Journal of Mathematical Behavior 60 (2020) 100805

Contents lists available at ScienceDirect

Journal of Mathematical Behavior

journal homepage: www.elsevier.com/locate/jmathb

Empirical recovery of learning progressions through the lens of educators

Leanne R. Ketterlin-Geller*, Yetunde Zannou, Anthony Sparks, Lindsey Perry

Southern Methodist University, United States

ARTICLE INFO

Keywords:
Mathematics
Learning progressions
Validity
Teacher perceptions
Survey methodology

ABSTRACT

Learning progressions represent the relationship between concepts within a domain and how students develop increasingly sophisticated thinking therein. Typical evidence sources used to validate theorized learning progressions are also used to validate the use and interpretation of assessments, such as student cognitive interviews and psychometric analyses of item responses on assessments (Alonzo, 2018; Duschl et al., 2011). However, evidence from student responses to assessment items may reflect an interaction with the assessment itself more so than students' domain-specific knowledge and understanding (Lai et al., 2017; Penuel et al., 2014). In this manuscript, we propose that educators' perspectives may serve as an independent source of evidence that can be integrated with traditional evidence sources (e.g., cognitive interviews with students, psychometric data) to overcome this shortcoming. This manuscript describes two studies that used surveys to draw on educator knowledge of students to identify upper and lower bounds of a learning progression (MMaRS study) and to understand the order of intermediary phases of learning (ESTAR study). For both studies, participants included mathematics educators who were classroom teachers or curriculum and assessment developers in relevant grades. Survey results yielded meaningful information to support or modify the hypothesized learning progressions for the respective studies, supporting the proposition that educators' perspectives can meaningfully complement commonly used evidence to validate the structure of learning progressions. Advantages and limitations to this approach are described.

1. Introduction

Learning progressions are intended to represent the nature of knowing in a domain. As representations of cognitive theories of learning, learning progressions describe the development of sophistication in students' thinking about discipline-specific topics (Confry, 2018; Corcoran, Mogat, & Rosher, 2009). Learning progressions extend beyond the value of content standards by specifying the intermediary phases of students' understanding that lead from foundational knowledge to more advanced thinking. The specificity of this developmental progression can meaningfully inform the teaching and learning process in a number of ways, such as by adding fine-grained learning targets to scope and sequence documents, helping teachers sequence instructional activities that are developmentally appropriate, and providing a blueprint for designing instructionally-relevant and sensitive formative assessment resources.

The value of learning progressions in the teaching and learning process can only be realized if the learning progressions represent well-established theories of learning. Theoretical assertions about the nature of knowing need to be supported by a robust body of

* Corresponding author at: Southern Methodist University, PO Box 750114, Dallas, TX 75275-0114, United States.
E-mail address: lkgeller@smu.edu (L.R. Ketterlin-Geller).

https://doi.org/10.1016/j.jmathb.2020.100805
Received 15 January 2020; Received in revised form 16 June 2020; Accepted 8 August 2020

empirical research that coalesces to form generalizable conclusions about the development of sophistication in students' thinking in the discipline. To fully explore the nature of knowing, evidence is needed from a variety of sources. In the current empirical research on validating learning progressions, much of the reported evidence comes from qualitative studies of students' articulated learning experiences (e.g., cognitive interviews, think aloud interviews) and psychometric analyses (Alonzo, 2018; Duschl, Maeng, & Sezen, 2011). Although these studies are necessary to support some hypotheses about the nature of knowing, additional evidence that is more closely associated with the teaching and learning process is warranted. Given their proximity to the teaching and learning process and their accumulated professional expertise from working with multiple children, teachers may be uniquely positioned to provide meaningful perspectives on students' development of sophistication in understanding.

The purpose of this manuscript is to describe sources of evidence that elicit teachers' perspectives on student learning. Two studies are reported in which teachers were asked to use their professional knowledge about how students' thinking develops to evaluate the plausibility of two elementary mathematics learning progressions. In both of these studies, teachers were presented with a theoretically-grounded learning progression and asked to consider the alignment between the learning progression and their professionally-informed experiences about how students develop understanding.

1.1. The structure of learning progressions

Most learning progressions focus on discipline-specific understandings and/or practices, and are structured to include a lower boundary from which intermediary steps commence, and an upper boundary marking the termination of the sequence (Confry, 2018; Corcoran et al., 2009). The lower and upper boundaries that flank the intermediary phases can be characterized as the benchmarks for entering and exiting the learning progression, respectively, and may inform decisions about instructional targets and/or summative student outcomes. The intermediary phases contribute to a coherent network of fine-grained steps that illuminate the sophistication and complexity in thinking that are required to progress from the lower toward the upper bounds. Although this characterization suggest may that learning progressions represent a rigid, hierarchical sequence for engaging in discipline-specific concepts and processes, the exact pathway may include inter- and intra-individual differences that vary based on the disciplinary focus, prior experiences and exposure to content, and/or instructional opportunities. Moreover, specific learning progressions do not exist in isolation; instead, they form a complex web of interrelated understandings and practices that can be meaningfully integrated through instructional experiences.

Articulating a learning progression typically begins by establishing the lower and upper boundaries within the focal topic. The foundational knowledge and skills that serve as the lower boundary are often defined by expectations of prior knowledge or exposure to disciplinary concepts (Corcoran et al., 2009), formal or informal study of child development (Clements, 2007), or intuitive accounts of everyday experiences or situations (Duschl et al., 2011). In science education, lower boundaries may reflect emerging understandings that include misconceptions or incomplete knowledge (Alonzo, 2018). The entry points to the learning progression should clearly establish the expected disciplinary preconceptions so as to make clear the starting place from which the intermediary phases commence (Korbin, Larson, Cromwell, & Garza, 2015). Because of the intertwined network of learning processes, the lower boundary of one learning progression may be the upper boundary of another and/or may inform the intermediary phases of still another.

Upper boundaries represent the end of the expected learning processes specified in a given learning progression. Although learning processes may continue beyond this point, the upper bounds are marked by advanced knowledge and skills that represent curricular expectations (e.g., content standards), canonical disciplinary knowledge, or societal expectations of competence that are influenced by new technologies and/or economic or workforce demands (Confry, 2018; Corcoran et al., 2009; Duschl et al., 2011). In some learning progressions, the upper boundary represents extensive knowledge and skills that reflect vast disciplinary expertise. These learning progressions often span multiple grade levels and may extend into college or career expectations or general scientific literacy needed for community citizenship (c.f., Gunckel, Covitt, Salinas, & Anderson, 2012). Other instantiations of learning progressions establish upper boundaries that are more proximal to the lower boundaries, thereby depicting shorter learning pathways (Confry, 2018).

For some hard-to-define constructs, the lower and upper boundaries are conjectural and may be based on theoretical assertions about predicted relationships between subdomains. For example, early mathematics is comprised of many topics, and not all topics are emphasized equally within the curriculum. Numeric relational reasoning (NRR), for example, is an essential component of early numeracy but is often not explicitly included in curricular standards, textbooks, or instruction. NRR is typically defined as the ability to mentally analyze relationships between numbers or expressions, often using knowledge of properties of operations, decomposition, and known facts (Baroody, Purpura, Eiland, Reid, & Paliwal, 2016; Carpenter, Levi, Franke, & Zeringue, 2005; Farrington-Flint, Canobi, Wood, & Faulkner, 2007). When students solve problems using NRR, they are integrating multiple early numeracy concepts through a "strategic" decision making process based on the relationships between numbers (Whitacre, Schoen, Champagne, & Goddard, 2017). Although progressions for some of the contributing numeracy concepts (c.f., number core in National Research Council, 2009) have been articulated, research is needed to more fully flesh out the foundational knowledge and skills that form the basis of students' numeric relational reasoning, and the expected uppermost knowledge and skills that would indicate advanced numeric relational reasoning. Once these anchors are established, the intermediary phases can be articulated. As this example illustrates, understanding the nature of the lower and upper boundaries is an essential step in articulating learning progressions for emerging constructs.

