JOURNAL OF MATHEMATICS EDUCATION AT TEACHERS COLLEGE

A Century of Leadership in Mathematics and Its Teaching

Standardization in Mathematics Education

TABLE OF CONTENTS

PREFACE

JOURNAL OF **MATHEMATICS EDUCATION** AT TEACHERS COLLEGE | SPRING 2018 | VOLUME 9, ISSUE 1

What Form of Mathematics are Assessments Assessing? The Case of Multiplication and Division in Fourth Grade NAEP Items

Karl W. Kosko
Kent State University

Rashmi Singh
Kent State University

ABSTRACT Multiplicative reasoning is a key concept in elementary school mathematics. Item statistics reported by the National Assessment of Educational Progress (NAEP) assessment provide the best current indicator for how well elementary students across the U.S. understand this, and other concepts. However, beyond expert reviews and statistical analysis, there is relatively little evidence on the validity of NAEP items for assessing mathematical reasoning. The present study examined the convergent validity of several public-release fourth grade NAEP multiplication items. Findings indicate that most items assessed lack sufficient validity for examining children's underlying conceptual knowledge. Rather, NAEP items may be less appropriate for assessing conceptual understanding and more appropriate for assessing procedural recall. Recommendations and implications are discussed.

KEYWORDS *multiplication and division, multiplicative reasoning, NAEP, assessment validity*

Introduction

Multiplicative reasoning is a key concept in elementary-level students' mathematical development. A student's understanding of multiplication and division relates to his or her engagement with rational numbers and fractions (Confrey & Harel, 1994; Norton et al., 2015), functions and algebraic reasoning (Russell et al., 2011), conception of equivalence (Singh & Kosko, 2016), and mathematical arguments (Kosko & Singh, 2016; Morris, 2009). Given the importance of multiplication and division in the development of students' mathematical thinking and learning, it is important to understand the current state of what elementary children can do with regards to multiplication and division, and the degree to which they understand these concepts. Various researchers examining elementary students' reasoning on multiplicative tasks suggest that most third- and fourth-grade students use variations of skip-counting strategies (Brickwedde, 2011; Mulligan & Mitchelmore, 1997). Additionally, Steffe (2007) conjectured that as many as half of students entering middle school operate with iterating multiplicative schemes (what some may refer to as skip-counting).

Since 1973, the National Assessment of Educational Progress (NAEP) has provided "the most comprehensive indicator available for student progress in the United States" (p. vii, Kloosterman et al., 2015). It is consistently used as a benchmark within and beyond the mathematics education community for what students know, and can do, mathematically. For example, recent NAEP data suggests that fourth-grade students can easily solve multiplication problems, but have difficulty with division (Mohr et al., 2015). Yet, students generally have less success when asked to explain their strategies or thinking. One potential reason for this is that some students may be able to perform procedures, but may lack a depth of understanding of those procedures.

In order to better understand observations such as those made by Mohr et al. (2015), the present study sought to differentiate whether NAEP items focusing on multiplication and division are accurate indicators of procedural recall or conceptual understanding. Specifically, we sought to examine what kinds of reasoning fourth-grade students generally demonstrate in relation to the responses they provide on multiple-choice items. To our knowledge, an examination of the validity of NAEP items in such a manner has not been conducted or presented to mathematics education researchers or practitioners. We believe such analysis is critically important given the weight that NAEP results have on mathematics education policy decisions in the U.S., and the importance of multiplication and division as topics in the elementary school curriculum.

Theoretical Framework

Multiplicative Reasoning

Multiplicative reasoning can be broadly described by the actions students take when engaging with multiplication and division tasks. Counting, subitizing, grouping, partitioning, and sharing have been identified as essential elements of multiplicative structure (Mulligan, 2004), with an initial step of such reasoning beginning with variations of skip-counting (Outhred & Mitchelmore, 2000). Further development of multiplicative reasoning requires the reconceptualization of units (Hiebert & Behr, 1988), including the understanding that a unit can be equivalent to 1 or be more than 1 (Chandler & Kamii, 2009). Although there are various theoretical frameworks for multiplicative reasoning, the present study incorporates the approach articulated by Scheme Theory (Hackenberg, 2010; Steffe 1992; Steffe, 1994). In general, multiplicative schemes require the coordination of at least two levels of units (i.e., a unit of units), and to develop multiplicative concepts, students require the anticipatory use of such schemes (Steffe, 1994). Following certain pre-multiplicative schemes are three multiplicative concepts. The first multiplicative concept (MC1) involves the coordination of two levels of units in activity for which students can take a composite unit as given and coordinate between two levels of units (e.g., solving 6×4 by counting 4, 8, 12, …). The second multiplicative concept (MC2) involves the coordination of three levels of units in activity, in which students can take two levels of units as given and coordinate between three levels of units in activity. For example, solving 14×4, a student at MC2 who has not learned the standard algorithm could use 6×4 as a starting point and skip count by 4s up from 24 until they have counted eight more 4s. The third multiplicative concept (MC3) involves the coordination of three levels of units, as does MC2, but for which schemes are used in an anticipatory manner (Hackenberg & Tillema, 2009; Norton et al., 2015; Olive, 2001). For example, solving 14×4, a student at MC3 who has not learned the standard algorithm may recognize that 14 groups of 4 can be separated into 10 and 4 groups of 4. Then, the student may solve the task by adding 10×4 and 4×4 together.

