The Mathematics Enthusiast

Volume 13
Number 1 *Numbers 1 & 2*

Article 5

2-2016

Assessing Mathematical Knowledge for Teaching: The Role of Teaching Context

Geoffrey Phelps

Heather Howell

Assessing Mathematical Knowledge for Teaching: The Role of Teaching Context

Geoffrey Phelps
Educational Testing Service

Heather Howell
Educational Testing Service

Abstract: Assessments of mathematical knowledge for teaching (MKT), which are often designed to measure specialized types of mathematical knowledge, typically include a representation of teaching practice in the assessment task. This analysis makes use of an existing, validated set of 10 assessment tasks to both describe and explore the function of the teaching contexts represented. We found that teaching context serves a variety of functions, some more critical than others. These context features play an important role in both the design of assessments of MKT and the types of mathematical knowledge assessed.

Keywords: teacher content knowledge, mathematical knowledge for teaching, teacher assessment, mathematics, assessment

Introduction

Mathematical Knowledge for Teaching (MKT) is the content knowledge used in recognizing, understanding, and responding to the mathematical problems and tasks encountered in teaching the subject (Ball & Bass, 2002; Ball, Thames & Phelps, 2008). Assessments of MKT are designed to measure the mathematical knowledge that teachers use in these teaching practices. A number of practice-based assessments of MKT have recently been developed for teachers of K-12 grades (Herbst & Kosko, 2014; Hill, Ball, & Schilling, 2008; Hill, Schilling, & Ball, 2004; Kersting, 2008; Krauss, Baumert, & Blum, 2008; McCrory, Floden, Ferrini-Mundy, Reckase, & Senk, 2012; Phelps, Weren, Croft, & Gitomer, 2014; Tatto et al., 2008).

We follow Ball, Thames, and Phelps (2008) in defining MKT to include the full range of mathematics content knowledge used in teaching. The most widely assessed component of MKT is the common content knowledge that is taught and learned as part of regular schooling and is familiar to most adults. There is a long history of assessing teachers' common mathematical knowledge (Hill, Sleep, Lewis, & Ball, 2007). Often these assessment tasks look identical to those on student assessments because the construct is essentially the content of the student curriculum, either at grade level or at a level above the assigned grade (Phelps, Howell, & Kirui, 2015).

MKT assessments have generally focused, however, on the specialized forms of content knowledge that only teachers need to use in the course of their day-to-day work (Ball et al., 2008). While definitions and focus vary in the literature, and the mapping of the MKT

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 13, no. 1&2**, pp. 52–70

construct is likely somewhat dependent on curriculum and culture, most studies share a focus on MKT as a form of applied knowledge that goes beyond common content knowledge (Krauss et al., 2008; McCrory et al., 2012; Thompson, 2015; Turner & Rowland, 2008). MKT assessments typically present teachers with content tasks that are encountered in teaching, such as interpreting student thinking and work, selecting materials for instruction, explaining concepts and procedures, or evaluating whether to use a representation for a particular instructional purpose (Ball & Bass, 2002; Hill et al., 2004). And since these tasks often occur in complex instructional contexts, MKT assessments typically also provide key information about the teaching context, such as the learning goals that direct the teaching, details about a student's prior academic work, or how students are grouped and organized (Phelps et al., 2015). Assessments of MKT differ in how teaching practice is represented. Some provide written descriptions, while others incorporate video or animations depicting mathematics teaching (see, for example, Herbst & Kosko, 2014; Hill et al., 2004; Kersting, 2008). These features of context support test takers in recognizing the relevant aspects of the content task, understanding the content problem, or providing a response to the assessment question.

This contextualization of MKT assessment tasks is in part theoretically motivated. Ball and Bass (2002) argue that *how* teachers encounter mathematics in their teaching directly shapes the nature of the mathematical knowledge that is needed. The context used in many MKT assessment tasks defines both *what* kinds of content knowledge teachers need to use and *how* they use this knowledge. Largely missing, however, from the current literature on MKT assessment are well-articulated design arguments that make clear the links between the construct and assessment task design (Mislevy & Haertel, 2006). Given the central role of teaching in MKT, it seems likely that any endeavor to assess MKT would require consideration of how context functions in the design of MKT tasks (Phelps et al., 2015).

