Developing Items to Assess Middle Grades Teachers

Identifying Attributes and Developing Items to Assess Middle Grades Teachers’ Multiplicative Reasoning

Andrew Izsák, Joanne Lobato
San Diego State University

Chandra Orrill
University of Massachusetts, Dartmouth

Erik Jacobson, Laine Bradshaw
University of Georgia

A growing number of projects, many funded by the National Science Foundation, have begun to combine research in mathematics education with psychometric models. Promises of such research include: (a) new insight about the distribution of mathematical knowledge in large populations, (b) empirical links between teachers’ mathematical knowledge and students’ achievement, (c) the development of measures that capture diagnostic information useful to school teachers or professional developers, and (d) demonstrations that more recently developed psychometric models can be harnessed to address challenges of assessing students’ and teachers’ knowledge of mathematics. Furthermore, the range of current and emerging psychometric models provides many possibilities for work and the intersection of mathematics education and psychometrics.

Over the past three decades, mathematics education has accumulated a research base on the teaching and learning of various content areas at a rapid pace. One measure of this growth is that new handbooks dedicated to research in mathematics education have been appearing every few years (e.g., English, 2003; Grouws, 1992; Lester, 2007). During this period, researchers have relied heavily on detailed case studies and other qualitative methods (e.g., Lesh & Kelly, 2000) to illuminate core phenomena, including the nature of mathematical thinking and problem solving, mechanisms of learning and transfer, students’ developmental trajectories in particular content areas, and knowledge that teachers need for their practice. Combining this research base with psychometric models could lead to new tools for mathematics education research, but challenges exist because it is not clear how well our understandings of mathematical cognition fit with currently available psychometric models.

Meanwhile, psychometric theory has moved from models based on test scores (e.g., classical test theory models), to models based on answers to individual test items (e.g., item response theory models), to models based on skills or components of reasoning required to answer particular test items (e.g., diagnostic classification models). Some researchers are pursuing applications of more established item response theory models to problems in mathematics education, while others are pursuing applications of more recent models, such as diagnostic classification models (DCMs). This paper describes work by the NSF-funded Diagnosing Teachers’ Multiplicative Reasoning (DTMR) project that is building a demonstration test form for middle grades teachers from the ground up within the DCM framework. The mathematical focus is fractions and proportions.

DO NOT CITE OR DISTRIBUTE WITHOUT PERMISSION
Two Examples of Multiplicative Reasoning with Quantities

The work reported here is motivated, in part, by standards for K-12 mathematics education articulated by the National Council of Teachers of Mathematics (NCTM; 2000). The NCTM standards emphasize the importance of both content strands—such as number and operations, algebra, and geometry—and processes of doing mathematics—such as problem solving, representing, and communicating. These standards are now reflected in several commercially available curricular materials, many of which were developed with support from the National Science Foundation. The examples below demonstrate that teaching and learning with these materials requires attention to more than algorithms for computing quickly and accurately.

The following examples illustrate some of the complexities that occur when multiplication is embedded in problem situations. The first example illustrates how the quantitative units become more complex as one moves from additive to multiplicative situations. Consider the following two problems:

1. Carrie has run 1/2 mile. If she walks another 1/3 mile, how far has she traveled?
2. John has 1/2 cup of flour for baking cookies. If each batch requires 1/3 cup of flour, how many batches can he make?

In the first problem, 1/2, 1/3, and the answer, 5/6, all refer to the same unit: one mile. In the second problem, 1/2 refers to the cups of flour John has, 1/3 refers to the cups of flour required per batch, and the answer, 3/2, refers to the number of batches that John can make. In contrast to Problem 1, each number in Problem 2 refers to a different unit. This implies that teachers must have strong command of referent units (the units to which numbers refer) if they are to use problem situations to teach arithmetic with whole numbers and with fractions.

The second example illustrates difficulties that can arise in classroom instruction when teachers do not have command of referent units. Izsák (2008) reported difficulties he observed as Ms. Archer introduced her sixth-grade students to fraction multiplication with a number line that showed a solution to 1/5 of 2/3 (see Figure 1). The drawing came from the teachers’ edition of the Bits and Pieces II unit in Connected Mathematics 2 (Lappan, Fey, Fitzgerald, Friel, & Phillips 2006). It showed the interval from 0 to 1 subdivided into thirds. Each third was further subdivided into 5 parts. Two parts were shaded as shown. Ms. Archer explained that the whole was divided into 3 equal parts, that the first third was divided into 5 equal parts, and that the whole was divided into 15 equal parts. She concluded that 1/5 of 1/3 was 1/15 and so 1/5 of 2/3 would be twice as much, or 2/15.

Figure 1. A number line model for showing 2/3 of 1/5.
One student’s questions about the demonstration suggested two sources of confusion. The first came from talking about 1/5 as 1/15: He thought 1/5 was equal to 3/15. The second came from shading two separate parts: He thought it made more sense to combine the two parts and think about 2/5 of 1/3.