Intermediary phases of learning connect the lower and upper boundaries of a learning progression. Conceptualized to illustrate the incremental progression of learning, the intermediary phases depict the increasing sophistication of students' thinking and represent the nature of knowing in the discipline. The intermediary phases are often articulated through a systematic process that includes an examination of disciplinary structures (e.g., expert task analysis), evidence about how children learn (e.g., cognitive interviews, teaching experiments), and large-scale studies that model students' thinking (e.g., psychometric analyses, longitudinal studies) (Clements, 2007; Confry, 2018; Duschl et al., 2011; Penuel, Confrey, Maloney, & Rupp, 2014; Sarama & Clements, 2019).

The intermediary phases, sometimes referred to as levels, are generally considered to be associated with specific conceptualizations (e.g., knowledge, skills, processes, strategies), ordered, and interconnected in some manner. For learning progressions in which the lower and upper boundaries are closely associated, the level of specificity (often referred to as grain size) of the intermediary phases will likely be precise (Ketterlin-Geller, Shivraj, Basaraba, & Yovanoff, 2019). These fine-grained intermediary phases denote the progression of understanding in a narrowly-defined range of content. In contrast, learning progressions that span multiple grades may have coarse-grained intermediary phases that focus on broad categories of content and represent discrete benchmarks in understanding. Moreover, a learning progression can be specified to include both coarse- and fine-grained details that are used for different purposes (Wilson, 2009). Regardless of grain size, the exact nature of these intermediary phases may be theoretically grounded and empirically validated across many students, but inter-individual differences almost certainly exist (Confry, 2018; Wright, 2014). As such, the intermediary phases should not be viewed as an intractable or fixed pathway, but instead as a (presumably) well-established theory of learning.

1.2. Sources of evidence to substantiate the structure

The sources and specification of evidence that are needed to establish a hypothetical learning progression as a theory of learning are at the center of discussion in the research literature. Tension centers on whether researchers perceive of learning progressions as existing independently from or integrally connected with assessments (Alonzo, 2018). In the view that learning progressions and assessments are intertwined, sources of evidence used to validate one may also serve as viable evidence for the other. However, in the former view where learning progressions and assessments are distinct, independent sources of evidence are needed to validate the distinct claims. Although these responses may contribute evidence to validate the learning progression as a theory of learning, these same responses should not be accumulated to simultaneously contribute to content- or construct-validity evidence for the tasks or items (Duschl et al., 2011). In this manuscript, we take the view that learning progressions serve as an overarching framework that exist independently from but may inform assessments. In this view, evidence collected to evaluate the hypothesized learning progression needs to be disentangled from evidence collected to validate the uses and interpretations of assessments.

As noted earlier, the most common forms of evidence to evaluate learning progressions found in the currently reported research include student interviews (e.g., cognitive interviews, think aloud interviews) and psychometric analyses. Inherent to the design of these studies, students respond to tasks or items on which their knowledge, skills, and/or strategies are captured (qualitatively and/or quantitatively) and assumptions about their learning processes are based. The intertwined nature of these sources of evidence makes it difficult to disentangle evidence about students' thinking processes from their responding behaviors, which may jeopardize the quality of the inferences made from the evidence.

Specifically, students' responding behaviors are influenced by task and/or item parameters that are designed (intentionally or not) to illuminate students' thinking in various ways. Some item features may obscure aspects of knowing that are not observable based on the selected assessment mode (Penuel et al., 2014). Similarly, on multiple-choice items, distractor characteristics can have a substantial impact on item difficulty (van Rijn, Graf, & Deane, 2014). Illustrating these concerns, Lai, Kobrin, DiCerbo, and Holland (2017) noted that some aspects of students' thinking that were specified in their learning progression were not observed during the think aloud interviews. Relatedly, the authors observed students' responses that contradicted their hypothesized learning progression in that some students demonstrated less sophisticated processes to solve more complex items. These findings point to possible context and/or task dependencies in students' responses that confound the evidence about students' thinking, particularly as it relates to defining the intermediary phases.

Teaching experiments have been explored as a possible compromise to this situation. Teaching experiments study students' interactions with instructional tasks to explore hypotheses about how students build on prior knowledge to deepen their conceptual understanding (Confry, 2018; Confrey, Maloney, & Corley, 2014), and have been used in various studies on learning trajectories (c.f. Kara, Simon, & Placa, 2018; Sarama & Clements, 2019; Simon, Placa, Avitzur, & Kara, 2018). Through an interview-based approach, the researcher takes the role of the teacher to intentionally examine the interaction between the student(s) and the tasks to determine if the task is eliciting the intended learning processes (Duschl et al., 2011). Evidence is accumulated through each successive researcher-student interaction to make inferences about the student's cognitive structures and how to use these inferences to promote future learning (Simon, 2018). However, because these studies span considerable time, include varied instructional actions, and represent a limited sample, they may pose practical constraints and have limited generalizability as sources of evidence for learning progressions. Other sources of evidence are needed that are feasible and, importantly, disentangle the effects of students' responding behaviors from assumptions about their thinking processes.

The purpose of this manuscript is to consider the role educators' perceptions of development of sophistication in students' thinking may play in accumulating evidence about the structure of learning progressions. Because these data do not rely on direct measurement of students' thinking, they may uniquely contribute to inferences about the structure of learning progressions. We propose that these data can supplement rather than supplant evidence from other sources.

1.3. The role of educator input

Research on the accuracy of teachers' perceptions of students is mixed. Much of the published research has focused on teachers' judgments of students' proficiency and the accuracy of their predictions of student performance (see a meta-analysis conducted by Sudkamp, Kaiser, & Moller (2012)), and paid relatively little attention to what teachers know about student thinking. Because many variables contribute to teachers' predictions of performance (e.g., teachers' knowledge of particular students, content knowledge, test characteristics and procedures), this body of research has limited utility for examining the contribution of teachers' knowledge about student thinking to validate learning progressions.

As stakeholders with robust in-situ knowledge of child development, teachers may be uniquely positioned to provide distinct and specialized information about how students develop sophistication in their knowledge and skills. Hill and Chin (2018) refer to this type of knowledge as teachers' *knowledge of students*, and it includes teachers' understanding of students' conceptions, preconceptions, and misconceptions; students' reactions to tasks; frequently used strategies; and typical pathways in which students develop understanding. Teachers' knowledge of students is different from but likely intertwined with other forms of teacher knowledge, such as pedagogical content knowledge. Pedagogical content knowledge may include aspects of teachers' knowledge of students, but extends into teachers' instructional responses to students' actions and behaviors (Kersting, Givvin, Thompson, Santagata, & Stigler, 2012). Although additional research is needed to confirm this presumption, emerging findings from Hill and Chin (2018) indicate that teachers' knowledge of students is a stable teacher-level trait that is transferable across different groups of students and may uniquely contribute to student outcomes independent of other forms of teachers' knowledge. Moreover, Martinez, Stecher, and Borko (2009) proposed that teachers' working knowledge of student learning may contribute to observed differences in the association between teachers' ratings of proficiency and standardized measures of student achievement. Combined, these sources of evidence suggest that teachers' ongoing and direct experience working with students may generate a stable understanding about how students learn.

