability concerns, (d) students' comments regarding item features, and (e) relevant information from the teacher survey about students' prior instruction, mastery, and any accommodations or English learner services.

Staff used Conrad and Blair's (1996) three-stage response model to note where students exhibited cognition during the process of answering the question. These stages include:

1. **Understanding the Question**. In the question stem, the student determined what information was requested and recognized the procedure to use to satisfy the request.
2. **Performing the Task**. The student performed the mental processes required by the task. The prototype items required students to use a math response process as specified by the item task templates.
3. **Response Formatting**. The participant mapped the response created by performing the task onto the available on-screen response options.

PIE staff used a charting approach (Willis, 2015) to code, display, and analyze the data. The coding was cross-checked by data analysis team members. Staff synthesized results across lab sessions by item and LP, using the chart data display to assist with summarizing results.

Results

PIE staff completed 17 cog lab sessions with fifth grade students at two schools, one less than the target sample size of 18. For the cog lab activities, staff provided one test form for each standard, with nine or 10 items per form; items covered all three learning pathway levels for one standard. Six students completed Forms 1 and 2, and five students completed Form 3. All students understood the process of thinking aloud before beginning the sessions.

Staff explored five research questions related to student cognitive processes and interpretations and interactions with PIE items and the PIE system. The results are organized by research question.

RQ 1: How much time is required to complete the prototype items?

Staff recorded response times for the test forms used in the cog lab sessions to better understand the relationship between the intended cognition and the time needed to formulate and indicate a response.

The elapsed time described in Table 4 includes the time the student spent thinking aloud in response to prompting from the facilitator; the elapsed time also includes the time the student spent interacting with the PIE assessment platform itself (e.g., scrolling, manipulating drag-and-drop items).

Table 4. Range of Test Times, by Form

Test Form	No. of forms	No. of items	Test Time range, in minutes
Form 1: PIE.5.DS.A.2 Data and Statistics (DS)	5	9	14:37–25:00
Form 2: PIE.5.GM.A.2 Geometry and Measurement (GM)	6	10	8:47–18:43
Form 3: PIE.5.RA.A.1a Relationships and Algebraic Thinking (RA)	6	9	16:36–38:32

Results indicate that students spent more time on the Form 3 questions (i.e., *Relationships and Algebraic Thinking*) than on the other test forms. This finding is consistent with the type of cognition (i.e., articulating mathematical reasoning) expected for the test items contained on Form 3. Many of the cognitive pathways associated with the items on Form 3 involved multiple calculations, likely contributing to longer completion times on this form. In contrast, items on Forms 1 and 2 involved identification, recall, and counting, which required less time for students to complete.

RQ 2: Do students interpret items as intended?

Staff sought to understand how students interpreted a range of test items, including constructed response, selected response, and technology enhanced items. For the pilot study, PIE staff developed 28 prototype items using the PIE task models (ATLAS, 2025). These items were aligned to a specific learning pathway level for each one of three Missouri mathematics standards. Informed by Conrad and Blair's (1996) three-stage model, for RQ2, staff focused on recognizing two potential difficulties students encountered in answering test items: difficulty understanding the question, and difficulty performing the task. While difficulties such as text comprehension, boundaries between concepts, difficulty carrying out the calculation, etc., may appear at any stage of the response process, RQ2 focused specifically on challenges with comprehending the question and carrying out the required task.

Overall and across all three test forms, staff noted 17 instances of difficulty understanding the question and 18 instances of difficulty performing the task. Through think-aloud activities and interactions with teaching assistants, students provided insight into their thinking as they navigated test items.

In instances where students had difficulty understanding the question, the most significant issue noted was the vocabulary and wording of the test item; some students expressed unfamiliarity with words like *outlier*, and others required an explanation of the specific meaning of *spread out* or *grouped together* in the context of the task they were asked to perform.

In instances where students struggled to perform the task, staff noted that many students had difficulty understanding how to create and manipulate graphs and plots. Some students demonstrated misconceptions (misunderstanding the necessary calculation, or selecting an inaccurate process for carrying out the task); others did not select any process, opting instead to guess.

Some examples of student interpretations of items are described below.

