



Construct maps and item frameworks: an example in proportional reasoning

Michele B. Carney¹ · Everett Smith² · Gwyneth Hughes³

Received: 14 September 2021 / Revised: 11 March 2022 / Accepted: 17 March 2022
© The Author(s) 2022

Abstract

Construct maps and related item type frameworks, which respectively describe and identify common patterns in student reasoning, provide teachers with tools to support instruction built upon students' intuitive thinking. We present a methodological approach—embedded within common assessment development processes—for developing such construct maps and item type frameworks which entails (1) articulating a construct map, (2) developing diagnostic item types, (3) comparing item type difficulties, and (4) modifying or providing evidence in support of the construct map. We present an example focused on students' proportional reasoning based on its centrality to mathematical success. The results further inform our construct map for students' proportional reasoning and provide a diagnostic item-type framework.

Keywords Proportional reasoning · Assessment · Diagnostic · Item types · Construct map

Introduction

Mathematics instruction should focus on building upon students' reasoning (Gravemeijer, 2004; Pellegrino et al., 2001; Simon, 1995). While, theoretically, this premise makes sense, identifying and using student reasoning to make instructional choices is complicated. A lack of teacher-ready models of student reasoning for different mathematics domains along with a paucity of associated assessment resources may inhibit teachers from using students' reasoning as a basis for instruction. Construct maps (Wilson, 2004) that describe common patterns in

✉ Michele B. Carney
michelecarney@boisestate.edu

¹ Boise State University, 1910 University Drive, Boise, ID, USA

² University of Illinois, Chicago, Chicago, IL, USA

³ Boise State University, Boise, ID, USA

students' thinking, in association with assessment item-type frameworks to diagnose patterns in students' thinking, could assist teachers in engaging in this practice.

Cognitively Guided Instruction (CGI) is an example of a project focused on articulating research findings in a manner that assists teachers in building upon students' mathematical reasoning. CGI materials (e.g., Carpenter et al., 2004) for addition and subtraction clearly articulate (1) a domain-specific construct map for students' solution strategies (i.e., counting all, counting on, and derived facts), and (2) an item-type framework (i.e., join, separate, part-part-whole, and compare problems) to assist in the diagnosis of student's understanding to inform instruction. While professional development around the CGI map and framework has demonstrated remarkable success (Carpenter et al., 1998; Fennema et al., 1996), relatively few other mathematics domains, with a focus on teachers as end users, have been similarly well articulated. Given the breadth of mathematics, the development of construct maps and item type frameworks should focus on core mathematics topics considered likely to be fundamental to students' future success; proportional reasoning is one such topic.

Students' ability to reason about situations involving ratios, rates, and proportions is key to future success in mathematics and science classes (e.g., Bush & Karp, 2013; Cetin & Ertekin, 2011; Johnson, 2015; Tucker et al., 2013). In particular, proportional reasoning is known to influence students' ability to conceptually understand more complex mathematics and science topics (Akutufba & Wallace, 1999; Ramful & Narod, 2014). Extensive research has been conducted around students' ability to reason proportionally (e.g., Karplus et al., 1983; Lamon, 2007; Lesh et al., 1988; Tourniaire & Pulos, 1985). These projects have developed relatively in-depth models for understanding students' proportional reasoning that distinguish among the different semantic structures of proportional reasoning problems (e.g., Harel & Behr, 1989; Lamon, 2005) and their relationship to student thinking. These models are invaluable for developing a nuanced understanding of students' proportional reasoning and for developing high-quality curricular resources. However, in our professional development work with in-service teachers, we find teachers want relatively simple models that can be broadly and quickly applied to classroom instruction, such as the CGI solution strategy map and problem type framework.

Based on our professional interest in supporting teachers' classroom practice, we focus on a relatively fine-grained model based on two ways students tend to reason about proportional relationships—the composed unit and multiplicative comparison conceptions (Cengiz & Rathouz, 2018; Ellis, 2013; Lobato et al., 2010). This model emphasizes the importance of students' fluently and flexibly making use of both conceptions with a focus on building a conceptual foundation for future topics, such as the meaning of slope within linear functions. Due to our use of a fine-grained measurement-based approach, we have adopted Wilson (2004) terminology of construct map to describe the articulation of a hierarchical model of cognition of increasingly sophisticated reasoning within a domain.

This paper has two main purposes. One is specific to proportional reasoning. Through the articulation of a construct map, we examine the hierarchical relationship between composed unit and multiplicative comparison conceptions to iteratively improve our understanding of student cognition. The second purpose is to describe a particular measurement-based methodological approach that could be

used generally across mathematics to develop domain-specific construct maps and interconnected item-type frameworks. Overall, this approach involves.

1. Construct map articulation: articulating a theorized progression for the construct under examination (Wilson, 2004).
2. Item type development: developing item types and associated items that target specific components within the construct map, including intentional analysis and control for potentially confounding factors.
3. Comparing item types: using Rasch item difficulty measures to examine patterns in hierarchical relationships between item types, as well as examine potential confounding factors.
4. Reexamination of the construct map: using results as a basis for map revision of or as supporting validity evidence for the construct map.

We situate the description of this methodological approach within the domain of proportional reasoning, making use of the composed unit and multiplicative comparison conceptions model of cognition (Lobato et al., 2010). We organize the remainder of the manuscript by presenting each of the above methodological steps, first described generally, followed by our example situated in proportional reasoning.

Construct map articulation

Construct maps represent the operationalization of a model of cognition, typically around reasoning processes. They are created at the beginning of the assessment development process to represent increasingly sophisticated levels of the construct under investigation in terms of respondents' reasoning and item responses (Wilson, 2004). They are similar to learning progressions (LPs) or hypothetical learning trajectories (HLTs) (Simon, 1995; Simon & Tzur, 2004) in that they are useful tools for constructs that have a hierarchical structure to their components (Wilson, 2009). However, construct maps focus on developing an assessment structure rather than activities for classroom instruction (as is typical of LPs and HLTs, cf. Duncan & Hmelo-Silver, 2009; Simon & Tzur, 2004, respectively). We also see construct maps as useful resources for teachers when formatively assessing student understanding to help determine where a student lies along a trajectory of reasoning processes. Similar to LPs or HLTs, construct maps start as hypothetical models of cognition, often informed by extant research when available. The methodological approach outlined in the introduction and described in the remainder of this article can provide quantitative evidence for the hierarchical ordering of qualitatively different components within the construct map. It is important to note we use the term hierarchical ordering to refer to experiencing more or less difficulty with the qualitatively different components within the construct map. This is different from a developmental hierarchy, which would specify that a particular aspect of cognition is necessary to achieve the next level. Further investigation would be needed to determine if the hierarchy revealed by a construct map is related to development.

Construct map for proportional reasoning

Wilson (2009) describes the need for “Clear definitions of what students are expected to learn [i.e., define the construct], and a theoretical framework for how that learning is expected to unfold [i.e., the construct map]...” (p. 4). Our examination of proportional reasoning and related construct map articulation focuses on students’ solution strategies and conceptions for missing-value problems and rate situations that are foundational to more sophisticated understandings, such as constant of proportionality, rate of change, and slope. We begin by explaining the connections between conceptions and common student solution strategies and the mathematical relationships involved in proportional situations. We use the term solution strategies to describe the commonalities in the observable mathematical processes students use to solve a problem. We use the term conceptions to describe how students perceive the relationships between quantities in proportional reasoning situations. Student conceptions inform their choice of solution strategies.

Mathematical relationships In order for students to reason proportionally, they need to be able to generalize their understanding of important mathematical relationships. More specifically, they need to fluently and flexibly make use of the scalar and functional relationships that exist in proportional situations. The scalar relationship describes the scale factor each quantity in the ratio can be multiplied by to generate an equivalent ratio (Lobato et al., 2010). The scale factor in the scalar relationship changes as the equivalent ratio to be generated changes. The functional relationship describes the constant multiplicative factor that exists between the two quantities in the ratio (Lobato et al., 2010). The multiplicative factor in the functional relationship remains constant in all equivalent ratios. There are several terms describing these relationships in the literature, each of which has its own specific mathematical significance and value to the larger community researching proportional reasoning. We have chosen to use the terms scalar and functional relationships because 1) they appear in texts teachers are likely to encounter in practice (e.g., Lobato et al., 2010) and 2) because the terms themselves describe how the relationships connect to larger mathematical concepts, namely scale factor and function. The top portion of Table 1 provides an example illustration of the scalar and functional relationships within the context of a missing-value problem.

Student conceptions Students’ fluent and flexible use of the scalar and functional relationships involves coordinating multiple understandings of ratios, proportions, and rates. In this paper, we differentiate between mathematical relationships inherent to all proportional relationships (scalar and functional) and students’ conceptions of proportional relationships—which depend on how a particular student understands the situation. A student may, for example, make use of the scalar or functional relationship as part of a taught procedure without being able to articulate or generalize the actual relationship. Our research interest is uncovering what mathematical conception(s) students actually hold, not simply which mathematical relationships they employ in their solution strategies.