In this manuscript, we build on the assumption that teachers' ongoing experiences working with students provide them with unique perspectives on how students develop knowledge that can meaningfully contribute to the establishment of the structure of learning progressions. In the following sections, we report on two examples in which educators' perspectives contribute to the body of evidence used to evaluate the structure of a learning progression. Within each section, we describe the methods used to collect and analyze educators' input and describe how their feedback was used to iteratively refine the learning progression.

2. Instantiated examples: evidence from two studies on learning progressions

The Elementary Students in Texas Algebra Ready (ESTAR; funded by the Texas Education Agency) and the Measuring Early Mathematical Reasoning Skills (MMaRS; funded by the National Science Foundation, #1721100) projects collected data through teacher surveys as a source of evidence to help substantiate each project's learning progressions. The ESTAR Learning Progressions illustrate how children in Grades 2–4 develop increasingly complex understanding of foundational algebra-related concepts, and include concepts of addition and subtraction, multiplication and division, and fractions. The MMaRS Learning Progressions represent how children in Grades K-2 develop an understanding of two critical early mathematics constructs, NRR and spatial reasoning. In both projects, the learning progressions were developed as part of ongoing efforts to develop instructionally-relevant classroom assessments to inform instruction. The learning progressions served as the content blueprint for these assessments and guided item and test specifications.

Both the ESTAR and MMaRS Learning Progressions were developed through an iterative process that included examining the extant literature, engaging with domain-specific experts, and collecting and analyzing empirical data from various sources (e.g., cognitive interviews and think aloud interviews which children, analyses of item responses). For the ESTAR Learning Progression, the lower and upper boundaries were established by grade-level expectations specified in the state content standards. Because the MMaRS Learning Progressions focus on early mathematics concepts that are not heavily emphasized in state content standards, the boundaries for the MMaRS Learning Progressions were established by examining the literature and in consultation with experts.

As part of the validation efforts for the learning progression, we integrated teachers' perceptions of student thinking. For the ESTAR Learning Progressions, we gathered educators' perceptions on the content and ordering of the intermediary phases as well as the plausibility of students' misconceptions. For the MMaRS Learning Progressions, we focused on teachers' perceptions of the importance and developmental appropriateness of the skill statements to evaluate the lower and upper boundaries. Collecting validity evidence for the ESTAR and MMaRS Learning Progressions lends credibility to the content and structure of the hypothesized learning progressions. The following sections describe each survey and how researchers used evidence from educators to inform the learning progressions.

Both projects use the following nomenclature. Each learning progression is identified with a label denoting the primary domain (e.g., addition and subtraction is labeled AS; numeric relational reasoning is labeled NRR). In instances where the domain has more than one targeted learning goal, a sequential label is provided (e.g., addition and subtraction has two learning goals, AS.A and AS.B). Targeted learning goals are composed of multiple core concepts. Each core concept is defined by a series of subcomponents. Following the notation identifying the targeted learning goal, a sequential numerical reference is used to identify the core concept (e.g., the first core concept in AS.A is AS.A.1). The subcomponents are sequentially referenced using lower case letters (e.g., the first subcomponent in AS.A.1 is labeled AS.A.1.a).

2.1. Study 1: ESTAR learning progressions survey

The purpose of this study was to investigate teachers' perceptions about the ESTAR Learning Progressions, focusing on the accuracy of the content and ordering of the intermediary phases as well as the plausibility of students' misconceptions. Five learning progressions were specified as part of the larger project. Each learning progression was divided into fine-grained core concepts that represented a sequential development of knowledge and skills. Within each core concept, greater specificity was provided by articulating specific subcomponents and misconceptions and errors that may commonly occur while learning these skills.

The following research questions guided data collection and analyses:

1 To what extent do elementary educators perceive the core concepts as comprehensive of the learning progressions? And subcomponents as comprehensive of the core concepts?
2 To what extent do elementary educators perceive the subcomponents as relevant to the core concepts?
3 To what extent do elementary educators agree with the hypothesized order in which students develop algebra-readiness knowledge and skills?
4 To what extent do elementary educators perceive hypothesized misconceptions and/or errors associated with each subcomponents as plausible?

2.1.1. Participants

The ESTAR Learning Progressions were developed as part of a larger algebra-readiness initiative in Texas. As a result, participants included Texas mathematics educators who were recent (within three years prior to completing the survey) classroom teachers, instructional support personnel, curriculum developers, or assessment developers.[1] Eligible participants had professional education experience in Texas in Grades 2–5. Researchers included Grade 5 educators because of their familiarity with the expected knowledge and skills of incoming fifth grade students.

Researchers used snowball sampling to create a contact list of potential participants. Initially, researchers sent a recruitment email to 50 educators in the authors' professional network in accordance with the Institutional Review Board. Contacts included Texas mathematics supervisors, instructional specialists, and K-12 classroom teachers. The recruitment email described the purpose and details of the study, participation eligibility requirements, an embedded link to express interest, and a request to forward the recruitment email to others in their network. The embedded link directed interested educators to answer eligibility questions and provide a contact email. An unknown number of educators received a forwarded copy of the recruitment email. As a result, the response rate is based on the number of educators who completed the interest form.

One-hundred and thirty-six educators met the eligibility criteria and were sent a link to complete the survey. Eighty completed the survey (response rate of 58.8 %), 23 began the survey but did not complete it, 29 never began the survey, and four declined to participate.[2] Table 1 summarizes the respondent sample.

Prior to completing the survey, participants were prompted to consent, complete a non-disclosure agreement, and watch a training video describing the learning progressions, organization of the ESTAR Learning Progression, and survey task. Respondents had 12 days to complete the survey. At the end of the survey, respondents were asked to provide demographic information.

2.1.2. Survey instrument

The ESTAR Learning Progressions Survey was administered electronically via Qualtrics. To prevent survey fatigue, educators evaluated one of the five ESTAR Learning Progressions on addition and subtraction (AS.A and AS.B), multiplication and division (MD.A and MD.B), or fractions (FR). Researchers used random assignment and quota features of the platform to ensure respondents were equally likely to review any of the learning progressions and that respondents were spread evenly across learning progressions. Incomplete surveys were discarded from the data set and resulted in varying numbers of responses per learning progression (see Table 2).

The survey focused on the intermediary phases (e.g., the concepts, knowledge, and skills that compose algebra-readiness) within the learning progression. Educators' responses contribute to the evidence about the appropriateness of the intermediary phases of the learning progression. First, respondents rated the relevance of each subcomponent within the core concepts and the comprehensiveness of the learning progression using a 4-point scale (1 = Not at all to 4 = Very). Ratings of 1 and 2 prompted respondents to provide additional feedback about their rating. Respondents could also provide additional comments and recommend additional subcomponents for inclusion. Next, respondents were asked to order the subcomponents according to how students' understanding develops. The subcomponents were presented in a random order to avoid inherent rank bias. Lastly, respondents reviewed hypothesized misconceptions and errors and rated how likely they were to be held or demonstrated by students on a 4-point scale

[1] Classroom teachers teach directly to students in the classroom setting; instructional support personnel directly support classroom teachers' instruction as an instructional coach or mentor; curriculum development personnel develop curriculum for mathematics instruction; and assessment development personnel create assessments for use in mathematics instruction outside of a classroom setting.

[2] In some instances, interested and eligible participants never received the automated email containing the survey link, possibly due to school district server filters. Researchers requested an alternate email address for respondents and extended the deadline. Four email addresses permanently failed; two were re-sent with corrections to misspelled email addresses. A total of 15 participants responded to this follow-up email, of which four declined to participate.

Table 1
Respondents and Years of Experience by Primary Educator Role.

