A WRITTEN, LARGE-SCALE ASSESSMENT MEASURING GRADATIONS IN STUDENTS' MULTIPLICATIVE REASONING

Conference Paper · July 2018

Heather Lynn Johnson
University of Colorado

Ron Tzur
University of Colorado

Nicola Hodkowski
University of Colorado

Cody Jorgensen
University of Colorado

A WRITTEN, LARGE-SCALE ASSESSMENT MEASURING GRADATIONS IN STUDENTS' MULTIPLICATIVE REASONING

Heather Lynn Johnson, Ron Tzur, Nicola Hodkowski, Cody Jorgensen, Bingqian Wei, Xin Wang, and Alan Davis

University of Colorado Denver

We examine a written, large-scale assessment that assessors can use to infer and measure gradations in students' scheme for whole number multiplicative reasoning. To design such an instrument we drew on Tzur's notion of fine grain assessment, which is used to distinguish two stages in the construction of a scheme: participatory and anticipatory. We briefly present the assessment items, the validation process, and reliability statistics–Cronbach's alpha, Rasch modeling, and student response patterns from students (N=492) in grades 3 and 4 (~ages 8-10), including distinctions in item difficulty levels. We discuss implications for large-scale assessment design and implementation.

In this study, we extend two recent studies from our large research project investigating elementary students' development of multiplicative reasoning (Hodkowski, Hornbein, Gardner, Jorgensen, Johnson, & Tzur, 2016)[1]. We report on a written assessment designed and implemented to infer into students' multiplicative reasoning. Such an assessment faces the challenge of finding an adequate alternative to the labor-intensive method of interviewing students. Whereas task-based, cognitive interviews afford inferring students' reasoning from their interactions with assessors, a large-scale assessment must allow valid and reliable inferences based solely on student responses. To face this challenge, we built on Norton and Wilkins' (2009) use of quantitative methods to measure students' reasoning based on models researchers obtained through interviews with small numbers of students. We expand the work of Norton and Wilkins by focusing on conceptual gradations that Tzur and colleagues (Tzur & Simon, 2004; Tzur, 2007) have postulated within students' reasoning—the *participatory and anticipatory* stages.

Researchers across the world have been studying students' challenges with whole numbers multiplicative reasoning (Lamon, 2007). In our work, we stress that such reasoning involves more than knowing multiplication facts and/or developing procedural skills. It includes students' meanings for multiplication (Steffe & Cobb, 1998), their insights into multiplicative relationships between numbers (Bakker, van den Heuvel-Panhuizen, & Robitzch, 2015), and their coordination of different kinds of units to form new units (Tzur, Johnson, McClintock, Xin, Si, Woodward, Hord, & Jin, 2013). In this study, we address the following problem: How can a written, large-scale assessment be used, in place of interviews, to infer gradations in students' scheme for whole number multiplicative reasoning?

2018. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.). *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 163-170). Umeå, Sweden: PME.

THEORETICAL AND CONCEPTUAL FRAMEWORK

Assessing assimilation (schemes)

As humans, we cannot directly observe the reasoning of others. Through interviewing or written methods, researchers can make inferences about others' reasoning. We draw on Piaget's (1985) core notion of *assimilation*—a cognitive intermediary between observable "stimuli" and "responses"—as a lens to make such inferences. Von Glasersfeld (1995) explained that assimilation, and reasoning, are made possible by a three-part cognitive building-block—a *scheme*. Schemes comprise: (1) a recognition template (situation) that triggers one's goal; (2) an activity to accomplish that goal; and (3) an effect that one mentally anticipates, or notices retroactively, to ensue from that goal-directed activity. We designed our assessment to measure gradations in students' mental use of schemes for multiplicative reasoning.