Table 1 The mathematical relationships - scalar and functional - and student conceptions - composed unit and multiplicative comparison - demonstrated with the following missing value task. If the 12 cookies for \$3 relationship remains the same no matter how many cookies you buy, how many cookies could you buy for \$45?

Mathematical Relationships		
Associated Student Conceptions		
Conception (Strategy)	Student Exemplar Response	Student Work Examples & Mathematical Relationship Used
Composed Unit (Scalar Multiplicative Strategy)	I need to multiply 3 by 15 to get 45, so I multiply 12 by 15 to get 180 cookies	Scalar relationship
Composed Unit (Unit Rate Strategy)	I divided 12 by 3 to get 4 cookies for \$1, if I have \$45 then I multiply 4 by 45 to get 180 cookies	Functional & scalar relationships
Multiplicative Comparison	Cookies are always 4 times more than the cost (\$), so I can multiply 45 by 4 to get 180 cookies	Functional relationship

In missing-value situations, students tend to conceive of ratios as a composed unit—two coordinated quantities—typically making use of the scalar relationship to generate equivalent ratios (Lobato et al., 2010). This conception of ratios is often initially demonstrated through strategies involving doubling and halving the quantities in the ratio or additive scaling through iterating and partitioning to generate equivalent ratios. Eventually, this conception can develop into strategies that make use of the scalar multiplicative relationship where the initial ratio is treated as a composed unit and each quantity in the ratio is scaled (multiplied or divided) by a single scale factor (see the row for composed unit: scalar multiplicative strategy in Table 1 for an example). Another strategy that makes use of the composed unit conception involves the generation of a unit rate by dividing one quantity in the ratio by the other to generate a per-one relationship. This unit rate can then be scaled to

generate additional equivalent ratios (see the row for composed unit: unit rate strategy in Table 1). These various approaches can all result in the correct answer to the problem but demonstrate different levels of fluency and flexibility in student conceptions and strategies related to the scalar relationship.

In contrast, the multiplicative comparison conception related to the functional relationship involves understanding and making use of the constant multiplicative relationship that exists between the two quantities in a ratio. There is some qualitative evidence students tend to have increased difficulty conceiving of the multiplicative comparison¹ as compared to the composed unit conception (Simon & Placa, 2012; Steinhorsdottir & Sriraman, 2009). Generation and use of a unit rate can mask students' lack of understanding of the multiplicative comparison relationship due to the resulting numerical quantities being the same. For example, for the ratio of 12 cookies for \$3, 4 cookies for \$1 would be a unit rate, and the related multiplicative comparison is the number of cookies is always 4 times the dollars. The number four is the same in the unit rate and multiplicative comparison, and the solution strategy—dividing 12 by 4—is also the same. The difference is in students' conceptions, distinguished by examining how students express what the 4 means in the ratio context, 4 cookies for \$1 (composed unit conception) or the number of cookies is always 4 times the number of dollars (multiplicative comparison conception). It is the multiplicative comparison conception that allows students to meaningfully generalize the relationship between the two variables in the equation, $y = 4x$.

In our own previous interview research, we found that students who used the strategy of dividing one quantity in the ratio by the other regularly expressed the unit rate as a composed unit (Carney & Crawford, 2016). Thus, while we knew examining solution strategies to missing-value problems provided teachers some insight into student reasoning, we realized we needed to go beyond this approach to understand which conceptions students held. Figure 1 visually represents the respondent reasoning portion of the construct map involving student conceptions. It includes the elements of solution strategies and mathematical relationships to highlight these connections. The relationships in our initial construct map were hypothesized from the extant literature (cf., Karplus et al., 1983; Lamon, 2007; Lobato et al., 2010; Misailidou & Williams, 2003a; Simon & Placa, 2012; Steinhorsdottir & Sriraman, 2009; Tourniaire & Pulos, 1985) and our previous assessment development iterations (Carney & Crawford, 2016; Carney et al., 2015, 2016). While there is some evidence that the composed unit conception is more accessible to students than the multiplicative comparison conception (e.g., Steinhorsdottir & Sriraman, 2009), as depicted in our construct map, we could not find any empirical studies that quantitatively examined the relationship between these two conceptions (Noelting conducted somewhat related hierarchical work with

¹ While the term multiplicative comparison has been used to describe the multiplicative relationships both within and between measure spaces, similar to Lobato et al. (2010), we use the term multiplicative comparison to describe the constant multiplicative comparison that exists between the two quantities involved in a rate situation due to the importance of conceiving of this relationship for developing an understanding of slope, rate of change, constant of proportionality, etc.

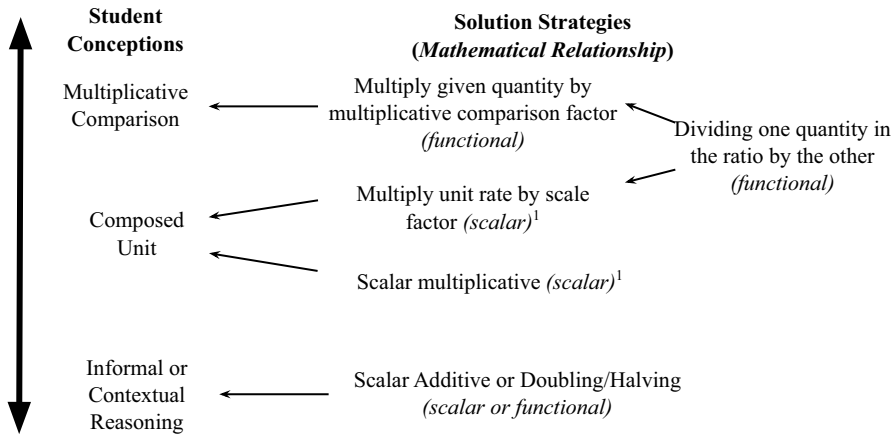


Fig. 1 The respondent (i.e., student) reasoning side of the construct map represents increasing sophistication in students' proportional reasoning conceptions. We have included student solution strategies and mathematical relationships to highlight these connections

proportions 1980a, b). Therefore, one aspect of our construct map we wanted to further investigate was the hierarchical relationship between the composed unit and multiplicative comparison conceptions as part of our efforts to iteratively improve our understanding of student cognition.

The construct map highlights what we found in our own prior research; examining student solution strategies to missing-value problems would not provide us with enough information to determine if a student held a multiplicative comparison conception (Carney & Crawford, 2016). We thus needed items that would let us differentiate students' use of a composed unit conception from that of the multiplicative comparison conception. Our approach involves operationalizing and isolating the conceptions in item types. The next section describes our development of item types that targeted specific components in the construct map, including intentionally controlling for likely confounding factors.

Item type development

In order for the results of assessments to inform classroom instruction, they need to situate the mathematics a student knows within a model of cognition that identifies critical components of conceptual understanding (Pellegrino et al., 2001). Despite the recommendations from the National Research Council Committee on the Foundations of Assessment (Pellegrino et al., 2001), the majority of mathematics assessments still use a mathematics domain sampling approach to assessment design. In this method, the designer develops a table of specifications for a particular mathematics domain and creates the assessment by sampling tasks from across different aspects of that domain. For example, when assessing the domain of fraction operations, a matrix may be built based on the four operations (addition, subtraction,

multiplication, and division) and fraction types (e.g., different denominators, use of mixed and improper fractions, etc.). Then tasks from the matrix are sampled to create an assessment. The results from these types of assessments usually indicate what mathematics topics within a domain a student knows and does not know, but they are not structured to provide insight into students' conceptual understanding (e.g., the meaning of the numerator and denominator, understanding fraction iteration and partitioning, etc.) as recommended by the National Research Council Committee on the Foundations of Assessment (Pellegrino et al., 2001). This makes such assessments powerful as large-scale summative assessments but less useful for formative purposes. For example, such a fraction assessment might tell you that a student can add fractions and not divide them but not be able to identify why.

In contrast, our assessment and item type development approach uses student cognition as the underlying model situated within a particular mathematics domain of interest. The construct map lays out the qualitatively different aspects of cognition that need to be assessed, and then item types are developed to isolate these aspects of cognition. Thus, the assessment results can be interpreted in relation to important aspects of student cognition situated within the mathematics domain, thus supporting and informing instructional choices that build upon student thinking. In addition, when item types are constructed to elicit critical conceptions specific to different construct map levels, we can examine the hierarchical relationships within the theorized construct map to better understand student cognition. Lastly, an important element of this work involves controlling for confounding factors known to change the difficulty of an item. This involves intentionally structuring items to ensure the difficulty level of a group of item types is primarily reflective of the aspect of cognition being assessed and extraneous factors known to impact item difficulty (e.g., whole number versus fractional answers) are either eliminated or when that is not possible, distributed equally across item types.