Ms. Archer had trouble responding to her student both during the lesson and during a subsequent interview in which a researcher used video excerpts to stimulate Ms. Archer’s memory. First, Ms. Archer did not discuss explicitly the different referent units for each fraction in the problem. In a correct solution, 2/3 and 2/15 refer to parts of the whole, while 1/5 refers to one part of 1/3. The student seemed unaware of the different referent units, and Ms. Archer added to the confusion when, at one point, she discussed 3/15 as if the 15ths referred to parts of 1/3. Second, Ms. Archer had trouble with visual instantiations of the distributive property. Considering the student’s suggestion, recognizing that a drawing showing 2/5 of 1/3 could also show 1/5 of 2/3, and recognizing that 1/5 of 2/3 was the same as 1/5 of each of two 1/3s (as shown in Figure 1) would require a chain of reasoning that could be symbolized formally as:

\[
\frac{2}{5} \times \frac{1}{3} = \frac{1}{5} \times \frac{2}{3} = \frac{1}{5} \times \frac{1}{3} + \frac{1}{5} \times \frac{1}{3}.
\]

During her interview, Ms. Archer had trouble responding to her student’s thinking and reported that she had not considered the distributive property during the lesson.

Questions one might ask about Ms. Archer’s capacities include the following: Did she struggle to identify appropriate referent units in other situations as well—for instance, in addition/subtraction situations? Did she struggle in the same way with drawn models for fraction multiplication that used areas rather than lengths? Did she recognize the distributive property in other situations? Answers to these questions would give important information about her ability to interpret students’ thinking and enable their learning.

Other questions that one might ask about the population of upper elementary and middle grades mathematics teachers include: Is Mrs. Archer an anomaly or are her difficulties common among teachers who teach fraction arithmetic? What other mathematical ideas, in addition to referent units, would be important (and feasible) to assess in large samples to gauge teachers’ capacities to reason about mathematics in ways needed to respond to students’ thinking? We are attempting to combine mathematics education research with DCMs to answer such questions.

**Background**

This section reviews research on teachers’ knowledge of rational numbers (the mathematical focus of the DTMR project) and ways that researchers have measured teachers’ knowledge, including recent efforts that have used item response theory models. A main point is that mathematics education research contains numerous reports of teachers’ difficulties understanding arithmetic with fractions when numbers refer to units embedded in problem situations, but currently available measures of teacher’s
knowledge are not well-suited for answering the questions we asked above about the population of upper elementary and middle grades teachers.

**Teachers’ Capacities to Explain Fraction and Decimal Multiplication**

Although most teachers can calculate correctly the product of two fractions or two decimals, several studies (e.g., Armstrong & Bezuk, 1995; Eisenhart et al., 1993; Izsák, 2008; Sowder, Philipp, Armstrong, & Schappelle, 1998; Tirosh, 2000; Tirosh & Graeber, 1990) have reported constraints on inservice and preservice teachers’ performance when using drawings to explain such products. The example of Ms. Archer above provides one example. Armstrong and Bezuk gave inservice middle school teachers a word problem for which computing 1/3 of 3/4 would be appropriate. The teachers recognized that the situation called for fraction multiplication but had a hard time explaining their thinking, drawing diagrams to match algorithmic solutions, and understanding the appropriate unit or whole for the problem. Sowder, Philipp et al. reported difficulties that a practicing middle grades teacher, Linda, had explaining fraction multiplication with an area model during interviews and classroom instruction (p. 83-87). Eisenhart et al. reported that a preservice K-8 teacher, Ms. Daniels, began but could not complete an explanation for .7 x 2.35 using a rectangular region. Ball, Lubienski, and Mewborn, (2001) illustrated a similar difficulty with a classroom vignette in which a teacher struggled to make sense of products of decimals using dimensions and areas of rectangles.

Other studies (Behr, Khoury, Harel, Post, & Lesh, 1997; Izsák, 2008) have not only documented but also explained constraints on teachers’ performance by examining their capacities to form and transform conceptual units. Behr et al. interviewed 30 preservice elementary teachers on the operator subconstruct, one of five rational number subconstructs investigated by members of the Rational Number Project (see Behr, Harel, Post, & Lesh, 1992, for a discussion of the five subconstructs). In the case of the operator construct, a fraction is thought of as acting on a number, object, or set. For instance, one might consider taking ¾ of a set. Behr et al. (1997) gave preservice teachers 8 bundles of 4 sticks and posed tasks that involved taking ¾ of the collection. The researchers sought to characterize the teachers’ strategies and to determine the conceptual units that they formed and transformed during their solutions. A main finding was that these teachers found it hard to distribute fractions as operators across their conceptual units—for instance, by taking 3/4 of each bundle. This suggested constraints on the flexibility with which the teachers could form and transform conceptual units when using sticks to find a fraction of a whole number. Note that Ms. Archer evidenced similar constraints in the example above (see Figure 1).