Multiplicative reasoning is essential to understanding contexts that involve multiplicative relationships. Within school mathematics, this is generally represented using specific algorithms. However, a student's ability to solve a multiplication or division problem through a conventional algorithm does not necessarily reflect conceptual understanding (Ebby, 2005; Kamii & Dominick, 1997; Leinwand, 1994; McNeal, 1995). Often, an over-emphasis on successfully performing the procedures of an algorithm neglects underlying mathematical concepts. For example, within the context of addition, Narode et al. (1993) observed that children's mathematical reasoning and confidence decreased after they memorized a standard algorithm. This indicates that although appropriately using algorithms is an important skill, a lack of understanding an algorithm can negate the algorithm's usefulness.

The NAEP Assessment

The U.S. Department of Education began administering the NAEP assessment in the 1972-1973 school year with the goal of administering the assessment periodically thereafter (Carpenter et al., 1975). Since 1990, the Main NAEP assessment includes items updated periodically and is administered to fourth and eighth-grade students every 2 years, and to twelfth-grade students every 4 years (Kloosterman, 2015). NAEP results are often reported in the press, along with other assessments, as indicators of what students know mathematically. In regard to multiplication and division, the first NAEP assessment indicated that less than a third of age 9 students could solve multiplication and division problems successfully (Carpenter et al., 1975). Five years later, this percentage increased to 60% (Carpenter et al., 1980). More recently, Mohr et al. (2015) reported that success on multiplication and division tasks generally ranges from 48% to 85%, with items assessing division showing more difficulty. In general, these statistics suggest that elementary students' understanding of multiplication and division has dramatically improved since the early 1970s. However, in examining item statistics for a division item [specifically, 15,336 ÷ 27 using the long division algorithm] Mohr et al. (2015) found that "36% incorrectly chose 17,605 as the quotient, suggesting that they entered the numbers wrong in the calculator and then ignored or did not know how to deal with the decimal value they saw" (pp.105–106). In other words, these students did not recognize a key feature of division. Such an observation suggests that, probabilistically, a portion of students who answered the item correctly also may have an incomplete understanding of division as well.

Prior study of the validity of NAEP items have predominately focused on item evaluation by experts in education. For example, Sugrue et al. (1995) found that the majority of NAEP items lack clarity of description in alignment with specific skills and understandings assessed. A later examination by Daro et al. (2007) found sufficient alignment between NAEP items and most states' standards. However, certain sub-constructs had too few items, and many items lacked complexity. More recently, Hughes et al. (2013) examined the alignment between NAEP items and the Common Core State Standards for Mathematics (CCSS-M) and overall alignment (apart from probability and statistics). These various examinations of item validity relied on the evaluation of items by experts. Lacking in the literature is an examination of students' reasoning in relation to their performance on NAEP items. Furthermore, no such reports are listed on the Department of Education's website, or on other major educational databases. To address this issue, this exploratory study focused on the convergent validity of fourth-grade NAEP multiplication items using student-level data.