Primary Educator Role	Respondents (N = 80)		Years of Experience
	N	(%)	M (SD)
Classroom Teacher	48	60	11.3 (6.47)
Instructional Specialist	23	28.8	16.3 (9.72)
Curriculum Specialist	9	11.3	34.3 (19.48)

Note. No respondents identified “Assessment Specialist” as their primary educator role and therefore, this role was excluded from the table.

Table 2
Distribution of Respondents by ESTAR Learning Progression and Core Concepts.

Learning Progressions	Respondents		Percentage of Respondents (N = 80)
	N	%	%
AS.A: Foundations of Addition and Subtraction	13	100	16.25
Core Concept 1	3	23.08	3.75
2	5	38.46	6.25
3	5	38.46	6.25
AS.B: Performing and Applying Addition and Subtraction	16	100	20
Core Concept 1	6	37.5	7.5
2	5	31.25	6.25
3	5	31.25	6.25
MD.A: Foundations of Multiplication and Division	18	100	22.5
Core Concept 1	5	27.78	6.25
2	4	22.22	5
3	4	22.22	5
4	5	27.78	6.25
MD.B: Performing and Applying Multiplication and Division	19	100	23.75
Core Concept 1	5	26.32	6.25
2	4	21.05	5
3	5	26.32	6.25
4	5	26.32	6.25
FR: Fractions as Numbers	14	100	17.5
Core Concept 1	5	35.71	6.25
2	5	35.71	6.25
3	4	28.57	5
Total	80	100	100

(1 = Very unlikely to 4 = Very likely).

2.1.3. *Analyses*

Researchers used several analyses to assess educators’ agreement with the hypothesized ordering of the ESTAR Learning Progression subcomponents and the likelihood that students would hold the listed misconceptions or errors. First, researchers calculated response frequencies, average ratings, and standard deviations for comprehensiveness, perceived relevance, preferred ordering of subcomponents, and likelihood of algebra-readiness misconceptions and/or errors. Second, to determine agreement between respondents on the ordering of the subcomponents, researchers calculated frequencies and a coefficient for concordance, Kendall’s W on a scale from 0 to 1 with the alpha value set at 0.05. Subcomponents for which there was great variability between respondents and the hypothesized learning progression underwent review for possible revision, taking into account whether the difference was statistically significant.

2.1.4. *Findings*

For the purpose of this manuscript, we present the findings for one learning progression: Concepts and Applications of Addition and Subtraction with Whole Numbers (AS.A). This progression is divided into three core concepts: A.1: Representing Addition and Subtraction, A.2: Composing Whole Numbers, and A.3: Inverse Relationships. Each of these core concepts is further divided into subcomponents.

Table 3 presents a summary of the respondents’ ratings for the comprehensiveness of the AS.A Learning Progression. Respondents rated the comprehensiveness of the core concepts for specifying the learning progression. Core concept AS.A.1 was rated as Somewhat Comprehensive by one respondent; the remaining respondents rated AS.A.1 as Mostly or Very Comprehensive. Two respondents rated AS.A.2 as Somewhat Comprehensive; the modal response was 4 (Very Comprehensive). Core concept AS.A.3 was rated by all respondents as Very Comprehensive. Respondents also reviewed the comprehensiveness of the subcomponents within each core

Table 3
Summary of Respondents' Ratings for Comprehensiveness of AS.A Learning Progression (n = 13).

AS.A. Core Concepts			Frequencies			
	M (SD)	Mode	1	2	3	4
AS.A.1	3.38 (0.65)	3		1	6	6
AS.A.2	3.54 (0.78)	4		2	2	9
AS.A.3	4.00 (0.00)	4				13

Note. Scale is 1 = Not at all; 2 = Somewhat; 3 = Mostly; 4 = Very.

concept (not included in Table 3). Modal responses indicate that respondents perceived all of the subcomponents as Mostly or Very Comprehensive; only two subcomponents (out of a total of 9) across three core concepts for AS.A received one rating of Somewhat Comprehensive.

To address the second research question, we asked respondents to rate the relevance of the subcomponents for each core concept. The modal response for all subcomponents was Mostly or Very Relevant. Four of nine subcomponents received a rating of Somewhat Relevant by at least one respondent (See Table 4).

To examine the ordering of the subcomponents in each core concept, we calculated the frequency with which respondents selected each ordering option. Table 5 provides a summary of the frequency of the ordering options selected by the respondents. Differences between the ordering options were statistically significant, indicating one option was selected most often. For core concepts A.1 and A.3, the most frequently selected ordering option aligned with the hypothesized order. For core concept A.2, the most frequently selected ordering option did not align with the hypothesized ordering.

To address the fourth research question, respondents rated the plausibility of all misconceptions and errors associated with the subcomponents. Sixty-three misconceptions and errors were identified for the AS.A Learning Progression. Seven (11 %) were rated as "Unlikely" or "Very Unlikely." Respondents suggested two additional misconceptions and/or errors. To illustrate these findings, all statements were ranked in order from strongest support—indicated by the highest mean and least variance—to the least supported—indicated by the lowest mean and greatest variance—for their appropriate use during item writing. Item writers were encouraged to draw on misconceptions and errors with stronger support, as identified by the survey respondents.

2.1.5. Conclusions

Findings from the ESTAR Learning Progressions Survey uniquely contributed to our understanding of the accuracy of the content and the ordering of the intermediary phases of the learning progression. Because the original learning progressions were hypothesized to reflect students' development of sophistication in these constructs, teachers' perspectives about the comprehensiveness and relevance of the content to student learning, the accuracy of the proposed order, and the plausibility of specific misconceptions and errors provide an important contribution to our evidentiary chain. Specifically, these data were combined with findings from pilot testing with children, reviews by experts in the field, and summaries of research conducted on students' thinking in early mathematics to develop a more comprehensive understanding of student learning.

Descriptive statistics were examined to understand the distribution of teachers' responses, and in most cases, the modal value was used to draw conclusions. A modal value of 3 or 4 was used to indicate agreement with the hypothesized learning progression; a modal value of 1 or 2 indicated disconfirming evidence. The research team carefully reviewed the data for both confirming and disconfirming evidence to better understand the context in which the data were obtained (e.g., number of respondents). A majority of the findings indicated agreement with the hypothesized learning progressions. These observations were noted in summary documents, which were used to track findings across all sources.

The purpose of collecting multiple sources of data was to further refine and update the learning progressions to more accurately reflected students' development of sophistication in the domain. Once data were collected across multiple sources, we reconciled the

Table 4
Summary of Respondents' Ratings for Relevance of the Subcomponents in AS.A Learning Progression (n = 13).

AS.A Subcomponents			Frequencies			
	M (SD)	Mode	1	2	3	4
AS.A.1.a	3.62 (0.65)	4		1	3	9
AS.A.1.b	3.38 (0.77)	4		2	4	7
AS.A.1.c	3.77 (0.44)	4			3	10
AS.A.2.a	3.15 (0.69)	3		2	7	4
AS.A.2.b	3.69 (0.63)	4		1	2	10
AS.A.2.c	3.77 (0.44)	4			3	10
AS.A.3.a	3.85 (0.38)	4			2	11
AS.A.3.b	4.00 (0.00)	4				13
AS.A.3.c	3.85 (0.38)	4			2	11

Note. Scale is 1 = Not at all; 2 = Somewhat; 3 = Mostly; 4 = Very.

Table 5
Summary of Respondents' Ordering of LP AS.A Subcomponents (n = 13).