Returning to the Cognitively Guided Instruction project, their work provides an example of using a model of cognition situated within a particular mathematics domain in conjunction with a set of item types to assess aspects of the model. The underlying model for cognition in CGI involves student ability to perform addition and subtraction operations via counting all, counting on, and using derived facts strategies. Particular item types were used to assist in determining where a student might lie within the model of cognition. For example, a join change unknown item type can be used to determine if a student is starting to use a count on strategy (versus count all) because a student using a counting all strategy will likely not be able to correctly solve a join change unknown problem (due to problem structure not providing the two parts to count all of). Correctly solving several join change unknown problems would be indicative of the child being capable of or nearly ready to apply counting on to strategies more generally. This example can also be used to highlight the importance of controlling for confounding factors. For instance, three-digit addends would likely increase the difficulty of a join change unknown item and students might answer incorrectly who actually do possess the ability to count on with simpler number sets.

Item type framework for proportional reasoning

Our item types were designed to elicit the three aspects of student conceptions highlighted in our construct map; informal or contextual reasoning, composed unit, and multiplicative comparison. Carney et al. (2022) conducted cognitive interviews with 33 students in grades 6–8 at three different schools using the item types described below to ensure student response processes supported the claim that students who solve particular item types correctly were likely to possess associated conceptions. The focus in designing the item types was to isolate students' use of particular conceptions to the extent possible. Our item type framework makes use of a contextual ratio situation and provides five different prompts (i.e., item types) related to this situation. Table 2 provides an example item block with the item types and their associated conceptions.

Below we describe each item type and how it assesses the associated student conception.

- **Small single-digit multiplier: informal reasoning.** A missing-value problem where the scalar or functional multiplier was 2, 3, or 4. These items are designed to be relatively easy for students and even included a picture of the initial ratio in a food cost context (see Appendix 1). Students who do not possess an explicit understanding of the mathematical relationships and/or composed unit and multiplicative comparison conceptions would still be able to solve these problems using more informal reasoning strategies. The intent is for the items to screen for students who primarily hold an incorrect additive perspective (Misailidou & Williams, 2003a).
- **Double-digit scalar multiplier: composed unit conception.** A missing-value problem where the scalar multiplier is between 15 and 18. The multiplier is large enough that it is unlikely a student would solve them through informal strategies alone but are still relatively accessible without a calculator. Success on these items would likely indicate some initial understanding of the scalar multiplicative relationship, which involves the use of a composed unit conception.
- **Unit rate situations: composed unit conception.** The unit rate items ask for the value of one of the quantities in the ratio when the other quantity value was one. Previous research indicated students who demonstrated understanding of unit rate did not, for the most part, demonstrate an understanding of the multiplicative comparison conception (Carney & Crawford, 2016). The intent for these items was to assist in the identification of individuals who possessed at least an initial understanding of generating a unit rate, which involves a composed unit conception.
- **Equations: multiplicative comparison.** Modeling a rate situation with a contextual equation is not something commonly found in the research literature but is a necessary skill for understanding algebraic representations of proportional and later linear relationships. The equation items involve filling in the blank of an equation, such as the *number of cookies* = ____ ● *cost*. These items are

Table 2 An example item block in association with the item types and student conceptions from the full construct map

Item block stem: If the relationship between cookies and cost (\$) remains the same no matter how many cookies you can buy: 12 sugar cookies for \$3 Complete the following statements about the relationships between cookies and cost (\$)			
Example item statements	Item types	Student conception	
Callie bought 12 sugar cookies for \$3. How many cookies can she buy with \$2?	Small single-digit multiplier	Informal reasoning	
It would cost _____ to buy 180 cookies	Double-digit scalar multiplier	Composed unit	
One cookie costs _____	Unit rate	Multiplicative comparison	
Number of cookies = _____ \times cost	Equation		
The cost is always _____ times the number of cookies	Generalizing		

designed to assist in the identification of individuals who could make use of the multiplicative comparison in an equation.

- Generalizing: multiplicative comparison. The generalizing items involve filling in the blank on a statement of the multiplicative comparison relationship, such as the cost is always _____ times the number of cookies. The intent is for these items to assist in the identification of students who could generalize the meaning behind the multiplicative comparison relationship (see Carraher et al., 2008) for more information on explicit statements of mathematical relations). Some students may generate an equation by guessing and checking or using procedural means. Asking students to articulate the relationship in sentence form tests their conception of what that number means in context.

Controlling for confounding factors

In designing assessment items, developers must ensure confounding factors are not inappropriately influencing the results (i.e., limit construct-irrelevant variance). As math topics become increasingly complex, it becomes more difficult to isolate the qualitatively different levels or components in the construct map due to confounding factors. For example, topics such as proportional reasoning necessarily become conflated with students' ability to understand and reason with fractions and decimals (Fernández et al., 2011; Rupley, 1981).

Our approach to addressing these issues was to identify the primary factors known to influence students' proportional reasoning from the extant literature. The attributes of ratios we identified were:

- (A) Same units (e.g., gallons blue paint to gallons yellow paint) versus different units (e.g., cookies to cost and miles to hours) (Lawton, 1993).
- (B) Continuous units (e.g., paint to paint and miles to hours) versus discrete units (e.g., cookies to cost) (Behr et al., 1992; Boyer et al., 2008; Jeong et al., 2007; Tourniaire & Pulos, 1985).
- (C) Difficulty of the numerical relationship (i.e., was the solution a whole number or fractional answer) (Fernández et al., 2011; Rupley, 1981; Saunders & Jesunathadas, 1988; Schwartz & Moore, 1998).

We attempted to control for these factors by consistently distributing them across the different item types. Thus, the effect of a particular confounding variable was relatively evenly distributed across the different item types. Therefore, we did not remove the potentially confounding factors but instead attempted to evenly distribute their influence across the item types. Appendix 2 provides the matrix of how these factors were distributed across the different item blocks that were used to make up the assessment forms.

However, some aspects of the numerical relationships could not be controlled. For example, the multipliers and answers for the double-digit scalar multiplier item

answers were all double- or triple-digit whole numbers, whereas the unit rate, equation, and generalizing items had a mix of a whole number and fractional answers. The impact of confounding variables is further discussed in the results and discussion sections.

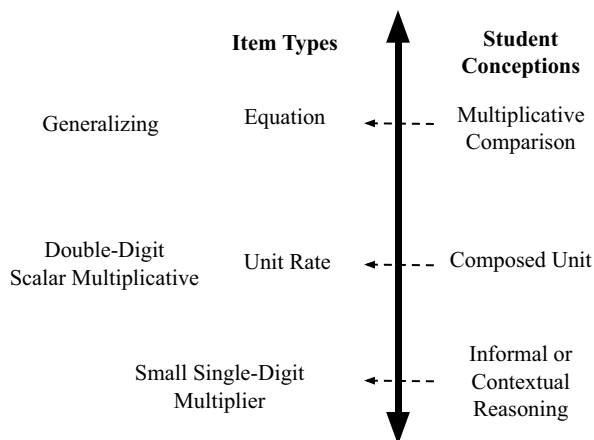
Existing proportional reasoning instruments

Several instruments have been built to better understand patterns in students' proportional reasoning using a similar analytic approach of examining associated Rasch item difficulties (e.g., Long et al., 2011; Misailidou & Williams, 2003b). For example, Long et al. (2011) built an instrument to assess the multiplicative conceptual field (Vergnaud, 1994) using a domain sampling approach to assessment construction. They used the subdomains of ratio, proportion, and percent from the TIMMS 2003 Mathematical Framework and selected 30 associated items from the TIMMS 2003 grade 8 item pool. Misailidou and Williams (2003b) also developed an instrument to broadly assess the proportional reasoning domain, primarily drawing from missing-value items selected or adapted from prior research articles related to proportional reasoning with a variety of numerical structures and contexts. In both cases, they conducted analyses related to student solution strategies and item difficulties. However, their lack of specification of an a priori construct map to inform the selection of their assessment items and the absence of a specific intent to control for confounding factors make it difficult to determine how well the assessment isolates particular aspects of cognition (in their case solution strategies).

Other projects, such as the Reframing Mathematical Futures II project (Siemon et al., 2017) and Mindmapper (Confrey et al., 2018), articulate hypothetical learning progressions, similar to a construct map. These types of assessments can be very useful to teachers to map individual students' knowledge within a mathematical domain along a trajectory of topics. Our work is similar but differs in specific ways. First is grain size because our work focuses on specific student conceptions, as opposed to knowledge of mathematical topics within a domain, it is necessarily smaller in scale. Another is our focus on the development of item types, as opposed to specific items and assessments. Item types offer the benefit of flexible use by teachers in conjunction with the construct map, but there is likely some tradeoff with rigorous validation efforts associated with fully formed assessments. Additionally, the use of multiple items within an item type also allows researchers to address/examine the impact of confounding variables on the placement of topics/conceptions along a difficulty hierarchy in a way the use of only a few items may not. This type of work serves as complementary to larger-scale learning progression assessments, and the two approaches can likely inform and improve each other.