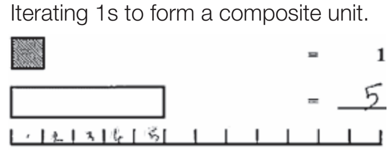
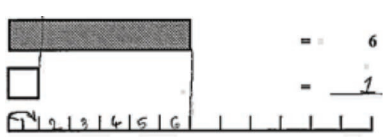
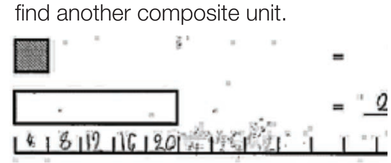
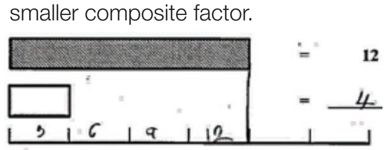
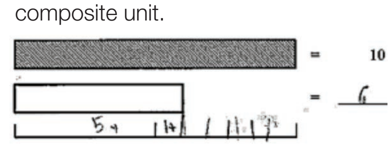
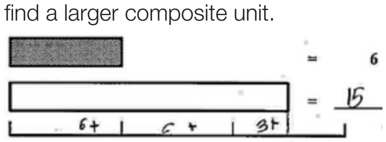
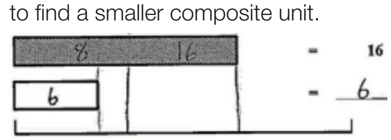
Method

Sample and Measures

Data were collected in May 2016 from 108 fourth-grade students in a Midwestern U.S. school district. Students completed an assessment packet that included the multiplicative reasoning assessment (MRA) and eight fourth-grade NAEP multiplication/division items. The MRA completed by participants is the second iteration of the instrument validated by Kosko and Singh (in press). The assessment includes 19 items that incorporate length models corresponding to the multiplicative concepts described by Hackenberg (2010). Table 1 presents example items from the MRA used in the present study. Rasch

Table 1

Example items from multiplicative reasoning assessment.

Category	Example items from multiplicative reasoning assessment	
Pre-Mult	Iterating 1s to form a composite unit. 	
MC1	Partitioning a composite unit into 1s. 	Iterating a composite unit to find another composite unit. 
MC2	Partitioning a composite unit into a smaller composite factor. 	Disembedding with 1s to find a composite unit. 
MC3	Disembedding with composites to find a larger composite unit. 	Disembedding with composites to find a smaller composite unit. 

modeling was used to estimate assessment scores, and to provide evidence for the reliability and validity of the assessment. The assessment demonstrated sufficient item reliability (0.96), suggesting that the MRA can distinguish between which items have lower or higher difficulty. The MRA also demonstrated sufficient person reliability (0.85), suggesting that the assessment can reliably distinguish between which individuals have higher or lower scores (Crocker & Algina, 2006; Wilson, 2005). Theta scores measuring the demonstrated ability levels of participants completing the MRA suggest scores above average for fourth-grade students (M = 0.95, SD = 2.47). Given that the MRA is designed to assess elementary students' multiplicative reasoning without the incorporation of canonical representations of multiplication and division (to avoid assessing memorized facts or procedures), it is the ideal assessment for examining convergent validity of fourth-grade NAEP multiplication items. Furthermore, evidence from student interviews and written work on the assessment indicates that the MRA does well in predicting the kinds of reasoning fourth grade students use in solving such canonical multiplication and division tasks (see Kosko, in review for a full description of the validation process for the MRA).

The eight NAEP items used in the present study are presented in Table 2 on page 5. All items included were multiple-choice items from released NAEP items from the online questions tool (NCES, 2016). Although including items from the same year assessment would provide a better comparison, NAEP releases only a subset of items used each year. To hedge against the potential variance introduced by items from different versions of the NAEP assessment, items were selected only from the 2011 and 2013 version of NAEP (the most recently released items at the time of this study). Items represent a cross-section of what NCES (2016) designated as "easy," "medium," and "hard" items. The items shown in Table 2 include the distribution of responses for the sample in normal text, and the national representative sample in adjacent italic text. Correct responses are designated with asterisks. Notably, for all but one item (NAEP_05), participants in the present study provide correct responses more frequently than most students nationally. However, students in the present sample did not use calculators in completing these items, whereas calculators were permitted for the national sample.

Analysis and Results

To assess convergent validity, Pearson correlation coefficients were calculated for the relationship between MRA scores and obtaining a correct response to each NAEP item examined. This provides validity evidence at the item level, but does not provide such evidence at the test level. Thus, implications of the current analysis are limited to item level characteristics.