Schemes for whole number multiplicative reasoning

Researchers explicitly distinguished multiplicative reasoning from successfully determining answers to multiplication problems (Bakker et al., 2015; Tzur et al., 2013). We infer that students engaging in multiplicative reasoning can use schemes to keep track of and coordinate different kinds of units. For example, consider this task: "Julia has 6 towers, each made from 3 stacking cubes. How many cubes did Julia use to make the towers?" A student may draw all cubes and correctly count them one-by-one. In contrast, a student engaging in multiplicative reasoning with an assimilatory scheme would coordinate three kinds of units: composite units (e.g., towers), the magnitude of each unit (e.g., cubes per tower), and units of 1 (e.g., total of individual cubes).

Tzur et al. (2013) identified six schemes for multiplicative reasoning. Our study focuses on assessing the first one, termed *multiplicative double counting* (mDC), which marks the shift from additive to multiplicative reasoning. A student having the mDC scheme could recognize a situation as consisting of two different kinds of units, set the goal to find the total of 1s in them, trigger the activity of simultaneously distributing and counting (keeping track of) accrual of 1s and of composite units (e.g., 1 tower is 3, 2 towers are 6, 3 towers are 9, … 6 towers are 18), and anticipate a new kind of unit as a result of her activity.

Gradations in students' schemes

When assessing students' reasoning, we do not mean that having a scheme is like flipping an "on-off" light switch. Rather, we distinguish gradations in schemes through Simon and Tzur's (2004) constructs of participatory and anticipatory stages, which differentiate a student's ability to bring forth a scheme. In the anticipatory stage, a student can independently, and spontaneously, do so. In the participatory stage, a student needs prompting to bring forth a goal-directed activity and its effect of an emerging scheme. We acknowledge that prompting can take different forms. In this paper, we focus on prompting involving additional supports, provided to a student through her sensory perception. For example, in the task involving Julia and the 6 towers, a hint could be a picture showing one completed tower and just a single cube for each of the remaining towers. If a student is at a participatory stage, such a hint may bring forth her activity of counting 1s and composite units, and thus may enable her to engage in multiplicative reasoning and to provide a correct response.

Fine grain assessment

To measure participatory-anticipatory gradations in students' mDC scheme, we adapted Tzur's (2007) fine grain assessment. In fine grain assessment, assessors begin with tasks that include no hints, then move to subsequent tasks including increasing levels of hints. We stress that the purpose of hints is *not* to funnel students to a certain solution method and/or correct answer. Rather, the purpose of hints is to provide students with opportunities to bring forth existing schemes. Including a hint-free task prior to other tasks allows assessors to infer the stage of a student's assimilatory scheme based on her or his solutions to tasks—first without and then with hints.

METHODS

We report on our methods to develop and implement a written assessment that targets gradations in a foundational scheme that indicates students' emerging multiplicative reasoning: multiplicative double counting (mDC). To design the assessment, we drew on the expertise of our large, diverse project team, which includes mathematics educators, a mathematician, research methodologists, and language experts.

The mDC assessment: Problems, items, and sub-items

The mDC assessment we developed contains five word problems. The first problem served as a screener problem (1-digit addition, 8+7), to foster initial success. The next four problems, together, allow inferring the stage of a student's mDC scheme (see multiplicative reasoning problems #2-5 below). To assess participatory-anticipatory gradations, each problem comprises at least three items. We designed the first item to be "hint-free." The subsequent items included increasing levels of hints. With each increasing level of hint, we intended to provide students opportunities to bring forth their mDC scheme. Therefore, hints provided increasingly specific information about the different kinds of units in the situation. Furthermore, to assess students' text comprehension, in each problem we included sub-items for which students filled in blanks with information given in a problem statement. For example, in Problem #3, students filled in this blank: "Alex put __ towers in the box." (see Figure 1).

In Problem #2, we focused on students' iteration of a composite unit (e.g., a tower of 3 cubes) to determine if it could constitute a larger composite unit (e.g., a tower of 24 cubes). In Problem #3, we intended for students to distribute items of one composite unit (3 cubes per tower) over another unit (6 towers) to find the total number of 1s in the compilation of composite units (total of 18 cubes). In Problem #4, we intended students to keep track of composite units (4 teams of 5 players each). We asked them to determine the correctness of a hypothetical student's (Joy) statement that, through "skip-counting: by 5, she found there were 35 players in all. In Problem #5, given a total number of items (28 cookies), we intended for students to segment this total by iterating a given composite unit (4 cookies per bag) to determine the total number of composite units (bags) that constitute the total.