- For two-part items, some students were able to determine the meaning of the first part of the question but did not use the appropriate process to answer the question and did not ask the facilitator for clarification.
- Some students requested an explanation of the second part of a two-part item; the facilitator repeated the question and offered clarification, but the student was still unable to determine the appropriate process to use.
- In a two-part item, one student required an explanation of the data table; although the facilitator provided prompting, the student did not connect the provided data with the response options.
- For two-part items, some students had difficulty performing the task, demonstrating misconceptions in which they summed values from a single list of numbers instead of comparing the values on the list to another list of numbers.
- Some students seemed to understand the question but were unable to determine a process to use and did not attempt to record a response.
- Some students expressed difficulty in gleaning information from a line plot, interpreting the meanings of values incorrectly or reversing the values shown on the *x*-axis and *y*-axis.
- Some students appeared to understand the question but then guessed at the answer.

RQ2 results suggest that modifications to vocabulary used in a question may help students comprehend the question, provided any modifications use construct-irrelevant terminology, as in some cases, vocabulary used in a question is construct-relevant and comprises part of the intended cognition. Additional modifications may include developing more succinct or more scaffolded instructions for the multiple-response items, including two-part items.

RQ 3: How do students interact with the features of the prototype items?

Staff were interested in understanding how students interacted with the prototype items in the PIE system and noted where item features may have affected students' cognitive processes. Informed by Conrad and Blair's (1996) three-stage model, for RQ3, staff focused on recognizing difficulties formatting a response.

Students were shown prototype items that contained various interactive features such as drag and drop, matrix interaction, constructed response, and "hot spots" (i.e., items with clickable zones located over specific values on a graph or line plot), requiring students to engage with the testing platform technology to complete the items. While observing student interactions with the items, staff noted areas where the features of the items may have caused difficulties with response formatting.

Overall, PIE staff identified 14 instances of difficulty with response formatting. These difficulties generally were related to manipulating the touch pad on computers that were unfamiliar to students and to the drag-and-drop functionality of some items. Other students encountered scrolling challenges while dragging item features to the appropriate area, and some clickable elements generated confusion. In these cases, most students ultimately were able to enter responses successfully.

Some examples of student interactions with items are described below:

- Completing drag-and-drop items with a touch pad instead of a mouse proved challenging for some students, but most were able to manipulate items after some prompting and explanation from the facilitator.
- A few students had difficulty determining which elements of the hot spot item types were clickable.
- Some students expressed difficulty in creating line plots using the PIE item functionality, specifically confusion about how to scroll through and add information to the plot.
- Some students were observed to drag numbers and letters to incorrect areas within the drop portion of the item and said they were confused about how to correct the error.
- A few students were unable to determine how to indicate their response until they were prompted by the facilitator.

RQ 4: Do students' responses represent the constructs the prototypes were designed to measure?

Multiple PIE item types were developed to elicit evidence of student mastery at specific levels for each learning pathway. To create prototype items, staff used task models, which are graphic organizers designed to guide item writers in creating items that are engaging and accessible for all students. PIE task models included a description of the aligned learning pathways and associated standards. The Targeted Cognition portion of the task model was intended to provide item writers with specific information about the knowledge and skills that students should demonstrate at each learning pathway level. For more information about the PIE task models, refer to the *Pathways for Instructionally Embedded Assessment: PIE Assessment Design and Development* technical report (ATLAS, 2025).

In this report, we discuss two standards and associated items as a sample of findings.

For RQ4, PIE staff created response process codes to categorize student interactions with the items during the cog labs. Staff identified intended response processes and then documented where students demonstrated misconceptions or unintended response processes. Staff were especially interested in identifying the nature of student misconceptions or other deviations from the intended cognition.

Using the item guidance provided and alignment with the intended response process, staff noted deviations from the expected response process. For example, an item that required a student to proceed through the order of operations would be expected to elicit the cognition needed to understand and execute this order to solve the problem; the student first must know the order of operations and then perform the required calculations in the correct order. A potential misconception in this process may involve a student completing calculations out of order (e.g., working backward).