A summary of the Pearson correlation coefficients is presented in Table 3 on page 6. Four of the eight statistics were found to be statistically significant. However, only one statistic met the minimum criteria for a meaningful effect size of .30 or above (Nunally & Bernstein, 1994). The item (NAEP_02) is illustrated in Table 2. On average, students who provided a correct response on this item had an MRA score of 1.97, which is considered representative of coordinating three levels of units. NAEP_02 effectively asks students to multiply 24 by 6, which conceptually represents a coordination of three levels of units (such as by adding $20\times6+4\times6$). Thus, the association with this level of multiplicative reasoning is sensible. The average Theta scores for other potential responses were all near zero (A = 0.20, B = 0.18, C = –0.50). This is particularly interesting since options B and C were both related to addition/subtraction, and option A involved division to obtain that response. Thus, students who used division to solve this task had a similar level of multiplicative reasoning to those who used addition or subtraction to solve it. Given its assessed validity and the corresponding statistics, such an item is potentially useful in examining the multiplicative reasoning of students nationwide.

By contrast, other items had insufficient indicators for validity. The item assessing solving a division task (NAEP_04) was found to have a near zero correlation with MRA scores. In examining average MRA scores per responses, an interesting pattern emerged that may provide some direction for improving the item validity (A = 0.26, B = 1.43, *C* = 1.63, D = 0.15). Specifically, students who answered "B" had similar levels of multiplicative reasoning as students who provided the correct response, "C." A primary difference between the responses for B and C is that one includes a decimal and the other does not. Thus, many students may have treated the comma as a decimal. Given that many students who

Table 2

First set of eight NAEP multiplication items for fourth grade.

Designation	Example items from multiplicative reasoning assessment			
NAEP_01 2013-4M3 #1	4 × 50 × 9 =			
	A 180	**B** 360	***C*** 1,800	***C*** 1,800
	2.3% *(7%)*	3.4% *(10%)*	92.0% *(75%)*	92.0% *(75%)*
NAEP_02 2011-4M8 #7	Patty expects that each tomato plant in her garden will bear 24 tomatoes. If there are 6 tomato plants in her garden, how many tomatoes does she expect?			
	A 4	**B** 18	**C** 30	***D*** 144
	19.3% *(26%)*	2.3% *(8%)*	1.1% *(12%)*	77.3% *(53%)*
NAEP_03 2013-4M6 #12	Which expression has the *least* value?			
	A 2 + 7 + 0 + 4	**B** (2 × 7) + 0 + 4	**C** 2 + (7 × 0) + 4	***D*** 2 × 7 × 0 × 4
	8.0% *(19%)*	2.3% *(9%)*	39.8% *(35%)*	50.0% *(36%)*
NAEP_04 2013-4M7 #4	Divide: $27 \sqrt{15,336}$			
	A 0.00176	**B** 56.78	***C*** 568	**D** 17,605
	3.4% *(13%)*	12.5% *(2%)*	78.4% *(48%)*	1.1% *(36%)*
NAEP_05 2011-4M8 #1	(47 x 75) ÷ 25 =			
	A 141	**B** 1,175	**C** 3,525	**D** 4,700
	54.5% *(83%)*	15.9% *(2%)*	23.9% *(11%)*	1.1% *(1%)*
NAEP_06 2011-4M12 #6	Multiply: 74 × 16			
	A 90	**B** 518	**C** 1,164	***D*** 1,184
	2.3% *(11%)*	4.5% *(17%)*	9.1% *(15%)*	80.7% *(52%)*
NAEP_07 2011-4M12 #10	Which factor of 12 is missing in this list of numbers? 1, 2, 3, 4, ___, 12			
	A 5	***B*** 6	**C** 8	**D** 10
	0.0% *(23%)*	94.3% *(47%)*	2.3% *(20%)*	1.1% *(8%)*
NAEP_08 2011-4M12 #13	Ms. Kim has 45 stickers that she wants to give out to 6 students. The students are sitting in a circle. Ms. Kim gives out one sticker at a time and keeps going around the circle until all the stickers are gone. How many of the students get more than 7 stickers?			
	A 2	***B*** 3	**C** 5	**D** 6
	11.4% *(12%)*	68.2% *(47%)*	6.8% *(17%)*	12.5% *(23%)*

Note: *Percentages in normal text represent the current sample and percentages in gray italics represent the national sample as reported by NCES (2016). Letters with asterisks indicate the designated correct response*

Table 3
First set of eight NAEP multiplication items for fourth grade.

NAEP Item Designations							
2013 4M3 #1	2011 4M8 #7	2013 4M6 #12	2013 4M7 #4	2011 4M8 #1	2011 4M12 #6	2011 4M12 #10	2011 4M12 #13
.23* n = 104	.37** n = 103	.28** n = 102	.08 n = 95	.10 n = 94	.11 n = 98	.27** n = 98	.18 n = 98
*p<.05, **p<.01							

responded with either B or C may be described as operating at least at MC2, and a conceptual understanding of division in this manner may be more representative of MC3 reasoning, this item may better assess procedural understanding of the long division algorithm and not the conceptual nature of division itself. Although we limit our descriptions to only the two aforementioned items, our item-level analysis yielded similar results across the eight NAEP items examined.

