DIAGNOSING REASONING TO MEASURE GROWTH IN PRE-SERVICE MIDDLE-GRADES TEACHERS' FACILITY WITH FRACTION ARITHMETIC

Andrew Izsák
The University of Georgia
izsak@uga.edu

Sybilla Beckmann
The University of Georgia
sybilla@math.uga.edu

Laine Bradshaw
The University of Georgia
laineb@uga.edu

The present study extends recent advances coordinating research on cognition and psychometric modeling around fractions. Recent research has demonstrated that the Diagnosing Teachers' Multiplicative Reasoning Fractions survey provides information about distinct components necessary for reasoning in terms of quantities when solving fraction arithmetic problems. The present study (a) adds a new component of validity for the survey and (b) examines the utility of the survey as a measure of growth in pre-service middle-grades teacher's facility with fraction arithmetic as they completed a 1-semester content course. Results provide an existence proof that that the survey is sensitive to shifts towards more proficient reasoning.

Keywords: Rational Numbers, Research Methods, Teacher Education-Preservice, Teacher Knowledge

One central challenge for mathematics education is fostering teachers' and students' capacities to reason about multiplicative relationships in terms of quantities. Relevant topics include whole-number multiplication and division, arithmetic with fractions, proportional relationships, and linear functions. Reasoning with quantities is emphasized in recent curriculum standards documents (e.g., Common Core State Standards Initiative, 2010; National Council of Teachers of Mathematics, 2000) and recommendations for teacher education (e.g., American Mathematical Society, 2010; Sowder et al., 1998). These standards and recommendations place high value on developing conceptual understanding by solving and reflecting on solutions to problems couched in quantities.

A second central challenge is developing ways to coordinate research on cognition with the increasing variety of available psychometric models (e.g., Izsák, Remillard, & Templin, in press). As explained below, most recent applications of psychometric models to mathematics education research, especially to research on teachers, have relied on well-established item response theory (IRT) models to measure knowledge. These models report single scores that locate examinees on uni-dimensional continuous scales. A recently developed family of psychometric models, called diagnostic classification models (DCMs), trade continuous for categorical variables in exchange for measuring multiple dimensions within practical testing conditions. Instead of reporting single scores, DCMs report profiles of strengths and weaknesses on several dimensions simultaneously. Bradshaw, Izsák, Templin, and Jacobson (2014) reported on the Diagnosing Teachers' Multiplicative Reasoning (DTMR) Fractions survey that was developed for use with DCMs and measures teachers' capacities to reason about multiplication and division of fractions in terms of quantities. That study reported on content and item-level validity of the survey and results of analyzing a national sample of 990 in-service middle grades teachers.

The present study extends research on the DTMR Fractions survey by applying it directly to mathematics teacher education. In particular, we administered the survey at the beginning and end of a 1-semester content course focused on arithmetic with rational numbers and asked two questions:

1. To what extent were the DTMR Fractions profiles consistent with pre-service teachers' reasoning across multiple survey items and related tasks?
2. Did the distribution of DTMR Fractions profiles shift after the 1-semester course on number and operations and, if so, how?

Wood, M. B., Turner, E. E., Civil, M., & Eli, J. A. (Eds.). (2016). *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Tucson, AZ: The University of Arizona.

The first question addresses additional aspects of validity not taken up by Bradshaw et al., and the second question asks what growth and change might be captured by the DTMR Fractions survey that would be obscured by measures designed for uni-dimensional IRT models.

Measuring Teachers' Mathematical Knowledge for Fraction Arithmetic

In the last 2 decades, researchers have made important advances conceptualizing teacher knowledge that supports student learning. Well-known examples include Shulman's (1986) knowledge categories, especially pedagogical content knowledge, and Ball and colleagues' (e.g., Ball, Thames, & Phelps, 2008) subsequent articulation of mathematical knowledge for teaching (MKT). New conceptualizations of teacher knowledge have spurred, in turn, new approaches to measuring teachers' mathematical knowledge (e.g., Baumert et al., 2010; Hill 2007; Kersting, Givven, Sotelo, & Stigler, 2010; Saderholm, Ronau, Brown, & Collins, 2010; Shechtman, Roschelle, Haertel, & Knudsen, 2010). Most of these measures have been developed for use with traditional, uni-dimensional IRT models.

In contrast to the teacher knowledge measures mentioned above, the DTMR Fractions survey captures information about four components for reasoning about fraction multiplication and division in terms of quantities. The solution outlined to the following problem illustrates what we mean by components of reasoning:

> A batch of brittle calls for ¼ of a cup of honey. Megan has 2/3 of a cup of honey. How many batches of brittle can Megan make?