**Teachers’ Capacities to Explain Fraction Division**

Most teachers can calculate correctly the quotient of two fractions or decimals, but several studies have reported constraints on teachers’ performance on fraction division tasks. A main finding reported across studies is that teachers confuse situations calling for division by a fraction with those calling for division by a whole number or multiplication by a fraction (Armstrong & Bezuk, 1995; Ball, 1990; Borko et al., 1992; Ma, 1999; Sowder et al., 1998). Ball asked 10 preservice elementary and 9 preservice secondary teachers to generate situations that would illustrate 1 3/4 divided by 1/2. Seventeen could compute
Developing Items to Assess Middle Grades Teachers - 5 -

the correct answer but only five were able to generate an appropriate word problem or situation. Five others generated situations that would illustrate $1 \frac{3}{4} \div 2$, one generated a situation that would illustrate $1 \frac{3}{4} \times 2$, and eight were unable to generate any situation, appropriate or inappropriate. Ma (1999) included the same fraction division task used by Ball (1990) in her investigation of 23 U.S. and 72 Chinese inservice teachers’ understandings of core mathematics topics taught in elementary grades. Ma confirmed constraints reported by Ball on U.S. teachers’ performance. Borko et al. reported a case in which the same Ms. Daniels was asked by a student to explain the invert and multiply rule for fraction division. The class had just computed the answer to $3/4 \div 1/2$. Ms. Daniels generated a situation and area representation that illustrated $3/4 \times 1/2$, realized that her example showed fraction multiplication instead of division, and was stumped. Sowder, Philipp et al. reported that a practicing middle-grades teacher, Cynthia, had trouble making appropriate connections between operations on fractional quantities and the arithmetic operations of multiplication and division (pp. 49-50).

**Written Assessments for Substantiating the Theory of Intuitive Models**

A further set of studies about teachers’ knowledge of arithmetic with rational numbers has used written assessments to substantiate a theory proposed by Fischbein, Deri, Nello, and Marino (1985) that people have intuitive models for arithmetic operations. The intuitive model for multiplication is repeated addition. The proposed models for division are based on sharing (partitive division) and measuring (quotative division).\(^{11}\) A main reason that researchers have been interested in intuitive models is that they support the operations of multiplication and division with whole numbers but are incongruous with applying these operations to fractions and decimals. As an example, based on their connections between multiplying whole numbers and repeated addition, students often think that multiplication always makes numbers larger. Fischbein et al. gathered evidence for their theory by administering word problems to 628 fifth-, seventh-, and ninth-grade students.

Other researchers (Graeber & Tirosh, 1988; Graeber, Tirosh, & Glover, 1989; Harel & Behr, 1995; Harel, Behr, Post, & Lesh, 1994; Harel, Post, & Behr, 1988; Post, Harel, Behr, & Lesh, 1991; Tirosh & Graeber, 1989) have adapted the word problems used by Fischbein et al. (1985) to investigate the extent to which elementary school teachers’ reasoning might be consistent with the same intuitive models. Main findings across these studies have been very consistent. For instance, with respect to multiplication, teachers have performed well on tasks where the multiplier was a whole number and almost as well on tasks where the multiplier was a decimal greater than one. In these same studies, teachers have found much harder word problems in which the multiplier was a decimal less than one. For instance, Graeber and colleagues (Graeber & Tirosh, 1988; Graeber et al., 1989) reported that teachers often used division inappropriately when working on problems like the following:

One kilogram of detergent is used in making 15 kilograms of soap. How much soap can be made from .75 kilograms of detergent? (Graeber & Tirosh, 1988, p. 264)
Similarly, teachers have performed well on partitive division word problems when the divisor was both a whole number and less than the dividend. Oftentimes, however, these same teachers have inverted the division (to divide a larger number by a smaller number) on problems like the following:

Twelve friends together bought 5 pounds of cookies. How many pounds did each friend get if they each got the same amount? (Graeber & Tirosh, 1988, p. 265)

**Written Assessments of Mathematical Knowledge for Teaching**

A second line of work that has developed further assessments for research purposes has roots in Shulman’s (1986) conceptualization of teachers’ knowledge. He proposed a set of knowledge categories that included subject-matter knowledge, pedagogical knowledge, and pedagogical content knowledge, among others. More recently, Ball and colleagues (Ball & Bass, 2000; Ball, Lubienski, & Mewborn, 2001; Ball, Thames, & Phelps, 2008) have developed a construct termed mathematical knowledge for teaching that refines some of the categories discussed by Shulman. Most relevant to the present report is the subdivision of subject-matter knowledge into common content knowledge and specialized content knowledge. Common content knowledge is knowledge of mathematics that many educated adults have and possibly use in a variety of professions—for instance, knowledge of procedures for computing with fractions. Specialized content knowledge is knowledge of mathematics that is used specifically in the work of teaching—for instance, knowledge that would support teachers’ efforts to analyze students’ novel approaches to computation and judge whether those approaches would generalize to other examples.

