

Elementary Children's Multiplicative Reasoning: Initial Validation of a Written Assessment

Karl W. Kosko and Rashmi Singh

Multiplicative reasoning is essential for students' engagement with various mathematical concepts. Although the field's understanding of children's multiplicative concepts has grown over the past 30 years, relatively few studies have examined the development of multiplicative concepts with whole numbers, and even fewer have studied this phenomenon at scale. The present study reports on the development of an assessment of elementary students' multiplicative concepts with whole numbers that can be used at a large scale. Findings suggest the initial version of the assessment has sufficient reliability and validity. Further, less than 20% of second grade students and approximately 50% of third grade students participating in the study engage in tasks with at least the first multiplicative concept.

Multiplicative reasoning is essential for students' development of a meaningful understanding of rational number (Confrey & Harel, 1994; Hackenberg & Tillema, 2009). In addition to other frameworks, it is often described in reference to Hackenberg's (2010) work on three multiplicative concepts in which each subsequent multiplicative concept is characterized by more sophisticated unit coordination. Multiplicative concepts are researchers' models for schemes constructed by students as a product of their prior counting schemes (Steffe, 1994). Each multiplicative concept is associated with a particular way of engaging in activities involving multiplicative relationships. For example, students who have constructed the third multiplicative concept (MC3) are generally more successful in working with

Dr. Karl Kosko is an assistant professor at Kent State University. His research is focused on how psychological and social factors relate to teacher facilitation of and student engagement in conveying mathematical meaning through discussions, argumentative writing, and other forms of engagement.

Rashmi Singh is a doctoral student at Kent State University. Her current research interests include (a) the study of students' and teachers' conceptions of mathematical ideas and (b) utilizing students' mathematical writing as a probe to identify potential areas for improvement.

more sophisticated fraction tasks than students who have constructed the second (MC2) or first (MC1) multiplicative concepts (Hackenberg & Tillema, 2009; Norton & Hackenberg, 2010).

Although research on multiplicative reasoning since the 1980s has seen many advances in the field's understanding of this topic (Confrey & Harel, 1994), the authors of the present study struggled to find an assessment that allowed for studying, at a larger scale, elementary students' holding of various multiplicative concepts (particularly with whole numbers). Specifically, there are studies examining whether, and in what ways, students respond to canonical representations of school-based mathematics such as word problems and symbolic representations (Brickwedde, 2011; English, 1991; Mulligan & Mitchelmore, 1997; Smith & Smith, 2006). Yet, such assessments generally require interviews with students to gauge multiplicative reasoning. Other assessments of unit coordination and/or multiplicative reasoning are being developed but typically have been used with middle grades students (i.e., Norton, Boyce, Phillips, et al., 2015). The lack of an instrument that can assess larger groups of elementary students leads to a lack of understanding the prevalence of different forms of multiplicative reasoning among students in these particular grade levels, as well as assessing such reasoning in relation to other mathematical constructs. To meet these needs, the present study reports on the development and initial piloting of an assessment of second- and third-grade students' unit coordination in association with demonstrated schemes for multiplicative reasoning in activity. Therefore, it is the purpose of the present study to report on efforts to create and validate an initial version of an assessment of elementary children's unit coordination with tasks eliciting multiplicative reasoning (hereafter, multiplicative tasks), and provide initial statistics regarding the prevalence of the different multiplicative concepts in elementary grades.

Theoretical Framework

Schemes, Operations, and Concepts

The present study is situated in scheme theory, with particular attention to students' whole number multiplicative concepts (Olive, 2001; Steffe, 1992, Steffe, 1994). A scheme is a set of mental actions, or operations, with a specific goal-oriented purpose (von Glasersfeld, 1995). For the purposes of the present study, the operations of unitizing, iterating, partitioning, and disembedding are of particular focus. Unitizing in multiplicative contexts involves conveying a collection or set of discrete countable objects as a countable unit in its own right (Steffe, 1994). Iterating involves the repetition of a unit, while partitioning involves the separation of a unit into iterable parts. Disembedding involves taking a composite unit (unit of units) from another larger unit in such a way as considering it as part of that larger unit (Confrey & Harel, 1994; Hackenberg, 2010; Steffe, 1992). According to Steffe (1992; 1994) the coordination of these actions in counting schemes can lead to the construction of schemes that are multiplicative in nature. When such schemes are reversible, in that an individual is able to return to the starting point of a scheme either by inversion or compensation, they are said to have become interiorized (Steffe, 1992; Hackenberg, 2010). Schemes that are interiorized are generally used in an anticipatory manner (i.e., anticipatory schemes) that allows for aspects of the constructed scheme to be taken as given and operated on without need for reconstruction in a similar context.

Multiplicative concepts involve the anticipatory use of multiplicative schemes (Hackenberg, 2010). Although schemes are not necessarily anticipatory, a multiplicative concept uses the taken-as-given involvement of prior constructed schemes and the application of such schemes to specific situations they are called for. Specifically, "concepts involve reversibility in that from the interiorized results of a scheme, one can always go back to the situation or activity of that scheme" (Hackenberg, 2010, p. 389). In this manner, concepts are considered an abstraction of an anticipatory scheme.

Multiplicative Concepts

Steffe (1994) noted that various pre-multiplication schemes may be used, in activity, prior to students' construction of multiplicative schemes. Such schemes include variations of count-by-1s strategies, in which abstracting a set of singular objects as representing countable composite units has not yet been done successfully (i.e., the schemes are not used in an anticipatory manner). Students may be able to solve problems considered as multiplicative by constructing such schemes in activity. Steffe (1994) described one such student who demonstrated counting a group of 18 by considering it as having six groups of three, but effectively counting by 1s in doing so. Steffe argued, however, that the student did not coordinate the unit of 3 in a multiplicative manner. Rather, the student created the composite unit of 3 in activity such that he did not anticipate or act upon 3 as a composite unit to find 18. According to Steffe (1994), this student's scheme is pre-multiplicative because, although he accounted for groups of 3 in his counting, he effectively counted by 1s in doing so. A second student described by Steffe (1994) also included what might appear as count-by-1 strategies, but did so in considering the composite unit as an organizing feature of how they counted (1, 2, 3; 4, 5, 6; 7, 8, 9; ... and so on). Such coordination anticipates a unit of units (i.e., two levels of units) and is therefore multiplicative.