In this paper, we focus on Problem #3. Figure 1 shows the first, "hint-free" item.

The picture to the right shows a box. Alex put ***6*** towers in the box. Alex made each tower with ***3*** cubes. The numbers on the picture show this.	6 Towers 3 Cubes in each Tower

A: Alex put _____ **towers** in the box.

B: Alex made **each tower** with _____ **cubes**.

C: How many **cubes in all** did Alex use to make ***6*** towers? (fill in blank): ________

Figure 1: mDC Assessment Problem #3, Item 1; Hint-free.

The second and third items of Problem #3 followed the same format as the first item. These items included increasing levels of hints. Figure 3 shows level 1 and level 2 hints for items 2 and 3, respectively. The level 1 hint included additional diagrammatic information about the activity of iterating a unit of "tower." The level 2 hint included additional diagrammatic information about the units composing the towers to be iterated, the "cubes per tower."

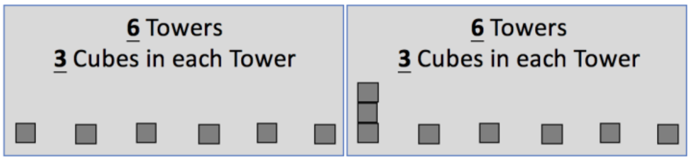

Figure 2: Problem #3, Items 2 and 3: Hint level 1 at left; Hint level 2 at right.

Validity and Reliability

We addressed construct validity through a five-phase process. Initially, Tzur created an expert draft for each problem. Second, he shared the draft with a content expert who gave feedback, with changes. Third, the project team worked on that version, leading to more revisions of language and diagrams. Fourth, this version went through an expert panel review. We gave this version to three experts in the field, who evaluated the problems and items, responding by: "keep," "change as follows," or "omit." The experts suggested a few revisions, but not any omissions. Fifth, Tzur conducted individual, cognitive interviews with five children to check the entire assessment. Issues arising from those interviews led our team to make further, finer revisions.

We addressed construct validity along three lenses: language, potential gender or cultural biases, and mathematical consistency with the multiplicative reasoning we intended to measure (Hodkowski et al., 2016). The mathematics educators and mathematician worked closely to address mathematical consistency. We drew on the language and literacy experts on our project team to develop word problem statements appropriate for students learning English as an additional language. In addition, we included situations familiar to the student population with whom we worked.

Tzur conducted cognitive interviews with 26 fourth-graders to determine the extent to which their responses to items on the mDC assessment and additional tasks he posed correlate with his inference into their mDC scheme. To determine if the mDC assessment could actually serve as a proxy for students' reasoning, as opposed to just the child's ability to correctly solve each item, we used the Kendall's Tau-b statistics to calculate agreement between Tzur's inferences and the score obtained from the written assessment items (Ktb=0.883, 2-tailed $p<.0005$). Thus, we claim the data of students' performance on items on the mDC items indicate students' engagement in multiplicative reasoning (mDC scheme).

Student population, student numbers, and assessment administration dates

Students participating in our study were from three different elementary schools in one small and one large public school district. Both districts were in the metropolitan area of a large US city. About 85% of the students in our study identified as students of color, and 70% were learning English as an additional language.

We report results from three administrations of the mDC assessment to students in 3rd and 4th grades: Spring 2016, Fall 2016, and Spring 2017. We report results from a total of 492 student assessments, produced by 404 unique students (some assessed twice or three times). Table 1 disaggregates the assessment totals by student grade and administration date. We analyzed data from all 492 available assessments, because they reflect students' assimilation (or lack thereof) of the problems into their mDC scheme in far-apart administrations. This larger number allowed us to further investigate gradations in students' reasoning.