In this study, we take the first steps in this direction by presenting arguments and illustrations for how context functions in a set of elementary-level MKT assessment tasks, with a particular focus on how context enables tasks to measure MKT that goes beyond common content knowledge. We do not take up the question of whether other sub-components of MKT are distinctly measureable, as other studies have done (see, for example, Hill et al. (2008) and Krauss et al. (2008) for different approaches to the measurement of PCK as a distinct domain). Our argument is simply that context matters in the assessment of some components of MKT more than others; in particular it matters more for components that go beyond common content knowledge. Because these types of knowledge have been the objects of intense interest in teacher education it is worth attending closely to how context matters in their assessment.

The paper is organized as follows. First, we discuss the role of context in establishing the construct validity of MKT assessments using illustrative examples. We follow Messick's (1989) view of construct validity, which helps to determine how relevant and representative the tasks are in measuring MKT. We begin with an example that includes three tasks that assess similar content focused on exponential expressions but vary in how teaching context is represented in the task. This set of tasks provides a concrete illustration of major differences in context and its function. Next, we discuss two task examples in detail to illustrate the design and content focus of MKT assessment tasks and to make clear our arguments about the role that context plays in these assessment tasks. Finally we present a summary of how context functions across the 10 tasks and discuss the implications for assessing MKT.

The Role of Teaching Context in Assessing MKT

The appropriate use of teaching context in the assessment of MKT can help avoid threats to construct validity, namely construct-irrelevant variance and construct under-representation. Construct-irrelevant variance occurs when an assessment represents dimensions that are irrelevant to the correct interpretation of the construct, and construct under-representation occurs when an assessment does not adequately represent the dimensions of the construct that are the focus of the assessment (Messick, 1989). In respect to MKT, many assessments are designed to measure the MKT that is specialized to the work of teaching. In cases where teaching context is critical to assessing particular aspects of MKT, the absence of teaching context could lead to construct under-representation.

We begin with an illustration designed to highlight the various roles that context can play in the measurement of MKT. We present three related example tasks in Figure 1. The example in panel C was developed for the Measures of Effective Teaching project (Phelps et al., 2014) and is one of the 10 tasks analyzed in this study. Task selection and analysis is addressed in more detail in the methods section. The examples in panels A and B of Figure 1 are variants created by the authors for illustrative purposes to demonstrate both when teaching context does and does not support the assessment of MKT.

A. Common Content Knowledge	B. Common Content Knowledge in a Teaching Context	C. Specialized Content Knowledge in a Teaching Context
Evaluate each of the following simple exponential expressions. $3^3 =$ $2^3 =$ $2^2 =$	Ms. Hupman is teaching an introductory lesson on exponents. She gives her students a set of problems to check their proficiency in evaluating simple exponential expressions. Ms. Hupman looks over the work from one of her students. For each of the answers, indicate if the student's evaluation is correct or incorrect. Correct / Incorrect $3^3 = 9$ ☐ ☐ $2^3 = 6$ ☐ ☐ $2^2 = 4$ ☐ ☐	Ms. Hupman is teaching an introductory lesson on exponents. She wants to give her students a quick problem at the end of class to check their proficiency in evaluating simple exponential expressions. Of the following expressions, which would be <u>least</u> useful in assessing student proficiency in evaluating simple exponential expressions? ○ 3^3 ○ 2^3 ○ 2^2
Key: 27, 8, 4	Key: incorrect, incorrect, correct	Key: 2^2

Figure 1. Tasks to illustrate differences in types of content knowledge assessment.

Each of these three tasks involves the same underlying mathematical content, but they differ in whether and how each is situated in teaching. Task A does not include a context and simply requires the test taker to evaluate three exponential expressions. This task is not situated in teaching other than representing mathematics that is part of the grade school curriculum. However, the absence of context in this task is construct relevant because no context is required to assess whether teachers can do the work of the student curriculum.

