Pathways for Instructionally Embedded Assessment

Pathways for Instructionally Embedded Assessment: PIE Assessment Design and Development

Technical Report

January 2025

Acknowledgments

Several ATLAS staff members made significant writing contributions to this report as listed below.

Kristine David
Allison Barkley
Eun Mi Kim
Auburn Andrew Jimenez
Breonna Yungeberg

We would also like to acknowledge several staff for their contributions to this report, Lisa Stephen, Candace Bond, David Whitcomb, and Jake Thompson.

Project Director: Shaun Bates, Missouri Department of Elementary and Secondary Education

Principal Investigator: Brooke Nash, ATLAS, University of Kansas

Mathematics Pathways Director: Mary Majerus, Missouri Department of Elementary and Secondary Education

Project Advisory Committee (PAC)

Brian Gong, Ph.D., Center of Assessment
Bob Henson, Ph.D., University of North Carolina-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
David Williamson, Ph.D., College Board
Phoebe Winter, Ph.D., Independent Consultant
Lisa Sireno, Missouri Department of Elementary and Secondary Education (DESE)

The project described in this report was developed under a grant from the U.S. Department of Education (S368A220019). The content does not necessarily represent the policies of the U.S. Department of Education, and no assumption of endorsement by the federal government should be made.

Table of Contents

Introduction 1
Content Framework 2
Missouri Priority Learning Standards 2
Learning Pathways 4
Assessment Design 5
Assessment Content 5
Instructionally Embedded Assessments 5
End-of-Year Assessment 5
Item Types 6
PIE Task Models 7
Description and Purpose 7
Task-Model Template Development 8
Prototype Items 9
Cognitive Labs 9
Assessment Development 10
Item Writing 10
Recruitment 10
Participants 10
Training 12
Process 12
Item Development 13
Item External Review 13
Item-Twin Development 18
Item Field-Testing 19
Recruitment 19
Form Development 19
Procedures 20
Results 21
Item Data-Review Process and Decisions 27
Next Steps for the System Pilot Study 27
Item-Twin Development 27

Preliminary Scoring-Model Calibration and Evaluation 27
Conclusions and Future Directions 28
References 31
Appendix A: Instructionally Embedded Assessment Pathway Levels 33
Appendix B: Task Model Example 37

Introduction

This technical report describes the steps and processes taken to design and develop assessments for the Pathways for Instructionally Embedded Assessment (PIE) project. The Missouri Department of Elementary and Secondary Education (Missouri DESE) leads the PIE project in partnership with Accessible Teaching, Learning, and Assessment Systems (ATLAS) at the University of Kansas.

Funded through a Competitive Grant for State Assessments, the PIE project aims to construct and evaluate a prototype assessment system built on cognitive models of student learning (named *learning pathways*) aligned with grade-level academic expectations. The prototype system includes assessments embedded throughout instruction, as well as at the end of the course or year. The Missouri Department of Elementary and Secondary Education (Missouri DESE) leads the PIE project in partnership with Accessible Teaching, Learning, and Assessment Systems (ATLAS) at the University of Kansas.

Instructionally embedded assessments give students the opportunity to demonstrate what they know and can do as they learn. Using learning pathways as the basis for assessments that are administered at instructionally relevant points in time, embedded assessments produce more fine-grained and timely information about student academic progress for teachers to inform their instructional decisions during the school year. The end-of-year assessment, in combination with the instructionally embedded assessments, supports a multiple-measures approach to assessment that has potential to meet the goals of traditional spring summative assessment models. The PIE project aims to evaluate the feasibility and technical adequacy of this integrated, prototype assessment system.

This technical report describes the steps and processes taken to design and develop assessments for the PIE project. The first section describes the content framework used for the assessment system, the learning pathways, as well as the scope of learning-pathway development for the PIE assessment system. The assessment design and assessment development sections describe the assessment design and steps taken to develop the PIE assessments. The PIE assessments are intended to be used in a full system pilot study conducted in fifth-grade mathematics classrooms during the 2024–2025 academic year.

The PIE assessment design and development phase of the project involved staff from several ATLAS teams and Missouri DESE. ATLAS project staff represented project management, content development, implementation, technology, and research and psychometric teams. In this report, “PIE project staff” refers to the collective efforts of the ATLAS staff and teams who contributed to the project. “MO-DESE project staff” refers to the Missouri DESE staff who contributed to the project.

Content Framework

Missouri Priority Learning Standards

The Missouri Learning Standards Grade- and Course-Level Expectations (MLS) identified priority standards to be used as a resource for Missouri educators (Missouri DESE, 2021). Priority standards focus on big ideas for each grade level. The mathematics standards are grouped by domain, clusters, standards, and substandards. Mathematical content is in a hierarchy: clusters are sublevels of domains, standards are sublevels of clusters, and substandards are sublevels of standards.

Missouri DESE selected 25 fifth-grade mathematics priority standards to develop learning pathways for the PIE project. A summary of the selected priority standards for the PIE project is shown in Table 1.

Table 1.

Summary of Selected Priority Standards

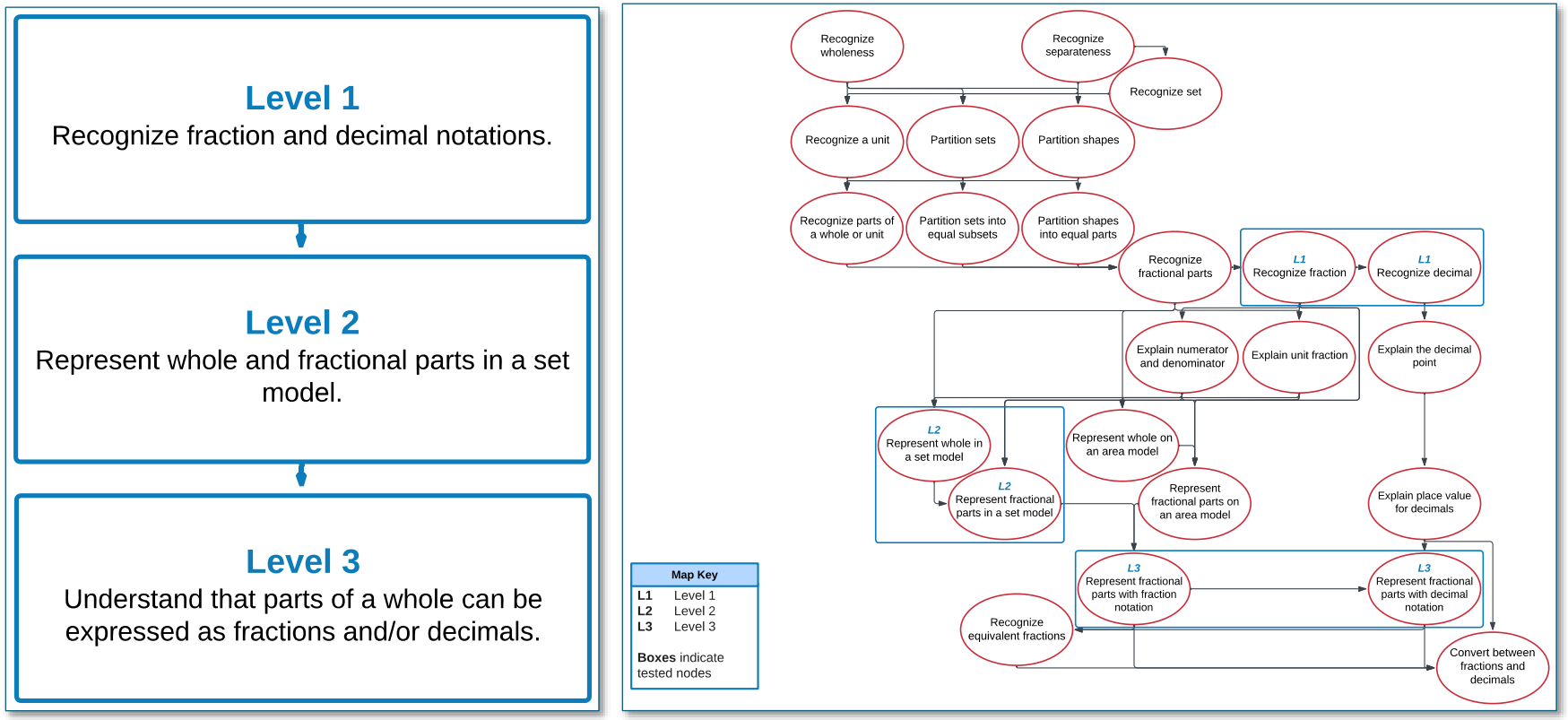
Domain	Cluster	Standard
NF: Number Sense and Operations in Fractions	Understand the relationship between fractions and decimals (denominators that are factors of 100).	5.NF.A.1
		5.NF.A.2
		5.NF.A.3
	Perform operations and solve problems with fractions and decimals.	5.NF.B.4
		5.NF.B.5a
		5.NF.B.5b
		5.NF.B.5c
		5.NF.B.5d
		5.NF.B.6
		5.NF.B.7a
		5.NF.B.7b
		5.NF.B.7c
		5.NF.B.8a
		5.NF.B.8b
RA: Relationships and Algebraic Thinking	Represent and analyze patterns and relationships.	5.RA.A.1a
		5.RA.A.1b
		5.RA.A.1c
		5.RA.A.1d
		5.RA.A.2
	Use the four operations to represent and solve problems.	5.RA.C.5
GM: Geometry and Measurement	Classify two- and three-dimensional geometric shapes.	5.GM.A.2
	Understand and compute volume.	5.GM.B.4a
		5.GM.B.4b
	Graph points on the Cartesian coordinate plane within the first quadrant to solve problems.	5.GM.C.6a
DS: Data and Statistics	Represent and analyze data.	5.DS.A.2

Learning Pathways

Learning pathways depict sequences of knowledge and skill acquisition that are aligned to grade-level mathematics content standards and are summarized within three vertically articulated levels (Kim et al., 2024). Targeting the Missouri DESE-selected priority standards for fifth-grade mathematics (Missouri DESE, 2021), 25 learning pathways (Kim et al., 2024) were constructed by adapting established processes and procedures of learning map development (e.g., Dynamic Learning Maps [DLM] Consortium, 2016; Swinburne Romine et al., 2018). The learning pathways include a more fine-grained map view of knowledge, skills and understandings leading up to and including the grade-level content standard, as well as a view of the three pathway levels generated from the maps (Kim et al., 2024), as shown in Figure 1. The learning pathways formed the foundational building blocks for the PIE project (Kim et al., 2024).

Figure 1.

Learning-Pathway Construct Example in a Level Diagram and a Map



Note. From "PIE Learning Pathways," by Pathways for Instructionally Embedded Assessment, 2024, pp. 1–2 (https://pie.atlas4learning.org/sites/default/files/documents/resources/PIE_Learning_Pathways.pdf). Copyright 2024 by University of Kansas, Accessible Teaching, Learning, and Assessment Systems (ATLAS),.