The first multiplicative concept (MC1) involves the coordination of two levels of units in activity (Hackenberg & Tillema, 2009). Specifically, students can take a composite unit as given and coordinate between two levels of units (Norton, Boyce, Ulrich, & Phillips, 2015). Solving 4×5 , a student at MC1 may count by 1s to 20 by keeping track of how many counts to 5 they made (e.g., "1, 2, 3, 4, 5; 6, 7, 8, 9, 10; ..."). Alternatively, some students may use skip counting approaches to count "5, 10, 15, 20" (Steffe, 1994). In both instances, the unit of 5 is taken as given and operated on, in activity, to coordinate these composite units.

MC2 involves the coordination of three levels of units in activity (Hackenberg & Tillema, 2009), in which students can take two levels of units as given and coordinate between three

levels of units in activity (Norton, Boyce, Ulrich, et al., 2015). Students at MC2 are able to operate upon composite units to form a unit of units of units and can disembed parts of a unit to operate on smaller units within a larger composite unit (Tillema, 2013). Students at MC2 can solve tasks similar to asking how many more sets of 5 does 50 have than 20. A student at MC2 may disembed 5 from 20, in activity, and operate on the remaining 30 ($50 - 20$) to determine there are six more groups of 5 in 50 than 20. This manipulation of the composite unit 5 takes as given at least two levels of units (coordinated between three levels of units). Specifically, MC2 is characterized by the anticipatory use of two levels of units. Although they can construct schemes in activity for coordination of three levels of units, as with any scheme constructed “in activity,” such schemes are not guaranteed to be constructed by a student at MC2 (i.e., they are probable).

MC3 involves the coordination of three levels of units, as does MC2 (Hackenberg & Tillema, 2009; Olive, 2001). However, all levels of units are interiorized. This allows for greater flexibility. For example, students may recognize, via anticipatory schemes, that 20 includes 4 groups of 5, and connect this with 60 making 12 groups of 5, or 3 groups of 4 groups of 5 units. Such flexibility enables students at MC3 to consider more than one composite unit to be used in given circumstances (i.e., is it more useful to consider 60 from the perspective of the 4 unit or the 12 unit?). A summary of all multiplicative concepts is provided in Table 1.

Each multiplicative concept involves reversibility of interiorized schemes (Hackenberg, 2010). However, students who have not interiorized schemes are capable of solving tasks via reversible schemes constructed in activity. For example, Steffe (1992) observed that Maya, a student able to count upwards by 3s (i.e., MC1), was able to use a counting-down-from strategy to determine how many 3s were in 12. Yet, this latter strategy was constructed in activity. Similarly, Hackenberg (2010) noticed that a sample of students at MC2 and MC3 were both able to solve reversible multiplicative tasks, but students at MC3 did so with anticipatory schemes while some students at MC2 did so by constructing schemes within the activity. By

accounting for whether such schemes are anticipatory or constructed in activity, various qualitative studies have distinguished between specific multiplicative schemes and concepts (Hackenberg, 2010; Norton, Boyce, Ulrich, et al., 2015; Steffe, 1992; Steffe, 1994). Given the focus of the present study (i.e., piloting a one-time written assessment at scale), the ability to distinguish between whether demonstrated schemes were anticipatory or constructed in activity could not be assessed directly or in as nuanced a manner as has been done in various teaching experiments. This issue is discussed in some detail in descriptions of item design.

Table 1

Levels of multiplicative reasoning aligned with multiplicative concepts and scheme theory

Construct	Description
Pre-multiplicative Schemes	Students rely on count-by-1s strategies and do not coordinate two levels of units in activity.
First Multiplicative Concept (MC1)	Students coordinate two levels of units in activity by anticipating one level of units (either via skip-counting or count-by-1s strategies).
Second Multiplicative Concept (MC2)	Students coordinate three levels of units in activity, but anticipate two levels of units.
Third Multiplicative Concept (MC3)	Students coordinate three levels of units via anticipatory schemes.

Need for and Needs of a Written Assessment

Conceptualization for Item Design

Unit coordination is fundamental for interpreting students' multiplicative concepts (Hackenberg, 2010; Steffe, 1994). Researchers studying elementary students' unit coordination in multiplicative tasks have found that most third- and fourth-grade students use variations of skip-counting strategies (Brickwedde, 2011; English, 1991; Mulligan & Mitchelmore, 1997). Further, Steffe (2017) has estimated that perhaps as many as half of students entering middle school operate with iterating multiplicative schemes. Such estimates are difficult to evaluate

because the vast majority of assessments for multiplicative reasoning include interview protocols, which are less likely to be administered to larger numbers of students (thereby excluding potential diverse populations with representative samples). Norton, Boyce, Phillips, et al. (2015) provide one exception regarding sample size, but even their sample size is limited ($n = 20$). Such assessments (Brickwedde, 2011; English, 1991; Mulligan & Mitchelmore, 1997; Smith & Smith, 2006) provide more information than a written assessment not incorporating an interview, but they are unable to collect data at scale pragmatically. A scalable, written assessment potentially allows for a more accurate estimate of the prevalence of students operating with particular schemes or concepts and enables researchers to examine effects associated with such operations at a larger scale. The latter was our own need, which led to the development of the present study. Specifically, we were interested in how elementary students' multiplicative unit coordination associated with other observable mathematical activities such as their conception of the equals sign (Singh & Kosko, 2016) and their engagement in mathematical argumentative writing (Kosko & Singh, 2016). Finding no readily available instrument that could be administered at scale, we constructed and piloted the one described in this paper. Yet, construction of items, as well as an overall assessment, of multiplicative concepts holds specific challenges, and a discussion of what information such items can provide is warranted.

The distinction of anticipatory schemes versus schemes constructed in activity allows for conceptually distinguishing between the multiplicative concepts (Hackenberg, 2010). Yet, such a distinction is less clear when examining students' work from only a paper-based assessment. Within a written assessment, one can ask students to show their work, but such work acts as an artifact and does not allow for an accurate distinction between a scheme that is anticipatory versus one constructed in activity. Further, follow-up prompts that might be asked during a face-to-face interview to clarify written work are necessarily absent from written assessments, and the detail provided by one child may not be provided by another. In the

end, the answer provided by various children completing a task on a written assessment administered at-scale is the most consistent indicator of their engagement with the mathematics in that task. If the tasks are designed properly, probabilistic indicators can estimate the schemes students most likely will have used. Further, when multiple tasks are provided that target specific schemes (similar to multiple sources of data), then the consistency of the child's responses serve as an indicator of whether the tasks are reliably assessing the schemes targeted. A necessary assumption, and one taken in the present paper, is to assume that such indications represent *at least* a scheme constructed in activity for a given task/item, but *not necessarily* an anticipatory scheme. In other words, a student may complete certain items successfully, but this success is considered as evidence of at least constructing the targeted scheme in activity. Evidence of schemes being anticipatory is provided through response to other items (i.e., items assessing reversibility of the targeted scheme). Additional evidence towards the validity of these interpretations was examined through students' written work and is discussed in depth in the methods section.