Ball and colleagues (e.g., Hill, 2007; Hill, Ball, & Schilling, 2008; Hill, Schilling, & Ball, 2004) have used subconstructs of mathematical knowledge for teaching, including common content knowledge and specialized content knowledge, to develop new measures of teachers’ mathematical knowledge intended for use with item response theory models. In contrast to the assessments discussed in the preceding section, these assessments use multiple-choice items and cover broader swathes of content. For instance, Hill, Schilling, et al. (2004) reported an instrument that reliably measures elementary teachers’ knowledge of number concepts; operations; and patterns, functions, and algebra. Subsequently, Hill (2007) reported a similar multiple-choice instrument for middle school teachers. One reason that mathematics education researchers have been interested in these instruments is that Hill, Rowan, and Ball (2005) demonstrated a positive correlation between teacher knowledge and student achievement. In particular, they reported that first- and third-grade teachers’ knowledge, as measured by their instruments, had an effect size comparable to SES on student gain scores on CTB/McGraw-Hill’s Terra Nova. The gains occurred over a one-year period in urban and suburban schools serving higher poverty populations. In essence, this result says that higher amounts of teachers’ mathematical knowledge for teaching are positively correlated with students’ test scores. Although the correlation is plausible, direct empirical evidence had eluded researchers for decades.
To summarize, many studies of teacher knowledge in this content area have reported constraints on teachers’ performance when numbers refer to units embedded in problem situations. Furthermore, current assessments of teachers’ knowledge in this content area do not concentrate on teachers’ capacities to reason with such units and do not provide information needed to answer questions about teachers’ reasoning like those we asked about Ms. Archer. The present study is a first attempt to develop an instrument that would provide such information.

**An Attribute-Based Approach to Assessing Middle Grades Teachers’ Multiplicative Reasoning**

This section describes our efforts to combine the mathematics education research base with DCMs to build measures of teachers’ knowledge of fractions and proportions. In contrast to the approach taken by Hill and colleagues (e.g., Hill 2007; Hill et al., 2004) in which researchers construct scales of broad knowledge categories—such as specialized content knowledge of number and operations—we are interested in measures that capture information about more fine-grained mathematical understandings. Such measures have potential to answer questions like the ones we asked about Mrs. Archer (see above). We also believe that such measures could be more sensitive to growth and change one might expect for teachers participating in professional development: One might see teachers learn to think about referent units in problem situations more consistently before they move along a scale of a broader knowledge category. With this goal in mind, we are developing a demonstration test form within the DCM framework.

DCMs are a family of recently developed psychometric models (see Leighton & Gierl, 2007; Rupp, Templin, & Henson, in press). What distinguishes DCMs from item response theory models is that DCMs estimate a “profile” of “attributes” that a person has “mastered.” A main point is that the profiles generated by DCMs have the potential to provide substantially more information than other psychometric models about teachers’ fine-grained understandings. When using DCMs to construct multiple-choice tests, content experts first specify “attributes” which are components of reasoning in a given domain. Test questions are then constructed around different subsets of attributes so that any given response, whether correct or incorrect, provides information about those attributes to which the teacher may be attending. Responses across all items on a test provide information about whether a teacher is or is not a “master” of each attribute. We interpret the statement that a teacher is a master of a particular attribute to mean that the teacher consistently uses that component of reasoning appropriately across situations. In cases where each attribute is treated as a dichotomous categorical variable (a person either is or is not a “master” of that attribute), a test built around $K$ attributes defines $2^K$ groups to which a teacher might possibly belong. Each group corresponds to a different “profile” that describes those attributes that teacher has and has not “mastered.” Thus, DCMs have potential to classify teachers in ways that provide information about multiple strengths and weaknesses in understanding.

**Identifying Attributes**

One challenge for developing attribute-based measures in mathematics education is the dearth of such measures: There are few models to follow for identifying attributes that
could serve as latent variables. The main example to date is de la Torre and Douglas’s (2004) application of DCMs to model data from Tatsuoka’s (1990) test about adding and subtracting mixed numbers. Sample attributes included (a) how to find common denominators and (b) how to subtract numerators. Whereas these attributes emphasize steps in numerical methods, the discussion of referent units and the example of Mrs. Archer (see above) demonstrates that facility with computation is only one aspect of proficiency with rational numbers.

Without clear models for identifying attributes, we had to devise our own method. We began with two basic criteria. First, we wanted attributes that captured important aspects of reasoning about fractions and proportions when numbers are embedded in problem situations. This is consistent with trying to measure knowledge that teachers need to respond to students’ reasoning when using standards-based circular materials. Second, we wanted attributes that cut across the typical delineation of topics, such as generating equivalent fractions; adding, subtracting, multiplying, and dividing fractions; and solving proportions. This is consistent with trying to measure more fundamental understandings that teachers need to reason across a variety of topics.