Grade	Spring 16	Fall 16	Spring 17	Total
3	81	146	117	344
4	26	83	39	148
Total	107	229	156	492

Table 1: Numbers of students taking the mDC assessment by grade and date.

Like researchers across the world (e.g., Bakker et al., 2015), we experienced challenges implementing this large-scale study with students in schools. One challenge included obtaining student and parent consent, which impacted our data analyses on sets of disaggregated data. To address challenges, we worked with teachers and school personnel to determine protocols and times beneficial to both parties.

Data entry

Six graduate research assistants (GRA) were trained to enter the student responses to the mDC assessments. To increase reliability, one GRA read the student's responses out loud and the other entered those into a spreadsheet as is. The first GRA could see and was asked to verify that responses entered correctly for every student. We coded no response as "9999" and an unreadable or incoherent answer as "5555."

Analysis: Cronbach's alpha, Rasch modeling, Student response patterns

To calculate overall item consistency of all four multiplicative reasoning problems (#2-5), we used the Cronbach's Alpha statistics for all items and sub-items. Moreover, we employed the principal component method of exploratory factor analysis to confirm that all items loaded onto a single principal component (here, the construct of mDC scheme), and that no additional factors could be extracted.

We conducted Rasch analysis to determine item difficulties and person measures. We tested if students who did not bring forth the mDC scheme on a hint-free item could do so on items containing hints. We used Rasch modeling with hint-free items and with items containing any form of hint (both level-1 and level-2 hints). Students bringing forth the mDC scheme on hint-free items would have an anticipatory stage of the scheme. Students having the participatory stage of the mDC scheme would bring forth the scheme after receiving the diagrammatic hints. We hypothesized that Rasch analysis would indicate items containing hints to be consistently less difficult than hint-free items. Next, we examined a Wright map, which organizes both persons and items by logits ranging from -3 to +3. Ideally, in a Wright map, the distribution of items should show a wide range of item difficulties, with more items in the middle than at the extremes, and each item on a unique difficulty level.

Besides Rasch modeling, we also examined students' response patterns for hint-free items and items containing hints. We grouped the data into four response patterns: (1) Hint-free Correct, Hint Correct; (2) Hint-free Correct, Hint Incorrect; (3) Hint-free Incorrect, Hint Correct; (4) Hint-free Incorrect, Hint Incorrect. We coded "correct" for items containing hints if students provided a correct response for any level of hint.

RESULTS

mDC assessment consistency and factor analysis

Chronbach's alpha for the 8 items in the mDC assessment (263 cases = 53.5% of all 492), 4 hint-free items and 4 items with hint, is 0.907. This value reflects excellent inter-item consistency. Rasch item analysis indicated a very high consistency (0.98).

Rasch modeling and Wright map

Our Wright map showed item difficulties ranging from -2.3 to 1.3 logit scores. In a few cases, two or three items had the same logit score. For three of the four problems, analysis of our Wright map showed that the hint-free items were more difficult than the items containing hints. The most difficult item was the hint-free sub-item C on Problem #3 (logit score = 1.3); second to it was the sub-item C on Problem #3 that contained a hint (logit score = 1.15). Although these logit scores were fairly close, our analysis confirms the hint-free item to be more difficult. Furthermore, the Rasch item reliability to be 0.98 indicates that items have a hierarchy of difficulty. We found more distinct results for Problems 3, 4, and 5.

Students' response patterns

For each of Problems 2-5, we compared the four groups of student response patterns in respect to students' overall performance on the remaining assessment. Table 2 shows students' response patterns to sub-item C of Problem #3 (See Figure 1). As expected, for each item, students' average scores in the Correct-Correct group (N=169) were highest, and students' average scores in the Incorrect-Incorrect group (N=264) were lowest. For the other two groups, Correct-Incorrect (N=20) and Incorrect-Correct (N=39), students' average scores were between the two extremes.

	Hint-Free Correct	Hint-Free Incorrect	Total
Hint Correct	169	39	208
Hint Incorrect	20	264	284
Total	189	303	492

Table 2: Students' response pattern to Situation 2, sub-item C.