Considering the response a child provides as an indicator of the scheme they used requires an assumption that some children who use schemes not targeted by a task may still find a correct solution. The literature on multiplicative reasoning implies that students at lower multiplicative concepts may solve tasks identified to align more with higher multiplicative concepts (Hackenberg, 2010; Steffe, 1994). This typically occurs with students that are at adjacent multiplicative concepts, in terms of hierarchy (i.e., a child at MC1 may successfully solve a task that targets MC2). In order to conceptualize the kinds of items needed for an assessment of multiplicative reasoning, we considered items in terms of whether students operating at different multiplicative concepts could, hypothetically, successfully solve particular tasks. Thus, we conceptualized that certain tasks should be successfully solvable using similar approaches by students at MC1, MC2, or MC3, but not all items solvable by a student at MC2 should be able to be solved by students at MC1, or not probabilistically so. For example, students at pre-multiplication or MC1 levels have generally not

been observed to disembed units (Hackenberg & Tillema, 2009). However, it is feasible that students at MC1 may enact strategies that allow them to solve tasks that are designed to include disembedding. While it may be possible for students at MC1 to solve such tasks, we hypothesize that it is less probable for a student at MC1 to do so than a student at MC2. Further, we consider such probabilistic tendencies to occur across multiplicative concepts regarding tasks of varying sophistication.

Figure 1 illustrates a generalization of our conception across multiplicative concepts. The circles in Figure 1 qualitatively represent students at varying levels capable of solving tasks at those levels, in activity. In other words, students at pre-multiplication levels of multiplicative reasoning may use counts of 1 in activity, and students at MC1 may coordinate two levels of units in activity. However, our conceptualization of items assesses students' completion of items from the standpoint of viewing successful completion as evidence towards *at least* a constructed scheme in activity, but *not necessarily* an anticipatory scheme.

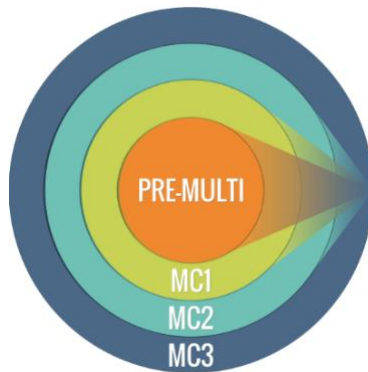


Figure 1. Item conceptualizations for evidence of students' multiplicative reasoning from particular concepts and/or schemes.

The students at the highest level (i.e., MC3) are hypothesized to be able to successfully solve all tasks aligned at MC3 and lower in the hierarchy, with a high degree of probability. In other words, we hypothesize that such students

may answer an item incorrectly, but it is much less likely than for a student at MC2 or MC1. On the other hand, students at lower levels may construct schemes in activity to successfully determine a solution for a task at a higher level. This particular probability is represented by the lightly shaded conical regions originating from the smaller circles and extending to the larger ones (see Figure 1). The areas covered by the conical regions are representative of the hierarchy of the probability students from less sophisticated levels would successfully solve a task at a higher sophisticated level (i.e., students at MC1 are more likely to solve an MC2 designated task than students at pre-multiplication). This conceptualization represents a growing body of evidence of how students might transition, or engage in tasks, across the multiplicative concepts (Norton, Boyce, Ulrich, et al., 2015; Tzur et al., 2012).

Important in conceptualizing how our items align with constructed schemes in activity, is the contexts associated with those items. We took inspiration from other written assessments of schemes and unit coordination (Izsák, Jacobson, & de Araujo, 2012; Norton, Boyce, Phillips, et al., 2015; Wilkins, Norton, & Boyce, 2013). Specifically, both Izsák et al. (2012) and Wilkins et al. (2013) incorporated a number of different visual models for fractions (i.e., length models and area models). While Izsák et al. (2012) and Wilkins et al. (2013) also included contexts with their tasks (e.g., drawing a fraction of a pie, or cutting a candy bar), we elected to design our multiplicative tasks without such contexts. Since our assessment targeted elementary aged students, we sought to create tasks that were as straightforward as possible. Similar to Norton, Boyce, Phillips, et al. (2015), we focused only on one particular visual representation: length models. However, the nature of our task prompts differ from Norton, Boyce, Phillips, et al.'s (2015) given the difference in targeted population (middle grades versus elementary grades students). I and Dougherty (2014) argued that length models representing continuous quantities allow for a multitude of multiplicative relationships and operations to be examined, thus allowing for versatile design of items assessing pre-multiplicative schemes, MC1, and MC2. Furthermore, because incorporation of different models may lead to different enacted

schemes by students (Eliustaoglu, 2016), we viewed the use of a singular model in our assessment as limiting the amount of unnecessary variance in item responses. The tradeoff in focusing on length models is that the schemes assessed are currently limited to this particular model and additional models should be included at some point in the assessment's development. Since we focused on the multiplicative concepts of second and third grade students, and the vast majority of studies have identified a high prevalence of iterating multiplicative approaches (Brickwedde, 2011; English, 1991; Mulligan & Mitchelmore, 1997), we chose not to design many items for MC3 (although some were included to test assumptions regarding reversibility multiplication tasks).¹

Design of the Items

The assessment was designed to determine if students demonstrated approaches to multiplication which were representative of certain concepts (i.e., pre-multiplication, MC1, MC2). We constructed six types of items to assess these concepts, with particular emphasis on assessing MC1 and MC2 (12 items total). All items were open-response in that we asked students to provide the correct response given a blank space. The first type of item required participants to iterate units of 1 to find a composite (see Figure 2). Successful completion of these items was hypothesized to align with at least those students at pre-multiplication who construct such schemes in activity (Steffe, 1994). Specifically, students using pre-multiplicative schemes may be able to count by 1s to find a longer length, but they also may not. In the item presented in Figure 2, we hypothesized that a student need use only a pre-multiplicative scheme to iterate 1s to find the total length of 5 (but they may also use more sophisticated schemes). Further, we hypothesized that this particular pre-multiplicative scheme is a scheme that develops prior to development of schemes for MC1.

¹ As a reminder to the reader, our focus on second and third grade relates to a separate study, for which the present assessment was designed.