As previously described, a construct map starts out as a theorized model where the ordering of elements is hypothesized based on previous research. Item types are then designed to assess the elements in the construct map. The next step involves conducting analyses on the item types to examine the relationships between them and to further develop and provide clarity to the elements in the construct map. Figure 2 presents our construct map for proportional reasoning including both the

Fig. 2 The full construct map representing the relationship between student conceptions and the item types, including the hypothesized difficulty of the item types



student conception elements and the item types in a hypothesized order of difficulty. For the purpose of iteratively improving the articulation of our construct map, we wanted to determine if the item types would consistently manifest themselves in a similar hierarchical pattern. In other words, would particular item types be easier or harder than others? If so, this could provide evidence of a hierarchical order amongst the conceptions within our construct map of students' proportional reasoning. In the next section, we present the methods and results related to our analyses. Our guiding research question is: What information do the item type difficulties provide to inform modification of or evidence in support of the construct map?

Comparing item types

Once an item type framework has been developed that addresses the qualitatively different components or levels within the construct map, assessments can be developed and administered, and Rasch analysis is used to generate item difficulties to compare across item types. Rasch analysis generates scores for both student ability and item difficulty. The estimates of student ability and item difficulty obtained from a Rasch analysis situate test takers' understanding (ability) and item difficulty along a common equal-interval scale when the data adhere to Rasch model requirements (Bond & Fox, 2013). While it is relatively common to compare student groups on their scale scores (cf., Foster et al., 2016), it is less common to compare item types on their difficulties but the analytic processes are the same.

Researchers (e.g., Andrich et al., 1997; Callingham & Siemon, 2021; Long et al., 2011) have examined individual item locations (along an interval scale) to assist in describing patterns in students' understanding of particular mathematics topics. Wolfe and Smith (2006) recommended the examination of means of similar types of items to provide valid evidence supporting an instrument's assessment of a particular construct. Similarly, Bolt et al. (2016) used item type means to describe relationships among construct components within an instrument. They use the term scale

anchoring to describe the identification of regions along a measurement scale where particular construct components tended to be assessed by related items. Siemon et al. (2018) describe these as zones and outline the steps in the process of identifying broad descriptions of behavior and teaching implications for each zone. Our approach seeks to intentionally isolate components within the construct map with our item types prior to assessment construction and administration. From a diagnostic perspective, these item types can then be used to assess students' understanding of different components within a construct. From the perspective of further articulating a construct map, the item type means can be compared to better understand hierarchical patterns in particular components within a construct.

The following proportional reasoning exemplar section for comparing item difficulties is formatted using a more traditional method and result structure.

Methods: comparing proportional reasoning item types

Participants

The goal of student sample selection was to have a wide range of person abilities, or student knowledge, related to proportional reasoning across the different item types. This helps to ensure each item has enough individuals who respond correctly and incorrectly to the item, improving the accuracy of item difficulty estimates. Therefore, we sought out students we expected to have limited exposure to proportional reasoning instruction (grade 6 students) all the way to students who should be done receiving classroom instruction on proportional reasoning (grade 9 students) and all the potential varying levels between.

In addition to our focus on a spread in abilities, we also wanted a relatively diverse sample across several variables that have the potential to influence mathematics achievement. Therefore, we secured samples from schools that differed across the following variables; instructional materials used (all four schools used different curricular resources), the socioeconomic status of students (schools ranged from 18 to 47% of students qualifying for free-and-reduced lunch), geographic setting-rural (1), suburban (2), and urban (1), and school designation-public charter (1) versus public traditional (3).

The assessment was administered to 1140 students in grades 6–9 from four different schools and school districts in the Pacific Northwest (grade six, 357; grade seven, 310; grade eight, 417; grade nine, 42; and unreported grade, 14). The majority of students at two middle schools took the assessments (school one, 531; school two, 378). The other two schools had a subset of students take the assessments (school 3, 88; school 4, 143). Students at schools three and four were targeted with the goal of broadening the participation of students who typically performed relatively well on the end-of-year assessments. This portion of the sample of students was obtained to ensure enough students scored in the upper range of the assessment to increase personal reliability (Linacre, 2016). It is important to understand that random sampling is not an assumption in the Rasch model. What is important is having a range of personal abilities (i.e., respondents that score from very low to very high on the

instrument). Item difficulty estimates are independent of the sampling distribution obtained if the data fits the Rasch model (Wright, 1977).

Setting and timeline

Meetings with teachers and school administrators were conducted to determine interest and discuss assessment administration procedures. The assessment packets were delivered to the schools within a week of these meetings. Teachers were asked to administer the assessments to students within the next 2 weeks. The administration directions requested that students not be allowed to use a calculator and that students could express their answer as either a fraction or a decimal if applicable. In addition, we communicated there was no time limit for the assessment but anticipated it would take students about 30 min.

Instruments

Six different forms of the assessment were created, each with four item blocks (see example item block in Table 2). There were a total of 20 items per form, with the first 10 items the same across all six forms and the remaining 10 items different. We needed data across a large collection of items, but there were too many items for each student to answer in a reasonable timeframe. Therefore, we used common item equating methods to put all data from each form on a common scale. This resulted in 68 distinct items (two items were repeated on two of the forms beyond the common 10 items). The items maintained a consistent format and spacing across all six forms. Appendix 1 provides the front side of an example assessment form.

Rasch data analysis

Researchers (e.g., Andrich et al., 1997; Callingham & Bond, 2006; Long et al., 2011) have argued for the use of Rasch methodology in mathematics education due to its diagnostic capabilities in describing and interpreting student and test performance in relation to the underlying construct. The dichotomous Rasch model can be represented by

$$L = \ln\left(\frac{P}{1 - P}\right) = B_n - D_i$$

where L is the natural logarithm of the probability of success over the probability of failure. B_n is a student's ability and D_i is an item's difficulty. The equation states that the log-likelihood for a student to answer an item correctly is a function of the difference between the item difficulty and the student's ability. As the positive difference between B and D becomes greater (indicating the student is more able than the item is difficult), the more likely a student is to respond correctly to an item. Conversely, the greater the negative difference, the more likely a student is to respond incorrectly to an item. In situations involving dichotomous scoring (0=incorrect,

1=correct), when student ability and item difficulty are the same, the model predicts a 50% probability that the student would respond correctly (or incorrectly).

The data were dichotomously scored (0=incorrect, 1=correct) and then analyzed using the Rasch model in the WINSTEPS version 3.70.0.5 (Linacre, 2016). There was no pattern to missing data, therefore, missing data was coded as missing, which does not bias estimates in Rasch measurement. Each form of the test was first analyzed independently with a focus on the examination of item fit for that form. Once specific fit issues were corrected (described below); the data from the six forms were combined into a single spreadsheet and analyzed through concurrent calibration.

Prior to examining the hierarchical functioning of the item type subgroups, we first determined if there was sufficient evidence of the technical quality of the items and their fit to Rasch model requirements. To do so, we examined the following item characteristics:

- (a) Do the item fit statistics support unidimensionality and local independence?
- (b) Do the item residuals support unidimensionality and local independence?

Item fit The technical quality of the items in the context of the Rasch model is examined through fit statistics. Fit statistics assist in the identification of items whose data do not conform to Rasch model requirements for unidimensionality and local independence via a comparison of what is observed and what is expected by the model. Given the size of our sample, we initially identified four items with Z-standardized (ZSTD) infit and outfit statistics greater than 2.58 ($p < .01$) for misfit, and these items were flagged for further investigation. The fit issues involved 18 of the 1140 students who had solved the majority of the problems either correctly or incorrectly, then incorrectly or correctly responding to a particularly easy or hard item, respectively. That is, their performance on that one item was not indicative of their performance on the overall assessment. For example, several students responded correctly to all of the assessment items except one extremely easy item. To examine the impact of these misfitting responses on the item measures, the data point representing an individual item (i.e., one of the 20 responses) was coded as missing for 18 of the 1140 students. This resulted in no items with Z-Standardized (ZSTD) infit and outfit statistics greater than 2.58 ($p < .01$). A comparison of the resulting item measures for the four (originally misfitting) items to their original item measures indicated very small changes in the overall item measures (-0.04 , -0.10 , -0.11 , and 0.14) that were not statistically significant ($p = 0.35$, 0.37 , 0.68 , 0.47 , respectively). Therefore, we opted to use the original item measures (i.e., the item measures prior to coding the 18 individual responses as missing).