Discussion

Analysis of the convergent validity of fourth-grade NAEP items for multiplication and division suggests that several NAEP items lack sufficient validity. In other words, these NAEP items in their current form are not good indicators of fourth-grade students' conceptual understanding of multiplication or division, but may be better indicators of procedural knowledge. Such a finding is highly significant considering the weight that NAEP data has on U.S. educational policy. Fortunately, item analysis suggests some potential revisions for some items that can not only improve the validity, but also provide more appropriate estimates of children's levels of multiplicative reasoning. The findings of this study suggest that appropriate steps in validating NAEP items may not be currently included in the drafting of items. Examination of the literature indicates many validation studies include ratings by experts in the field, but no other study was identified that included student-level data (i.e., response processes). Thus, a primary implication of the present study is that NAEP data be interpreted with caution, as the validity of some items may be insufficiently assessed. However, the findings presented here are from a subset of items on the NAEP assessment with a non-representative sample. Although the findings have important implications for test design, they should be interpreted with these limitations in mind.

There are two significant implications of findings from the present study. The first is that the present study is limited to a sub-sample of public-release NAEP items, and a further study may be needed to determine if the validity of other NAEP items should be brought into question. However, this is a particularly difficult task given the restrictions placed on access to NAEP items (Kloosterman, 2015). Furthermore, should the findings in this study be confirmed and also shown to transfer to other items across topics and grade levels for NAEP, there is a fundamental question of "what should be done next?" Thus, the second major implication of the present study is for those within mathematics education to examine the nature of items that assess conceptual understanding, and not merely procedural recall. The MRA provides one meaningful starting point for the concepts of multiplication and division (Kosko, in review; Kosko & Singh, in press), but other concepts need similar study. Further, by better understanding the nature of items assessing conceptual understanding, both researchers and practitioners can apply such lessons directly to mathematical tasks used within the classroom. Both implications for research provide two possible paths for those seeking to improve how assessments affect educational policy and teachers' classrooms: one seeks to clarify the nature of the problem and the other assumes the problem while seeking to clarify potential solutions. Regardless of which path researchers and mathematics educators take, the results of the present study suggest a strong need for both.

References

Brickwedde, J. (2011). *Transition from additive to multiplicative thinking: A design and teaching experiment with third through fifth graders* (doctoral dissertation). Minneapolis, MN: University of Minnesota.

Carpenter, T. P., Coburn, T. G., Reys, R. E., & Wilson, J. W. (1975). Results and implications of the NAEP mathematics assessment: elementary school. *The Arithmetic Teacher, 22*(6), 438–450.

Carpenter, T. P., Kepner, H., Corbitt, M. K., Lindquist, M. M., & Reys, R. E. (1980). Results and implications of the second NAEP mathematics assessments: Elementary school. *The Arithmetic Teacher, 27*(8), 10–12, 44-47.

Chandler, C. C., & Kamii, C. (2009). Giving change when payment is made with a dime: The difficulty of tens and ones. *Journal for Research in Mathematics Education,* 97-118.

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. (1986). *Introduction to classical and modern test theory.* Mason, OH: Thomson Wadsworth.

Daro, P., Stancavage, F., Ortega, M., DeStefano, L. & Linn, R. (2007). *Validity study of the NAEP mathematics assessment: Grades 4 and 8* (research report). Palo Alto, CA: American Institutes for Research.

Ebby, C. B. (2005). The powers and pitfalls of algorithmic knowledge: a case study. *The Journal of Mathematical Behavior, 24*(1), 73–87.

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.

Hiebert, J., & Behr, M. (1988). *Number Concepts and Operations in the Middle Grades, Volume 2.* Research Agenda for Mathematics Education. National Council of Teachers of Mathematics, 1906 Association Drive, Reston, VA 22091.

Hughes, G. B., Daro, P., Holtzman, D., & Middleton, K. (2013). *A study of the alignment between the NAEP mathematics framework and the Common Core State Standards for Mathematics (CCSS-M)* (research report). Palo Alto, CA: American Institutes for Research.

Kamii, C., & Dominick, A. (1998). The harmful effects of algorithms in grades 1–4. *The teaching and learning of algorithms in school mathematics,19,* 130–140.