The second item type involved reversibility of the first item type by finding a unit of 1 from a composite length. Successful completion of the second item type was hypothesized to be evidence of students at MC1. Specifically, such a student would need to consider the item in Figure 2 as illustrating a length of 4 (one unit) but also simultaneously illustrating a length of four 1s (second unit). We hypothesized Item Type 2 as eliciting an anticipatory scheme of one level of units (i.e., consider the given unit as a composite of 1s in order to partition it into 1s). The third item type was hypothesized as providing evidence of in-activity construction for students at least at MC1, as it requires taking a composite non-1 unit as given and iterating that unit to find the total length. Figure 2 illustrates the example task requiring the iteration of length 5 three times to find the total length of 15. A student could, hypothetically, consider five 1s for each length 5, but this is also considered as evidence towards MC1. It is also possible, but less likely, that a student could find the total length of 15 by using a pre-multiplicative scheme (not accounting for iterable 5s), and we account for this less probable possibility in our test design (see Figure 1).

The fourth and fifth item types shown in Figure 2 were hypothesized as providing evidence for at least MC2. The reversibility variant (item type four) requires considering the given rod is 8 long as a unit of units that can be partitioned into parts with lengths other than 1 (i.e., partitioning into four parts to find length 2). We hypothesized this as requiring an anticipatory scheme for coordinating two levels of units, with coordination of three levels being potentially constructed in activity. The fifth item type example in Figure 2 allows for a student to successfully complete the item by disembedding a unit of 1 from 2 to find the total length of 3 for the second length. Specifically, we hypothesized that the student must anticipate 2 as a unit, and also anticipate that 2 contains two 1s that can be operated on.

The sixth item type was a variant of the disembedding item type for MC2 (Item Type 5). The example provided in Figure 2 requires anticipating 24 as a unit that can contain units of units. Thus, one approach to solving the task is to partition 24 into two equal parts (12 and 12), and then compare the 12 partition to the

unshaded length. Recognizing that 12 is four parts of a five part length requires at least two levels of unit coordination (disembedding similar to Item Type 4), and therefore we hypothesized such items would require anticipatory unit coordination of three levels of units. Although we were not seeking to assess MC3 with the present assessment, we included such items to determine the prevalence of students we hypothesized to be at MC2 (from responses to the assessment) able to successfully complete such items. Additionally, responses for MC3 items are useful in conceptualizing future versions of the assessment for higher grade levels that would assess MC3. Two forms (A and B) of the assessment were created with the only difference being the order of the items presented.

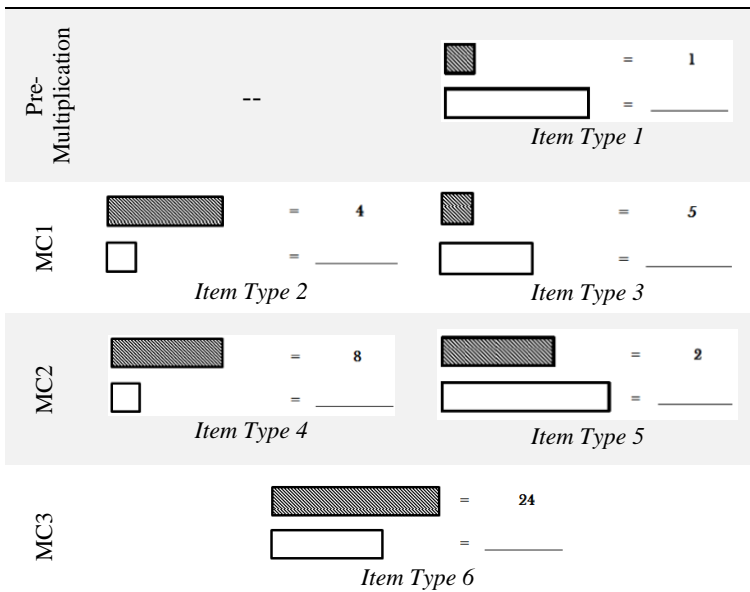


Figure 2. Example item types for the assessment following the prompt: “Given how long the shaded shape is, tell how much the unshaded shape is.”

The design of the items and their hypothesized conceptual placements facilitate the purpose of the present study: to create and pilot an assessment of elementary children’s unit

coordination with multiplicative tasks. The design principles hold some similarities to other assessments that have examined unit coordination, such as Norton, Boyce, Phillips, et al.'s (2015) assessment of middle grades students' multiplicative unit coordination, Wilkins et al.'s (2013) assessment of middle grade students' unit coordination with fractions, and Izsák et al.'s (2012) assessment of teachers' unit coordination with fractions. Important in the framing of the present study is an understanding of how various projects on assessment development (particularly those which focus on unit coordination) have reported different stages of their work. For example, Izsák et al. (2012) used mixture Rasch models for studying different subgroups within the targeted population. Izsák et al.'s (2012) work built upon several prior studies of teachers' mathematical knowledge for teaching, as well as a validated version of the instrument by Izsák, Orrill, Cohen, and Brown (2010). By contrast, Norton, Boyce, Phillips, et al. (2015) report on the validation of a rubric to be used in scoring a written assessment, which is arguably an initial step before piloting with a larger sample. The present study is closer in stage of development to that of Norton, Boyce, Phillips, et al. (2015), with a focus on classical test theory. However, future work on the assessment will likely resemble the kinds of work done by others at later stages of development (i.e., Izsák et al., 2012). In the next section, the work of initially examining validity and reliability of the assessment are discussed.

Methods

Sample

In May 2015, 163 second- and third-grade students in two different school districts completed the assessment. Both suburban school districts are located in the Midwestern U.S. The sample is comparable regarding gender (49.1% male; 50.9% female), grade level (44.2% second-grade; 55.8% third-grade), and form administered (49.7% Form A; 50.3% Form B). Data for the assessment was coded in two ways. First, we coded student responses for correctness (0=incorrect; 1=correct).

Second, we coded students' demonstrated written strategies so that we could examine how such demonstrated schemes corresponded to correct responses by item. In the next section, we describe the second coding in detail.

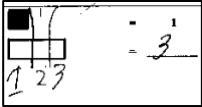
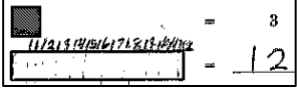

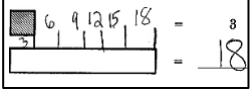

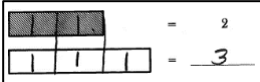
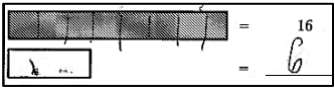


Coded Schemes

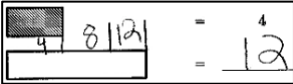

We used an iterative coding process to develop a rubric assessing the demonstrated work of students completing the assessment. Each author surveyed students' responses independently to identify reoccurring student strategies in the data. The authors discussed identified strategies and created an initial rubric describing different applications of schemes. Although numerous specific strategies were identified through students' written work, we consolidated several strategies as applying specific schemes and this resulted in eleven specific codes including six targeted schemes and five "other" categorized codings. All codings are listed in Table 2, with illustrated examples provided for the targeted schemes.