The solution we present presumes a teacher providing opportunities for students to solve fraction division problems before learning a general numeric method, such as multiplying by the reciprocal of the divisor. Thus, the teacher and students have to reason with quantities directly, which is consistent with the curriculum standards and recommendations for teacher education mentioned above.

First, to see the opportunity for discussing division, a teacher would have to recognize that the Brittle problem asks the signature how-many-groups question for measurement division. To support students' reasoning with quantities in the problem, some drawn model would be useful. The rest of the solution we present makes use of double number lines, but our larger points are not dependent on this choice of drawn model. Figure 1a shows a double number line that uses lengths to depict cups of honey. One number line represents cups subdivided into thirds, and one represents cups divided into fourths. Juxtaposing the two number lines highlights the challenge that fourths and thirds do not subdivide one another evenly. Figure 1b illustrates how partitioning thirds into 4 parts and fourths into 3 parts creates a finer unit, twelfths, which subdivide both thirds and fourths. The final challenge is to interpret the mini-pieces in terms of the given situation. There are multiple candidates, including interpreting one mini-piece as a twelfth of 1 cup, as a fourth of 1/3 cup, and as a third of ¼ cup. Because the problem asks about the number of ¼ cups in 2/3 cups, ¼ cup is the appropriate referent unit: There are 8/3 ¼-cups in 2/3 cups.

This is not the only solution to the Brittle problem, but it does illustrate that constructing a solution requires multiple, constituent components of reasoning. The given solution highlights the ability to (a) recognize the appropriateness of an arithmetic operation for modeling a given problem situation, (b) use whole-number factor-product combinations as a resource for partitioning quantities, and (c) identify appropriate referent units for each number. A measure of knowledge for this domain designed for traditional, uni-dimensional IRT models would collapse information about these different components into a single score.

Wood, M. B., Turner, E. E., Civil, M., & Eli, J. A. (Eds.). (2016). *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Tucson, AZ: The University of Arizona.

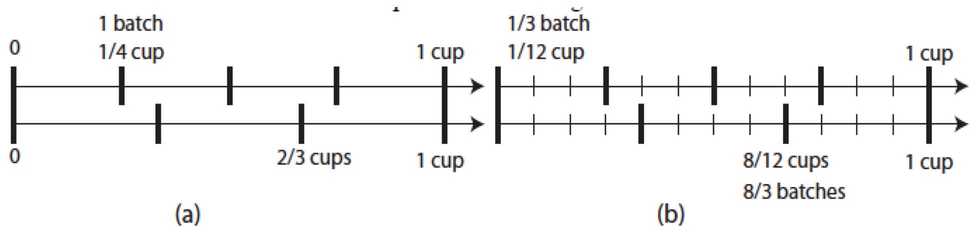

Figure 1. Reasoning with a double number line.

The DTMR Fractions survey consists of 27 items that measure these three components (termed appropriateness, partitioning and iterating, and referent unit, respectively) and one more, reversibility, which is important for solving partitive division problems. Some items measure just one of the four components, referred to as attributes. Other items measure more than one attribute at the same time. Nineteen of the items are multiple choice, and eight are constructed response. Figure 2 shows a released item that measures two attributes simultaneously: Selecting the correct answer choice, b, requires identifying the correct referent unit for 1/8 and partitioning intervals appropriately. The DTMR Fractions survey systematically elicited combinations of the four target attributes and was designed for use with DCMs, which report profiles of strengths and weaknesses on the multiple attributes instead of single scores on one overall dimension.

Ms. Roland gave her students the following problem to solve:

> *Candice has 4/5 of a meter of cloth. She uses 1/8 of a meter for a project. How much cloth does she have left after the project?*

She had students use the number line so that they could draw the lengths. Which of the following diagrams shows the solution? Assume all intervals are subdivided equally.