Residual analysis As the item fit statistics described previously are not useful in all situations for detecting model violations, a principal component analysis (PCA) of standardized residuals is also recommended. This combination of methods has

proven effective at detecting departures from the unidimensionality and local independence requirements of the Rasch model employed in this evaluation (Linacre, 1998; Smith, 2002).

The Rasch model assumes data are unidimensional. Data should be checked for this assumption. Data are composed of the Rasch dimension (the construct under investigation) and unexplained variance. The unexplained variance is composed of random variance and other effects (e.g., secondary dimensions, a strand within the Rasch dimension, or random correlations). A PCA of the standardized residuals is run in the WINSTEPS program to decompose the unexplained variance. If there are contrasts with eigenvalues greater than 2.0, then you need to examine the item clusters for these contrasts to determine if there is potentially a meaningful second dimension. In addition, disattenuated person measure correlations less than +0.3 may also indicate a secondary dimension.

The PCA of the residuals for our data indicated we had one contrast barely above an eigenvalue of 2.0 at 2.07. Given that this was 1.6% of the observed variance, we suspected this was likely random variation. However, we proceeded to check the item clusters to determine if a meaningful secondary dimension could be seen based on the item groupings. There did not appear to be an underlying structure when examining the clusters of items in the first contrast. However, to double-check, we also examined the disattenuated person measure correlations. They were 1.0 across the three clusters indicating there was not a secondary dimension. Therefore, our data can be assumed to fit the Rasch model assumption of unidimensionality and local independence.

Results: comparing proportional reasoning item types

A one-way ANOVA was conducted to compare the effect of item type on Rasch item difficulties for small single-digit multiplier, double-digit scalar multiplier, unit rate, generalizing, and equation items. There was a significant effect of item types on item difficulties at the $p < .05$ level for the five item types [$F(4, 63) = 36.11$, $p < .001$]. Post hoc comparisons using Tukey's HSD test were conducted to determine which pairs of the five item type means differed significantly. These results are given in Table 3 and indicate there was a significant difference between all item types except double-digit scalar multiplier to unit rate and double-digit scalar multiplier to generalizing. Box plots of the item measures by item type are provided in Fig. 3.

Examination of the box plots in conjunction with the item type means in Table 3 helps to illuminate the differences between the five item types designed to assess student understanding and student ability to make use of particular solution strategies. Small single-digit multiplier items are the easiest item types, followed by the two item types focused on the composed unit conception—unit rate and double-digit scalar multiplier. There was not a significant difference between these two item types. The two item types designed to press for the multiplicative comparison conception—generalizing and equations—had the highest item difficulties, indicating

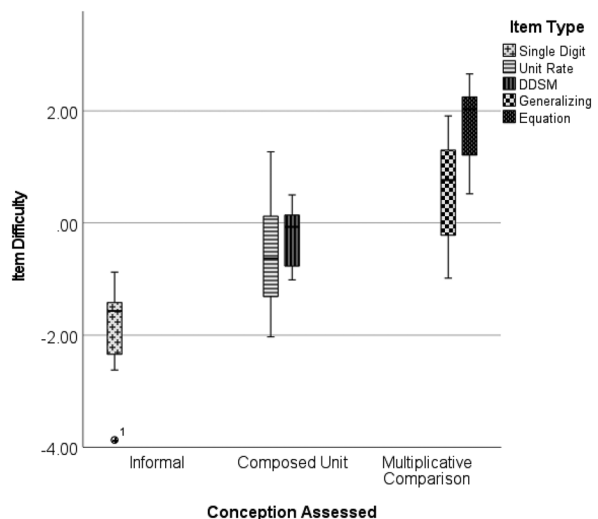
Table 3 Mean and standard deviation of the item difficulties for each item type and Dunnett T3 post hoc comparison results between item types

	<i>n</i>	SD	DDSM	UR	G	Eq
Small single digit multiplier (SD)	12	–				
Double digit scalar multiplier (DDSM)	14	SD < DDSM**	–			
Unit rate (UR)	14	SD < UR*	Not significant	–		
Generalizing (G)	14	SD < G**	Not significant	UR < G*	–	
Equation (Eq)	14	SD < Eq**	DDSM < Eq**	UR < Eq**	G < Eq*	–
Mean		–1.91	–0.21	–0.55	0.60	1.79
S.D		0.80	0.49	1.03	0.98	0.70

$p < .001^{**}$; $p < .01^{*}$

these items were the most difficult for students to answer correctly. Within these two item types, the equation item type was significantly harder than the generalizing item type.

When the item types are grouped within the conception they assess, as in Fig. 3, they highlight the trajectory of proportional reasoning conceptions from informal to the composed unit to multiplicative comparison. It is important to note that the unit rate, generalizing, and equation item type means were differentially impacted by items that had whole number answers (see the placement of whole number answer types in Fig. 4 indicated by the triangles). That is, as much as we tried to spread out the impact of confounding variables, answer type appeared to have differentially impacted the item type means of the multiplicative comparison item types compared to the unit rate item type. There were 14 unit rate items, 2 with whole-number answers, if removed the unit rate item difficulty mean increases by 0.16 ($M = -0.39$,

Fig. 3 Box plot of item difficulties by item type grouped by conception assessed

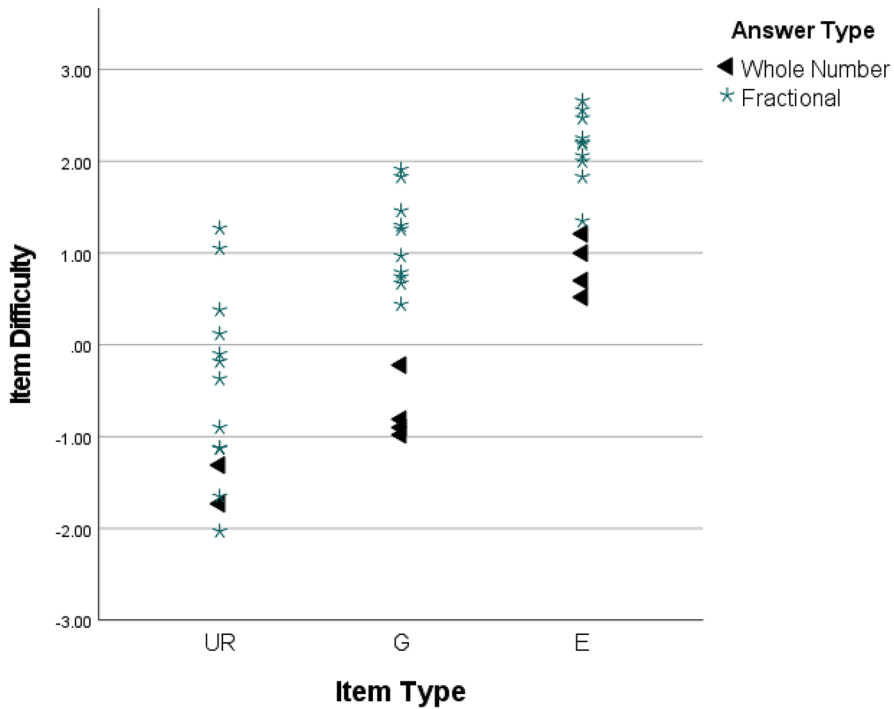


Fig. 4 Item difficulties by item type, highlighting the unequal impact of the whole number answer type items on the means of the three item types

SD=1.02). There were 14 generalizing items, 4 with whole-number answers, if these were removed the generalizing item difficulty mean increases by 0.54, and the standard deviation decreases by 0.48 ($M=1.14$, $SD=0.50$). Lastly, there were 14 equation items, 4 with whole-number answers, if these were removed, the equation item difficulty mean increases by 0.37, and the standard deviation decreases by 0.32 ($M=2.16$, $SD=0.38$). If the 10 items with whole-number answers are removed from the ANOVA, there is a significant effect of item types on item difficulties at the $p < .5$ level for the five item types [$F(4, 63)=55.49$, $p < .001$]. Post hoc comparisons using Tukey's HSD test found the significant differences between the item types mirroring Table 3 with the addition of a significant difference between the unit rate ($M=-0.39$, $SD=1.02$) and generalizing ($M=1.14$, $SD=0.50$) item types. (These 10 whole number answer type items are indicated in the grayed-out boxes in Appendix 2). The implications of answer type impact are further discussed in the next section.

Discussion

The research reported here articulates a methodological approach to examining hierarchical relationships between qualitatively different components in a construct map, with a focus on iteratively improving the construct map specifications. The fourth and final step in this approach is reexamining the construct map, which we do in the next section first generally and then situated in our example of proportional reasoning. This is followed by implications for classroom instruction and assessment development.