We identified 10 attributes by drawing on both the extant mathematics education research literature and our own research on students’ and teachers’ understandings of fractions and proportions. These 10 attributes span many of the important understandings that research suggests are critical for reasoning about rational numbers when they are embedded in problem situations. Although we are developing a test for teachers, we made use of research on students as well for two main reasons. First, research on students’ understandings in this domain is more extensive than research on teachers’ understandings. Thus, research on students provides a more elaborate picture of the necessary building blocks (and hence potential attributes) in our chosen domain. In essence, we used research on students to fill gaps in research on teachers. Second, teachers need to be able to reason about fractions and proportions in ways that are accessible to their students. Thus, it makes sense to assess the extent to which teachers can use ideas that are accessible to students when working with fractions and proportions.

Figure 2 shows that we organized the 10 attributes into three clusters. We emphasize that these 10 were a set of initial attributes on which to base our item development. As a result of the study described below we are learning more about characteristics that make attributes practical for measures development.

We have already described the Referent Units attribute and illustrated it with examples from Mrs. Archer’s classroom. We describe three more attributes in more detail, Norming, A one-Bths of M, and Reasoning Proportionally by Operating with Composed Units. These four attributes will suffice to illustrate lessons we have learned about identifying attributes for measures development. A main lesson that we illustrate in the results section is that translating mathematics education research into attributes that serve as latent variables is a very difficult task.
**Section** | **Attribute**
--- | ---
Reasoning with Units | 1 Norming
| 2 Referent Units
| 3 Nested Units
Interpreting Rational Numbers & Proportional Reasoning | 4 A one-Bths of M
| 5 Reasoning Proportionally by Operating with Composed Units
| 6 Reasoning Proportionally by Using Multiplicative Comparisons
| 7 Reciprocal Relations of Relative Size
Creating Relationships & Determining Appropriateness | 8 Connections among Fractions, Ratios, Decimals, and Quotients
| 9 Equivalence
| 10 Appropriateness

*Figure 2. Initial set of attributes*

**Norming.** This attribute refers to establishing the whole that serves as the unit of measurement (see Lamon, 1994, 2007 p. 644, for discussions of norming). We are interested in two particular cases. The first case involves selecting the whole from alternative choices. One common instructional activity for which teachers have to be skilled at norming is using base-10 blocks (see Figure 2) to represent decimal numbers. To represent 6.54, one could let the flat represent 1, the rod represent one tenth, and the cube represent one hundredth (6 flats, 5 rods, and 4 cubes). Alternatively, one could let the block represent 1, the flat represent one tenth, and the rod represent one hundredth (6 blocks, 5 flats, and 4 rods). Sometimes teachers are overly restricted in their interpretations of base-10 blocks and think that only the cube can represent 1.

![Base-10 blocks](image)

*Figure 2. Base-10 blocks.*

The second case has to do with making alternate choices for the whole. The following common type of problem involves shifting explicitly from one choice for the whole to a second:

Sam and Morgan are comparing the amount of liquid in their beakers as shown in the diagram below. Sam claims that Morgan has 20% less than she has. Morgan claims that Sam has 25% more than she has. Who is
right?

To see that both Sam and Morgan can be right requires first using Sam’s amount of liquid as the whole and then using Morgan’s amount of liquid as the whole.

**A one-Bths of M.** One common way that elementary and middle school teachers and students interpret the fraction A/B is as pairs of whole numbers where B indicates the cardinality of a set and A indicates the cardinality of a subset. A teacher might illustrate this meaning to students using the example of cutting a pizza into 8 equal-sized parts and taking 3 of them. This interpretation of fractions can make it hard for students to make sense of improper fractions: How could you take 9 slices from the same pizza? A one-Bths of M refers to an alternative interpretation of fractions that emphasizes A copies of the unit fraction one-Bth of some quantity M (e.g., Thompson & Saldanha, 2003). Using this interpretation, one can interpret 3/8 as 3 one-eighths and 9/8 as 9 one-eighths. The M underscores that the one-Bths might not be of the whole—for instance, one could think of 9 one-eighths of 2 acres. Also, in contrast to emphasizing the cardinality of sets, the A one-Bths of M interpretation emphasizes the interpretation of fractions as statements about multiplicative relationships between parts and wholes. Teachers need to be able to reason with the A one-Bths of M interpretation of fractions so that they can facilitate the development of this fundamental way of thinking in their students.