Order	N	%
AS.A.1: Representing Addition and Subtraction		
AS.A.1.a, AS.A.1.b, AS.A.1.c	11	85
AS.A.1.a, AS.A.1.c, AS.A.1.b	1	8
AS.A.1.b, AS.A.1.a, AS.A.1.c	1	8
Kendall's W = 0.852*	$X^2 = 22.15$	$p < 0.001$
AS.A.2: Composing Whole Numbers		
AS.A.2.a, AS.A.2.b, AS.A.2.c	5	39
AS.A.2.a, AS.A.2.c, AS.A.2.b	7	54
AS.A.2.c, AS.A.2.b, AS.A.2.a	1	8
Kendall's W = 0.55*	$X^2 = 14.31$	$p = 0.001$
AS.A.3: Inverse Relationships		
AS.A.3.a, AS.A.3.b, AS.A.3.c	8	62
AS.A.3.a, AS.A.3.c, AS.A.3.b	2	15
AS.A.3.b, AS.A.3.a, AS.A.3.c	2	15
AS.A.3.c, AS.A.3.b, AS.A.3.a	1	8
Kendall's W = 0.479*	$X^2 = 12.46$	$p = 0.002$

Note. Ordering arrangements that had a frequency of 0 are not shown.

findings by considering whether a preponderance of trustworthy data provided confirming or disconfirming evidence. We evaluated trustworthiness based on the sample size from which the data were drawn. Random assignment of respondents to targeted learning goals helped to limit bias in the sample; however, the small and inconsistent sample size within and across subcomponents was a limitation of this study. The small number of participants who completed the survey resulted in some core concepts and subcomponents being reviewed by at minimum three educators, while others were more often reviewed by at least five educators (see Table 2 above).

To illustrate how we reconciled data across multiple sources, we examine core concept A.2 (Composing Whole Numbers). When reviewing the ordering of this core concept, the survey data indicated that respondents did not agree with the original ordering. Seven out of 12 teachers supported a revised order, and the concordance coefficient was statistically significant. However, five teachers provided evidence confirming the original ordering. To reconcile these findings, we consulted the literature to re-examine the rationale for our original ordering. Finding areas of disagreement in the literature, we decided to revise the ordering to align with the evidence obtained from the survey.

After all such revisions were made, we reconvened the experts who supported the initial development of the learning progressions. We shared the changes we made and our rationale, and solicited their input on the comprehensiveness and relevance of the content, accuracy of ordering, and plausibility of the misconceptions and errors. Although outside of the scope of this manuscript, these data were further integrated into our summary documents and used to make final revisions to the learning progressions. In instances where the experts disagreed with the revisions, we discussed their concerns, consulted the literature, and came to a mutually agreeable decision.

2.2. Study 2: MMaRS learning progressions survey

The purpose of this study was to investigate teachers' perceptions about the developmental appropriateness of the subcomponents in the MMaRS Learning Progression focused on NRR, particularly considering the boundary of each subcomponent across Grades K-2. The learning progression was divided into three targeted learning goals: A: Relations, B: Composition and Decomposition, and C: Properties of Operations. Each targeted learning goal has three to four core concepts (sequentially numbered). Multiple subcomponents were articulated for each core concept (referenced by lower case letters). Please note that the nomenclature is consistent with that used for the ESTAR Learning Progressions.

The following research questions guided data collection and analyses:

1 To what extent do K-3 teachers rate the MMaRS Learning Progression subcomponents as developmentally appropriate for the grade they teach? Does this differ by grade level?
2 To what extent do K-3 teachers rate the subcomponents as important focal skills for the grade they teach? Does this differ by grade level?
3 What is the relationship between teachers' ratings of the subcomponents as developmentally appropriate and a focal topic for the grade level?

2.2.1. Participants

MMaRS participants are representative of a national sample because the MMaRS Learning Progressions were not developed to align to a specific state's content standards. Participants included general education teachers, special education teachers, and other education professionals who work with children in Grades K-3. Even though the MMaRS Learning Progressions span Grades K-2, Grade 3 teachers were eligible to participate in the survey to provide a unique perspective about the longitudinal appropriateness of

Table 6
MMaRS Survey Respondent Demographics (n = 274).

	Total	%
Gender		
Male	31	11.31
Female	238	86.86
Prefer not to answer	5	1.82
Race		
Asian American/Pacific Islander	8	2.92
Black/African American	26	9.49
Hispanic/Latino American	66	24.09
Native American	3	1.09
White/European American	155	56.57
Multiracial	5	1.82
Other	4	1.46
Prefer not to answer	7	2.55
Age		
20 – 29	51	18.61
30 – 39	88	32.12
40 – 49	77	27.10
50 – 59	44	16.06
60 years or greater	13	4.74
Prefer not to answer	1	0.36

the progressions.

Researchers used a variety of methods to recruit participants including social media (e.g., Twitter), newsletters from the authors' research unit to over 3000 educators, and district-level and university colleagues who were asked to send the survey to their networks. Institutional Review Board procedures were followed. A total of 274 educators consented and completed the survey. Participants were largely Grades K-3 general education teachers (84 %) from around the United States. The remaining respondents were special education teachers, mathematics coaches, and interventionists (16 %). Table 6 summarizes the respondent sample.

2.2.2. Survey instrument

The MMaRS Learning Progressions Survey was administered electronically via Qualtrics. Similar to the ESTAR Learning Progressions survey, each educator was randomly assigned to answer questions about one of the learning progressions (i.e., Relations, Composition and Decomposition, or Properties of Operations). Thus, the number of educators who responded to questions about each targeted learning goal is approximately balanced–92 for Relations, 90 for Composition and Decomposition, and 92 for Properties of Operations.

Teachers were shown each subcomponent, without any reference to the hypothesized grade level, and asked five questions about each skill statement:

1 How much time do you spend teaching and reviewing this topic? (4-point scale, 1 = Not taught to 4 = Taught 2–3 times per week)
2 When during the school year, do you teach this topic? (Categorial responses: Not taught, Fall, Winter, Spring)
3 How important is this topic as a prerequisite skill in the grade you teach? (4-point scale, 0 = Not important to 3 = Very important)
4 How important is this topic as a focal skill in the grade you teach? (4-point scale, 0 = Not important to 3 = Very important)
5 How developmentally appropriate is this topic for the grade you teach? (4-point scale, 0 = Not appropriate to 3 = Very appropriate)

2.2.3. Analyses

For the purpose of the current manuscript, we report teachers' perceptions about (1) the importance of the subcomponent as a focal skill in the grade they teach and (2) the developmental appropriateness of the subcomponents for the grade they teach. With each subcomponent, we tested the independence of the reported developmental appropriateness and the reported importance as a focal skill by grade level. Significant dependence by grade level would indicate that teachers' perceptions in these domains depended on their target grade level. For example, significant dependence by grade level in developmental appropriateness may indicate that teachers in certain grades find the subcomponents more developmentally appropriate than teachers in other grades. Furthermore, we visualized the relation between the reported importance as a focal skill with the developmental appropriateness by grade with modified bubble plots. Modified bubble plots chart the frequency of responses in two domains. Larger bubbles indicate higher agreement with the rating in the domains. For example, if nine teachers responded somewhat important and somewhat appropriate in Grade 1 and five teachers responded similarly in Grade 2, then the bubble would be proportionally larger for Grade 1 because of the higher frequency. We conducted these analyses for each subcomponent to assess the appropriateness of the lower and upper boundaries of the subcomponents within the learning progression.

<table>
<tr><th rowspan="2">Code</th><th colspan="3">Kindergarten</th><th colspan="3">Grade 1</th><th colspan="3">Grade 2</th></tr>
<tr><th>Beginning</th><th>Middle</th><th>End</th><th>Beginning</th><th>Middle</th><th>End</th><th>Beginning</th><th>Middle</th><th>End</th></tr>
<tr><td>A.4.a</td><td colspan="4">Find how much more/less between two quantities using counting strategies.</td><td colspan="5"></td></tr>
</table>

Fig. 1. Bounds for Skill Statement A.4.a. This figure illustrates the initial hypothesized boundaries (in various parts of the year) of student skill development to “Find how much more/less between two quantities using counting strategies”.