Reexamine construct map

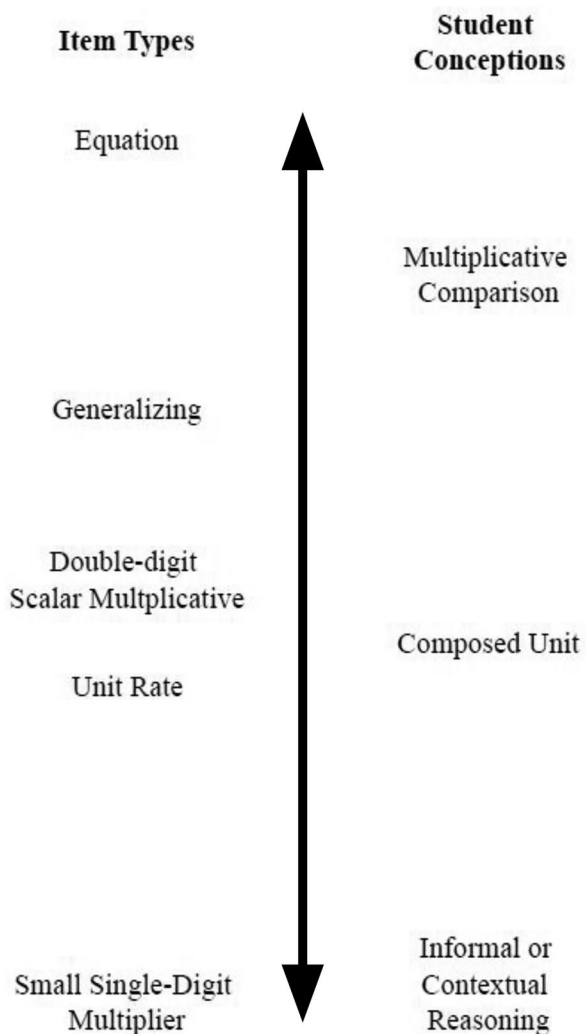
Reexamining the construct map involves determining how well the initial construct map aligns with the results of the item type analysis to determine if the proposed hierarchy aligns with the empirical findings. When findings differ, it is important to try to determine the cause. Is it an issue with the hypothetical model for cognition? Are the item types meaningfully assessing the conception? Or, are there confounding factors impacting the item type means? Determining the cause will help the developer understand if the construct map needs to be changed or if there is an issue with the items. This is a critical aspect of the methodological process and may result in further cycles.

For our proportional reasoning example, the student conceptions side of the construct map stayed the same in terms of ordering, but the item type side was revised to reflect the item type means (see the revised construct map in Fig. 5²). The item type portion of the construct map is further discussed in implications for classroom instruction.

Our findings provide supporting evidence that item types designed to assess students' understanding of the multiplicative comparison conception are more difficult than item types designed to assess students' understanding of the composed unit conception. This provides confirmatory evidence of the placement of these conceptions on the construct map. This finding is not surprising given the general acceptance in the literature of this hierarchical relationship (e.g., Simon & Placa, 2012; Steinhorsdottir & Sriraman, 2009). However, this study is the first we know of to operationalize these conceptions into item types and then examine the associated item difficulties. Given our previous research findings indicating no difference in item difficulties between problems designed to press for use of the scalar and functional relationships in missing-value situations (Carney et al., 2016), the current findings provide further evidence that the difference in difficulty is likely related to students' conceptions—not the mathematical relationships themselves.

² It is important to note our construct map is only one of several potential maps of proportional reasoning that could be developed depending upon the perspective and purpose of the developer. Our model may limit attention to other important aspects of proportional reasoning. We hope that others will attempt to verify aspects, modify the existing map and item type framework, and/or develop their own in the future.

Fig. 5 The final construct map representing the relationship between student conceptions and item types based on the empirical difficulties of the item types



Implications for classroom instruction

Generally, we see the combination of the construct map and item types as useful tools for teachers and curriculum and assessment developers. The hierarchical ordering of the qualitatively different components within the construct map helps to address Shepard's (2018) call for qualitatively meaningful interpretation of quantitative test scores. If used as a formative assessment tool, these item types can assist teachers in determining (roughly) where students are within the map and allow them to track student progress. Curriculum and assessment developers could incorporate the item types and construct map information into the materials, as called for in reviews of these materials (e.g., Dole & Shield, 2008).

More specifically for proportional reasoning, the finding of a significant difference between unit rate and item types designed to measure students' understanding of the

multiplicative comparison conception (generalizing and equation), once answer type bias is controlled for, provides evidence that unit rate solution strategies—while mathematically resulting in the same answer—do not indicate an understanding of the multiplicative comparison conception as others have also suspected, “...we do not assume that the per-one notion of intensive quantities brings with it other important ideas, such as the invariant multiplicative relationship between covarying quantities (Simon & Placa, 2012, p. 39).” Additionally, the significant difference between the generalizing and equation item types, indicates stating the multiplicative comparison relationship in a generalized manner was easier for students than making use of the relationship in an equation. This provides some evidence that the ability to understand slope in an equation from both a composed unit and multiplicative comparison conception, may be contingent upon first meaningfully generalizing the multiplicative comparison relationship.

The findings related to the hierarchical relationships for students’ conceptions and item types have potential implications for classroom instruction that need to be further explored. For example, is the hierarchical relationship between the composed unit and multiplicative comparison conceptions a result of development or instruction? Lo et al. (2004) describe an extensive focus on the multiplicative comparison conception in Asian textbooks. It is possible that for students who experience a different initial instructional experience around ratios, rates, and proportions the conception or item type hierarchies would shift. This would be an important area of research to help delineate the boundaries for application of the construct map and item type framework that also has instructional implications. What does the presence of a hierarchical relationship indicate in terms of classroom instruction? Do teachers need to focus first on one conception and then the other, or is the issue that the multiplicative comparison is not a significant enough focus of instruction (at least in some countries), and therefore, our instruction should more strongly focus on the articulation of the multiplicative comparison conception from the beginning of proportional reasoning instruction? These types of questions are not answered by the present research, but our work could support further research needed to assist teachers in making these types of instructional decisions.

Implications for assessment development

The impact of confounding factors is an important component of this work and has implications for assessment developers. Our results indicate that careful attention to potential confounding variables—such as answer type—is an important component to determining where a topic or conception is “placed” along a score continuum. In other words, if confounding variables are not carefully attended to, the mean item difficulties of items assessing a topic can be impacted. Thus, impacting where the topic is considered to lie along the continuum. It makes one wonder whether large scale, online assessment systems that assign mathematical competencies to student scale scores (e.g., this student can add and subtract rational numbers) have controlled for confounding variables in ways that support their score interpretation claims. We hope that this work encourages assessment developers to carefully attend to this aspect of validity related to score interpretation.

Limitations and future research

The first item presented for each item block was the item type with the easiest item difficulty (small single-digit multiplier) if reading from left to right. The remaining items were placed to the right in a vertical format and could be done in any order. However, if a student moved from the top down within an item block, they would experience more difficult item types in the 2nd and 3rd items (see Appendix 1 for an example). It is possible that presenting the more difficult items earlier in the sequence may have impacted test takers. A future investigation could administer the items in different orders based on predicted difficulty to examine the potential effects of item type order as a confounding variable.

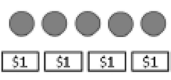
We maintained the contextual nature of the item block by using words instead of letters/symbols within the equation, but it is still possible the more formal/mathematical structure of the equation item types impacted their difficulty. This progression in formality is similar to what occurs in mathematics classrooms. However, it may be useful to develop additional item types that assess students' multiplicative comparison conceptions around the functional relationship to better understand the impact of formal mathematical structures on reasoning.

Conclusion

The National Research Council Committee on the Foundations of Assessment (Pellegrino et al., 2001) called for the development of assessments focused on student cognition several years ago. More recently, we have seen a rise in the use of psychometric models in mathematics education as a tool for both assessment development and to investigate student cognition (cf. Izsak et al. 2016). As noted by Izsak et al. (2016), doing this combined work well requires the pairing of extensive expertise in understanding students' cognition as it relates to particular mathematics domains with measurement expertise. We see this combined expertise as key to the development of high-quality information and materials applicable for use at the classroom level. We also see the need to be pragmatic in the choices we make when conducting this work if the focus is on classroom applicability. High levels of expertise in students' mathematical cognition and/or measurement can lead to relatively detailed or precise models that may be ideal for research purposes but could go beyond what is desired by classroom teachers. Instructional usefulness should be the priority of this work (Pellegrino et al., 2016). For example, while our proportional reasoning domain articulation is relatively simple compared to the extensive research that has been conducted in this area, its focus is on generating information and materials useful to classroom teachers. Therefore, we made some pragmatic choices to simplify our domain articulation for that purpose, such as using language in our construct map that aligns with materials designed for use by middle school teachers around proportional reasoning (Lobato et al., 2010; Olson et al., 2015). Similarly, while our measurement model is relatively simple compared to others, its simplicity makes the interpretation of the construct map and item type framework accessible to a broad audience. Therefore, we see a combination of mathematics education and psychometric evidence as critical to this approach, while at the same time we see a need for a strong focus on instructional relevance.