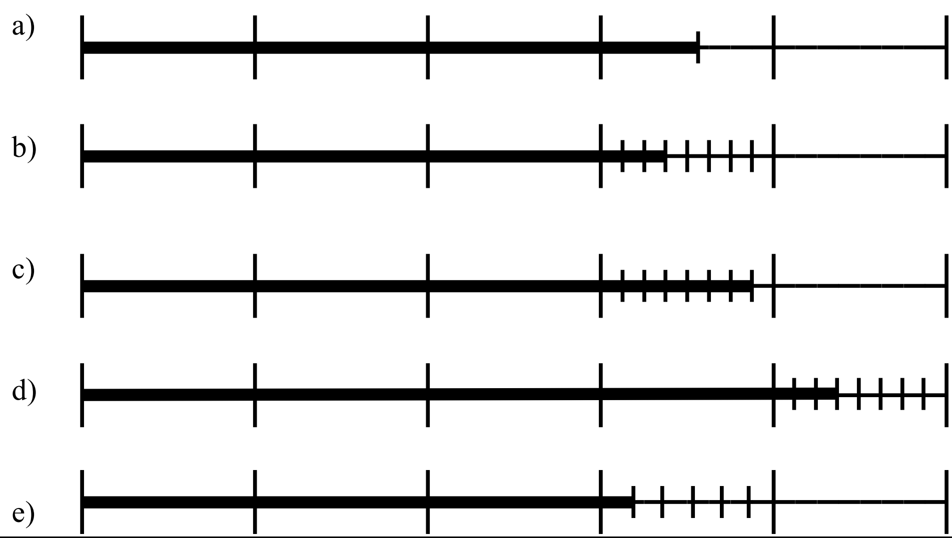

Figure 2. DTMR Fractions item. From Izsák, Jacobson, de Araujo, and Orrill (2010). © 2010 by University of Chicago. All rights reserved.

Methods

Data for the present report come from a larger on-going study of pre-service teachers' reasoning about multiplication and division, fractions, and proportional relationships. We administered the DTMR Fractions survey to a cohort of 22 pre-service middle-grades mathematics teachers before and

Wood, M. B., Turner, E. E., Civil, M., & Eli, J. A. (Eds.). (2016). *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Tucson, AZ: The University of Arizona.

after a number and operations content course offered in Fall 2014. The course was part of a teacher education program in a large, public university in the Southeast United States and emphasized reasoning with quantities to develop conceptual understanding of multiplication and division with whole numbers and with fractions. A following algebra course, offered in Spring 2015, focused on proportional relationships and linear equations. Beckmann taught both courses. For the present report, we focus on data from the number and operations course.

The pre-service teachers completed the DTMR Fractions survey for the first time as the first homework assignment in the number and operations course (August 2014). They were told that the survey was a formative assessment and that they should simply do their best. Bradshaw analyzed the survey data (item responses) to estimate DTMR Fractions profiles for all 22 pre-service teachers. Given the size of the national sample (n = 990), the DCM analysis was limited to treating each attribute as a dichotomous variable. Bradshaw then selected six pre-service teachers with a variety of initial profiles to be focal participants for this study. She revealed the overall distribution of DTMR profiles to Izsák and Beckmann but not profiles for particular individuals. Thus, Izsák and Beckmann were blind to the profiles until the end of the study.

Izsák and Beckmann then conducted a series of six individual video-recorded interviews with each of the six focal participants. The interviews began during the number and operations course and continued into the algebra course. In addition to the interviews, we collected written artifacts (e.g., homeworks, quizzes, and tests) from all pre-service teachers in the course. The pre-service teachers completed the DTMR Fractions survey for the second time as the first homework assignment in the algebra course (January 2015). Bradshaw analyzed the item responses to estimate DTMR Fractions profiles again for all 22 pre-service teachers, but still reported only the overall distribution of profiles, not those for specific individuals.

Within 1 week after the pre-service teachers completed the DTMR Fractions survey for the second time, Izsák interviewed each of the six focal participants about their item responses. This was the third in the series of six interviews. Because the DTMR Fractions survey takes about 1.5 hours to complete, Izsák and Beckmann selected a subset of the items to address each of the four attributes, referent unit, partitioning and iterating, appropriateness, and reversibility. During the first half of the interview, Izsák went through the selected items and limited follow-up questions to ones like *How did you interpret the problem?*, *Why did you select the choice you did?*, and *For what reasons did you reject alternate choices?* He made clear that the pre-service teachers could change their answers if they wanted. Thus, the interview data were more reliable for inferring the focal participants' thinking at the time of the interview, not when they completed the survey for homework. (At the same time, the pre-service teachers did not experience instruction between completing the survey for homework and the interview that would have provided additional practice with the four attributes.) During the second half of the interview, Izsák took a second pass through the same items and asked more extensive follow-up questions. He also presented additional tasks not part of the DTMR Fractions survey to gather further information about the participants' use of the four attributes. Each interview lasted about 80 minutes.