Assessment Design

Assessment items were developed for two components of the PIE system: instructionally embedded assessments and an end-of-year assessment. These assessments are described below, including item-type information, assessment blueprints, and task-model development.

Assessment Content

There are four fifth-grade domains to be assessed within the PIE blueprints: Number Sense and Operations in Fractions; Relationships and Algebraic Thinking; Geometry and Measurement; and Data and Statistics. PIE assessments measure students' KSUs at the learning-pathway levels; each item measures one pathway level for one standard. Assessment content (i.e., learning pathway levels) is provided in Appendix A.

Instructionally Embedded Assessments

During the instructionally embedded assessment window of the pilot study, teachers will create approximately seven to eight groupings of two to five content standards (see Appendix A) as the basis for instruction and assessment. Each learning pathway will be measured with three to four items (see Xu, 2019). For each content grouping, teachers will administer short assessments at the beginning, middle, and end of the instructional unit on the grouping of standards. The assessments are dynamically generated to measure the knowledge and skills aligned to the standards selected by the teacher for the grouping. By the end of the instructionally embedded window, all standards are assessed.

The instructionally embedded assessments consist of 1-point and 2-point items. One-point items measure a primary KSU. Two-point items measure multiple KSUs and may require multiple student responses, within the same item, to arrive at the correct answer.

Additionally, 34% of the items are in context. These items may require students to apply their understanding of the KSUs in a real-world context or to interpret the results within the context of the items.

End-of-Year Assessment

For the purposes of the pilot study a fixed, 50-item, end-of-year assessment form will be administered to students. Assessment content is noted in Table 2. Because the pilot study is based on a subset of the Missouri priority mathematics standards, the specific content standards are also included as a column in Table 2.

Table 2.

End-of-Year Assessment Content

Domain	Cluster	Standard	Cluster point total	Domain point total
Number Sense & Operations in Fractions (NF)	Extend understanding of fraction equivalence and ordering.	5.NF.A.1 5.NF.A.2 5.NF.A.3	9	37
	Extend understanding of operations on whole numbers to fraction operations.	5.NF.B.4 5.NF.B.5a 5.NF.B.5b 5.NF.B.5c 5.NF.B.5d 5.NF.B.6 5.NF.B.7a 5.NF.B.7b 5.NF.B.7c 5.NF.B.8a 5.NF.B.8b	28	
Relationships & Algebraic Thinking (RA)	Use the four operations with whole numbers to solve problems.	5.RA.A.1a 5.RA.A.1b 5.RA.A.1c 5.RA.A.1d 5.RA.A.2	10	13
	Generate and analyze patterns.	5.RA.C.5	3	
Geometry & Measurement (GM)	GM.A: Classify 2-dimenstional shapes by properties of their lines and angles.	5.GM.A.2	2	8
	GM.B: Understand the concepts of angle and measure angles.	5.GM.B.4a 5.GM.B.4b	4	
	GM.C: Solve problems involving measurement and conversion of measurements from a larger unit to a smaller unit.	5.GM.C.6a	2	
Data & Statistics (DS)	DS.A: Represent and analyze data.	5.DS.A.2	3	3
Point totals			61	61

Item Types

Multiple item types are used to appropriately elicit evidence of mastery for each learning-pathway level. For example, a multiple-choice item may best measure one of the learning pathways, whereas a short-answer item may better elicit evidence of a different learning-

pathway level. Table 3 displays all item types used in the PIE assessments. During the construction of the task models, appropriate item types were evaluated and selected according to the demands required for each learning pathway.

Table 3.

PIE Item Types

Item category	Specific item type
Selected response	Multiple choice (single select and multiple select)
Constructed response	Short answer (quantitative)
	Drop down
	Hot spot
	Matrix interaction (including multiple select)
Technology enhanced	Drag and drop (graphic, text, or both)
	Matching lines
	Placing points
	Composite item types

In addition to the item types, item blocks were included in available item types. A block of items includes two items that appear together and have a shared stimulus.

PIE Task Models

The test-design plan includes task models that are based on evidence-centered design. These task models provide guidelines for test developers to design assessments items that engage all students, including students with disabilities and English learners, and that provide all students with access to the assessments. In total, 75 task models were developed, one for each of the learning-pathway levels. An example of a PIE task model is in Appendix B.

Description and Purpose

The task models used for PIE serve as graphic organizers that guide test developers (Wine & Hoffman, 2024). Each of the three learning-pathway levels aligned to the 25 content standards has a unique task model. The first section of the task model displays item-level metadata, including the level statement, Missouri standard code, domain, cluster, and standard language. Each level statement aligned to the standard (i.e., level 1, level 2, level 3) is included on the task-model front page to illustrate how the level statements build to level 3, the grade-level expectation. Additionally, related standards are noted.

The second section of the task model, Targeted Cognition, includes specific information regarding the specific knowledge, skills, and abilities demanded at each learning-pathway level to be used for assessment purposes. This includes the types of evidence considered outside the targeted cognition, disambiguation from similar standards of the targeted cognition, grade-level-specific considerations for the targeted cognition, and assessment challenges (including but not limited to issues that may arise when assessing the targeted cognition with existing tools available with computer assessments).

The third section of the task model is Item-Level Considerations.

- Recommended item types
- Students' misconceptions/bases for distractors
- Terminology that may be required, recommended, or verboten
- UDL options
- Accessibility/test-administration customizations (e.g., for multiple-choice items, consider response options that contain long phrases or complete sentences should be written using parallel structure to ease the reading comprehension burden)
- Graphics guidance
- Other math requirements, such as calculator usage
- Depth of knowledge ceiling
- Context or scenario ideas
- Recently overused ideas (e.g., contexts that have been used with assessment items in the past and should be avoided to offer greater diversity of topics for student engagement)
- Unsuccessful approaches when building items

Additionally, Universal Design for Learning (CAST, 2018) is considered, and options are provided to guide item writers to construct items that allow all students to access the assessment item content. Specifically, the Universal Design for Learning guidelines that are applicable to the content will be included to guide item writers in developing items that are engaging and accessible, and that avoid introducing barriers to students' abilities to show what they know and can do. This specific item-level information guides item writers to construct items and components, including characteristics and features that should be considered in the item construction to allow students to provide evidence of the mastery of the learning pathway.

The fourth section of the task model, Models/Item Skeletons, provides models to be referenced by item writers. The models offer explanations and annotations to further guide item writers by providing rationales that they may apply in the construction of new items.

The last section of the task model, Nodes, articulates both the node numbers and names and the node descriptions from the learning-pathway maps (see Kim et al., 2024, for a map example). The learning-pathway nodes depict the domain KSUs that display the learning targets and their precursors (DLM Consortium, 2016; Swinburne Romine et al., 2018; see also Kim et al., 2024).

Task-Model Template Development

The PIE task-model template was based on a modified version of the Dynamic Learning Maps (DLM) Assessment System task models and the work of Wine & Hoffman (2024). The task-model template provided the target of the level statement for assessment purposes as well as any specific boundaries that needed to be applied to items (e.g., grade-level limitations). Item writers were encouraged to be creative in constructing items to provide evidence of the targeted cognition within these boundaries, as well as using established approaches. The task models provide a foundation for item drafting and development.

Before item development, the PIE project team worked in groups of two or three to complete the task models for all 75 level statements. A kick-off meeting was held before constructing the task-model content to discuss the process for construction and to provide an overview of the sections. During development, a high-level review of the task models was done by a senior math content development team member to check for consistency, and an editing check was done to ensure there were no grammatical or style errors.

Prototype Items

Using the PIE task models, the project team developed 28 prototype items, including three to four items aligned to a learning-pathway level for each of three standards. The prototype items were written to elicit evidence of the targeted cognition, as described in the task models. The items used a variety of item types, including selected-response, constructed-response, and technology-enhanced items. After the initial development, the items were reviewed by the project team to ensure alignment to the level statement. The development of prototype items also involved graphics creation, a fact-checking review, and a fairness review to identify any accessibility, bias, or sensitivity concerns. To ensure all aspects of the items functioned appropriately, items were copyedited and reviewed in Student Portal, the platform students use to take assessments.

The item prototypes were shared with the Project Advisory Committee (PAC) on June 23, 2023. PAC members suggested that the grain size of the task models support the development of items that include all the necessary evidence required in the learning pathway level statement, while ensuring that items written to the same pathway level were fungible (interchangeable; meaning they are expected to perform the same). This feedback informed final task-model development before item writing.

Cognitive Labs

The purpose of the cognitive labs was to gather evidence of student response processes and interactions with the PIE prototype items. The cognitive labs were held at two sites in Missouri in December 2023 and January 2024. Between the two sites, four teachers and 17 students engaged in the cognitive labs.

There were three forms for cognitive labs, one form for each standard. The items on each form covered all three levels for one of the standards. Each form had nine or 10 items. The number of items associated with each standard and learning pathway level are displayed in Table 4.

Table 4.

Number of Items per Standard and Learning-Pathway Level

Standard and learning-pathway level	No. of items
5.DS.A.2	
Level 1	3
Level 2	3
Level 3	3
5.GM.A.2	
Level 1	4
Level 2	3
Level 3	3
5.RA.A.1a	
Level 1	3
Level 2	3
Level 3	3

Students clearly articulated their thinking during the labs. It was noted that students at one of the sites seemed less comfortable with the technology and item types than at the other site. Additionally, staff observed some misconceptions and gaps in content knowledge even when students had previously learned the content. This information will be used to further inform educators on the learning pathways and assist in more-targeted instruction as students progress through the instructionally embedded assessments during the pilot study.

Assessment Development

Item Writing

Educators were convened to write items during a workshop held in Columbia, Missouri, from September 20–22, 2023. Twenty-eight Missouri educators wrote 402 items to assess the learning-pathway levels aligned to 25 priority fifth-grade Missouri mathematics standards.

Recruitment

Educators were recruited for the item-writing workshops via Missouri listservs and recruitment fliers. Seventy-two educators submitted an interest survey for the workshop, including demographic information, education, and teaching experience. The PIE project team reviewed the submissions and invited participants based on years of experience in teaching K–12 students, years of experience teaching mathematics, and to ensure a diverse geographical representation, to the extent possible.

Participants

Twenty-eight item writers attended the item-writing workshop. All participants were female and non-Hispanic. Ninety-three percent (n = 26) self-identified as White, and 7% (n = 2) self-

identified as Black. Table 5, Table 6, Table 7, Table 8, and Table 9 provide additional information about the item writers.

Table 5.

*Item Writers' K–12 Teaching Experience (*N *= 28)*

K–12 teaching experience, in years	*n*	%
0–5	5	17.9
6–10	4	14.2
11–15	6	21.4
16–20	4	14.2
Over 20	9	32.1

Table 6.