**Reasoning Proportionally by Operating with Composed Units.** Proportional reasoning involves reasoning with equivalent ratios, and teachers need ways to think about such ratios that are accessible to students. One way to conceive of a ratio that is accessible to students is operating on composed units (e.g., Lamon, 1994, 1995; Lobato & Thanheiser, 2002; Olive & Lobato, 2008). As an example, consider a situation in which a character travels 10 centimeters in 4 seconds. The distance and time can be joined in a new unit, denoted here as a (10:4) unit. Students can generate equivalent speeds by joining replicates (e.g., joining three copies of a (10:4) unit to produce a (30:12) unit) or by partitioning the original unit into equal-sized parts (e.g., partitioning a (10:4) unit into four (2.5:1) units). One can also combine a (10:4) unit with a (2.5:1) unit to create a (12.5:5) unit that also describes the same speed. Composed unit reasoning preserves multiplicative comparisons while remaining accessible to students who are just learning to reason proportionally.

**Item Development**
Once we had identified our initial set of 10 attributes, we spent 3 months drafting multiple-choice items intended to target either individual attributes or combinations of
attributes. For each item we wrote alternate choices so that they, too, would provide information about the attributes. We planned to score items dichotomously, initially, but anticipated using nominal scoring as an option in the future. We met weekly to discuss ways to strengthen draft items and to brainstorm ideas for further items on which project members could work in between meetings. As is often the case in test development, we set aside many items that we drafted because of flaws that we could see even before trying them out with teachers.

The following two examples illustrate our approach to item design. They are not actual items that we used in the pilot test forms, but they are very similar. The first item (Figure 4) is intended to measure the Norming and A one-Bths of M attributes. The second item (Figure 5) is intended to measure the Referent Units and Composed Unit Reasoning attributes.

Below is a group of 5 circles followed by three interpretations. Which of the interpretations are sensible?

![Diagram of circles](image)

I. The diagram can show $\frac{3}{5}$.
II. The diagram can show $1\frac{2}{3}$.
III. The diagram can show $\frac{5}{2}$.

Choices:
A  I only.
B  I and II only.
C  I, II, and III.
D  None of the above.

*Figure 4. A sample item intended to measure Norming and A one-Bths of M.*

The correct answer to this item is choice C. To see this, a teacher would have to first take the entire collection of 5 circles as the whole (to see $\frac{3}{5}$), then take the 3 black circles as
Developing Items to Assess Middle Grades Teachers - 12 -

the whole (to see $1\frac{2}{3}$), and then take the 2 white circles as the whole (to see $\frac{5}{2}$). Because interpreting statement III requires thinking about 5 copies of $\frac{1}{2}$ of 2 circles, selecting this choice indicates both the Norming and the A one-Bths of M attributes. A teacher who selected choice A would not demonstrate the norming attribute because only one choice for the whole was made. A teacher who selected choice B would demonstrate norming, but A one-Bths of M is not required for interpreting mixed numbers. We often included choices like choice D so that teachers were not forced into choosing an answer that, in fact, they did not understand.

Milo is going to make a batch of his favorite chocolate brownies. He wants to make a batch that is $\frac{12}{5}$ the amount of the original recipe. To make $\frac{2}{3}$ of the batch he wants to make, he knows he needs 16 ounces of water. How much water is needed for the original brownie recipe?

![Diagram](image.png)

Choices:
A 8 ounces.
B 10 ounces.
C 24 ounces.
D None of the above.

*Figure 5.* A sample item intended to measure Referent Units and Composed Unit Reasoning.

The correct answer for this item is choice B. To see this, a teacher might reason from a unit that composed 16 oz. with $\frac{2}{3}$ batch, denoted here as (16 oz.:2/3 batch). The teacher might partition this unit to form an (8 oz.:1/3 batch) unit and then make copies of the result to form a (24 oz.:1 batch) unit. The teacher would then have to combine this composed unit reasoning with attention to referent units: The 1 refers to one batch, not one original recipe. To complete the solution, the teacher could partition the (24 oz.:1 batch) unit into 12 parts to create a (2 oz.:1/5 recipe) unit, and then make copies to form a
Developing Items to Assess Middle Grades Teachers - 13 -

(10 oz.:1 recipe) unit. Choices A and C correspond to different points in this correct solution, and therefore indicate composed unit reasoning, but not appropriate attention to the original recipe as the appropriate referent unit. Again, we included choice D so that teachers were not forced into choosing an answer that, in fact, they did not understand.

![Figure 6. A Q-matrix for a Pilot Form](image)

After 3 months, we had generated enough items to construct two pilot forms that concentrated on just the first six attributes. We wanted to see how well our items functioned and if course corrections were necessary before writing items that covered the full set of 10 attributes. Because we were interested in piloting as many items as possible, we did not include any common items. Rather we tried to create two forms so that the numbers of items that loaded onto each attribute were similar. For each form we generated a Q-matrix (see Figure 6). Q-matrices summarize how test items load onto
attributes and play a central role in the DCM framework. Numbers indicate items, and letters indicate part of a testlet. An “x” indicates a place where an item does load onto an attribute.