For items aligned to standard **5.DS.A.2**, which asked students to interpret a line plot and use drag and drop to sort the line plots according to how the data were organized (grouped together or spread out), the expected response process is for students to count the total number of values to categorize the line plot. A noted misconception for this standard was that, when asked to determine the frequency of values on the line plot, students counted the number of unique values or used the sum of the values instead of counting the total number of values.

A sample of cognitive lab protocol questions are indicated below. To understand students' cognitive processes while they interacted with the items, facilitators were provided with multiple prompts to observe and monitor student thinking and potential misconceptions for the same test item.

- For Item 3 in Level 1 of standard **5.RA.A.1** ("Generate two numeric patterns given two rules") facilitators were directed to consider two questions:
  - *Do students determine a rule for each pattern before considering the pattern rules provided?*
  - *Are the students processing the patterns from right to left or left to right?*
- For Item 3 in Level 1 of standard **5.DS.A.2** ("Create a line plot to represent a given or generated data set, and analyze the data to answer questions and solve problems, recognizing the outliers and generating the median") facilitators were directed to observe for the following:
  - *The language in this item is slightly different than the typical wording. Do students understand the task clearly?*
  - *Do students understand how to work with dragging a compound object such as a line plot?*
  - *Do students reflect on their initial categorization to see if there are indeed similarities in the objects in each category?*
  - *Do students scroll to see all line plots?*

Staff coded the findings according to the intended response processes for the items, as shown:

1. Uses intended response process (or alternate correct response process)
2. Uses a misconception
3. Uses an incorrect response process (not a misconception; uses process of elimination, or guesses)
4. Did not attempt or facilitator provided support to student to answer question

Table 5 summarizes the number of times a given response was recorded for each item type aggregated across test forms.

Table 5. Response Codes by Item Type

Item Type	Items (*n*)	Total Number of coded responses (*n*)	Intended response process	Unintended response process	Mis-conception	Admin support or no attempt
Constructed response	8	60	28 (47%)	12 (20%)	15 (25%)	5 (8%)
Hot spot	6	34	10 (30%)	3 (9%)	18 (53%)	3 (9%)
Matrix interaction	5	39	17 (44%)	5 (13%)	14 (36%)	3 (8%)
Multiple drop bucket	3	17	5 (30%)	2 (12%)	7 (41%)	3 (18%)
Background graphic	2	10	8 (80%)	2 (20%)	0 (0%)	0 (0%)
Matching lines	1	9	4 (44%)	1 (11%)	2 (22%)	2 (22%)
Drop down	1	6	1 (17%)	3 (50%)	1 (17%)	1 (17%)
Multiple choice	1	5	3 (60%)	2 (40%)	0 (0%)	0 (0%)
Gap match	1	6	5 (83%)	0 (0%)	1 (17%)	0 (0%)

Note. Multiple response codes may have been recorded for some student–item interactions.

The drop-down item type appeared to elicit unintended response processes more frequently than intended processes; however, because of the small number (*n* = 1) of these items on test forms, any interpretations should be made with caution. While the background graphic and gap match items appeared to elicit intended responses appropriately, the same caution applies because of the small number (*n* = 1) of these items on the test forms. A mixture of response types was observed for the constructed response items, with 47% intended response processes, 25% misconceptions, and 20% unintended response processes. Hot spot items appeared to elicit misconceptions more frequently than intended or unintended response processes. Some item types, like matching lines, drop down, and multiple drop bucket, demonstrated higher non-response rates or were more likely to need support.

Facilitators also noted the characteristics of the misconceptions students exhibited. For example, a frequent misconception seen in Item 3 in Level 1 of **5.RA.A.1a** was “the student may only compare two numbers in the pattern to determine a rule, and the rule may not work for all numbers in the pattern.” Similarly, for the Level 2 task associated with standard **5.RA.A.1a**, “Extend a numeric pattern by applying the rule,” facilitators noted a misconception in which "the student may confuse growing/shrinking patterns with repeating patterns; therefore, a student may think the numbers in the pattern should repeat."