Kloosterman, P. (2015). An introduction to NAEP. In P. Kloosterman, D. Mohr, & C. Walcott (Eds.), What mathematics do students know and how is that knowledge changing?: Evidence from the National Assessment of Educational Progress (pp. 1–18). Charlotte, NC: Information Age Publishing.

Kloosterman, P., Mohr, D., & Walcott, C. (2015). Preface. In P. Kloosterman, D. Mohr, & C. Walcott (Eds.), *What mathematics do students know and how is that knowledge changing?: Evidence from the National Assessment of Educational Progress* (pp. vii–xii). Charlotte, NC: Information Age Publishing.

Kosko, K. W. (in review). *A multiplicative reasoning assessment for fourth and fifth grade students.*

Kosko, K. W., & Singh, R. (in press). Elementary children's multiplicative reasoning: Initial validation of a written assessment. *The Mathematics Educator.*

Kosko, K. W., & Singh, R., (2016). Colligation and unit coordination in mathematical argumentative writing. In Adolphson, K. V. and Olson, T. M. (Eds.), *Proceedings of the 43rd annual meeting of the Research Council of Mathematics Learning* (pp. 18–25), Orlando, FL: RCML.

Leinwand, S. (1994). Four teacher-friendly postulates for thriving in a sea of change. *The Mathematics Teacher, 87*(6), 392–393.

McNeal, B. (1995). Learning not to think in a textbook-based mathematics class. *The Journal of Mathematical Behavior, 14*(2), 205–234.

Mohr, D., Roach, M., & Hudson, R. A. (2015). Student performance in number and operations. In P. Kloosterman, D. Mohr, & C. Walcott (Eds.), *What mathematics do students know and how is that knowledge changing?: Evidence from the National Assessment of Educational Progress* (pp. 81–136). Charlotte, NC: Information Age Publishing.

Morris, A. K. (2009). Representations that enable children to engage in deductive argument. In D. A. Stylianou, M. L. Blanton, & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K–16 perspective* (pp. 87–101). New York: Routledge.

Mulligan, J. (2004) Key Concept of early number learning. In A. McIntosh, & L. Sparrow (Eds), *Beyond written computation* (pp.17–29). MASTEC, Edith Cowan University, Perth, Western Australia.

Mulligan, J., & Mitchelmore, M. (1997). Young children's intuitive models of multiplication and division. *Journal for Research in Mathematics Education, 28*(3), 309–330.

Narode, R., Board, J. & Davenport, L. (1993). Algorithms supplant understanding: Case studies of primary students' strategies for double-digit addition and subtraction. In J. R. Becker & B. J. Preece (Eds), *Proceedings of the Fifteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 254–260). San Jose, CA: Center for Mathematics and Computer Science Education, San Jose State University.

National Center for Education Statistics [NCES]. (2016). NAEP data explorer: Online questions tool. Retrieved from: http://nces.ed.gov/nationsreport card/itmrlsx/

Norton, A., Boyce, S., Phillips, N., Anwyll, T., Ulrich, C., & Wilkins, J. (2015). A written instrument for assessing students' units coordination structures. *Journal of Mathematics Education, 10*(2), 111–136.

Nunally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd Ed.). New York: McGraw-Hill.

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.

Outhred, L. N., & Mitchelmore, M. C. (2000). Young children's intuitive understanding of rectangular area measurement. *Journal for Research in Mathematics Education, 31*(2), 144–167.

Russell, S. J., Schifter, D., & Bastable, V. (2011). Developing algebraic thinking in the context of arithmetic. In J. Cai & E. Knuth (Eds.), *Early algebraization: A global dialogue from multiple perspectives* (pp. 43–70). Berlin: Springer-Verlag.

Singh, R., & Kosko, K. W. (2016). *Elementary students' understanding of equivalence and multiplicative reasoning.* Presented at the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Tuscan, AZ.

Sugrue, B., Novak, J., Burstein, L., Lewis, E., Koretz, D. M., & Linn, R. L. (1995). *Mapping test items to the 1992 NAEP mathematics achievement level descriptions: Mathematics educators' interpretations and their relationship to student performance* (research report). Washington, D.C.: National Center for Research on Evalaution.

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. (2007). *Problems in mathematics education.* Paper presented for the Senior Scholar Award for the Special Interest Group for Research in Mathematics Education (SIG-RME) at the annual conference of the American Educational Research Association, Chicago, IL.

Wilson, M. (2004). *Constructing measures: An item response modeling approach.* Mahwah, IL: Lawrence Erlbaum.