Izsák and Beckmann then analyzed the interview data to diagnose profiles. For each participant, they analyzed the interview data item-by-item listing evidence for and against mastery of each attribute. (The term *mastery* comes from the psychometric literature, which we take as a synonym for *proficiency*.) They also included a category called "other comments." Izsák and Beckmann included evidence for and against mastery wherever they found it both in the first and the second halves of the interviews. Izsák and Beckmann first watched the interviews separately and then together to compare notes. Discrepant interpretations were discussed until resolved. Once Izsák and Beckmann had discussed the entire video, they looked at all evidence for and against mastery of each attribute. They used a 1 to indicate mastery, a 0 to indicate non-mastery, and X to indicate cases where there was too little evidence or there was a balance of contrary evidence for and against mastery of a particular

Wood, M. B., Turner, E. E., Civil, M., & Eli, J. A. (Eds.). (2016). *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Tucson, AZ: The University of Arizona.

attribute. Thus, Izsák and Beckmann allowed themselves more flexibility in diagnosing profiles than was permitted in the DCM analysis. Izsák and Beckmann gave their expert diagnoses for all six focal participants to Bradshaw, who then revealed the profiles as determined by the DCM analysis.

Finally, we underscore that we take the diagnosis of attribute mastery (indicated by a 1) to mean that the person uses that particular attribute fairly consistently across situations. In contrast, we take a diagnosis of non-mastery (indicated by a 0) to simply mean that there was a lack of evidence for a particular attribute. This is not the same as saying a person does not “have” the attribute: A person might have but not demonstrate a particular piece of knowledge, even across multiple tasks, for a variety of reasons.

Results

Our first research question asked to what extent were the DTMR Fractions profiles consistent with pre-service teachers’ reasoning across multiple survey items and related tasks? Answering this question extends prior research that focused on content and item-attribute validity of the DTMR Fractions survey. Content validity, that is confirming that the DTMR Fractions items did indeed address important aspects of reasoning about fraction arithmetic in terms of quantities, was established by six external reviewers (four mathematicians and two mathematics education researchers) who examined the items and a document explaining the four attributes. Item-attribute validity, that is confirming that each item addressed the intended attributes, was established through interviews with teachers as part of the item development process and confirmed by DCM analysis of item response data from the large national sample. Because teacher interviews focused only on a subset of items as part of the item development process, and before the national sample was collected, there was no way diagnose their profiles either through expert analysis of interview data or statistical analysis of item response data.

The interview data afforded comparison of profiles as diagnosed by experts (Izsák and Beckmann) reviewing performance across multiple tasks and as diagnosed by analyzing item responses using a DCM. When considering places where expert and DCM diagnoses did and did not align, it is important to remember that each was based on overlapping but not identical information. Expert diagnoses were based on a subset of item responses, item responses revised during the interview that might indicate false positives or negatives, and information from the second half of the interviews that contained additional tasks not on the DTMR Fractions survey. DCM diagnoses were based on responses to the complete set of survey items and statistical information about item characteristics.

Table 1 shows profiles for the six focal participants as diagnosed with the DCM analysis before and after the number and operations content course and as diagnosed by Izsák and Beckmann based on the interview described above. Each string indicates a diagnosis of mastery (1), non-mastery (0), or uncertain mastery status (X) for the four attributes in the following order: referent unit, partitioning and iterating, appropriateness, and reversibility. Given six participants and four attributes, Izsák and Beckmann considered 24 diagnoses. They assigned a 1 in 10 cases, a 0 in 7 cases, and an X in the remaining 7 cases. Of the 17 cases where they assigned a 1 or 0, their diagnoses and the profiles Bradshaw generated with the DCM analysis agreed in 13. The four discrepancies were concentrated in two pre-service teachers, Alex and Diana, and we explain them as follows. Based on explanations Alex provided during her interview, Izsák and Beckmann judged several of her incorrect responses to partitioning and iterating items and reversibility items to be false negatives. Thus, they attributed mastery of these attributes where the DCM analysis did not. For Diana, both the expert and the DCM diagnoses with respect to referent unit were ambivalent. The DCM diagnosis indicated a probability for mastery of about .6 (probabilities greater than .5 are assigned mastery in the DCM model) and tilted one direction; the expert diagnosis tilted the other direction. Finally, Izsák and Beckmann judged several of Diana’s incorrect responses to partitioning

Wood, M. B., Turner, E. E., Civil, M., & Eli, J. A. (Eds.). (2016). *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Tucson, AZ: The University of Arizona.

and iterating items to be false negatives, leading to an expert diagnosis (mastery) that disagreed with the DCM diagnosis (non-mastery). Of the 7 cases where Izsák and Beckmann assigned an X, indicating uncertain mastery status, the DCM analysis diagnosed mastery in 5 cases and non-mastery in 2 cases. In part, these differences reflect the fact that Izsák and Beckmann were conservative in diagnosing mastery.