2.2.4. Findings

For the purpose of this manuscript, we present the data from one subcomponent of the NRR Learning Progression (“Find how much more/less between two quantities using counting strategies”). The subcomponent and its bounds are described in Fig. 1. We hypothesized that instruction of the skill begins in Kindergarten and continues to the beginning of first grade. As indicated by the nomenclature A.4.a, this statement is the first subcomponent (a) in the fourth core concept (4) in the Relations (A) learning goal.

Table 7 highlights teachers’ responses on the developmental appropriateness and importance as a focal skill for this subcomponent by grade level. Percentages are calculated for the grade level reported. Examining the modal value and the distribution of responses suggests that a majority of teachers in Grades K-3 indicate the subcomponent is developmentally appropriate and important as a focal skill.

Furthermore, we inspected the modified bubble plots for “Find how much more/less between two quantities using counting strategies” (Fig. 2) and found high agreement across grades between teachers’ perception of the developmental appropriateness and importance as a focal skill. This is evidenced by 11 (73 %) Kindergarten teachers, 17 (81 %) first grade teachers, and 16 (80 %) second grade teachers that responded either Appropriate/Important or Very Appropriate/Very Important.

2.2.5. Conclusions

Data obtained from the MMaRS Learning Progressions Survey were intended to inform the placement of the lower and upper boundaries of the NRR learning progression. To that end, we solicited teachers’ perceptions about the developmental appropriateness and importance as a focal skill for each subcomponent by grade level. To form a comprehensive perspective about the entry and exit knowledge and skills for NRR, these data were combined with evidence gathered through cognitive interviews with children in grades K-3, reviews by experts in the field, and summaries of research conducted on students’ numeric relational reasoning.

Similar to the analytic approach used with the ESTAR Learning Progressions Survey, we examined the descriptive statistics to draw conclusions about the teachers’ perspectives. Because teachers were unaware of the originally hypothesized boundaries, the modal value was not used as confirmatory evidence, but instead to provide evidence about the developmental appropriateness and/or importance of the subcomponents as a focal skill for each grade level. Modal values of 1 or 2 were interpreted to mean that teachers did not perceive the subcomponent as developmentally appropriate or important as a focal skill for that grade. Modal values of 3 or 4 provided evidence that teachers perceive the subcomponent as developmentally appropriate or important as a focal skill for that grade. For the subcomponent “Find how much more/less between two quantities using counting strategies,” the modal values for grades K- 3 indicate that teachers perceive this skill as being developmentally appropriate and important as a focal skill across grades. The bubble plots illustrate relatively strong agreement across these dimensions. Just as was described for the ESTAR Learning Progressions survey, these observations were recorded in summary documents that combined data from other sources for reconciliation.

Table 7
Teachers’ Responses for “Find how much more/less between two quantities using counting strategies” by Grade Level.

	Grade Level			
	K (n = 15)	1 (n = 20)	2 (n = 21)[a]	3 (n = 21)
Developmental appropriateness				
Not Appropriate	0 (0%)	0 (0%)	1 (5%)	1 (4.7 %)
Somewhat Appropriate	4 (27 %)	3 (15 %)	4 (20 %)	3 (14 %)
Appropriate	8 (53 %)	9 (60 %)	9 (60 %)	9 (43 %)
Very Appropriate	3 (20 %)	8 (40 %)	7 (35 %)	8 (38 %)
Focal skill importance				
Not Important	1 (6.7 %)	0 (0%)	1 (4.7 %)	2 (9.5 %)
Somewhat Important	3 (20 %)	3 (15 %)	4 (19 %)	4 (19 %)
Important	8 (53 %)	8 (40 %)	10 (48 %)	8 (38 %)
Very Important	3 (20 %)	9 (60 %)	6 (29 %)	7 (33 %)

[a] For Developmental appropriateness, Grade 2, n = 20.

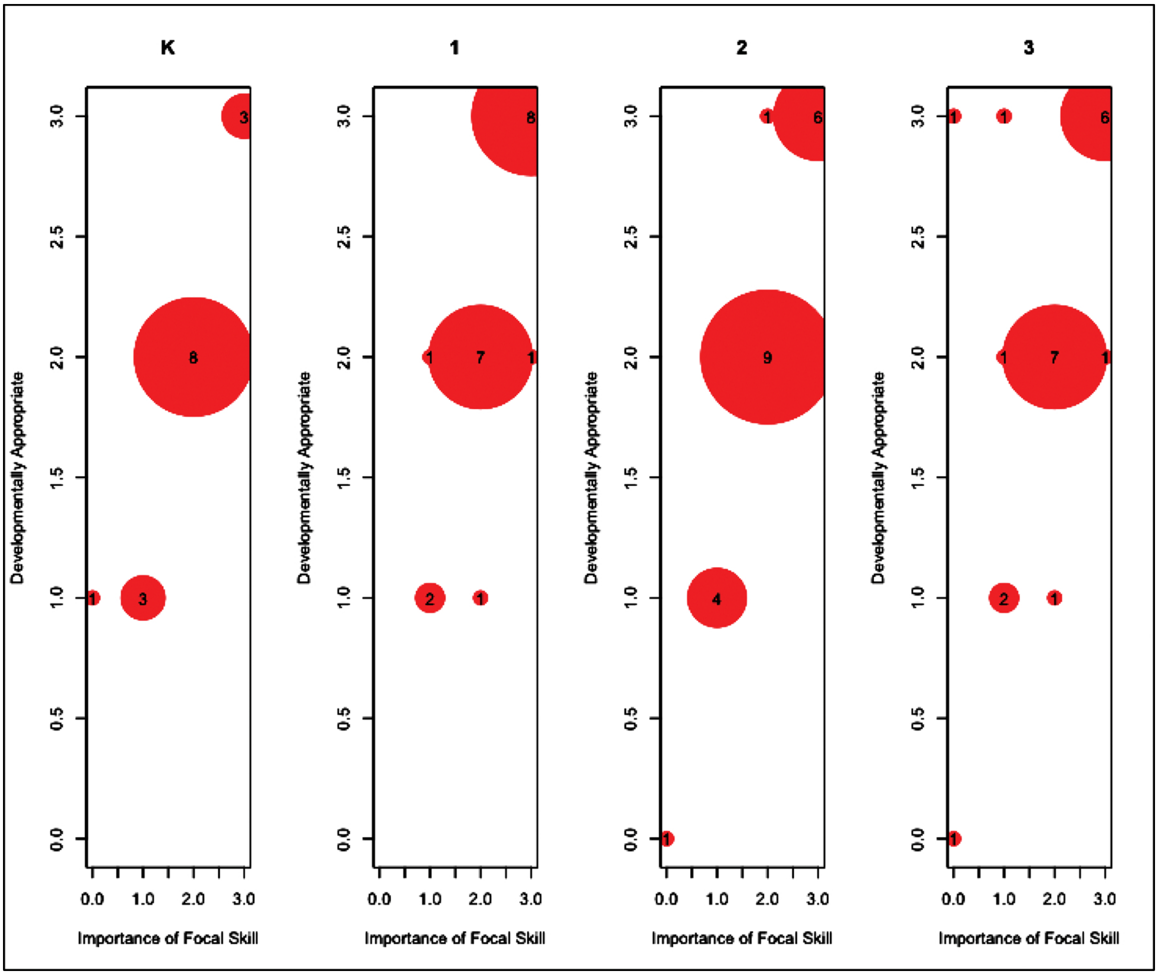

Fig. 2. Bubble Plot for Subcomponent A.4.a. This figure illustrates agreement across between teachers' perception of the developmental appropriateness and importance across grades for the subcomponent "Find how much more/less between two quantities using counting strategies".