Data
Once we had assembled the two pilot forms, we contacted a convenience sample of 14 middle grades teachers in California and Georgia. We asked the teachers to complete the test and a survey of professional history. The survey included questions about certification, grades taught, years of teaching experience, and hours spent during the previous year in professional development. We mailed the pilot test forms to teachers and asked them to find a block of about 1.5 hours to complete the test. As an incentive, we paid teachers $100 as a token of appreciation for the time it took them to complete the test and an interview described next.

After teachers completed the test, we conducted Item Response interviews with each. The interviews took place 1–3 days after teachers completed the test. During the interviews, we asked teachers how they interpreted each item and how they selected their answer choice. Typical questions that we asked included “Can you tell me why you selected choice A?,” “Why did you not select choice C?,” and “Was there something in the wording of the question that was confusing to you?” We also videotaped the interviews because, as the sample items above illustrate, many of our items asked teachers to interpret drawn models (e.g., length and area models) and, in order to understand their explanations, we needed to record their hand gestures and any writing that they used. These interviews allowed us to see if teachers interpreted the question statements as we intended, were confused by the wording of a question or a particular drawing, and used our intended attributes when selecting their choices.

Analysis and Results
We analyzed the interview data item-by-item, compiling the responses of all 14 teachers. We used talk, gesture, and inscription as evidence for the teachers’ reasoning and, in particular, whether they used the attributes we intended to select an answer choice or, instead, used some other understandings. The main thing we wanted to learn was whether answer choices accurately reflected teachers’ “mastery” of our intended attributes. We found that some of our items performed well, while others either underestimated or overestimated teachers’ use of our intended attributes. For some of our items, teachers demonstrated a range of strategies that we did not fully anticipate. We are using these results to identify characteristics of items and attributes that could lead to more consistent success in our next iteration of item writing and piloting. That work is currently under way.

Norming. Some items intended to target the Norming attribute worked better than others. Tasks that asked teachers to reason about base-10 blocks (see Figure 2) did seem to accurately assess teachers’ use of the Norming attribute. That is, teachers would indicate different choices for the whole through statements like “if the rod is the whole, then… if the block is the whole, then….” Alternatively, they might make comments like “if this is the whole, then… but if this is the whole then…” while using hand gestures or a pencil to
indicate which shape they were taking as the whole. When teachers got such items wrong, they typically demonstrated inflexible choices for the whole.

Other items worked less well, including items that resembled the Circles item (see Figure 4). Teachers could readily see \( \frac{3}{5} \) in the diagram, but they sometimes had trouble taking the two white circles as a whole. A main source of difficulty was that some teachers interpreted these circles as being “empty.” The immediate problem with the item is fixed easily by adding shading to the two white circles.

The larger lesson is that teachers may make a different set of assumptions about (or use a different set of conventions for) interpreting drawn models than those we use. Furthermore, these differences may interfere with accurately assessing teachers. We have uncovered other instances of this general issues in further items. Our goal is to minimize difficulties interpreting drawings, but we do not expect to eliminate them entirely.

**Referent Units.** Items intended to target the Referent Units attribute were the most successful at diagnosing teachers. That is, when teachers selected correct choices, the explanations they gave made explicit the units that they were attaching to numbers. When teachers selected incorrect choices, they typically did so either because they attached the wrong units to a number or expressed uncertainty about which units were appropriate. Consistent with research on teachers’ knowledge reviewed above, the teachers we interviewed had particular difficulty understanding referent units in division situations. One reason why these items tended to be successful was that we were able to craft them in such a way that computing answers using numeric methods did not help teachers select among the choices.

**A one-Bths of M.** Items intended to target A one-Bths of M were less successful than items intended to target the Norming or Referent Units attributes. One way that a person can demonstrate A one-Bths of M is by partitioning a quantity M to find one-Bth of that quantity and then iterating the one-Bth to concatenate A copies. The evidence is particularly strong if the person partitions and iterates to construct an improper fraction. We wrote items that we thought would require the use of unit fractions to answer correctly, but we were not always successful.

Two related phenomena obscured our access to teachers understanding of fractions as multiplicative relationships between parts and the whole. First, teachers we interviewed rarely found unit fractions (one-Bth) as an intermediate step when working on our items. Instead, they often had other means of answering items correctly. Second, we observed several cases where teachers reasoned about improper fractions as mixed numbers. To illustrate, a teacher might reason about \( \frac{7}{4} \) pizzas by converting to \( 1\frac{3}{4} \) pizzas, saying that there is one whole pizza and three out of four slices of a second pizza. In such cases, teachers appeared to interpret fractions as pairs of whole numbers in which the denominator indicated the cardinality of the set and the numerator indicated the
cardinality of the subset. This approach circumvented the need to interpret fractions as statements about multiplicative relationships between parts and the whole.