*Item Writers' Mathematics Teaching Experience (*N *= 28)*

Mathematics teaching experience, in years	*n*	%
0–5	6	21.4
6–10	6	21.4
11–15	6	21.4
16–20	3	10.7
Over 20	7	25.0

The 28 item writers represented a highly qualified group of educators with mathematics teaching experience. The degrees held by item writers are shown in Table 7; all item writers held at least a bachelor's degree. Most item writers (*n* = 18; 64.3%) also held a master's degree.

Table 7.

*Item Writers' Education (*N *= 28)*

Degree	*n*	%
Bachelor's degree	10	35.7
Master's degree	18	64.3

The professional roles reported by item writers are shown in Table 8. Although item writers had a range of professional roles, they were primarily classroom teachers.

Table 8.

*Item Writers' Roles (*N *= 28)*

Role	*n*	%
Classroom teacher	22	78.6
Instructional coach	5	17.9
Elementary math coordinator	1	3.6
Curriculum coordinator	1	3.6
Interventionist	1	3.6

Note. Item writers holding multiple roles are included in the category for each role they serve.

Item writers came from across the state of Missouri. There are nine supervisory districts in Missouri (A–I), and representation from each district is reported.

Table 9.

*Supervisory Districts (*N *= 28)*

Supervisory district	*n*	%
A	5	17.9
B	4	14.2
C	7	25.0
D	4	14.2
E	2	7.1
F	1	3.6
G	1	3.6
H	2	7.1
I	2	7.1

Training

On the first day of the workshop, PIE project staff provided an overview of the PIE project and an orientation to item-writing best practices. Item writers engaged in identifying common errors when writing items as part of this orientation. Task models were explained and reviewed with item writers to provide specific information (e.g., metadata including standards, levels, item types, etc.) for consideration when writing their items. Fairness considerations were noted within the training, including universal design principles, system supports that are offered to students while engaging with items, accommodations that are provided to students per their personal needs profile, bias concerns, sensitivity concerns, and language considerations. and Item writers noted these considerations when constructing items.

Process

After orientation and training, item writers produced and peer-reviewed their items with other participants. The PIE project team provided ongoing guidance to the item writers, provided input while the writers were constructing their items, and were available to answer questions

from the item writers during the workshop. In total, 443 items were written during the workshop.

Item Development

After the item-writing workshop, the draft items were further developed by the PIE project team. The development of items included revisions to ensure alignment to the level statement/standard, fact-checking review, the addition of graphics when appropriate, accessibility review to identify any accessibility, bias, or sensitivity concerns, and editorial review. Finally, MO-DESE project staff reviewed the items prior to an external review of items.

Item External Review

The purpose of the external review was to evaluate whether the PIE items measured the intended cognition. The items were reviewed by two separate panels of educators. The first panel considered the content of the items and ensured that the items aligned to the level statement, were free of errors, and were keyed correctly. The second panel focused on identifying any fairness or accessibility concerns in the items. The content and fairness reviews occurred asynchronously from November 13, 2023, to January 15, 2024.

Recruitment

Missouri educators were recruited for the external review via Missouri listservs and contacted by MO-DESE project staff. Sixty-two educators submitted an interest survey for the workshop, including demographic information, education, years of teaching experience and experience working with a diverse population of students, including but not limited to students with disabilities, English learners, and students who have experienced trauma. Educators who participated in the item-writing workshop were not eligible to participate in the external review.

PIE project staff reviewed the pool of potential external reviewers to identify a list of diverse panel of item reviewers. MO-DESE project staff reviewed the list and provided additional recommendations. External reviewers then received invitations via email to participate in the external review of items.

Participants

Sixteen educators participated in the external review of items. Most panelists were female, White, and non-Hispanic; had more than 16 years of K–12 teaching experience and 16 years of teaching mathematics; and held at least a master's degree. Table 10, Table 11, Table 12, Table 13, Table 14, and Table 15 give additional information about the external reviewers.

Table 10.

*Demographics of External Reviewers (*N *= 16)*

Demographic	Content panel		Fairness panel	
	N	%	*n*	%
Gender				
Female	8	100	7	87.5
Male	0	0.0	1	12.5
Race				
White	8	100	8	100.0
Black	0	0.0	0	0.0
Hispanic ethnicity				
Non-Hispanic	8	100	8	100
Hispanic	0	0.0	0	0.0

Table 11.

*External Reviewers' Experience Teaching Grades K–12 (*N *= 16)*

K–12 teaching experience, in years	Content panel		Fairness panel	
	n	%	*n*	%
0–5	0	0.0	0	0.0
6–10	1	12.5	1	12.5
11–15	2	25.0	3	37.5
16–20	3	37.5	3	37.5
Over 20	2	25.0	1	12.5

Table 12.

*External Reviewers' Experience Teaching Mathematics (*N *= 16)*

Mathematics teaching experience, in years	Content panel		Fairness panel	
	n	%	*n*	%
0–5	0	0.0	1	12.5
6–10	1	12.5	1	25.5
11–15	2	25.0	3	37.5
16–20	3	37.5	3	37.5
Over 20	2	25.0	0	0.0

The degrees held by external reviewers are shown below. All external reviewers held at least a master's degree.

Table 13.

*Item Writers' Degree Types (*N *= 16)*

Degree	Content panel		Fairness panel	
	n	%	*n*	%
Master's degree	7	87.5	5	62.5
Other advanced degree	1	12.5	3	37.5

Most panelists reported that they were teachers.

Table 14.

*Item Writers' Roles (*N *= 16)*

Role	Content panel		Fairness panel	
	n	%	*n*	%
Teacher	6	75.0	3	37.5
Instructional coach	3	37.5	4	50.0
Building principal	0	0.0	1	12.5

Note. External reviewers holding multiple roles are included in the category for each role they serve.

External reviewers were from across the state of Missouri. There are nine supervisory districts in Missouri and representation from each district is reported in Table 15.

Table 15.

*Item Writers' Supervisory Districts (*N *= 16)*

Supervisory district	Content panel		Fairness panel	
	n	%	*n*	%
A	0	0.0	1	12.5
B	1	12.5	2	25.0
C	3	37.5	0	0.0
D	1	12.5	2	25.0
E	1	12.5	0	0.0
F	0	0.0	0	0.0
G	1	12.5	1	12.5
H	1	12.5	2	25.0
I	0	0.0	0	0.0

Fairness panelists reported their experiences with students with diverse national origins (25%), trauma experiences (75%), various socioeconomic statuses (88%), disabilities (63%), various orientations (38%), English language learners (50%), and various religious backgrounds (38%).

Training

The external review training began with an virtual orientation for all reviewers via Zoom. During the orientation, participants were given background information on the PIE assessment system design and an overview of the external-review process. Participants engaged in practice sessions to confirm their understanding of the process and to ask any questions before beginning their external review of items. The participants were trained in using External Review, the secure online system used to display the items and metadata, and on capturing their feedback on each item.

Criteria

Content panelists focused their review on the content of the items, including alignment of items to the level statements, grade-level appropriateness, and ensuring the items were free of content errors. Fairness panelists focused their review on any bias, sensitivity, and accessibility concerns. The following guiding questions were used by panelists during their reviews.

Content review

EVIDENCE AND TARGETED COGNITION

- Does the work product (i.e., the answer or answers) contain the information required by the Learning Pathway Level-Specific Evidence Statement?
- Does the item provide an opportunity for the student to engage in the targeted cognition listed in the Targeted Cognition Explanation for Assessment Purposes?

ITEM CONSIDERATIONS

- Is there a clear question or statement?
- Is the identified key correct?
- Is each distractor plausible, and does it represent an error in the targeted cognition (if applicable)?

QUESTIONS TO GUIDE FEEDBACK

For places where reviewers anticipated where, in the item, a disruption relative to the targeted cognition may occur, they were asked to answer the following questions:

- Where in the item does the problem appear?
- Which group of test takers are at risk?
- How will the student's cognitive path be inappropriately disrupted?
- Which skills will be impacted?

Fairness review

ACCESSIBILITY

- Does the item allow for all students in the tested population to interact with the content?
- Does the graphic (if included) represent information in a clear manner?

BIAS

- Are there any features in the item that unfairly advantage or disadvantage a particular group of students?

SENSITIVITY

- Are there any features in the item that may evoke a strong emotional reaction that may interfere with the performance of a particular group of students?

QUESTIONS TO GUIDE FEEDBACK

For places where there may be accessibility, bias, or sensitivity concerns, answer the following questions:

- Where in the item does the problem appear?
- Which group of test takers are at risk?

Process

After the training session, reviewers were given three batches of eight to 15 items to review asynchronously during each week between October 2, 2023, and February 12, 2024. In total, 402 items were reviewed by three or four content reviewers and three or four fairness reviewers. To ensure that unique content was reviewed by reviewers in each batch, the distribution of content reviewers and fairness reviewers was altered with each batch.

Ratings

As external reviewers reviewed their items, they were prompted to accept the item, accept the item with revisions, or reject the item. If the reviewer accepted the item with revisions or rejected it, the reviewer was required to provide a rationale and comment about their decision.

The external-review ratings and comments were collected from each content and fairness panelist. The external-review ratings of 402 items by content and fairness panels are shown.

- 30% of the items were accepted by all reviewers in one or both panels.
- 40% of the items were accepted with revisions by one or both panels.
- 24% of the items had two or more revise ratings by one or both panels.
- 7% of the items had one or more reject rating by one or both panels.

Items rated by individual panels are listed in Table 16.

Table 16.

*Item Ratings by Panel Type (*N *= 402)*

Panel type and rating	Items (*n*)	Items (%)
Content		
All Accept ratings	227	56
1 Revise rating	132	33
≥2 or more Revise ratings	36	9
≥1 or more Reject ratings	7	2
Fairness		
All Accept ratings	187	47
1 Revise rating	119	30
≥2 Revise ratings	73	18
≥1 or more Reject ratings	23	6

Decisions

After the external review of the items was completed, the PIE project team reviewed the panel recommendations and revised items as necessary. PIE project staff accepted and made no revisions to 47% of the items and revised 53% of the items. No items were rejected.

The revised items went through additional internal-review processes, including content, accessibility, fact-checking, and editorial reviews. The edited items, along with the original feedback received from external-review participants, were given to MO-DESE project staff for review before field-testing.

Item-Twin Development

Item twins are content analogs (Karvonen et al., 2020) to items that were developed by item writers. Item twins were constructed using the task models and were created for PIE assessments for retest use. A subset of item twins were field-tested to confirm that item statistics from the original item could be applied to the item twin.