Each code is considered as signifying at least a scheme constructed in activity, but not necessarily an anticipatory scheme (although such schemes are possibly anticipatory). Important to note is that while coding of schemes for individual items was considered evidence of it being constructed at least in activity, we describe the relation between different items as evidence towards establishing whether schemes are anticipatory later in this paper in our description of indices. Code 1 signifies a pre-multiplication scheme where students demonstrated iterating units of $1/n$ times to find a length/whole (a composite). As shown in Table 2, students might use such a scheme in a context more probable to result in a correct response (the first example), but may also apply it in cases where other strategies are more probable to lead to success. The second example for code 1 shows a child iterating 1 twelve times to find the total length. The length is actually 18, but the design of the item makes application of this pre-multiplication scheme have a less probable chance for success (in finding the correct length). Although we do not provide such examples for all codes, we noted such use of strategies for various items, and we discuss

their alignment (or non-alignment) with targeted strategies and schemes for items later in this paper.

Table 2
Examples of coding for demonstrated schemes

Construct	Code	Example
Pre-Mult	1. Iterating 1 units n times	
		
MC1	2. Partitioning into n parts to find 1 units	
	3. Iterating non-1 units n times	
MC2	4. Partitioning into n parts to find non-1 units	
	5. Disembedding a unit to iterate n times	
MC3	6. Decompose partitions into non-1 units (may include coordination of partitions in both length models)	
Other	7. Finding a 1 unit	
	8. Using an inappropriate non-1 unit	

9. Disembedding a unit but not iterating appropriately	
10. Equipartitioning, halving, or doubling	
11. Not enough information	N/A

For MC1, code 2 signifies reversibility of the scheme for code 1 where an individual demonstrated partitioning into n parts to find 1-units (i.e., partitioning a given length 8 into eight equal parts to designate a unit of 1). Code 3 designated iterating non-1 units n times (i.e., iterating 3 six times to find length 18). For MC2, code 4 designated partitioning into n parts to find non-1 units (i.e., reversibility of code 3). Code 5 designated disembedding a unit from another unit and iterating it n times. The reversibility of code 5 was designated as code 6 and hypothesized to align with MC3. Code 6 designated disembedding partitions into non-1 units that would match a whole. Additional schemes observed included: identifying a length as 1, regardless of partitioning or iterating (code 7); identifying a non-1 unit that was used in a previous item (code 8); disembedding to find an appropriate unit, but not iterating appropriately (code 9); variations of equipartitioning approaches (code 10); and a general code for when no, or insufficient, student work was provided to assign a code (code 11). Although strategies considered to be associated with schemes for specific items were often observed along with correct responses (i.e., code 5 with items I and J), it was not necessary for a correct solution to be associated with evidence for a scheme to be coded.

Both authors coded 10% of the data together to improve the rubric and solicit examples such as the ones presented in Table 2 for each identified scheme. Next, the remaining 90% of data were coded independently to determine interrater reliability. A Cohen’s Kappa statistic was calculated ($K = 0.70, p < 0.001$), indicating substantial reliability (see Landis & Koch, 1977). Following this result, both authors reconciled codes before continuing with data analysis.

Results and Findings

Overall Test Validity and Reliability

We used classical item analysis (CIA) to examine response data (correct/incorrect) as well as the coded schemes (see Crocker & Algina, 2006). CIA uses traditional statistical measures and approaches (associated with classical test theory) to examine the reliability and validity of items within an assessment and the overall assessment score. For the response data, CIA involved determining item difficulty (mean of correct responses), item discrimination (how well the item discriminated between lower and higher scoring participants), and the overall statistical reliability of the responses for measuring the test score (via Cronbach's alpha). A Cronbach's alpha coefficient of .79 was calculated for the overall response score for all 12 items ($M = 4.88$, $SD = 2.88$, $Range = 0 - 11$). All items had sufficient item-total correlations of .30 or higher, indicating that, for each item, a correct response had a meaningful correlation with a higher overall score, and vice versa. Next, we examined the discrimination statistic for each item. We used point-biserial correlation statistics to determine how well each item discriminated between lower and higher scoring participants. An item discrimination correlation of at least .30 is generally considered satisfactory, while .40 or higher is considered to have very good discriminatory power (Crocker & Algina, 2006). Presented in Table 3, all items were found to have sufficient discriminatory power, as designated by the D statistic in the Item Response Statistics. Also, presented in the Item Response Statistics column of Table 3 are the item difficulty statistics. In general, it is good practice to have a range of difficulty among items in an assessment. For this particular assessment, we further expected item difficulty to increase from pre-multiplication to MC1 items to MC2 items in a manner that aligned with our item conceptualizations (see Figure 1). As hypothesized, item difficulty appeared to generally increase from one classification of items to another. This, along with the reliability coefficient and various discrimination statistics

reported, suggests that students' correct/incorrect responses to the assessment have sufficient statistical reliability.

Table 3
Item Statistics for the Assessment Response Data and Coded Schemes

Item	Response Statistics		Student Work <i>Targeted</i> <i>Scheme Coded</i>	Correlation <i>Response &</i> <i>Scheme</i>	
	<i>M (SD)</i>	<i>D</i>	<i>M (SD)</i>	ϕ	<i>n</i>
Pre-Mult.					
A	.65 (.48)	.41	.97 (.18)	.12	149
B	.86 (.35)	.38	.95 (.22)	.23**	147
MC1					
C ^a	.61 (.49)	.49	.57 (.50)	.81***	143
D ^a	.61 (.49)	.47	.60 (.49)	.91***	143
E	.29 (.46)	.39	.60 (.49)	.55***	138
F	.46 (.50)	.52	.56 (.50)	.83***	140
MC2					
G ^a	.33 (.47)	.61	.36 (.48)	.91***	137
H ^a	.40 (.49)	.35	.33 (.47)	.74***	137
I	.14 (.34)	.44	.18 (.39)	.90***	128
J	.30 (.46)	.46	.32 (.47)	.82***	125
MC3					
K ^a	.09 (.28)	.31	.17 (.38)	.65***	108
L ^a	.15 (.36)	.37	.18 (.39)	.72***	106

Note: Lower sample size for correlations is due to insufficient student evidence for schemes.