For Item 3 in Level 1 of **5.RA.A.1a**, facilitators noted it was possible for students to apply the correct process to determine the pattern rule and then engage in a misconception where they performed calculations out of sequence according to the order of operations. Therefore, this student's interactions with the item were coded both 1 ("Uses intended response process") and 2 ("Uses a misconception"). For **5.DS.A.2**, some students demonstrated challenges in understanding the meaning of item vocabulary (e.g., *outliers*, *spread out*) and required clarification from the facilitator, leading coders to code these item interactions as 4 ("Did not attempt or facilitator provided support to student to answer question").

When prompted to elaborate on their thinking for Item 3 in Level 1 of **5.RA.A.1a**, one student discussed it with the facilitator.

> Facilitator: "*You seem pretty sure about those. Were those easy?*"
>
> Student: "*A little bit. I know 6 + 4 gives me 10, and 24 − 3 gives me 21.*"
>
> Facilitator [after prompting the student to scroll down]: "*Do any of these seem more reasonable?*"
>
> Student: "*Yes.*" [student selects **Multiply by 4**]
>
> Facilitator: "*You feel like multiply by 4 works better?*"
>
> Student: "*If it's division . . . division gives us a smaller number.*"

The information from the task models and the guiding prompts for observations gave staff insight into how students worked through the cognition that the item constructs were intended to measure.

RQ 5: What are students' perceptions of the item prototypes?

For RQ5, staff collected information from students on their perceptions of and experiences with the PIE system and test items. Students were asked a series of questions and invited to provide their thoughts on the items and their anticipated performance.

Interview Question 1: Did you like this test? Why or why not?

Overall, students appreciated many of the features of PIE assessments, including the ability to return to previously answered questions, and one student spoke favorably of PIE in comparison to the *I-Ready* assessment they were accustomed to. One student disliked PIE because they dislike all tests. Another student enjoyed the test because they liked math and felt appropriately challenged by and interested in the different types of items. One student appreciated that the cog lab gave them the opportunity to explain their thinking. Students also spoke about the drag-and-drop functionality of some of the items; two students described positive experiences, while one student expressed negative feelings about the drag-and-drop items. Another student mentioned that it was difficult for them to understand where to click, and yet another student suggested that all items be displayed on the screen without having to scroll to view them all.

"*I liked the way I was able to put the X's where I felt like they were meant, and no, I don't like the way they asked the question because it was hard for me to understand what to click on and how to solve the question.*"

"*I like to drag and drop on certain things. . . but if it's reading or math and it says how many people blah blah blah, then I like to do the multiple choice.*"

"*I don't like any test. . . It gives my finger a workout.*"

Interview Question 2: Were the questions hard? Easy? Why?

Students had mixed opinions on the difficulty level of the PIE items, but most students stated that the items were moderately difficult. One student noted feeling confused about some of the items. This student also struggled with the meaning of the word *outliers* in one item but felt that the other questions were of moderate difficulty.

Interview Question 3: Did they [the questions] become easier or harder as you went along?

In general, students offered mixed opinions on the difficulty level as they proceeded through test items. Several students felt that the difficulty did not change throughout the test, one student indicated that the test became easier for them, and two students felt that the items became more difficult.

As noted for Interview Question 2, one student described confusion about some of the vocabulary in the questions and associated this confusion with increasing difficulty:

Student: "*Some [questions] were easy, and then when I got to question[s] 5 . . . and 7, it was a little confusing because . . . I didn't understand how to answer the question.*"

Facilitator: "*Are you talking about the one [where] you weren't sure how to sort them into groups?*"

Student [agreeing]: "*And I didn't know what to click on.*"

Facilitator: "*And was the other one—you didn't know how to answer the question, or you weren't sure what an outlier was?*"

Student: "*I wasn't sure what an outlier was.*"

Interview Question 4: Have you learned about these questions in math already?

A majority of students indicated that they had previously learned the concepts in the PIE prototype items. Two students mentioned that most of the concepts had been covered in the previous school year, while one student stated they had not covered the concepts. Two students indicated that they were familiar with most of the concepts, with one student mentioning a specific component they were not familiar with:

"*I learned a lot about line plot and how to create my own line plot . . . but I never learned about outliers. Everything else on the test I already learned about.*"

Interview Question 5: Do you think you got the answers right?