Nineteen item twins were developed by the PIE project team. The item twins were developed using level-1 and level-2 items and their respective task models. Each item twin measured the same KSUs as the original item and was written using the specifications described in the task model. Item twins used the same item type and overall structure as the original item. Some features that were modified from the original items for the item twins were context and values. For example, an item that asks students to determine the volume of a given rectangular prism may have an item twin with the same student prompt, but the rectangular prism in the item twin would have different dimensions than the prism used in the original item.

Each field-tested item twin was placed on the same form as the original item. The purpose of field-testing the item twins with its original-item counterpart was to determine whether differences between the item twin and the original item resulted in a change in item performance. After field-testing, the data from the item twin and original item were reviewed

together. On the basis of this review, item twins were designated for potential use in the pilot study or rejected. Of the 19 field-tested item twins, 11 items were eligible for pilot use; the other eight items were rejected. The retained item twins informed subsequent item-twin development for eventual use on the pilot study. Specifically, the types of changes between the original item and item twins field-tested were reviewed, and those changes were replicated for subsequent item-twin development.

Item Field-Testing

The PIE item field test was delivered to students March 4–22, 2024. The purpose of the field test was to evaluate the technical quality of the new items prior to form development for the instructionally embedded and end-of-year assessments. Additional level-3 items were developed and represented on the field-test forms to populate both the instructionally embedded assessments and end-of-year assessments. End-of-year assessments include only level-3 items, so additional items at this level were needed.

Recruitment

PIE project staff distributed recruitment messages to the MO-DESE project staff. Using the inclusion criteria (i.e., fifth-grade students in Missouri who participate in general education programs of study and their teachers), MO-DESE project staff distributed the recruitment message to districts that were likely to be interested with schools willing to participate, including those who had participated in prior PIE project activities (e.g., external review of learning maps, item writing, etc.). Districts then confirmed their commitment to data-management tasks and identified their test coordinator to MO-DESE project staff. Schools verified they had eligible teachers and obtained approval to participate.

School and district leaders were asked to forward recruitment information to their current fifth-grade math teachers. After recruitment was completed, PIE project staff sent emails to districts and schools outlining the next steps for participating in the item field test.

The goal was to recruit approximately 65 teachers from across Missouri districts and regions who were currently teaching fifth-grade students. The target number of students to participate in the item field test was approximately 1,625.

Form Development

Before field-test form development, PIE project staff created a test map to ensure all form requirements were met. A total of 19 forms were developed for the item field test; each form measured the three learning-pathway levels for two content standards, and each pathway level was measured by three to four items. Additional level-3 items measuring any content standard were also included on each form, for a total of 22 items per form.

A total of 411 items were field-tested, including 19 item twins. Some items were repeated on two forms to meet the required minimum of three to four items per level for each standard. PIE project staff considered the items and level statements for each content standard when populating the forms. The test form maps balanced items across the 19 forms, maximizing the number of items to field test and meet psychometric requirements. Level statements with more cognitively complex content (e.g., 2 point items) were balanced across forms in order to

minimize student fatigue and to try to balance testing time across the forms. Standards that contained similar tasks were balanced across the forms to avoid student frustration from repetitive tasks. The field-test forms were reviewed before field-testing by PIE project staff to mitigate any issues with clueing and clang and to ensure the items were scored correctly.

Procedures

Teacher training

To prepare for the PIE item field test, educators administering the PIE item field test were given a brief test-administration manual, quick guide, and an informational video.

Administration

Before students were administered in the PIE item field test, a practice test was available for students to engage with the testing platform and the item types available during the assessment. The practice test was deployed on the Kite® platform, and students were provided with a universal, generic login and password to enter the practice test. The practice test included 12 practice-test items, including multiple-choice, constructed-response, and technology-enhanced items. The practice-test items did not include math content; rather, they provided students an opportunity to engage with the item types and the functionality of the Kite system.

For the item field test, generic user IDs were generated and provided to each teacher for each student in their classroom, and no personal identifiable information was collected. Each student randomly received one of 19 available forms on the Kite platform. Because generic IDs were used, all students had access to all designated features and accommodations typically available to them through their personal profiles, including those listed in Table 17.

Table 17.

Designated Features and Accommodations

Embedded designated features and accommodations	Local-level designated features and accommodations
Color contrast	Bilingual word-to-word dictionary
Color overlay	Oral reading by teacher in native language for English learners
Reverse contrast	Signing
Masking	Testing individually
Text-to-speech/read aloud/spoken	Administration of assessment in several sessions or at a specific time of day
Single switch	Use of assistive devices

Braille was not available for the PIE field test.

Kite Student Portal also provided several tools for all students during the PIE field test. Students could choose the following tools to use during the assessment:

- Pointer

- Highlighter
- Eraser
- Notes
- Sketch pad
- Striker
- Magnification

Results

District and school participants

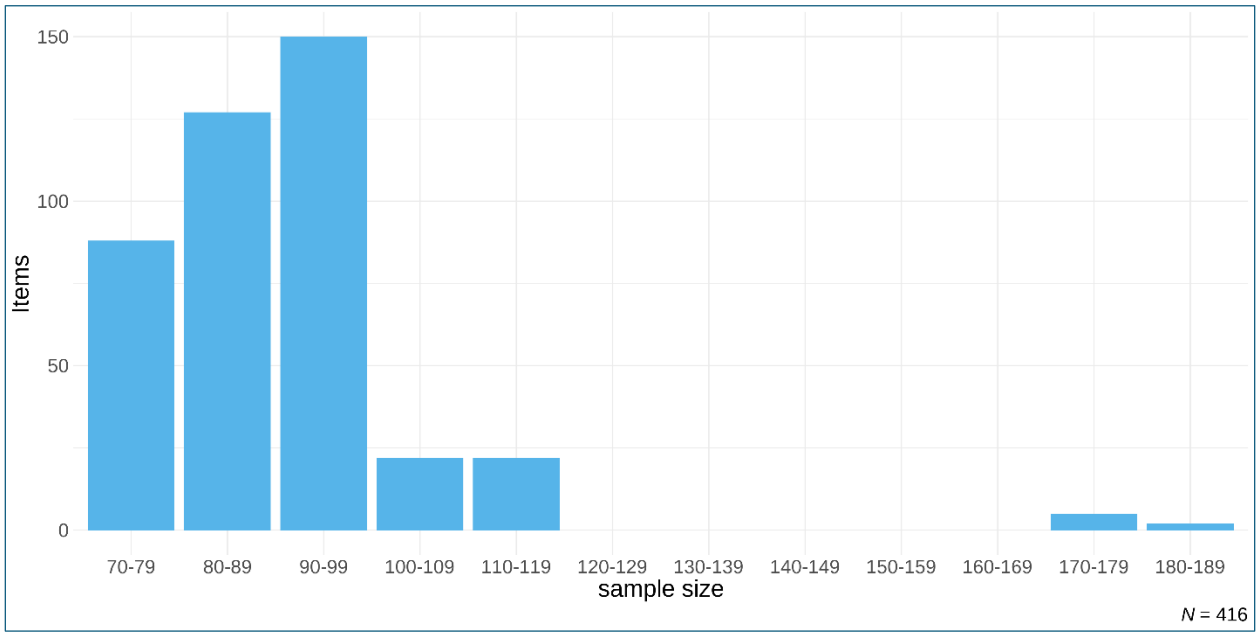
The final participation statistics for the PIE item field test included 1,708 fifth-grade students, representing 71 teachers, 36 schools, and 32 districts.

Sample sizes

Figure 2 shows the student sample size distribution for the field-tested items.

Figure 2.

Student Sample Size Distribution for Math Items



Item-data summaries

Students' response data from the item field-test administration window was used to calculate item statistics and evaluate item quality. Item statistics included overall item difficulty, conditional item difficulty, and item-discrimination indices for each item.

The *item-difficulty index* denotes the average difficulty of an item. For 1-point items, the item-difficulty index represents the overall probability (i.e., *p* value) that students provide a correct response to an item. For 2-point items, the item-difficulty index represents the average score of students on an item, ranging from 0 to 2. PIE project staff flagged 1- and 2-point items as being too easy if item-difficulty indices were greater than 0.95 or 1.9, respectively. Items were also

flagged as too challenging if item-difficulty indices were less than 0.25 or 0.5, respectively. In total, 363 items (87.3%) were within the acceptable range defined by the flagging criteria. Figure 3 and Figure 4 summarize the overall item-difficulty indices for 1- and 2-point items.

Figure 3.