Task B includes a context that shows a student's evaluation of three simple exponential expressions. The student has answered two problems incorrectly and one correctly. The test taker does not need to draw conclusions about *why* the student answered each correctly or incorrectly. He only needs to evaluate each problem and check the correct answer against the student's answer to determine whether the student's answer is correct or incorrect. While the look and feel of Tasks A and B are different, the mathematical work and knowledge required to answer is essentially the same. Both measure a test taker's ability to evaluate expressions. The context in Task B is arguably *construct-irrelevant* (Messick, 1989), meaning that its presence or absence does not relate directly to the skill of exponent arithmetic. However, the longer text included in Task B increases the reading burden on the test taker, raising the possibility that the task might unintentionally measure reading ability in addition to the skill of exponent arithmetic. Reading load is not necessarily problematic; the text is not excessive in length and the level of reading required may be well within the abilities of the tested population. But to the extent that such a task measures something unintended (in this case, reading ability), it can be a source of *construct-irrelevant variance* in the test scores (Messick, 1989).

Task C, like Task B, includes a written teaching scenario. But in this case, the context serves to direct the test taker to consider which expression would be a poor choice for teachers to use in understanding whether students know how to evaluate expressions. To respond to this task, the test taker needs to already know, or know how to figure out, what kinds of confusion students are likely to exhibit (e.g., confusion about which number is the base or exponent or confusion around what kind of operation is required to evaluate the expression). The test taker then needs to anticipate what the solution to each of the problems would be using the incorrect methods students might apply and from this figure out which problems reveal these confusions. The mathematical knowledge involved in responding to this task goes beyond the common content knowledge of how to evaluate exponents. The context that is included in this task is relatively minimal but clearly necessary; without the context the test taker lacks key information for comparing the problem choices. Unlike Task B, where the context is *irrelevant* to the content assessed, in Task C the context is *relevant* and arguably critical to the content knowledge that is being assessed.

The three tasks shown in Figure 1 are intended to illustrate a number of key points in the design of MKT assessment tasks. First, context is not always needed. Most notably, as illustrated in panel A, when teachers are simply doing the math that their students are learning, there is likely no need for context (Phelps, Howell, Schilling, & Liu, 2015). The context in Task B illustrates an authentic situation in teaching that requires the teacher to have common content knowledge. But from an assessment perspective, when measuring this type of MKT, it will often be more efficient to present the task without a context, as illustrated in Task A. A basic principle of assessment design is that irrelevant context should be avoided to the greatest extent possible so that only the intended construct is measured (Messick, 1989).

Second, as illustrated by task C, context can play a critical and relevant role in assessing the construct when the goal is to assess the components of MKT that go beyond the mathematics that students are expected to master (i.e., SCK or PCK in the Ball et al. (2008) model). In such cases, eliminating the context might shift the focus of the task in ways that leave the test taker unsure what is being asked or might fundamentally change the content assessed. Eliminating context entirely could reduce tests of MKT to assessing only the types of common content knowledge illustrated by task A, which would lead to tests that suffered from threats of construct under-representation (Messick, 1989).

Figure 1 also illustrates that it is not always simple to determine whether context is relevant. At first glance, Tasks B and C seem quite similar. It is only through analysis of the work that each task requires of the test taker and consideration of the measured construct that such a determination can be made. Consequently, from an assessment design perspective, it is critical to clarify how context that is included in an assessment task is relevant to the particular features of the construct being assessed (Mislevy & Haertel, 2006).

Methods

Our goal in this study was to systematically investigate the ways in which teaching context can function in tasks designed to elicit the types of MKT that are particular to the work of teaching mathematics. While we follow general procedures for qualitative coding, our method differs from typical qualitative work in two key ways. First, our 'data' are the tasks themselves. We selected a set of tasks for which we have a large set of ancillary data showing that they perform well as measures of CKT and that context matters in how respondents reason through each task. We did not, however, examine teachers' actual response data in this particular study. Our claims therefore are built on arguments about task design and not on empirical data comprised of test takers' responses. Therefore, our results are the categories and associated characteristics of task design that emerged in the course of the close analysis of the MKT tasks. We think this type of close, rigorous analysis helps to call attention to aspects of task design that are otherwise largely invisible, even to test designers. We describe the process in some detail to help the reader follow our logic.