We use the subcomponent "Find how much more/less between two quantities using counting strategies" to illustrate how we reconciled findings across sources. As shown in Fig. 1, the originally specified boundaries for this subcomponent ranged from the beginning of Kindergarten through the beginning of Grade 1. Findings from the survey indicate that teachers perceive of this skill as developmentally appropriate and important as a focal skill through Grade 3. Given these data, the researchers examined other sources of evidence to determine if the boundaries of the subcomponent should be modified. It is beyond the scope of this article to describe all data sources; however, results from cognitive interviews indicated that only two out of the four Grade 2 students were able to answer the question associated with this subcomponent correctly within 99. Conversely, experts in the field of mathematics education noted that the upper boundary should not extend beyond Grade 1. Because of the conflicting results, the boundaries for this subcomponent require additional evidence. Future studies, such as pilot tests of experimental items, are needed to adequately resolve this discrepancy.

3. Discussion

Educators are uniquely positioned within the teaching and learning process to provide meaningful input to inform the specification of learning progressions. Integrating their content knowledge and pedagogical content knowledge with knowledge of students (Hill & Chin, 2018) may provide teachers with a deep understanding of the mechanisms by which students develop sophistication in the domain, which can contribute to specifying the structure of learning progressions. Specifically, educators know the grade-level curriculum—and often those preceding and following a specific grade level. With this knowledge, they are familiar with specific points in the curriculum where students tend to struggle and the nature of those difficulties. Many teachers are able to anticipate students' errors because of their recurrence over time and across groups of students, and have likely developed an understanding of why students struggle with certain content through careful observation and engagement with students during the learning process. As such, educators may have insights into student learning that can help verify the structure of learning progressions.

In this manuscript, we illustrate how teachers' perspectives on student learning can be collected, analyzed, and used to help validate learning progressions. Across both projects, we designed a survey to capture teachers' input on specific aspects of the structure of learning progressions. With the ESTAR Learning Progressions, teachers provided input on the intermediary phases, helping to evaluate the ordering of the subcomponents and providing input on the plausibility of the specific misconceptions. In the MMaRS Learning Progression, teachers contributed meaningful information about the lower and upper boundaries of the learning processes across grades by considering the developmental appropriateness and relative importance of specific subcomponents as a focal skill. These data were analyzed in relation to the hypothesized structure of the learning progression. Trends were recorded and verified with other sources of data to inform revisions to the learning progressions. Modifications were made to the hypothesized learning progressions when findings from multiple sources of evidence (e.g., teacher survey, cognitive interviews, expert review) converged.

Central to the process of validating our learning progressions was synthesizing findings across multiple data sources, classifying evidence as confirming or disconfirming of the hypothesized learning progression, and reconciling competing findings that led to different hypotheses about how children develop sophistication in the content. Just as in validation efforts that examine the trustworthiness and meaningfulness of the uses and interpretations of test results (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014), validation of learning progressions is an iterative process that requires contribution of evidence from multiple sources, consideration of counter arguments, and an overall evaluation of the coherence and plausibility of evidence culminating in a judgment about the degree to which the original claims are supported or refuted by the data. In our research, we considered two primary factors when making this evaluative judgment. First, we considered the trustworthiness of the data obtained from each source. Although we designed each study to provide meaningful and reliable data, some limitations arose during implementation that impacted the internal validity of the studies. For example, we had smaller than anticipated sample sizes for the ESTAR Learning Progressions survey. Second, we considered whether the preponderance of evidence supported or refuted the originally hypothesized structure of the learning progressions. In some cases, such as with the ESTAR Learning Progressions survey, multiple sources of evidence converged that supported revising the ordering of the A.2. However, in other cases, such as the boundaries for A.4.a in the MMaRS Learning Progressions, the evidence was mixed, indicating that additional data are needed to draw conclusions.

As with test validity, validation of learning progressions can be considered as a never-ending cycle of gathering evidence to evaluate the claims about student learning. As such, mechanisms need to be in place to assess the sufficiency of the evidence and areas in which counter claims may be plausible. For the ESTAR and MMaRS Learning Progressions, we continue to gather additional evidence about student learning. For ESTAR, some questions that remain unanswered include how the learning progressions for one targeted learning goal intersect with those of another for individual students, the nature of the dependency between the structure of the learning progressions and instructional sequencing, and whether students with difficulties in mathematics follow similar or different pathways than the specified learning progression. Related to the ESTAR Learning Progressions survey, we wonder if our results will be replicable with another sample of teachers and if specific teacher characteristics (e.g., number of years teaching, curricula taught) impact their perceptions. For MMaRS, we continue to evaluate the lower and upper boundaries of the learning progression through pilot testing and psychometric analyses; these data will be combined with existing evidence to further update the structure of the learning progression. Also, because this project focuses on students in grades K-2, we have questions about the role of prior exposure to early mathematics concepts at home and/or in other early childhood settings in students' progression of learning. We also question the extent to which the lower and upper boundaries are influenced by the number range of the content standards. These and other questions will contribute to our understanding of the factors that influence student learning.

3.1. Implications for learning progressions research

The value of including teachers' perspectives in the validation process can be characterized as contributing both substantively and methodologically diverse input. Substantively, as previously stated, teachers bring unique and specialized knowledge and perspectives about the teaching and learning process that can meaningfully contribute to the body of evidence. Not only do they bring knowledge about students, but they also bring knowledge of curricular expectations and students' learning opportunities, which may impact the specification of the lower and upper boundaries as well as the ordering of the intermediary steps. Although it is likely impossible to deduce which form of teachers' knowledge they draw upon when responding to survey questions, their integrated perspectives of the content, pedagogy, and students contribute to their professional judgment about student learning.

Methodologically speaking, including teachers' voices as a source of evidence provides another lens from which to view the nature of knowing in a domain. As discussed earlier, some sources of data that are currently being used to validate learning progressions rely on students' responses to tasks or items that elicit their thinking. Although an important source of validity evidence, these data are inherently intertwined with task or item features that may impact students' responses. Incorporating data that are distinct from students' responses to tasks or items, such as the surveys used in the studies reported in this manuscript, may help disentangle students' thinking from other responding behaviors. Moreover, survey methodology is an efficient mechanism for gathering data. When survey data are used to triangulate findings from qualitative studies with small sample sizes, such as cognitive interviews or teaching experiments, they may increase the generalizability of the results because the data are obtained from teachers who are basing their responses on the cumulation of knowledge gained from working with many students. Although we had small sample sizes in the research reported in this manuscript, surveys represent an approach to collecting data that can improve the generalizability of the results.

3.2. Limitations

Both of the studies described above have several noted limitations. There are additional limitations that may impact the utility of integrating teachers' perspectives with the body of evidence to evaluate the validity of learning progressions. First, there are a number of factors that might impact teachers' responses to the surveys and our ability to rely on them as useful sources of information. For example, the number of years and role (e.g., special or general education settings) in which teachers have been working with children may impact the depth of their understanding of how children learn. Relatedly, teachers may be committed to the scope and sequence that guides their instruction, regardless of how this aligns with the processes of learning. To capture the value of teachers' in-situ knowledge of the learning process, it is important to frame their analysis from the perspective of how children learn as opposed to how the content standards, textbook, or mandated scope and sequence documents structure the content. To mitigate this limitation, follow-up interviews can be conducted with teachers to better understand what motivated their responses and what prior knowledge was elicited as they considered the questions. In addition, surveys can include demographic questions to isolate factors during analysis that may contribute to educator responses in unintended ways.