Item-Difficulty Indices for 1-Point Math Items

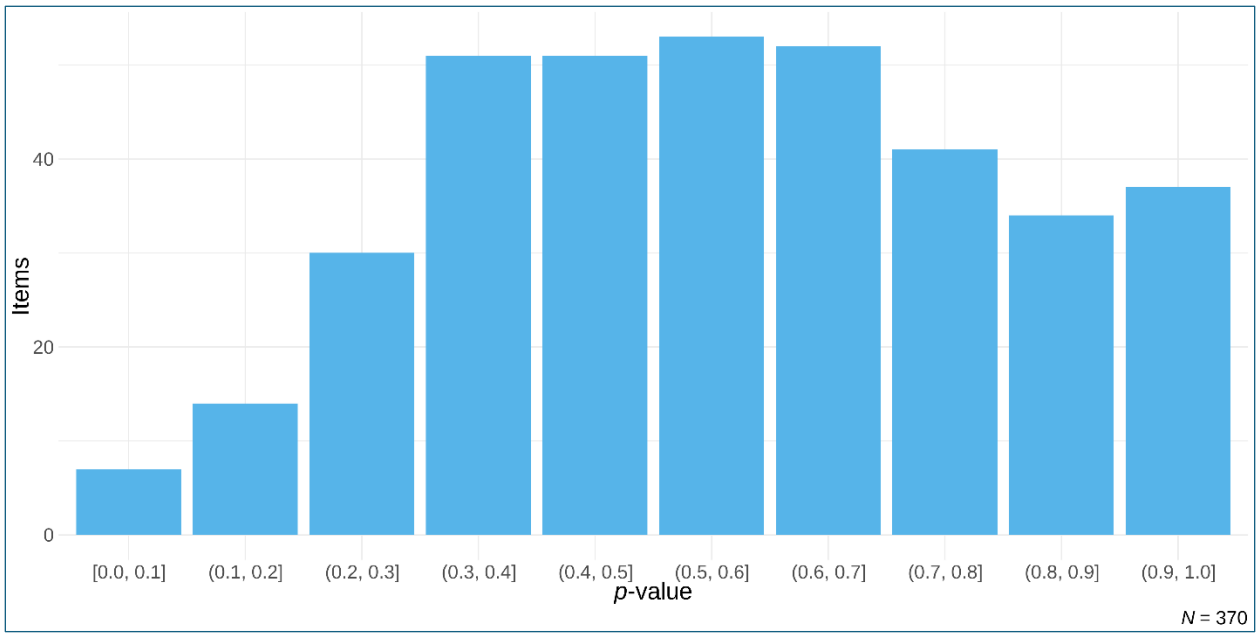
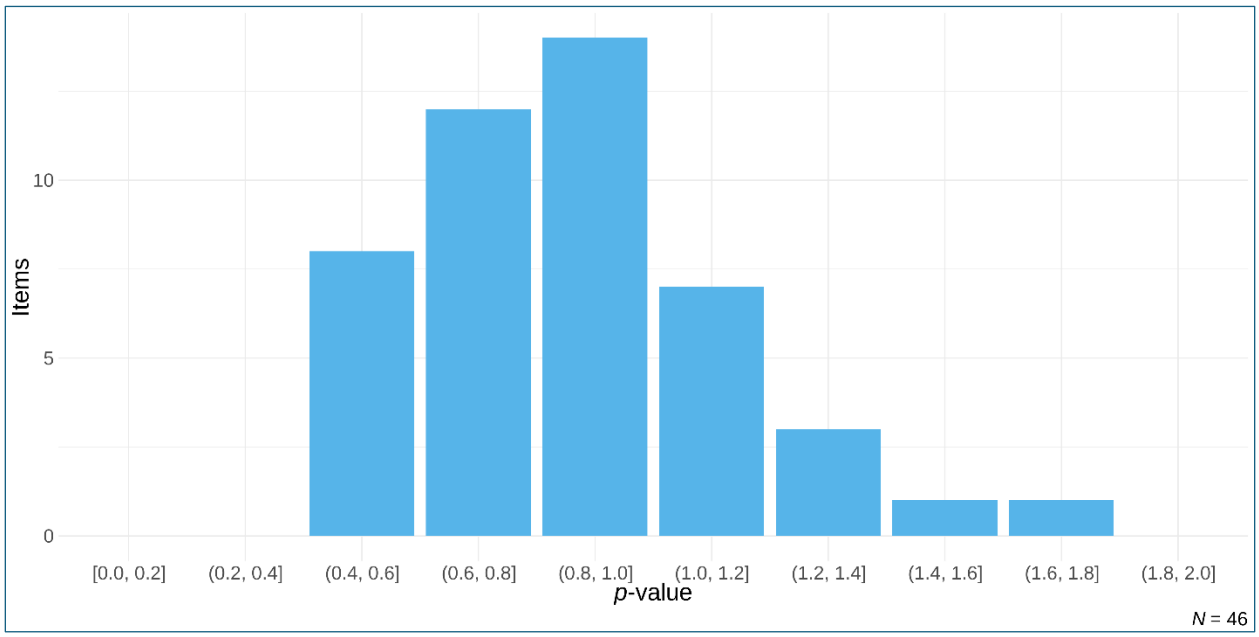


Figure 4.

Item-Difficulty Indices for 2-Point Math Items



The *conditional item-difficulty index* (e.g., Thompson, 2019) denotes the average difficulty of an item for masters and nonmasters within each learning-pathway level. Students were designated as masters or nonmasters using the 80% correct rule. Each item measures a single pathway level of a content standard. Students who received 80% or more of the available points associated with a particular level of a content standard were designated as masters of that level, and students who received less than 80% of the available points were designated as nonmasters of that level. For 1-point items, the conditional item-difficulty index represents the conditional probability (i.e., conditional *p* value) that masters and nonmasters provide a correct response to an item. For 2-point items, the conditional item-difficulty index represents the average score of masters and nonmasters for an item, with values ranging from 0 to 2. PIE project staff flagged 1- and 2-point items as potentially problematic under two conditions. An item was flagged if masters had a conditional *p* value lower than 0.4 or an average item score lower than 0.8. An item was also flagged if nonmasters had a conditional *p* value greater than 0.6 or average item score greater than or 1.2. Four-hundred ten items (98.6%) were above the flagging threshold for masters, and 328 items (78.8%) were below the flagging threshold for nonmasters. Figure 5 and Figure 6 summarize the item-difficulty indices for 1-point items for masters and nonmasters, respectively. Five items (1.2%) were omitted from Figure 5 because of insufficient sample sizes.

Figure 5.

Item-Difficulty Indices for Masters for 1-Point Math Items

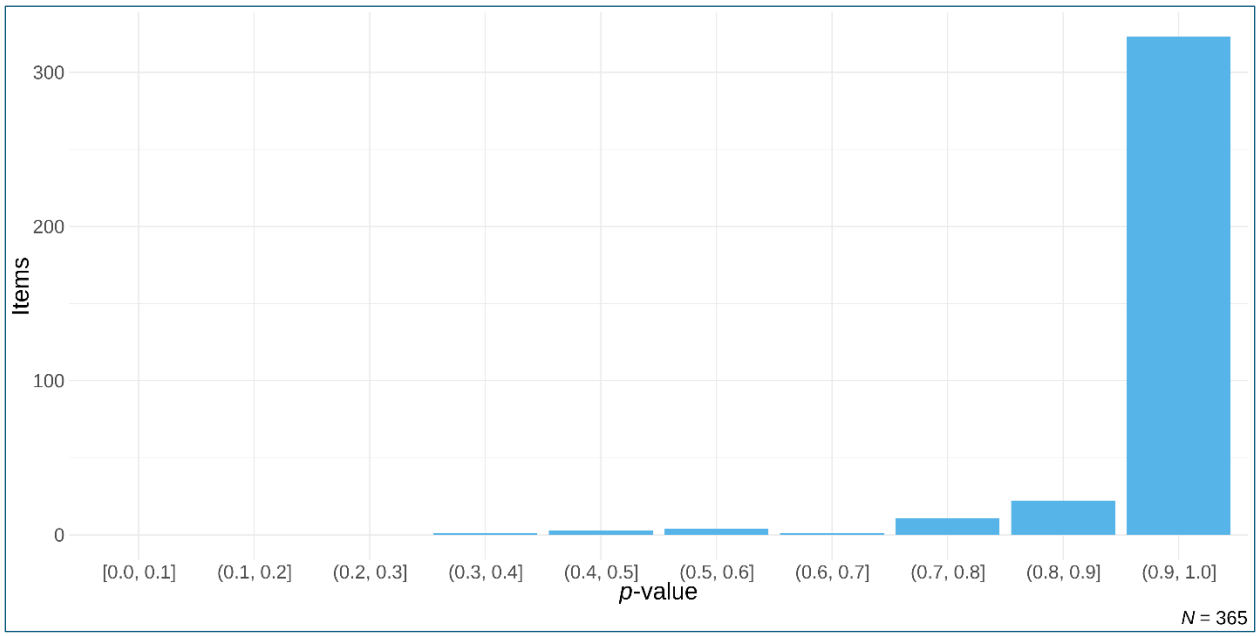


Figure 6.

Item-Difficulty Indices for Nonmasters for 1-Point Math Items

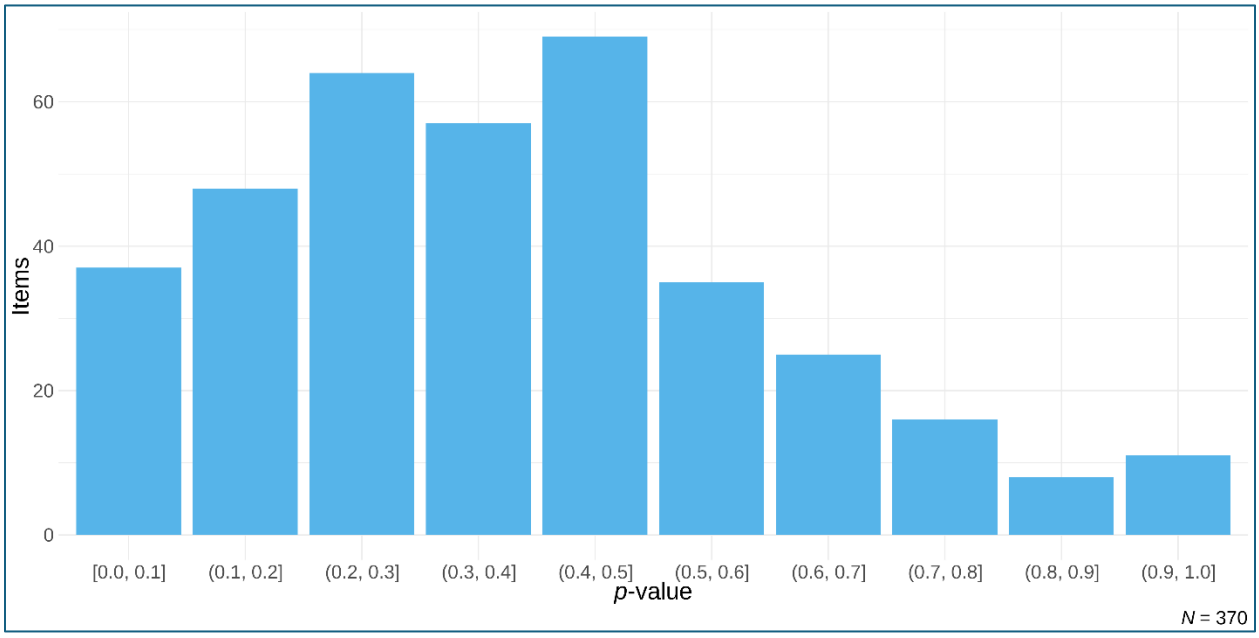


Figure 7 and Figure 8 summarize the item-difficulty indices for 2-point items for masters and nonmasters, respectively.

Figure 7.