These results indicate significant consistency between the expert and DCM diagnoses, especially when, as stated above, we interpret non-mastery to mean lack of evidence for a particular attribute. The consistency gives us increased confidence that DTMR Fractions profiles do, in fact, reflect relative strengths and weaknesses in reasoning about fraction arithmetic in terms of quantities. Confidence in profiles was critical for our second research question.

Table 1: Focal Profiles Before and After the Number and Operations Course

	DCM Diagnosis Pre: August 2014	DCM Diagnosis Post: January 2015	Expert Diagnosis (Post)
Alice	[0000]	[0000]	[0101]
Jack	[0000]	[0000]	[000X]
Linda	[0111]	[1111]	[11X1]
Claire	[1111]	[1111]	[X111]
Diana	[0001]	[1001]	[01X1]
Kelly	[0000]	[1011]	[X0XX]

Our second research question asked whether the distribution of DTMR Fractions profiles shifted after the 1-semester course on number and operations and, if so, how? This question is important because a primary motivation for developing the DTMR Fractions survey was to create an instrument sensitive to diversity in reasoning and shifts in that reasoning. Figure 3 shows a significant shift in the distribution of profiles before (Fall 2014) and after (Spring 2015) the number and operations course. In Fall 2014, 9 of the 22 pre-service teachers were diagnosed as masters of none of the attributes and only three were diagnosed as masters of referent unit. In Spring 2015, while 6 of the pre-service teachers still did not demonstrate mastery of any one of the four attributes, 7 more demonstrated mastery of all four attributes. Furthermore, comparing profiles before and after for each of 22 pre-service teachers revealed that 13 moved to profiles with additional mastered attributes and that all pre-service teachers who demonstrated mastery of at least two attributes at pretest demonstrated mastery of all four attributes at posttest.

Furthermore, the acquisition of additional attributes reflected what was emphasized to greater and lesser extent in the number and operations content course. In particular, Beckmann placed less emphasis on using whole-number factor-product combinations as a resource for partitioning and iterating. Of the 16 pre-service teachers who had not mastered partitioning and iterating at the inception of the course, 5 (~31%) had mastered and 11 had not mastered this attribute at posttest. Meanwhile, overall gains on the remaining three attributes, which were given more attention in the course, were higher. Of the 19 pre-service teachers who had not mastered referent unit at the inception of the course, 10 (~53%) had mastered and 9 and not mastered this attribute at posttest. Of the 16 pre-service teachers who had not mastered appropriateness at the inception of the course, 7 (~43%) had mastered and 9 had not mastered this attribute at posttest. Of the 9 pre-service teachers who had not mastered reversibility at the inception of the course, 4 (~44%) had mastered and 5 had not mastered this attribute at posttest.

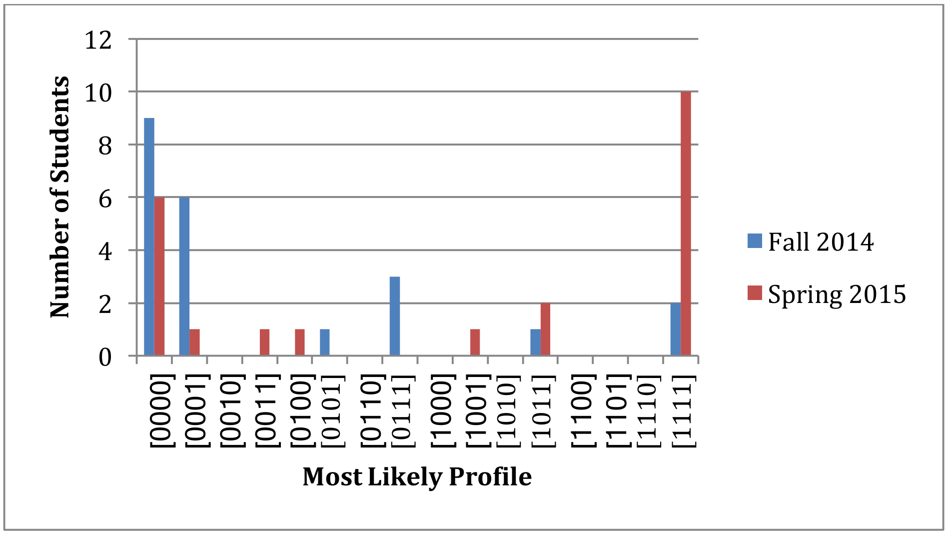

Figure 3. Distribution of DCM profiles before and after the Number and Operations content course.