Another limitation that is endemic to survey methodology should be recognized. Namely, because surveys are typically completed without the researchers present, respondents may misinterpret the question, be distracted while responding, or only complete part of the survey. These actions may compromise the integrity of the data. Phrasing the questions for maximum understandability, providing an introduction to the survey, and offering incentives may improve the quality of the data. Also, larger sample sizes may help minimize the impact of spurious data on the findings.

Third, capturing teachers' perceptions about student learning may only address a subset of research questions needed to specify learning progressions. For example, teacher surveys may not be useful for understanding how students will respond to specific instructional approaches designed to deepen students' understanding; teaching experiments are uniquely suited to address these types of questions. Similarly, think aloud interviews with children may generate the most relevant data for understanding how specific tasks elicit students' thinking. As such, conducting surveys of teachers' perceptions may be best viewed as providing a breadth of evidence about student learning and not as a mechanism to supplant other data sources.

3.3. Conclusions

Traditional evidence used to validate the lower and upper bounds and intermediary phases of learning progressions may conflate student response to assessment items with their understanding of concepts within a domain. In this manuscript, we illustrate how educator experience and knowledge of students can provide another source of evidence to validate these aspects of hypothesized learning progressions. Two studies demonstrated how surveys were used to effectively and efficiently solicit educator perspectives of how students learn concepts—ordering of concepts, process of learning, and misconceptions—to inform focal points, ordering of learning progression concepts, and evaluate the comprehensiveness of ideas. In reporting on both studies, researchers provided examples of how educator input was analyzed in terms of the extent of agreement with the hypothesized structure to confirm the proposed structure of concepts within a domain or to highlight concepts, misconceptions, or orderings for further review alongside other sources of evidence.

As a complementary source of evidence, electronic surveys are feasible, generalizable, and help to separate assumptions about student learning processes from their responses on assessments. Surveys are cost-effective to administer and can employ features such as random assignment to strengthens the interpretation of resulting evidence. These studies indicate the value of incorporating educators' perspectives throughout the process and as a means to triangulate results from context dependent traditional validation sources.

Funding

Research conducted for this manuscript is based upon work supported by the (1) National Science Foundation under Grant No. 1721100 and (2) Texas Education Agency. Any opinions, findings, and conclusions or recommendations expressed in these materials are those of the authors and do not necessarily reflect the views of the National Science Foundation and the Texas Education Agency.

CRediT authorship contribution statement

Leanne R. Ketterlin-Geller: Conceptualization, Methodology, Supervision, Project administration, Funding acquisition. **Yetunde Zannou:** Conceptualization, Methodology, Formal analysis, Investigation, Visualization. **Anthony Sparks:** Formal analysis, Investigation, Data curation, Writing - review & editing, Visualization. **Lindsey Perry:** Conceptualization, Methodology, Writing - original draft, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors report no declarations of interest.

References

Alonzo, A. C. (2018). An argument for formative assessment with science learning progressions. *Applied Measurement in Education, 31*(2), 104–112.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Baroody, A. J., Purpura, D. J., Eiland, M. D., Reid, E. E., & Paliwal, V. (2016). Does fostering reasoning strategies for relatively difficult basic combinations promote transfer by K-3 students? *Journal of Educational Psychology, 108*(4), 576–591.

Carpenter, T. P., Levi, L., Franke, M. L., & Zeringue, J. K. (2005). Algebra in elementary school: Developing relational thinking. *ZDM, 37*(1), 53–59.

Clements, D. H. (2007). Curriculum research: Toward a framework for "Research-based Curricula". *Journal for Research in Mathematics Education, 38*(1), 35–70.

Confrey, J., Maloney, A. P., & Corley, A. K. (2014). Learning trajectories: A framework for connecting standards with curriculum. *ZDM Mathematics Education, 46*, 719–733.

Confry, J. (2018). *Future of education and skills 2030: Curriculum analysis – A synthesis of research on learning trajectories/progressions in mathematics (EDU/EDPC[2018]44/ANN3).* Organisation for Economic Co-operation and Development.

Corcoran, T., Mogat, F. A., & Rosher, A. (2009). *Learning progressions in science: An evidence-based approach to reform (CPRE Research Report #RR-63)*Philadelphia, PA: Consortium for Policy Research in Education Philadelphia, PA: Consortium for Policy Research in Education.

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education, 47*(2), 123–182.

Farrington-Flint, L., Canobi, K. H., Wood, C., & Faulkner, D. (2007). The role of relational reasoning in children's addition concepts. *British Journal of Developmental Psychology, 25*, 227–246.

Gunckel, K. L., Covitt, B. A., Salinas, I., & Anderson, C. W. (2012). A learning progression for water in socio-ecological systems. *Journal of Research in Science Teaching, 49*(7), 843–868.

Hill, H. C., & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal, 55*(5), 1076–1112.

Kara, M., Simon, M. A., & Placa, N. (2018). An empirically-based trajectory for fostering abstraction of equivalent-fraction concepts: A study of the Learning through Activity research program. *Journal of Mathematical Behavior, 52*, 134–150.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*(3), 568–589.

Ketterlin-Geller, L. R., Shivraj, P., Basaraba, D., & Yovanoff, P. (2019). Using mathematical learning progressions to design diagnostic assessments. *Measurement: Interdisciplinary Research and Perspectives, 17*(1), 1–22. https://doi.org/10.1080/15366367.2018.1479087.

Korbin, J. L., Larson, S., Cromwell, A., & Garza, P. (2015). A framework for evaluating learning progressions on features related to their intended uses. *Journal of Educational Research and Practice, 5*(1), 58–73.

Lai, E. R., Kobrin, J. L., DiCerbo, K. E., & Holland, L. R. (2017). Tracing the assessment triangle with learning progression-aligned assessments in mathematics. *Measurement: Interdisciplinary Research and Perspectives, 15*(3–4), 143–162.

Martinez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment, 14*, 78–102.

National Research Council (2009). Mathematics learning in early childhood: Paths toward excellence and equity. In C. T. Cross, T. A. Woods, & H. Schweingruber (Eds.). *Committee on Early Childhood Mathematics*.

Penuel, W. R., Confrey, J., Maloney, A., & Rupp, A. A. (2014). Design decisions in developing learning trajectories-based assessments in mathematics: A case study. *Journal of the Learning Sciences, 23*(1), 47–95. https://doi.org/10.1080/10508406.2013.866118.

Sarama, J., & Clements, D. H. (2019). Learning trajectories in early mathematics education. In D. Siemon (Ed.). *Researching and using progressions (trajectories) in mathematics education* (pp. 32–55). Leiden, Netherlands: Brill.

Simon, M. A. (2018). An emerging methodology for studying mathematics concept learning and instructional design. *Journal of Mathematical Behavior, 52*, 113–121.

Simon, M. A., Placa, N., Avitzur, A., & Kara, M. (2018). Promoting a concept of fraction-as-measure: A study of the learning through activity research program. *Journal of Mathematical Behavior, 52*, 122–133.

Sudkamp, A., Kaiser, J., & Moller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743–762.

van Rijn, P. W., Graf, E. A., & Deane, P. (2014). Empirical recovery of argumentation learning progression assessments of English language arts. *Psicologia Educativa, 20*, 109–115.

Whitacre, I., Schoen, R. C., Champagne, Z., & Goddard, A. (2017). Relational thinking: What's the difference? *Teaching Children Mathematics, 23*(5), 303–309.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching, 46*(6), 716–730.

Wright, V. (2014). Toward a hypothetical learning trajectory for rational number. *Mathematics Education Research Journal, 26*, 635–657.