Item-Difficulty Indices for Masters for 2-Point Math Items

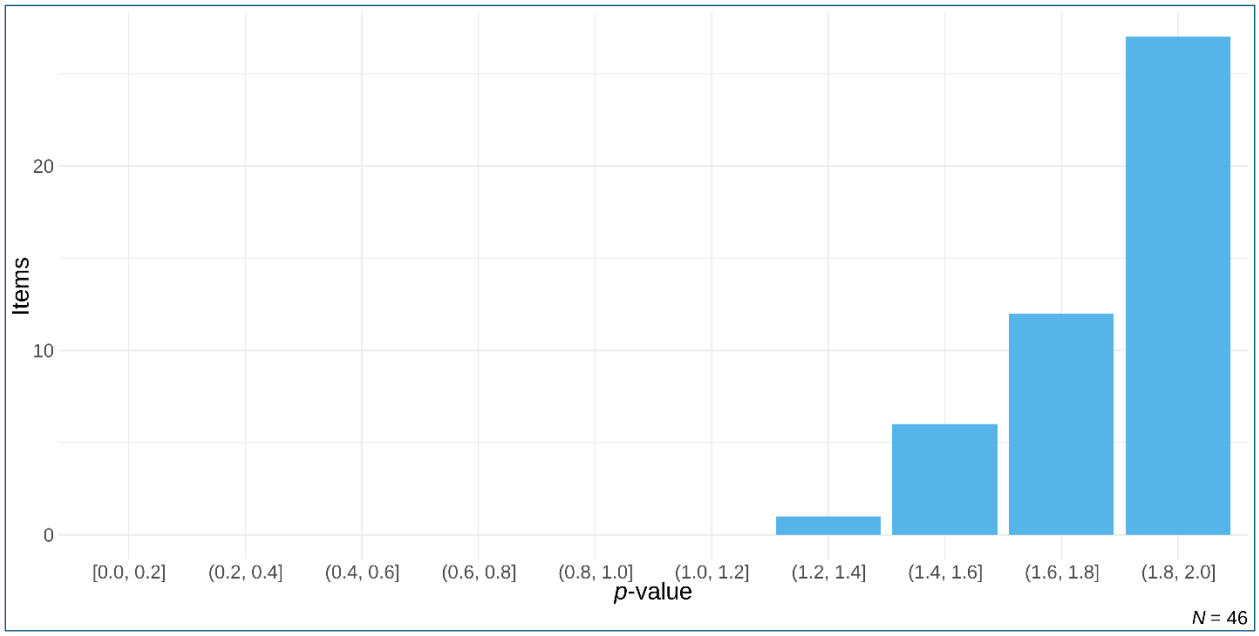
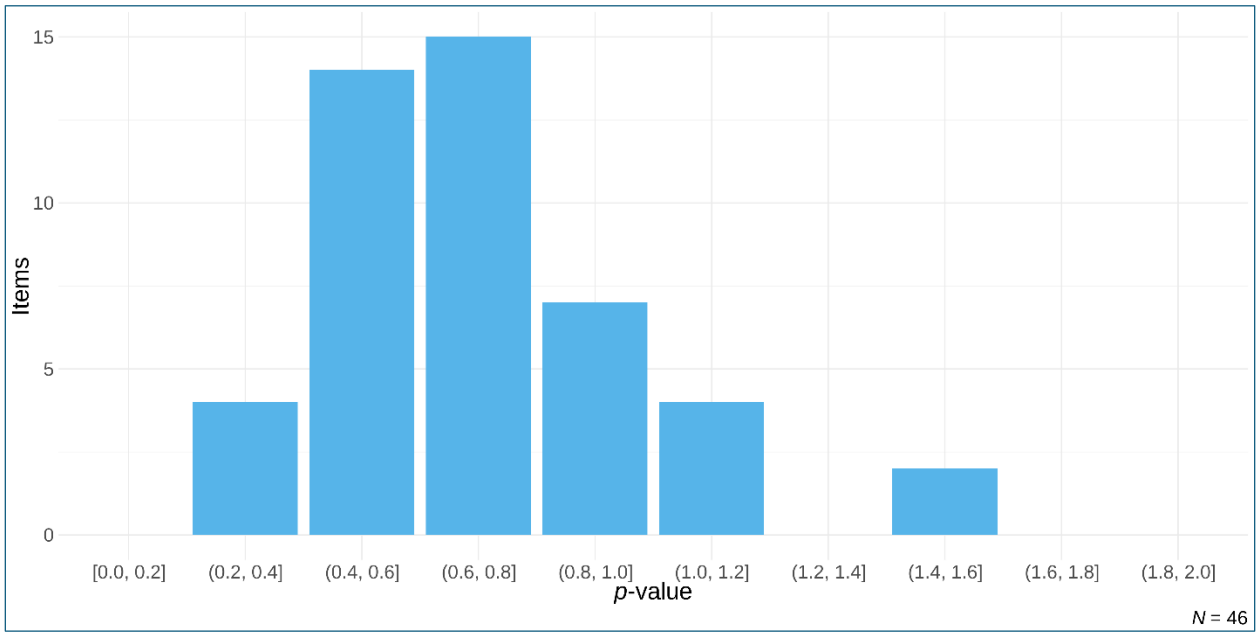


Figure 8.

Item-Difficulty Indices for Nonmasters for 2-Point Math Items



The *item-discrimination index* (e.g., de la Torre, 2008) denotes the difference in item-difficulty indices between masters and nonmasters within each learning-pathway level. For 1-point items, the item-discrimination index was calculated as the difference in the conditional *p* values between masters and nonmasters. For 2-point items, the item-discrimination index was calculated as the difference in average item scores between masters and nonmasters. If masters and nonmasters have high and low item-difficulty indices for an item, then an item is said to have *high discrimination*. However, if there is little difference between item-difficulty indices for masters and nonmasters, then an item is said to have *low discrimination*. PIE project staff flagged items as potentially problematic if that item had a negative discrimination index (e.g., nonmasters demonstrated higher item-response probabilities or average item scores compared to masters). Four-hundred ten items (98.6%) were above the flagging threshold. Figure 9 and Figure 10 summarize the item-discrimination indices for 1- and 2-point items. One item (0.2%) was below the flagging threshold and therefore omitted from Figure 9. Five items (1.2%) were omitted from Figure 9 because of missing statistics.

Figure 9.

Item-Discrimination Indices for 1-Point Math Items

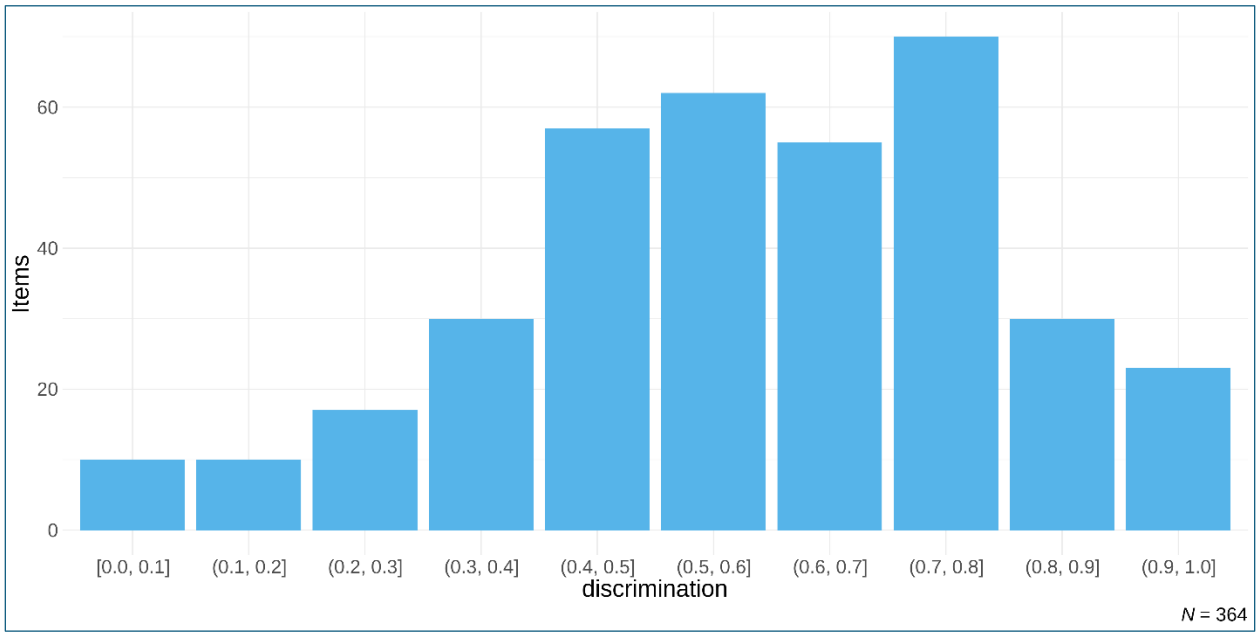
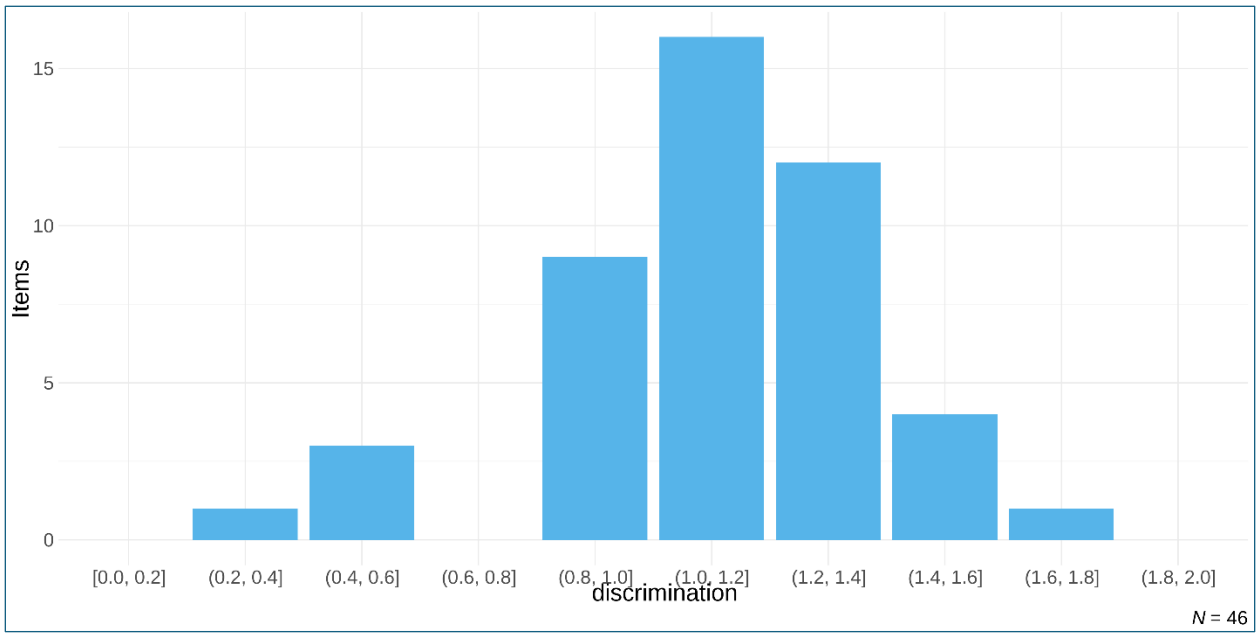


Figure 10.

Item-Discrimination Indices for 2-Point Math Items



Item Data-Review Process and Decisions

PIE project staff reviewed each item and the associated item statistics to provide recommendations for their use. Five blocks of items were included in the field test and although the field test data statistics include 416 items, for the purpose of decisions there are 411 items. MO-DESE project staff reviewed the recommendations and collaborated with the ATLAS team to make final decisions on item usage, including items for use in the pilot assessments, items available for release for other uses, and items rejected. After fulfilling the blueprint and retest needs, 55 additional items remained in the pool. They are noted as reserved. Item decisions are listed in Table 18.

Table 18.