One lesson one might draw from our progress thus far is that we simply need to be more ingenious when writing multiple-choice items so that the correct answer will more clearly separate teachers who “have” the A one-Bths of M attribute from those who do not. Another lesson one might draw is that the evidence needed to make this separation can only be found in the process by which teachers answer an item. Currently, we are using constructed response items in a second round of interviews with teachers and are finding that evidence for whether teachers do or do not use A one-Bths of M is subtle. First, teachers do not talk about the A one-Bths of M interpretation of fractions as explicitly as they do the pairs of whole numbers interpretation. Teachers often discuss the pairs of whole numbers explanation with their students, so it makes senses that they articulate this understanding readily. The teachers we have interviewed are less articulate about finding and interpreting unit fractions. Nevertheless, we sometimes see evidence of iterating unit fractions by watching how teachers produce drawings and use hand gestures. The challenge then is devising items that allow teachers to demonstrate more tacit understandings of fractions. Furthermore, this suggests that constructed response items might be better suited for this attribute (even though that means that scoring takes more time).

**Composed Units.** Items intended to target the Composed Units attribute have also been less successful. The main challenge is that many of the teachers we interviewed relied heavily on numeric computation to make sense of proportional situations. To illustrate, in the case of items similar of the Brownie item (see Figure 5), teachers tend to set up and solve an equation similar to \( \frac{12}{5} = \frac{24}{x} \). Teachers seem to have at least as much difficulty (if not more difficulty) reasoning about proportions as they do reasoning about fractions when numbers refer to units embedded in problem situations.

When interviewing teachers, we can see more clearly if teachers can or cannot reason with composed units either by asking teachers not to compute answers using numeric methods or by allowing them to compute answers first and then asking if they have other ways to answer the same item.

Currently, we are conducting a second round of interviews with teachers using constructed response items and are finding that teachers can only reason with composed units in very limited ways. They can form a composed unit and halve, double, or triple it, but they are not able to reason about more complicated relationships such as \( \frac{5}{8} \) of a composed unit.

**A Fundamental Challenge.** A fundamental challenge that we have encountered is writing items that cannot be answered by simple numeric calculation. We know that teachers often rely on calculation to confirm correct answers and, in some cases, they seem to have few other resources for answering our items. In other cases, teachers seem
to need to know the numerical answer to a computation before they feel comfortable reasoning directly with drawn models included in our items. This is a problem because it obscures our view of teachers’ capacities to reason with units embedded in problem situations.

The primary strategy we have tried to date for addressing this challenge is to extend what worked well for our referent unit items. Recall that those items were successful, in part, because we were able to write items for which computing answers using numeric methods did not help teachers select among the choices. One version of this strategy has been to write items for which all of the answer choices refer to the same, correct numeric answer. A second version of this strategy has been to write items for which all of the answer choices refer to different, incorrect numeric answers. Oftentimes, to accomplish this, we have written answer choices that describe students’ strategies and have asked teachers to evaluate those strategies. To select among the choices, teachers have to comprehend students’ thinking and judge whether the described mathematical thinking is fundamentally sound. As an example, a student might begin to use correct composed unit reasoning but make a computation error along the way to an incorrect answer. Unfortunately, this approach to item writing has at least two downsides. First, the reading demands increase considerably. Second, we have found that teachers can be reluctant to judge students’ strategies as incorrect. This can occur when teachers do not follow a students’ strategy, and it can occur when teachers adopt a nurturing stance toward students in which they encourage students’ efforts regardless of the soundness of the underlying mathematical reasoning.

In work currently underway, we are conducting further interviews with teachers to get a better handle on understandings that seem to separate teachers (especially with respect to proportional reasoning), the attributes on which we are basing our item writing, and the item formats that we are using. Such experimentation seems essential to figuring out how to harness the mathematics education research base to develop reliable and valid measures that produce profiles within the DCM framework. Again, the promise of this work is that, if successful, such profiles could provide important information about teachers’ capacities to reason about rational numbers embedded in problem situations as required by standards-based curricular materials.

Acknowledgement
The research reported in this article was supported by the National Science Foundation under Grant No. DRL-0903411. The opinions expressed are those of the authors and do not necessarily reflect the views of NSF. Correspondence should be addressed to Andrew Izsák, Center for Research in Mathematics and Science Education, San Diego State University, 6475 Alvarado Road, Suite 206, San Diego, CA 92120.
Developing Items to Assess Middle Grades Teachers - 18 -

References


Developing Items to Assess Middle Grades Teachers - 19 -


---

i Connected Mathematics is one of the NSF-funded standards-based curricular materials for middle grades mathematics.

ii The operation of division can answer two types of questions when numbers are embedded in situations. Quotative (sometimes called measurement) division answers the question how much of one quantity is in a second quantity—for instance, if each car can take 5 people and 20 people want to go on a field trip, how many cars are needed (how
many 5’s are in 20)? Partitive (sometimes called sharing) division answers the question how many units of one quantity are associated with one unit of a second quantity—for instance, if 20 people want to go on a field trip and there are 5 cars, how many people should get into each car (how many people in each of 5 cars)? See Fischbein et al. (1985) or Greer (1992) for further discussion of the distinction between these two forms of division.