With regard to the development and articulation of construct maps and item type frameworks, while these may be relatively common processes as steps in an assessment development cycle (Wilson, 2004), their articulation for use by teachers is less well defined. The materials generated by the CGI professional development project provide one example (e.g., Carpenter et al., 2004). We see the need for further investigation related to how well-developed construct maps and item type frameworks can be articulated and presented to ensure their usefulness and accessibility for the classroom teacher and how this type of work can be scaled to large numbers of teachers. A focus on developing and scaling these types of instructional and curricular resources could assist teachers in answering the call to provide instruction focused on building upon students' thinking.

Appendix 1

<p>1a) Callie bought 5 sugar cookies for \$4. How many cookies can she buy with \$8?</p>  <p>Answer: _____</p>	<p>If the relationship between cookies and cost (\$) remains the same no matter how many cookies you buy:</p> <p style="text-align: center;"><i>5 sugar cookies for \$4</i></p> <p>Complete the following statements about the relationship between cookies and cost (\$).</p> <p>1b) number of cookies = _____ * cost * indicates multiplication</p> <p>1c) The cost is always _____ times the number of cookies.</p> <p>1d) One cookie costs _____</p> <p>1e) It would cost _____ to buy 75 cookies.</p>
<p>2a) Jason can hike 5 miles in 2 hours. How far can Jason hike in 6 hours?</p> <p>Answer: _____</p>	<p>If the relationship between miles and hours remains the same no matter how many miles Jason hikes:</p> <p style="text-align: center;"><i>5 miles in 2 hours</i></p> <p>Complete the following statements about the relationship between miles and hours.</p> <p>2b) hours = _____ * number of miles * indicates multiplication</p> <p>2c) The number of miles is always _____ times the hours.</p> <p>2d) In one hour, Jason hikes _____ miles.</p> <p>2e) It would take _____ hours to hike 80 miles.</p>

Appendix 2

The table below presents the four different ratios and three different units used in the construction of the items from the item types. Each row in the table represents one item block.

Items Types										
Ratio	Units	Small Single-Digit Multiplier			Double-Digit Scalar Multiplier			Unit Rate	Equation	Generalizing
5:4	Discrete: Cookies & Cost	Cookies (c)	5		Cookies (c)	5	75	C÷\$ = 1.25	C= 1.25*\$	\$ is 0.8of C
		Cost (\$)	4	8	Cost (\$)	4				
	Scalar x2			Scalar x15						
	Continuous Different Units: Miles & Hours	Miles (m)	5		Miles (m)	5	90	M÷H = 1.25	H= 0.8*M	M is 1.25of H
		Hours (h)	4	16	Hours (h)	4				
	Scalar x4			Scalar x18						
5:2	Continuous Same Units: Red and White Paint	Red (r)	5	20	Red (r)	5	75	R÷W = 1.25	W = 0.8*R	R is 1.25of W
		White (w)	4		White (w)	4				
	Scalar x4			Scalar x15						
	Discrete: Cookies & Cost	Cookies (c)	5	20	Cookies (c)	5	90	\$÷C= 0.4	\$= 0.4*C	C is 2.5of \$
		Cost (\$)	2		Cost (\$)	2				
	Scalar x4			Scalar x18						
6:3	Continuous Different Units: Miles & Hours	Miles (m)	5		Miles (m)	5	80	M÷H= 2.5	H= 0.4*M	M is 2.5of H
		Hours (h)	2	6	Hours (h)	2				
	Scalar x3			Scalar x16						
	Continuous Same Units: Red and White Paint	Red (r)	5	20	Red (r)	5	90	R÷W = 2.5	W= 0.4*R	R is 2.5of W
		White (w)	2		White (w)	2				
	Scalar x4			Scalar x18						
Ratio	Units	Small Single-Digit Multiplier			Double-Digit Scalar Multiplier			Unit Rate	Equation	Generalizing
6:3	Discrete: Cookies & Cost	Cookies (c)	6		Cookies (c)	6	90	\$÷C= 0.5	C=2*\$	\$ is 0.5of C
		Cost (\$)	3	2	Cost (\$)	3				
	Functional x2			Scalar x15						
	Continuous Different Units: Miles & Hours	Miles (m)	6	8	Miles (m)	6	90	M÷H=2	H= 0.5*M	M is 2 of H
		Hours (h)	3		Hours (h)	3				
	Functional ÷2			Scalar x18						
12:3	Continuous Same Units: Red and White Paint	Red (r)	6	4	Red (r)	6	108	W÷R= 0.5	W= 0.5*R	R is 2 of W
		White (w)	3		White (w)	3				
	Functional x2			Scalar x18						
	Discrete: Cookies & Cost	Cookies (c)	12		Cookies (c)	12		\$÷C= 0.25	C=4*\$	\$ is 0.25of C
		Cost (\$)	3	2	Cost (\$)	3	45			
	Functional x4			Scalar x15						
12:3	Continuous Different Units: Miles & Hours	Miles (m)	12		Miles (m)	12		H÷M= 0.25	H= 0.25*M	M is 4 of H
		Hours (h)	3	2	Hours (h)	3	45			
	Functional x4			Scalar x15						
	Continuous Same Units: Red and White Paint	Red (r)	12		Red (r)	12		W÷R= 0.25	R=4*W	W is 0.25of R
		White (w)	3	2	White (w)	3	45			
	Functional x4			Scalar x15						
							R÷W=4	W= 0.25*R	R is 4 of W	

The grayed-out portions of the table indicate those items that were removed from analysis due to their whole number answers.

Availability of data and material Not available due to IRB limitations.

Code availability Not applicable.

Declarations

Ethics approval Reviewed and exempted from further review by our institutional review board.

Consent to participate Informed consent is not required due to data not including specific identifiers. Not required by our institutional review board.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akatufba, A. H., & Wallace, J. (1999). Mathematical dimensions of students' use of proportional reasoning in high school physics. *School Science & Mathematics*, 99(1), 31–41.
- Andrich, D., De Jong, J., & Sheridan, B. E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 59–70). New York, NY: Waxmann.
- Behr, M. J., Harel, G., Post, T., & Lesh, R. (1992). Rational number, ratio, and proportion. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 296–333). McMillan.
- Bolt, D. M., Kim, J.-S., Blanton, M., & Knuth, E. (2016). Applications of item response theory in mathematics education research. In A. Izsak, J. Remillard, & J. Templin (Eds.), *Psychometric methods in mathematics education: Opportunities, challenges, and interdisciplinary collaborations* (pp. 31–52). National Council of Teachers of Mathematics.
- Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.
- Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, 44(5), 1478–1490.
- Bush, S. B., & Karp, K. S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior*, 32(3), 613–632.
- Callingham, R., & Bond, T. (2006). Research in mathematics education and Rasch measurement. *Mathematics Education Research Journal*, 18(2), 1–10.
- Callingham, R., & Siemon, D. (2021). Connecting multiplicative thinking and mathematical reasoning in the middle years. *The Journal of Mathematical Behavior*, 61, 1–12.
- Carney, M., & Crawford, A. (2016). Students' Reasoning Around The functional relationship in proportional situations. In M. Wood, E. Turner, M. Civil, & J. Eli (Eds.), *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 187–190). Tucson, AZ: The University of Arizona.
- Carney, M., Hughes, G., Brendefur, J., Crawford, A., & Totorica, T. (2015). Conducting classroom assessment based in cognition: The case of proportional reasoning. In C. Suurtamm (Ed.), *NCTM*