Item Decisions

Decision	Items (*N* = 411)	
	n	%
For use during pilot assessment	334	81.7
Reserved	55	13.4
Released	9	2.2
Rejected	13	3.2

Next Steps for the System Pilot Study

Item-Twin Development

Subsequent item-twin construction development was informed by the field test data and initial item-twins. After the field-test data review, PIE project staff implemented the same development process as described above to develop 141 additional item twins in preparation for the full-system pilot study.

Preliminary Scoring-Model Calibration and Evaluation

After the item field test, the data were used to estimate and evaluate two diagnostic classification scoring models for future use in the full-system pilot study: the loglinear cognitive diagnostic model (LCDM; Henson et al., 2009) and the hierarchical diagnostic classification model (HDCM; Templin & Bradshaw, 2014). During the model-calibration phase, LCDM and HDCM models were estimated separately for each content standard, resulting in a total of 50 models (25 LCDM, 25 HDCM). Each model was defined to measure three attributes (i.e., level 1, level 2, level 3). The LCDM classified each student into one of eight mastery classes. The HDCM classified each student into one of four mastery classes, representing a linear hierarchy among the levels. The models were estimated using a Bayesian estimation approach, which permits more-robust methods for evaluating model fit than traditional maximum-likelihood-estimation approaches.

The LCDM and HDCM models were evaluated and compared for each content standard using likelihood-based, relative model-fit indices (e.g., PSIS-LOO, WAIC; Vehtari et al., 2017; Watanabe & Opper, 2010) and distributions of absolute-fit model indices (e.g., posterior

predictive model checks). Posterior predictive model checks were calculated at the model level for raw-score distributions using the posterior distributions of the model parameters (see Thompson, 2019), and a posterior predictive *p* (ppp) value was obtained for each model for each content standard. Adjusted ppp values were calculated using the Holm correction to control for family-wise error rates associated with testing of multiple models across content standards, and content standards were flagged for model misfit if the adjusted ppp value was less than 0.05. The HDCM was identified as the preferred model across 84% of the content standards using adjusted ppp values as the decision criteria and was recommended for use instead of the LCDM for the pilot study.

Conclusions and Future Directions

The assessment design and development objectives for the Goal 1 PIE project were met including design and development of: (1) test blueprints to inform item development for instructionally embedded and end of year assessments, (2) task models to support content developers in creating high-quality, aligned content that engages all students, (3) item prototypes to guide the full item-development process, and (4) a set of field-tested assessments ready for use for the subsequent (Goal 2) PIE pilot study.

The design and development phases of the PIE assessments resulted in several successes and lessons learned that may be applied to future work. First, task models based in ECD principles were used by item writers to develop 443 items. Items that followed the level statement descriptions and item guidelines were overall well-aligned to their associated level statements. Feedback from item writers included appreciation of how the level statements were unpacked and explained and that they would appreciate access to the task models to plan instruction. Although the task models were designed for assessment purposes, future considerations should include deciding which elements of the task models may be refined to inform instructional resources, such as unpacking cognition of the level statement or standard.

The goal for the item writing event was to produce 392 items. ATLAS staff used prior test development experience with in-person writing events to derive this estimate. At the end of the event, 443 items were developed, which surprised the team considering the amount of time used for training on how to use task models. However, the time spent on training appeared to be effective given that 402 of the 443 items were subsequently moved forward for field testing (after additional reviews and revisions). The low attrition rate was likely due to the documentation of the targeted cognition and characteristics of items that provide evidence of the targeted cognition. Future considerations should include exploring how further refinement of task models after the first round of development based on lessons learned could further increase efficiency and lower attrition rates during development.

Furthermore, ATLAS test developers used the newly developed items and their corresponding task models to create item twins. These twins included identical KSUs that were in the original items and met the item constraints described in the task models. Around 20 twins were included in the field test. Overall, the field-tested twins met the expectation of statistical fungibility and therefore could be used for retest purposes. Future considerations should

include how to use this process of creating item twins to efficiently develop well-aligned, high-quality item banks.

In the next phase of development, the external review of items was conducted in a virtual environment. The orientation was held synchronously to allow reviewers the opportunity to engage in guided practice with their peers to ensure reviewers were comfortable with the expectations of the review. Batches of items were provided for the reviewers over several weeks to provide flexibility in the development timeline and to accommodate reviewers' schedules. However, the virtual environment of the review did not allow for discussions amongst reviewers to discuss their feedback collectively. Therefore, PIE project staff and MO-DESE project staff were accountable for deciphering the comments to apply and enhance the quality of the items. In future work, we may consider an on-site external review to allow for synchronous discussions of the collective feedback to allow both the external reviewers, PIE project staff, and MO-DESE project staff to improve item quality.

Technology-enhanced items were included in both the cognitive labs and the field test. During the cognitive labs, students understood how to interact with the item types and appropriately responded to the items. Based on this information, PIE project staff did not limit which item types could be used during the field test. Although students were successful using various item types, it was recognized that other factors in the items should be considered to minimize any barriers for students. For example, because students struggled to understand where to place drop options when there were multiple open-ended drop zones, PIE project staff limited the number of drop zones and drop options provided in drag-and-drop items for the field test. Additionally, because students took more testing time to determine how to engage with less familiar item types, items developed for the field test included more specific and direct instructions to provide guidance about interacting and responding to the item types. In future work, we may consider including all potential item types in the cognitive labs and vary the number of interactions needed in those items. This may allow us to compare how students interact with various item types while also considering the number of required interactions within each item.

Furthermore, the item field test was successfully administered and met its intended goals of collecting student response data to evaluate the technical quality of the new assessment items and conduct preliminary model calibrations in preparation for the subsequent system pilot study. While the item field test was administered during a relatively busy time of the school year, recruitment efforts and the field test administration design resulted in surpassing our recruitment goals. Specifically, the administration design relied on providing teachers with generic student logins which meant that teachers did not need to access the Kite system directly or upload student roster data. The administration design was intended to be as easy and burden-free on teachers as possible. While this approach positively impacted recruitment and participation efforts, the limitation was a lack of student demographic data that could be used for additional item analyses, such as differential item functioning. In future work, we may consider ways to collect student demographic data while maintaining the streamlined administration approach for teachers (e.g., including demographic questions on the test form).

Finally, from a psychometric perspective, the field test design was successful in supporting the estimation of DCMs for scoring the pilot. By requiring students to test on all pathway levels within a standard during the field test, we were able to collect data that allowed us to estimate the relationships between each of the pathway levels. This data collection design also allowed us to conduct preliminary analyses evaluating the hierarchy of pathway levels within each standard. These analyses supported our scoring plan for the subsequent (Goal 2) pilot study, where the hierarchical assumptions of student mastery were enforced. Had students instead been assessed on more standards at fewer levels (e.g., only level 3 on 5-6 standards), these analyses would not have been possible. Future projects may consider field test designs that more closely mimic the intended operational administration (i.e., levels are assessed at multiple time points), as this would allow for the estimation of a longitudinal model that may provide better insights into students' acquisition of KSUs.

References

CAST. (2018). Universal Design for Learning guidelines version 2.2. http://udlguidelines.cast.org

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, *45*, *343–362.* https://doi.org/10.1111/j.1745-3984.2008.00069.x

Dynamic Learning Maps Consortium. (2016). *2014–2015 Technical manual–Integrated model*. University of Kansas, Center for Educational Testing and Evaluation. https://dynamiclearningmaps.org/sites/default/files/documents/publication/Technical_Manual_IM_2014-15.pdf

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log linear models with latent variables. *Psychometrika*, *74*, 191–210. https://doi.org/10.1007/s11336-008-9089-5

Karvonen, M., Ruhter, L., Shipman, M., Swinburne Romine, R., & Tiemann, G. (2020). *Designing, developing, and evaluating innovative science assessments.* University of Kansas, Accessible Teaching, Learning, and Assessment Systems (ATLAS). https://ismart.works/sites/default/files/documents/Publications/I-SMART_Goal2_TestDev_TechnicalReport_FINAL.pdf

Kim, E. M., Nash, B., & Swinburne Romine, R. (2024). *Pathways for instructionally embedded assessment (PIE): Developing learning pathways for the PIE assessment system.* University of Kansas, Accessible Teaching, Learning, and Assessment Systems (ATLAS). https://pie.atlas4learning.org/sites/default/files/documents/resources/PIE_LP_Tech_Report.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

Swinburne Romine, R., Andersen, L., Schuster, J., & Karvonen, M. (2018). *Developing and evaluating learning map models in science: Evidence from the I-SMART project*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems (ATLAS). https://ismart.works/sites/default/files/documents/Publications/ISMART_Goal_1_Technical_Report_FINAL_0.pdf

Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79(2), 317–339. https://doi.org/10.1007/s11336-013-9362-0

Thompson, W. J. (2019). *Bayesian psychometrics for diagnostic assessments: A proof of concept* (Research Report No. 19-01). University of Kansas, Accessible Teaching, Learning, and Assessment Systems. https://osf.io/preprints/edarxiv/jzqs8

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Wine, M., & Hoffman, A. (2024, October 8–10). *RTD task models: Addressing item development challenges through organizational learning*. Paper presented at the annual meeting of the Northeast Educational Research Association, Trumbull, CT, United States. https://www.researchgate.net/publication/384564472_Wine_W_Hoffman_A_2024_RTD_Task_Models_Addressing_Item_Development_Challenges_Through_Organizational_Learning

Watanabe, S., & Opper, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*(12), 3571–3594. https://www.jmlr.org/papers/volume11/watanabe10a/watanabe10a.pdf

Xu, G. (2019). Identifiability and cognitive diagnosis models. In M. von Davier & Y.-S. Lee (Eds.), Handbook of diagnostic classification models (pp. 333–357). Springer. https://doi.org/10.1007/978-3-030-05584-4_16

Appendix A: Instructionally Embedded Assessment Pathway Levels

Domain	Cluster	Standard	Pathway level	Description of pathway level
NF	A	5.NF.A.1	1	Recognize fraction and decimal notations.
			2	Represent whole and fractional parts in a set model.
			3	Understand that parts of a whole can be expressed as fractions and/or decimals.
		5.NF.A.2	1	Explain numerator, denominator, and decimal point.
			2	Explain place value for decimals, and recognize equivalent fractions.
			3	Convert decimals to fractions and fractions to decimals.
		5.NF.A.3	1	Compare unit fractions.
			2	Compare fractions with like denominators and like decimals.
			3	Compare and order fractions and/or decimals up to the thousandths place using symbols (<, >, =) and justify the solution.
	B	5.NF.B.4	1	Compare fractional parts.
			2	Compare fractions and decimals to benchmarks.
			3	Estimate results of sums, differences, and products with fractions and decimals up to thousandths.
		5.NF.B.5a	1	Represent fractional parts on a length model.
			2	Represent equivalent fractions with length models.
			3	Estimate the size of the product of two fractions based on the size of the factors.
		5.NF.B.5b	1	Explain unit fraction (1/n), and count by fractional parts.
			2	Express fraction as a product of a whole number and a unit fraction.
			3	Explain why multiplying a given number by a fraction greater than one results in a product larger than the given number.
		5.NF.B.5c	1	Express fraction as division.
			2	Express division as fraction multiplication.
			3	Explain why multiplying a given number by a fraction less than one results in a product smaller than the given number.
		5.NF.B.5d	1	Represent equivalent fractions by multiplying or dividing the fraction by a fraction form of one.
			2	Express fraction in equivalent forms.