^a*Targeted scheme for this item is the reversibility for item set targeted concept.*

* $p < .05$, ** $p < .01$, *** $p < .001$

Our next step was to examine the construct validity of our items. To do this, we examined the relationship between response score per item (correct/incorrect) and whether students' demonstrated work on the item was coded as the targeted scheme for that item or not. As shown in Table 2, code 1 aligns with items PreMult_A and PreMult_B in Table 3, code 2 aligns with items MC1_C and MC1_D, and so forth. To examine whether a correct response to an item generally corresponded to the targeted scheme for that item, we dichotomized the coding for each items (0 = strategies do not correspond to targeted item scheme; 1 = strategies correspond to targeted item scheme). The

means for each item are shown in the Student Work column of Table 3. We examined correlations between presence of targeted schemes and correct responses using the ϕ statistic, shown in the last column of Table 3. The vast majority of these correlations were found to be strong and were statistically significant from chance (a strong correlation is considered above .50). This indicates that students who provided the correct solution for specific items tended to also show evidence (via their written work on the task) demonstrating the scheme hypothesized to lead to a correct response. We do not claim that correct solutions were necessarily or always the result of the use of a targeted scheme for a particular item, but that it is much more probable than not. These findings provide support for our hypothesized relationship between the design of an item eliciting specific schemes and associating with a correct item response for MC1, MC2, and the two MC3 items. However, the pre-multiplication items appeared to have lower associations between correct response and targeted scheme. Closer inspection of these items revealed that this was primarily due to incorrect responses that were associated with attempts to use the targeted scheme (i.e., iterating 1s to form a composite unit). Specifically, students who provided incorrect responses typically attempted to count by 1s, but did so adding onto the already presented length.

Multiplicative Concepts in Elementary Grades

We examined whether the assessment distinguished between second and third grade students. Since multiplication is an emphasized topic in third grade, the test should show some differences in response scores and coded schemes. An independent samples t test found a statistically significant difference between grade-level scores ($t = 4.20, p < .001$) indicating that third grade students had more correct responses to items ($M = 5.67, SD = 2.95$) than second grade students ($M = 3.89, SD = 2.46$). Thus, on average, third grade students gave the correct response to two additional items than the average second grade student. We interpreted these findings as evidence towards further supporting the validity of the assessment. To determine if these differences were due to potential differences in

multiplicative concepts, we examined the differences in percentages of demonstrated target schemes by item and item type (see Figure 3). Recall that the coded schemes for students did not necessarily translate into correct responses. With this in mind, two interesting trends are noticeable in Figure 3. First, there are similar percentages of using the identified pre-multiplication scheme (iterating units of 1 to form a composite unit), as well as for code 2 on items MC1_C and MC1_D (partitioning a unit into 1s). However, third grade students were more likely than second grade students to construct schemes that involved iterating a non-1 unit (see items MC1_E & MC1_F in Figure 3). From these items onward, third grade students tended to use the targeted scheme more consistently than second grade students. The divergence in using this scheme is associated with the mean difference of two items in score between grades observed in the t test.

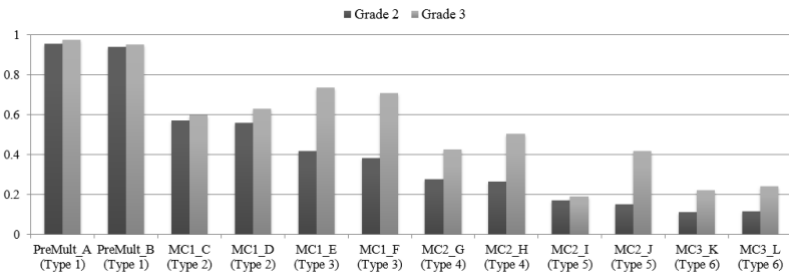


Figure 3. Targeted schemes demonstrated per item and differentiated by grade. Grade level is differentiated by dark and light gray bars (grades 2 and 3 respectively).

Following a comparison of item responses and coded schemes between grade levels, we constructed preliminary indices for pre-multiplication, MC1 and MC2. Each index was determined by completion of greater than 50% correct response for items targeting specific schemes or concepts, as well as greater than 50% correct response on any item sets of lower difficulty.² For example, a student aligned with the MC1 index

² The use of 50% as a benchmark was selected as a conservative threshold for success, loosely resembling interpretations of discrimination indices for IRT models. However, percentages representing mastery are often based on

would need to have more than half of the responses correct for PreMult items (PreMult_A and PreMult_B), as well as MC1 items (MC1_C, MC1_D, MC1_E, and MC1_F). Since only one student in the sample completed more than half of the MC3 items successfully, we exclude the classification in our report. The indices were meant to align with our item conceptualization (see Figure 3), as well as more general descriptions of multiplicative concepts in the literature. Thus, our indices represent the hypothesized minimum multiplicative concept with which a student is aligned. Next, the multiplication indices were cross-checked with coded multiplicative schemes for evidence of validity. Students classified at the pre-multiplication level tended to respond to items PreMult_A and PreMult_B with the targeted scheme (Table 4). However, these students used targeted schemes for MC1 items roughly 40% of the time, and such use was not consistent across students. In other words, on items MC1_C and MC1_D some students in the pre-multiplication index likely constructed schemes in activity for partitioning a composite unit into 1s, and on items MC1_E and MC1_F, other pre-multiplication students demonstrated iterating by non-1 units with schemes likely constructed in activity. This is consistent with descriptions by Steffe (1992) that such students can, at times, complete more sophisticated tasks with schemes constructed in activity. However, the inconsistent application of such schemes suggests use of these schemes is probabilistically unlikely. We noticed similar trends for the other indices in regards to schemes not hypothesized to be constructed in activity for students at those levels (see Table 4). Therefore, the descriptive statistics in Table 4 provide evidence that although students operating with less sophisticated concepts or schemes may successfully complete items hypothesized to align with more sophisticated concepts or

collective feedback from panels of teachers or other experts (Rupp, Templin, & Henson, 2010). Continued development of the assessment will require revision of this benchmark. Its use here should be interpreted as tentative pending further study.

schemes, it is less probable than for students with more sophisticated concepts. An additional facet of the statistics presented in Table 4 is that students who were assigned to a particular index based on their response scores had a high probability of demonstrating schemes aligned with that level and lower levels. We believe that this provides evidence of using anticipatory schemes and constructing schemes in activity that are described for those respective concepts (see Table 1). The descriptive statistics in Table 4 provide initial support for this relationship, additional support for the validity of the assessment, and support for our conceptualization of how students may generally respond to such items (see Figure 1).