These results suggest we can measure gradations in students' multiplicative reasoning. Yet, gradations were not entirely clear-cut. Some students responded correctly to a hint-free item, then incorrectly to an item containing a hint (Correct-Incorrect, N=20). This seems to run counter to our conjecture that hints could provide students opportunities to bring forth their schemes. Despite this seeming discrepancy, the Correct-Incorrect group (N=20) scored lower on the overall mDC assessment than the Incorrect-Correct group (N=39), who responded incorrectly to the hint-free item. We infer that other factors, such as guessing, accounted for this result.

DISCUSSION

Based on our results, assessors can use the mDC assessment to measure gradations in students' mDC scheme for whole number multiplicative reasoning. Gradations include two stages, anticipatory and participatory, indicated by whether students demonstrated evidence of bringing forth a scheme before or after being given a hint.

To date, researchers have used small scale, labor intensive interview methods to distinguish students' anticipatory and participatory stages of conceptual development (e.g., Simon et al., 2016; Simon & Tzur, 2004). Our mDC assessment is a step toward measuring gradations in students' reasoning on a large scale. Although researchers have identified finer grained distinctions at the participatory and anticipatory stages (e.g., Simon et al., 2016), currently our assessment is only sensitive enough for researchers to use to measure distinctions between anticipatory and participatory stages. We would need further refinement to make more nuanced distinctions.

Distinguishing between anticipatory and participatory stages is useful for explaining a challenge common to teachers, termed "the next day phenomenon." For example, a student at a participatory stage may bring forth her mDC scheme to determine a total number of cubes, given 6 towers with 3 cubes in each. Yet, the same student may not bring forth her mDC scheme on a seemingly similar task. Students' participatory stage is a crucial and vulnerable stage in learning, and can explain, in part, why students may not yet be able to engage in multiplicative reasoning without additional support.

[1]Acknowledgment

This study was supported by the US National Science Foundation (grant 1503206). The opinions expressed do not necessarily reflect the views of the Foundation.

References

Bakker, M., van den Heuvel-Panhuizen, M., & Robitzsch, A. (2015). Effects of playing mathematics computer games on primary school students' multiplicative reasoning ability. *Contemporary Educational Psychology*, *40*, 55–71.

Hodkowski , N. M., Hornbein, P., Gardner, A., Johnson, H. L., Jorgensen, C., & Tzur, R. 2016, November). Designing a stage-sensitive written assessment of elementary students' scheme for 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 of the International Group for the Psychology of Mathematics Education* (pp. 1581-1587). Tucson, AZ: The University of Arizona.

Lamon, S. J. (2007). Rational numbers and proportional reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 629-667). Charlotte, NC: Information Age Publishing.

Norton, A., & Wilkins, J. L. M. (2009). A quantitative analysis of children's splitting operations and fraction schemes. *The Journal of Mathematical Behavior*, *28*(2), 150–161.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago: University of Chicago Press.

Simon, M. A., Placa, N., & Avitzur, A. (2016). Participatory and anticipatory stages of mathematical concept learning: Further empirical and theoretical development. *Journal for Research in Mathematics Education*, *47*(1), 63–93.

Steffe, L. P., & Cobb, P. (1998). Multiplicative and Divisional Schemes. *Focus on Learning Problems in Mathematics*, *20*(1), 45–61.

Tzur, R. (2007). Fine grain assessment of students' mathematical understanding: participatory and anticipatory stages in learning a new mathematical conception. *Educational Studies in Mathematics*, *66*(3), 273–291.

Tzur, R., Johnson, H. L., McClintock, E., Xin, Y. P., Si, L., Woodward, J., Hord, C., & Jin, X. (2013). Distinguishing schemes and tasks in children's development of multiplicative reasoning. *PNA, 7*(3), 85-101.

Tzur, R., & Simon, M. (2004). Distinguishing two stages of mathematics conceptual learning. *International Journal of Mathematical Educ. in Science and Technology*, *2*(2), 287–304.

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. Washington, D.C.: Falmer.