Domain	Cluster	Standard	Pathway level	Description of pathway level
			3	Explain why multiplying the numerator and denominator by the same number is equivalent to multiplying the fraction by one.
		5.NF.B.6	1	Add and subtract fractional parts.
			2	Solve problems involving fraction addition and subtraction with like denominators.
			3	Solve problems involving addition and subtraction of fractions, including mixed numbers, with unlike denominators, and justify the solution.
		5.NF.B.7a	1	Represent fractional parts on an area model.
			2	Determine the area of a rectangle with fractional side lengths.
			3	Recognize the relationship between multiplying fractions and finding the areas of rectangles with fractional side lengths.
		5.NF.B.7b	1	Represent fractional parts on an area model.
			2	Represent whole number and fraction multiplication using area models.
			3	Calculate and interpret the product of a fraction by a whole number and a whole number by a fraction.
		5.NF.B.7c	1	Represent fractional parts on an area model.
			2	Represent whole number and fraction multiplication using area models.
			3	Calculate and interpret the product of two fractions less than one.
		5.NF.B.8a	1	Represent fractional parts on a length model.
			2	Represent fraction division using length models.
			3	Calculate and interpret the quotient of a unit fraction by a non-zero whole number.
		5.NF.B.8b	1	Decompose fractions into sums of unit fractions.
			2	Explain division of whole numbers by unit fractions.
			3	Calculate and interpret the quotient of a whole number by a unit fraction.
RA	A	5.RA.A.1a	1	Recognize the rule in a numeric pattern.
			2	Extend a numeric pattern by applying the rule.
			3	Generate two numeric patterns given the rules.
		5.RA.A.1b	1	Recognize the order of elements in a repeating pattern.
			2	Organize two numeric patterns in a table.
			3	Translate two numeric patterns into ordered pairs.
		5.RA.A.1c	1	Recognize the origin of a Cartesian coordinate plane.
			2	Explain coordinate pairs.

Domain	Cluster	Standard	Pathway level	Description of pathway level
			3	Graph two numeric patterns on a Cartesian coordinate plane.
		5.RA.A.1d	1	Organize a numeric pattern in a table.
			2	Translate a table of values into ordered pairs.
			3	Identify the relationship between the terms of two numeric patterns.
		5.RA.A.2	1	Recognize growing and shrinking patterns.
			2	Generate a numeric pattern given a rule.
			3	Write a rule to describe or explain a given numeric pattern.
	C	5.RA.C.5	1	Explain the relationship between addition and subtraction and the relationship between multiplication and division.
			2	Solve one-step problems involving addition, subtraction, multiplication, and division.
			3	Solve and justify multi-step problems involving variables, whole numbers, fractions, and decimals.
GM	A	5.GM.A.2	1	Recognize the properties of two- or three-dimensional figures.
			2	Classify two- or three-dimensional figures based on their properties.
			3	Classify two- or three-dimensional figures in a hierarchy based on their properties.
	B	5.GM.B.4a	1	Recognize measurable attribute and unit.
			2	Recognize unit of measurement, and explain volume.
			3	Describe a unit cube as a cube with edge lengths of 1 unit, volume of 1 cubic unit, and can be used to measure volume.
		5.GM.B.4b	1	Compare and order volumes by direct comparison.
			2	Determine volume by counting unit cubes.
			3	Understand that the volume of a right rectangular prism can be found by stacking multiple layers of the base.
	C	5.GM.C.6a	1	Recognize the structure of a number line.
			2	Explain the first quadrant and origin of a Cartesian coordinate plane.
			3	Represent the first quadrant of a Cartesian coordinate plane with scaled perpendicular number lines that intersect at (0, 0), the origin.
DS	A	5.DS.A.2	1	Organize real-world data and answer questions about the data.

Domain	Cluster	Standard	Pathway level	Description of pathway level
			2	Recognize the structure of a line plot (dot plot) and use the graph to read the data.
			3	Create a line plot (dot plot) to represent a data set, and analyze the data to answer questions and solve problems, including recognizing the outliers and generating the median.

Appendix B: Task Model Example

L1 (G5) PIE Task Model (PTM)

PIE Task Model

Overview	Targeted Cognition	Item Considerations	Models	Nodes

Level Statement: Organize real-world data and answer questions about the data.
Missouri Standard: 5.DS.A.2
Domain: Data and Statistics
Cluster: Represent and analyze data.
Standard: 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.

Level 1 (L1) Statement: Organize real-world data and answer questions about the data.

Level 2 (L2) Statement: Recognize the structure of a line plot (dot plot) and use the graph to read the data.

Level 3 (L3) Statement: Create a line plot (dot plot) to represent a data set, and analyze the data to answer questions and solve problems, including recognizing the outliers and generating the median.

Level-Specific Evidence Statements

1. The work contains the data values in numerical order.
2. The work contains the data values organized in a table.
3. The work contains whether the data set is spread out or grouped together given a line plot.

Related Standards

5.DS.A.1			

Notable Other Levels at This or Other Grades

This page of the PIE Task Model (PTM) provides an overview of the specified level in the context of other levels of the standard.

Targeted Cognition

Overview	Targeted Cognition	Item Considerations	Models	Nodes

Level Statement: Organize real-world data and answer questions about the data.

Targeted Cognition Explanation for Assessment Purposes

Organize:

- The student evaluates the given data.
- The student begins with the smallest value.
- The student repeats each value as many times as it appears.
- The student orders the data from least to greatest.
- The student evaluates the data.
- The student identifies the descriptions in the table.
- The student counts the number of instances each data point appears.
- The student enters the value into the table next to its corresponding description.

Answer Questions:

- The student evaluates the line plot and observes the placement of the data points.
- The student determines if the points are grouped together or spread out.

Outside the Targeted Cognition

Disambiguation from Similar Standards of the Targeted Cognition

- 5.DS.A.1 uses line graphs instead of line plots.
- Level 2 for 5.DS.A.1 asks the student to read the graph and identify the parts of the graph.
- Level 3 for 5.DS.A.1 asks the student to answer questions about the data that are limited to median, outlier, most or least, or more or less.

Grade-Level-Specific Considerations for the Targeted Cognition

Assessment Challenges

This page explains the meaning of the targeted cognition, i.e., how ATLAS assessment views the core of the knowledge, skills, and/or understandings (KSUs) of which it aims to elicit evidence. Likewise, it explains consideration for the targeted cognition arising from level-specific issues. To do this, it highlights common points of confusion about distinctions between different levels within and across standards and even grade levels.

Item Level Considerations

Overview | Targeted Cognition | Item Considerations | Models | Nodes

Level Statement: Organize real-world data and answer questions about the data.

Recommended Item Type(s)

- Drag and Drop (Background Graphic or Labeling)
- Drag and Drop (Multiple Drop Bucket)
- Drag and Drop (Ordering)
- Matrix
- Multiple Choice (Single Select)
- Short Answer (Constructed Response)

Student Misconceptions/Bases for Distractors

- When determining frequency, the student may count the number of unique values or use the sum of the values instead of counting the total number of values.
- When ordering the data, the student may not repeat values.
- The student may confuse the idea of "spread out" and "grouped together".

Terminology

Required	Recommended	Verboten
	data grouped together line plot order spread out	

UDL Options

- Ensure relevance to gain student attention and authenticity to help the student connect to the content.
- Simplify vocabulary that is not content specific and necessary to assess the standard.
- Use the simplest possible syntax and structure.
- If pictorial information is provided, key information (such as dimensions, etc.) should be provided in both the stem and the graphic.

Accessibility/Test Administration Customizations

Matrix:

- The columns in a matrix should answer the mathematical question being asked (such as greater, less, a number, a type of angle or figure, etc.). Answers such as "true" and "false" can confuse students as to what is being asked.
- Keep in mind the size of the total matrix. The smaller the components, the more difficult it will be for students to navigate.

Multiple Choice (Single Select):

- When response options contain long phrases or complete sentences, these should be written using parallel structure to ease the reading comprehension burden.

- When multiple-choice responses are long, consider whether the common part of the sentence could be provided in the stem rather than repeated in each response option.

Drag and Drop (Multiple Drop Buckets):

- Consider what students are being asked to process in determining what the elements are and what the targets are. The labels on the buckets need to be clear and understandable without examples.

Graphics Guidance

Other Math Requirements	**Calculator: Yes**	
	No	X
	N/A	

Depth of Knowledge (DOK) Ceiling
1

Context/Scenario Ideas
Context is required.

Recently (Over) Used Ideas

Unsuccessful Approaches (i.e., Do Not Do This)

This page of the PTM focuses on guidance for writing individual items/questions. This is the most granular level of guidance provided for those attempting slightly more creative approaches to item development. Those attempting such an approach should still refer to the models below.

Models/Item Skeletons

Overview	Targeted Cognition	Item Considerations	Models	Nodes

Level Statement: Organize real-world data and answer questions about the data.

Items	Explanation/Annotations
Provide at least three line plots. Prompt the student to determine whether each data set is spread out or grouped together.	Drag and Drop (Multiple Drop Bucket) • The buckets will be the categories: "Spread Out", and "Grouped Together". Matrix • Provide one line plot in each row of the matrix. • Provide two labeled columns ("Grouped Together" or "Spread Out") • The student must select the correct explanation for each expression using radio buttons. Drag and Drop (Background Graphic or Labeling) • Put an empty box next to each line plot. • The student must drag and drop either "Spread Out" or "Grouped Together" next to each line plot.
Items Provide a data set with ten values in context. Prompt the student to put the values in numerical order or into the table.	**Explanation/Annotations** Drag and Drop (Ordering) • Lists **must** include repeating values. Multiple Choice (Single Select) Short Answer (Constructed Response) • Provide a labeled table. • The student enters the data into the right side of the table.

These models serve as examples that item writers can base new items upon. These models should *not* be copied rigidly, but rather should serve as bases from which item writers may creatively come up with new items—so long as they elicit evidence of the targeted cognition, as described in the second page of this PTM. They may be modified or added to over time.

Node(s)				
Overview	Targeted Cognition	Item Considerations	Models	Nodes
Level Statement: Organize real-world data and answer questions about the data.				

Node Number/Name	Node Description
PIE-M-2	Organize real-world data into groups
PIE-M-3	Describe features of a data distribution