Table 4
Percent of Students at Each Index using Targeted Schemes on Items

Index	PreMult		MC1				MC2				MC3	
	A	B	C	D	E	F	G	H	I	J	K	L
PreM	99.0	96.8	40.9	45.2	40.7	36.0	14.9	12.1	6.2	15.4	7.2	8.5
MC1	100	96.3	88.5	92.3	96.2	96.2	50.0	57.1	18.2	27.3	10.5	23.5
MC2	100	100	100	100	100	100	100	90.5	65.0	90.5	50.0	60.0

Table 5 presents the descriptive statistics for the constructed indices from the assessment. Apparent in the distribution are the differences between second and third grade students classified as demonstrating MC1 and MC2, and this was found to be statistically independent from chance ($\chi^2(df=2) = 15.22, p < .001$), providing support for the validity of these indices. These findings indicate that the assessment reliably discriminates between students at different grade levels, demonstrating operations with different multiplicative concepts.

Table 5

Estimated Percent of Students Demonstrating Various Multiplicative Indices by Grade Level

Grade	<i>n</i>	Pre-Multiplication	MC1	MC2
2	72	83.3	11.1	5.6
3	91	54.3	23.9	20.7
Total	163	67.5	18.4	14.1

Note. With exception of pre-multiplication schemes, estimates are based on more than 50% of items completed successfully for item sets.

Discussion

The findings from this study provide useful information for those interested in studying the multiplicative reasoning of elementary students in the realm of multiplication with whole numbers. Results from the pilot of this assessment suggest a sound and useful metric for second and third grade students' multiplicative reasoning. Items show statistical reliability and ability to discriminate between lower and higher scorers. Additionally, evidence from coding of student work suggests a strong correlation between hypothesized multiplicative schemes and successfully completing specific items (MC1, MC2, MC3). Results provide evidence towards the assessments' validity from multiple perspectives, including alignment between response scores and coded schemes, validity associated with differences in second and third grade, and verification of qualitative work on schemes that suggests individuals operating at lower multiplicative concepts can sometimes successfully complete more complex tasks (though probabilistically less likely).

Our results suggest clear differences between second- and third-grade students' demonstrated multiplicative concepts. Recall that the indices represent a minimum threshold (i.e., some students designated as at MC1 may have used anticipatory schemes that suggest they are at MC2). Although it is not expected that most second grade students would demonstrate MC1 or MC2, our findings indicate that roughly half of third grade students may be operating with variations of pre-

multiplication schemes on the assessment. These findings should be considered preliminary, given both the stage of development of the assessment and that the sample may not be representative of the population of U.S. second- and third-grade students at large. However, should similar findings be replicated at a larger scale with a more representative sample, then there are some clear and troubling implications. Specifically, the body of research examining multiplicative concepts has focused on its relationship to developing understandings of, and operations with, fractions (Hackenberg, 2010; Hackenberg & Tillema, 2009; Norton et al., 2015). According to the Common Core Standards for Mathematics (CCSSI, 2010), third grade is when fractions are first formally introduced. If nearly half of third grade students in this study are not using multiplicative reasoning in specific tasks such as those included in the assessment, then it is likely that their understanding of, and operations with, fractions is similarly restricted. Yet, as discussed earlier, these findings are preliminary with regard to the prevalence of multiplicative reasoning in these grade levels, and further study is needed to examine the potential relationships that scores on this assessment may have on other activities in school mathematics.

The assessment presented in this paper provides a useful and reliable instrument for assessing elementary students' multiplicative reasoning up to operating at least at MC2. Although the present version of the assessment shows both sufficient reliability and validity, additional work is needed to further improve the assessment, both for its use in second and third grade, as well as its potential use in later grades. First, the correlation between demonstrated strategy and correct response for the pre-multiplication items, as well as the pattern of responses for these students throughout the assessment, suggests that a small portion of students may not have interpreted the directions for the assessment as we intended. Inclusion of sample items with revised instructions for facilitating the assessment may reduce this. Another issue identified through comparison of coded strategies and response was that some students demonstrated appropriate schemes but, due to slight misjudgments of length, provided an incorrect response.

Although minimal, a potential means of reducing this error may be to include reference lines for each item. For example, a reference line shown for a potential revision of Item MC3_L is presented in Figure 4. The interval spacing for the reference line suggests a potential partition of length 24, but to successfully apply this partition a student must still engage in disembedding with composite numbers. In making such a revisions, one would need to be mindful of the intervals for the reference line given the targeted scheme for specific items.

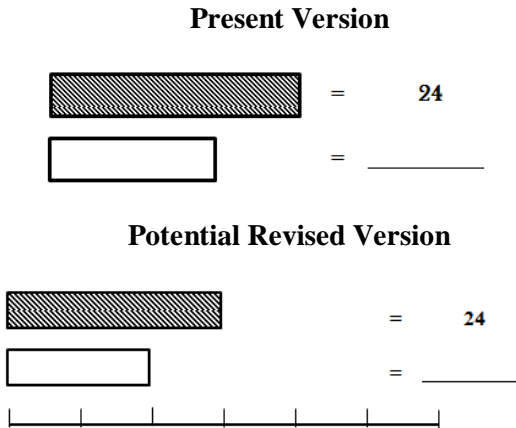


Figure 4. Potential revision of items using Item L as an example.

A critical revision for future versions of the assessment is to include additional items, particularly in regards to assessing MC3. Further, it may be useful to include additional items for each scheme to further examine the reliability of the item design criteria. For example, Item MC2_I performed statistically in ways similar to MC3 items. Although we believe inclusion of a reference line may help alleviate this trend, additional items of the same format may help illustrate potential traits of items currently not observed. Finally, additional means of validating the assessment are needed including clinical interviews with students following their completion of the assessment, and also examination of how responses to this assessment correspond with responses to other kinds of multiplicative tasks. Such efforts at validation are not likely to signify the assessment as

“good” or “bad” but will likely allow for more appropriate interpretations of what various scores on the assessment mean in regards to a child’s mathematical activity.