We draw three main conclusions from these results. First, the shift in profiles from Fall 2014 to Spring 2015 behaved appropriately in the sense that in all but two cases when profiles shifted, pre-service teachers gained attributes. (In the two exceptions, pre-service teachers who demonstrated mastery only of reversibility on the pretest demonstrated mastery of no attributes on posttest). Second, the shift in profiles suggest that many pre-service teachers who entered the numbers and operations course demonstrating mastery of no attributes faced significant challenges developing all four components of reasoning with quantities. Teachers who had facility with at least some of these components tended to master the remaining ones during the course. Third, that shifts in profiles were consistent with what was given more and less emphasis in the course suggests that shifts in profiles reflected the pre-service teachers' opportunities to learn or develop different components of reasoning measured by the DTMR Fractions survey.

Discussion

The present study contributes to the development, validation, and application of measures that capture information about moment-to-moment mathematical reasoning, and does so in the critical mathematical domain of fraction arithmetic. We demonstrated (a) considerable agreement, or convergent evidence, between expert and DCM diagnoses of the four DTMR fractions attributes, and (b) that when pre-service teachers' profiles shifted, they added attributes in ways that reflected the mathematical emphases of their number and operations course. These results demonstrate the possibility of measurement that gets closer to moment-to-moment reasoning than can be captured by a single, summative score. In particular, we have used a survey consisting primarily of multiple-choice items to compare teachers in ways more nuanced than ordering them along a single dimension from less to more able or knowledgeable. Approaches like the one we report have the realistic potential of providing diagnostic information to teacher educators about those components of reasoning with which teachers are more and less facile. This can be useful formative and summative information. Future studies will be needed to see if our results generalize to other cohorts of pre-service teachers and other number and operations courses. Finally, although the present study is situated in teacher education, the potential benefits of multi-dimensional measurement extends to students as well.

Wood, M. B., Turner, E. E., Civil, M., & Eli, J. A. (Eds.). (2016). *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Tucson, AZ: The University of Arizona.

Acknowledgements

Research reported in this chapter was supported by the National Science Foundation under Grants Nos. DRL-0903411and DRL-1420307. The opinions expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

American Mathematical Society. (2010). *The mathematical education of teachers II*. Conference Board of Mathematical Sciences, Issues in Mathematics Education, Vol. 17. Washington, DC: Author.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389–407.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . & Tsai Y. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180.

Bradshaw, L., Izsák, A., Templin, J., & Jacobson, E. (2014). Diagnosing teachers' understandings of rational number: Building a multidimensional test within the diagnostic classification framework. *Educational Measurement: Issues and Practice*, *33*(1), pp. 2–14.

Common Core State Standards Initiative (2010). *The common core state standards for mathematics*. Washington, D.C.: Author.

Hill, H. C. (2007). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind policy initiative. *Educational Evaluation and Policy Analysis*, *29*(2), 95–114.

Izsák, A., Jacobson, E., de Araujo, Z., & Orrill, C. H. (2012) Measuring mathematical knowledge for teaching fractions with drawn quantities. *Journal for Research in Mathematics Education*, *43*(4), 391–427.

Izsák, A., Remillard, J., & Templin, J. (Eds). (in press) Psychometric methods in mathematics education: Opportunities, challenges, and interdisciplinary collaborations. *Journal for Research in Mathematics Education.*

Kersting, N., Givvin, K., Sotelo, F., & Stigler, J. (2010). Teachers' analysis of classroom video predicts student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, *61*(1–2), 172–181.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Saderholm, J., Ronau, R., Brown, E. T., & Collins, G. (2010). Validation of the Diagnostic Teacher Assessment of Mathematics and Science (DTAMS) instrument. *School Science and Mathematics*, *110*(4), 180–192.

Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J. (2010). Investigating links form teacher knowledge, to classroom practice, to student learning in the instructional system of the middle-school mathematics classroom. *Cognition and Instruction*, *28*(3), 317–359.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Sowder, J. T., Armstrong, B., Lamon, S., Simon, M., Sowder, L., & Thompson, A. (1998). Educating teachers to teach multiplicative structures in the middle grades. *Journal of Mathematics Teacher Education*, *1*(2), 127–155.