Some students expressed some uncertainty about their performance, while one student seemed relatively certain they had not answered all of the questions correctly and mentioned the specific test item (Item 2) they thought they answered incorrectly. Several students felt confident they had answered all questions correctly. When probed, one of the uncertain students felt like they knew what to do (i.e., they understood the correct process for solving the problem), but they were unsure about their overall performance.

Interview Question 6: Thinking of how you take tests in school, do you sometimes use extra tools? Something like bigger print, glossaries, read aloud, extra time?

Overall, students described multiple types of additional supports. Two students noted that they received extra time, while one student mentioned that they were able to zoom in on (i.e., enlarge) the test item. Another student mentioned that they sometimes received extra time, but that PIE was shorter than other tests they had taken. One student noted that sometimes computer-based tests provided them with immediate feedback that "*tells me something I don't know or something I haven't learned yet*," whereas paper-based tests only showed them the test question. One student did not use additional tools or supports, and one student was not asked this question.

Interview Question 7: How could we make this test better?

In general, students requested more clarity from the PIE test items; one student requested more information about what they should do. Another student requested definitions of unfamiliar words. One student wanted the test to be harder, because it "*makes [me] work hard so I can get a scholarship.*" One student was not asked for input on how to improve the test.

Discussion

Findings from the PIE cognitive labs shed light on how students perceived and interacted with the prototype items. The findings will be incorporated into the body of validity evidence for the PIE assessment and used to inform future item-design features, vocabulary, directions, and presentation as the PIE system continues to be refined.

Overall, students had mixed experiences with the prototype items; some students struggled with some of the vocabulary and wording used and described difficulties that stemmed from how the items were presented and described. Some students found the items relatively easy, while others described them as difficult. A few students observed that the items grew progressively more difficult as they proceeded through content. Some students wanted the instructions to be clearer, and one student suggested adding a glossary of terms to aid in understanding the necessary vocabulary.

The interactive features of some items caused confusion for some students; a few students required support from the facilitator to determine how to drag and drop responses to the appropriate area on the item. Hot spot items, which prompted students to click in specific zones in the testing platform, appeared to generate more misconceptions than did intended or unintended response processes. At times, the vocabulary of items, such as the terminology associated with line or dot plots, presented challenges; some students said they were unfamiliar with words like *outlier*, and others required an explanation of the specific meaning of *spread out* or *grouped together* in the context of the task they were asked to perform. Internal research conducted by Missouri DESE on drop down items has suggested that unintended response processes may be related to accessibility and usability for some students. These items may introduce unintended cognitive burdens when they are not adequately designed.

Analysis of students' item responses identified some additional misconceptions in student thinking, including misidentification of the appropriate process to use in solving the problem and confusion (i.e., inversion) of values and labels of line plots. In some instances, facilitators prompted students or clarified instructions or both, although these supports did not always result in students entering a response.

PIE staff will use the evidence gathered from the cog lab sessions to explore ongoing improvements to the PIE item types, including identifying item types that students may require practice interacting with. Using more accessible language may correct misconceptions related to vocabulary used in questions as long as vocabulary changes do not invalidate terminology that is part of the targeted cognition. A technology practice test was provided to teachers and students during the pilot study to help students gain familiarity with technology enhanced item types. Cog lab findings will also be used to inform future design and development efforts when new items are created.

References

Accessible Teaching, Learning, and Assessment Systems. (2025). *Pathways for Instructionally Embedded Assessment: PIE Assessment Design and Development* [technical report]. https://pie.atlas4learning.org/sites/default/files/documents/resources/PIE_Assessment_Design_Development_Technical_Report.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* American Educational Research Association. https://www.testingstandards.net/open-access-files.html

Conrad, F., & Blair, J. (1996). From impressions to data: Increasing the objectivity of cognitive interviews. In *Proceedings of the Section on Survey Research Methods, Annual meetings of the American Statistical Association* (Vol. 1, No. 10). https://www.bls.gov/osmr/research-papers/1996/pdf/st960080.pdf

Missouri Department of Elementary and Secondary Education. (2021). *Priority standards for leveraging learning in mathematics: Grades K–12*. https://dese.mo.gov/college-career-readiness/curriculum/academic-standards/priority-standards

Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford University Press.