This study represents the initial piloting of an assessment of multiplicative reasoning. As such, it represents a successful effort at building an explanatory model of students’ responses to multiplicative tasks based upon prior qualitative work on this topic. Further, the study moves towards answering a call made by Kilpatrick (2001) for connecting qualitative and quantitative approaches to researching similar phenomenon to build a body of evidence. As stated here, further study is necessary to both confirm and extend the findings of the present analysis, and to improve the assessment. In such manner, assessments similar to the one discussed in the present study may be modified in presentation to provide useful diagnostic feedback for teachers and researchers, and inform those making practical decisions to improve the mathematics education of children. Lastly, the process for developing, aligning, and initially piloting an assessment based on data from prior qualitative work may be useful for others endeavoring to design such assessments.

Authors’ Note

The authors thank Anderson (Andy) Norton and Jesse L. M. (Jay) Wilkins for comments and feedback on earlier versions of this paper. We also thank Lynde Kosko for rendering Figure 1 from our conceptual sketches. This paper is based on an earlier version presented at AERA 2016.

References

- Brickwedde, J. (2011). Transition from additive to multiplicative thinking: A design and teaching experiment with third through fifth graders (Doctoral dissertation, University of Minnesota). Retrieved from <https://conservancy.umn.edu/>
- Confrey, J., & Harel, G. (1994). Introduction. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. vii-xxvii). Albany, NY: State University of New York.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Mason, OH: Thomson Wadsworth.

Elementary Children's Multiplicative Concepts

- Eliustaoglu, E. (2016). Investigation of fraction schemes and models as a means to understand how sixth grade students make sense of fractions (Masters thesis, Kent State University). Retrieved from <http://www.ohiolink.edu/etd/>
- English, L. D. (1991). Young children's combinatoric strategies. *Educational Studies in Mathematics*, 22(5), 451-474.
- Hackenberg, A. J. (2010). Students' reasoning with reversible multiplicative relationships. *Cognition and Instruction*, 28(4), 383-432.
- Hackenberg, A. J., & Tillema, E. S. (2009). Students' whole number multiplicative concepts: A critical constructive resource for fraction composition schemes. *Journal of Mathematical Behavior*, 28(1), 1-18.
- I, J. Y., & Dougherty, B. J. (January, 2014). Linking multiplication models to conceptual understanding in measurement approach. Paper presented at the 11th Hawaii International Conference of Education, Honolulu, HI. Retrieved at https://www.researchgate.net/profile/Ji_Yeong_I/publication/270279409_Linking_Multiplication_Models_to_Conceptual_Understanding_in_Measurement_Approach/links/54a5e2870cf267bdb9082af5.pdf
- Izsák, A., Jacobson, E., & de Araujo, Z. (2012). Measuring mathematical knowledge for teaching fractions with drawn quantities. *Journal for Research in Mathematics Education*, 43(4), 391-427.
- Izsák, A., Orrill, C. H., Cohen, A. S., & Brown, R. E. (2010). Using the mixture Rasch model to assess middle grades teachers' reasoning about rational numbers. *Elementary School Journal*, 110, 279-300.
- Kilpatrick, J. (2001). Where's the evidence? *Journal for Research in Mathematics Education*, 32(4), 421-427.
- Kosko, K. W., & Singh, R. (2016). Children's linguistic and numeric unit coordination. In M. B. Wood, E. E. Turner, M. Civil, & J. A. Eli (Eds.), *Proceedings of the 38th annual meeting of the North American Chapter for the Psychology of Mathematics Education* (pp. 605-612). Tucson, AZ: The University of Arizona.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Mulligan, J., & Mitchelmore, M. (1997). Young children's intuitive models of multiplication and division. *Journal for Research in Mathematics Education*, 28(3), 309-330.
- National Governors Association Center for Best Practices, & Council of Chief State School Officers (2010). *Common core state standards for mathematics*. Washington, DC: Author.

- Norton, A., Boyce, S., Phillips, N., Anwyll, T., Ulrich, C., & Wilkins, J. (2015). A written instrument for assessing students' units coordination structures. *IEJME - Mathematics Education*, 10(2), 111-136.
- Norton, A., Boyce, S., Ulrich, C., & Phillips, N. (2015). Students' units coordination activity: A cross-sectional analysis. *Journal of Mathematical Behavior*, 39, 51-66.
- Norton, A., & Hackenberg, A. J. (2010). Continuing research on students' fraction schemes. In L. P. Steffe & J. Olive (Eds.), *Children's fractional knowledge* (pp. 341-352). New York: Springer.
- Olive, J. (2001). Children's number sequences: An explanation of Steffe's constructs and an extrapolation to rational numbers of arithmetic. *The Mathematics Educator*, 11(1),4-9.
- Singh, R., & Kosko, K. W. (2016). Elementary students' understanding of equivalence and multiplicative reasoning. In M. B. Wood, E. E. Turner, M. Civil, & J. A. Eli (Eds.), *Proceedings of the 38th annual meeting of the North American Chapter for the Psychology of Mathematics Education* (p. 212). Tucson, AZ: The University of Arizona.
- Smith, S. Z., & Smith, M. E. (2006). Assessing elementary understanding of multiplication concepts. *School Science and Mathematics*, 106(3), 140-149.
- Steffe, L. P. (1992). Schemes of action and operation involving composite units. *Learning and Individual Differences*, 4(3), 259-309.
- Steffe, L. P. (1994). Children's multiplying schemes. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 3-39). Albany, NY: State University of New York.
- Steffe, L. P. (2002). A new hypothesis concerning children's fractional knowledge. *Journal of Mathematical Behavior*, 20, 267-307.
- Steffe, L. P. (2017). Psychology in mathematics education: Past, present, and future. Plenary Talk given at the 39th annual meeting of the North American Chapter for the Psychology of Mathematics Education, Indianapolis, IN.
- Tillema, E. S. (2013). Relating one and two-dimensional quantities: Students' multiplicative reasoning in combinatorial and spatial contexts. *Journal of Mathematical Behavior*, 32(3), 331-352.
- Tzur, R., Johnson, H., McClintock, E., Xin, Y. P., Si, L., Kenney, R., Woodward, J., Hord, C., & Jin, X. (2012). Children's development of multiplicative reasoning: A scheme and tasks framework. In T. Y. Tso (Ed.), *Proceeding of the 36th conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 155-162). Taipei: PME.

Elementary Children's Multiplicative Concepts

- Wilkins, J. L. M., Norton, A., & Boyce, S. J. (2013). Validating a written instrument for assessing students' fractions schemes and operations. *The Mathematics Educator*, 22(2), 31-54.
- von Glasersfeld, E. (1995). Sensory experience, abstraction, and teaching. In L. P. Steffe and J. Gale (Eds.), *Constructivism in Education* (pp. 369-383). Hillsdale, NJ: Lawrence Erlbaum Associates.