- Annual Perspectives in Mathematics Education 2015: Assessment to enhance learning and teaching* (pp. 209–218). Reston, VA: National Council for Teachers of Mathematics.
- Carney, M., Paulding, K., & Champion, J. (2022). Efficient assessment of students' proportional reasoning. *Applied Measurement in Education*, 35(1), 1–17.
- Carney, M., Smith, E., Hughes, G., Brendefur, J., & Crawford, A. (2016). Influence of proportional number relationships on item accessibility and students strategies. *Mathematics Education Research Journal*, 28(4), 503–522.
- Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. (2004). *Children's mathematics: Cognitively Guided Instruction*: Heinemann.
- Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E., & Empson, S. B. (1998). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education*, 29(1), 3–20.
- Carraher, D. W., Martinez, M. V., & Schliemann, A. D. (2008). Early Algebra and Mathematical Generalization. *ZDM Mathematics Education*, 40(1), 3–22.
- Cengiz, N., & Rathouz, M. (2018). Making sense of equivalent ratios. *Mathematics Teaching in the Middle School*, 24(3), 148–155.
- Cetin, H., & Ertekin, E. (2011). The relationship between eighth grade primary school students' proportional reasoning skills and success in solving equations. *International Journal of Instruction*, 4(1), 47–62.
- Confrey, J., McGowan, W., Shah, M., Belcher, M., Hennessey, M., & Maloney, A. (2018). Using digital diagnostic classroom assessments based on learning trajectories to drive instruction and deepen teacher knowledge. In D. Siemon, T. Barkatas, & R. Seah (Eds.), *Researching and using learning progressions (trajectories) in mathematics education in global education in the 21st century series*. Leiden, the Netherlands: SENSE Publishers.
- Dole, S., & Shield, M. (2008). The capacity of two Australian eighth-grade textbooks for promoting proportional reasoning. *Research in Mathematics Education*, 10(1), 19–35.
- Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching*, 46(6), 606–609.
- Ellis, A. (2013). Research brief: Teaching ratio and proportion in the middle grades. Retrieved from Reston, VA: <http://www.nctm.org/news/content.aspx?id=35822>
- Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V., & Emson, S. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, 27(4), 404–434.
- Fernández, C., Llinares, S., Van Dooren, W., De Bock, D., & Verschaffel, L. (2011). Effect of number structure and nature of quantities on secondary school students' proportional reasoning. *Studia Psychologica*, 53(1), 69–81.
- Foster, M. E., Anthony, J. L., Clements, D. H., Sarama, J., & Williams, J. M. (2016). Improving mathematics learning of kindergarten students through computer-assisted instruction. *Journal for Research in Mathematics Education*, 47(3), 206–232.
- Gravemeijer, K. (2004). Local instruction theories as means of support for teachers in reform mathematics education. *Mathematical Thinking and Learning*, 6(2), 105–128.
- Harel, G., & Behr, M. (1989). Structure and Hierarchy of Missing Value Proportion Problems and Their Representations. *Journal of Mathematical Behavior*, 8(1), 77–119.
- Izsak, A., Remillard, J., & Templin, J. (Eds.). (2016). *Psychometric methods in mathematics education : opportunities, challenges, and interdisciplinary collaborations*. National Council of Teachers of Mathematics.
- Jeong, Y., Levine, S. C., & Huttenlocher, J. (2007). The development of proportional reasoning: Effect of continuous versus discrete quantities. *Journal of Cognition and Development*, 8(2), 237–256.
- Johnson, H. L. (2015). Secondary students' quantification of ratio and rate: A framework for reasoning about change in covarying quantities. *Mathematical Thinking and Learning: An International Journal*, 17(1), 64–90.
- Karplus, R., Pulos, S., & Stage, E. K. (1983). Early adolescents' proportional reasoning on 'rate' problems. *Educational Studies in Mathematics*, 14(3), 219–233.
- Lamon, S. (2005). *Teaching Fractions and Ratios for Understanding: Essential Content Knowledge and Instructional Strategies for Teacher, Second Edition*. Taylor & Francis.
- Lamon, S. (2007). Rational numbers and proportional reasoning: Toward a theoretical framework for research. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 629–667).

- Lawton, C. A. (1993). Contextual factors affecting errors in proportional reasoning. *Journal for Research in Mathematics Education*, 24(5), 460–466.
- Lesh, R., Post, T., & Behr, M. (1988). Proportional reasoning. In J. E. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (Vol. 2, pp. 93–118).
- Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best?. *Journal of Outcome Measurement*, 2, 266–283.
- Linacre, J. M. (2016). Winsteps Rasch Measurement (Version 3.92.1). Retrieved from www.winsteps.com
- Lo, J.-J., Watanabe, T., & Cai, J. (2004). Developing ratio concepts: An Asian perspective. *Mathematics Teaching in the Middle School*, 9(7), 362–367.
- Lobato, J., Ellis, A. B., & Charles, R. I. (2010). *Developing essential understanding of ratios, proportions, and proportional reasoning for teaching mathematics in grades 6–8*. Reston, VA: National Council of Teachers of Mathematics.
- Long, C., Wendt, H., & Dunne, T. (2011). Applying Rasch measurement in mathematics education research: Steps towards a triangulated investigation into proficiency in the multiplicative conceptual field. *Educational Research and Evaluation: An International Journal of Theory and Practice*, 17(5), 387–407.
- Misailidou, C., & Williams, J. (2003a). Children's proportional reasoning and tendency for additive strategy: The role of models. *Research in Mathematics Education*, 5(1), 215–247.
- Misailidou, C., & Williams, J. (2003b). Diagnostic assessment of children's proportional reasoning. *The Journal of Mathematical Behavior*, 22(3), 335–368.
- Noelting, G. (1980a). The development of proportional reasoning and the ratio concept Part I–Differentiation of stages. *Educational Studies in Mathematics*, 11(2), 217–253.
- Noelting, G. (1980b). The development of proportional reasoning and the ratio concept Part II–Problem-structure at successive stages; problem-solving strategies and the mechanism of adaptive restructuring. *Educational Studies in Mathematics*, 11(3), 331–363.
- Olson, T., Olson, M., & Slovin, H. (2015). *Putting essential understanding of ratios and proportions into practice in grades 6–8*. National Council for Teachers of Mathematics.
- Pellegrino, J. W., Chudowsky, N., Glaser, R., & National Research Council Committee on the Foundations of Assessment. (2001). *Knowing what students know : the science and design of educational assessment*. National Academy Press. Retrieved March 2, 2022, from <https://www.nap.edu/download/10019>
- Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist*, 51(1), 59–81.
- Ramful, A., & Narod, F. (2014). Proportional reasoning in the learning of chemistry: Levels of complexity. *Mathematics Education Research Journal*, 26(1), 25–46. <https://doi.org/10.1007/s13394-013-0110-7>
- Rupley, W. H. (1981). *The effects of numerical characteristics on the difficulty of proportion problems (dissertation)*. University of California.
- Saunders, W. L., & Jesunathadas, J. (1988). The effect of task content upon proportional reasoning. *Journal of Research in Science Teaching*, 25(1), 59–67.
- Schwartz, D. L., & Moore, J. L. (1998). On the role of mathematics in explaining the material world: Mental models for proportional reasoning. *Cognitive Science*, 22(4), 471–516.
- Shepard, L. A. (2018). Learning progressions as tools for assessment and learning. *Applied Measurement in Education*, 31(2), 165–174.
- Siemon, D., Callingham, R., Day, L., Horne, M., Seah, R., Stephens, M., & Watson, J. (2018). From research to practice: The case of mathematical reasoning. In Hunter, J., Perger, P., & Darragh, L. (Eds.). *Making waves, opening spaces: Proceedings of the 41st annual conference of the Mathematics Education Research Group of Australasia*. (pp. 40–49). Mathematics Education Research Group of Australasia Inc.
- Siemon, D., Day, L., Stephens, M., Horne, M., Callingham, R., Watson, J., & Seah, R. (2017). Reframing mathematical futures: Using learning progressions to support mathematical thinking in the middle years. In Downton, A., & Hall, J. (Eds.), *40 years on: We are still Learning! Proceedings of the 40th Annual Conference of the Mathematics Education Research Group of Australia* (p. 650). Melbourne, AU.
- Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, 26, 114–145.

- Simon, M. A., & Tzur, R. (2004). Explicating the role of mathematical tasks in conceptual learning : An elaboration of the hypothetical learning trajectory. *Mathematical Thinking and Learning*, 6(2), 91–104.
- Simon, M., & Placa, N. (2012). Reasoning about intensive quantities in whole-number multiplication? A possible basis for ratio understanding. *For the Learning of Mathematics*, 32(2), 35–41.
- Smith, R. (2002). Understanding Rasch measurement: Detecting and evaluating the impact of multi-dimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 5(4), 430–448.
- Steinhorsdottir, O. B., & Sriraman, B. (2009). Icelandic 5th-grade girls' developmental trajectories in proportional reasoning. *Mathematics Education Research Journal*, 21(1), 6–30.
- Tourniaire, F., & Pulos, S. (1985). Proportional reasoning: A review of the literature. *Educational Studies in Mathematics*, 16(2), 181–204.
- Tucker, M., Daro, P., Snow, C., Pellegrino, J., Everson, H., Glasper, R., Mannes, K., Fain, P. (2013). What does it really mean to be college and work ready?. The mathematics and English literacy required of first year community college students. Retrieved from Washington, DC.
- Vergnaud, G. (1994). Multiplicative conceptual field: What and why. In J. Confrey & G. Harel (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 41–59). State University of New York Press.
- Wilson, M. (2004). *Constructing measures: An item response modeling approach*. New York, Routledge.
- Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, 46(6), 716–730.
- Wolfe, E. W., & Smith, E. V. (2006). Instrument development tools and activities for measure validation using Rasch models: Part II–validation activities. *Journal of Applied Measurement*, 8(2), 204–234.
- Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14, 97–116.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.