Selection of MKT Tasks for Analysis

The analysis that follows focuses on a set of 10 mathematics tasks that were developed as part of the Measures of Effective Teaching project to measure elementary level MKT (Phelps et al., 2014). These tasks were chosen because we had strong evidence from a prior cognitive interview study that they situated test takers in teaching practice as designed (Gitomer, Phelps, Weren, Howell, & Croft, 2014; Howell, Phelps, Croft, Kirui, & Gitomer, 2013). As part of that study, we wrote rationales detailing the embedded assumptions about how context would function and about the construct each task measured. The study established that the alignment of participant reasoning to these rationales was strongly related to answering correctly or incorrectly. Across all mathematics task level interview responses (n = 640), 88% showed the desired pattern in which correct answers matched pre-specified correct reasoning and incorrect answers did not match that correct reasoning. For 97% of responses, participants reported that the task was an authentic representation of actual teaching practice. The study also found no evidence that reading load introduced construct-irrelevant variance by interfering with test takers' interaction with the assessment tasks (Gitomer et al., 2014; Howell et al., 2013). These response patterns led us to conclude that knowledgeable teachers were situated in context as specified by the task design.

It is worth clarifying that our goal was not to generalize to all MKT tasks or other such practice-based items. Instead we used strong tasks from a prior study with the goal of using this selection as a site for naming and defining important task design characteristics. Specifically, in order to understand how context can function, we required a set of tasks that measure more than common content knowledge, in which context is available to be analyzed, and for which we have some evidence that the context serves a function.

Analytic Method

As a first step in the analysis we expanded the written rationales used in the prior study to account more explicitly for context features and to understand better the role that context played in these tasks (Howell et al., 2013). We started by simply describing the context and its role in shaping how the test taker interacts with the content problem. These descriptions constituted the first step in our qualitative analysis and subsequently became objects of the second step of analysis. A summary of such a description is provided below for the task shown in figure 2.

To assess her students' prior knowledge about evaluating arithmetic expressions, Ms. Santiago assigned a worksheet of problems. She noticed that Alexis answered the first two incorrectly and the next two correctly.

1) $7 \times 2 - 6 + 3$ = 5

2) $9 - 5 + (16 \div 8)$ = 2

3) $9 + 24 \div 3 - 1$ = 16

4) $17 - (3 + 7 \times 2)$ = 0

Which of the remaining problems is Alexis likely to answer incorrectly?

- O $8 + 7 - 12 \div 3$
- O $13 - 3 \times 2 + 5$
- O $(27 \div 3 - 4) + 8$
- O $(16 - 12) \times 5 + 10$

Figure 2. The Santiago task.

To respond to this task, the test taker needs to analyze the four examples of Alexis' work, determine what she did to get the first two problems wrong, and then test any hypotheses about her confusion to see if they are consistent with answering the other problems correctly. The test taker needs to select an option that Alexis would answer incorrectly, assuming Alexis persists in the same error. However, the underlying, important task is to figure out what Alexis is misunderstanding. The assessment task is focused on the recurrent teaching practice of diagnosing student understandings or misunderstandings based on the written work they produce.

Analysis of the given problems reveals that in each of the incorrect problems, Alexis has added before subtracting. In the first problem she added 6 and 3 first and then subtracted the total of 9 rather than subtracting 6 and then adding 3. In the second problem she added 5 and 2 (where 2 is the result of 16 divided by 8) and then subtracted the total of 7 rather than subtracting 5 then adding 2. However, in the third and fourth problems this particular error does not lead to an incorrect answer. In the third problem, the ordering of the operations happens to be such that adding before subtracting is appropriate. In the fourth problem, the parentheses indicate that the expression inside should be added first before subtracting. There is not enough evidence to know *why* Alexis is making this error, although experienced teachers may recognize it as a possible overgeneralization of the use of the mnemonic PEMDAS[1] to dictate the order of operations. If we assume that Alexis will persist in the same error, the second answer option is the only option she would answer incorrectly because for each of the others, like the third and fourth given problems, adding before subtracting happens to be correct.

The scenario only specifies "arithmetic expressions" as the content topic under study, but the form in which the mathematics problems are written provides a great deal of subtle contextual information about the level of the students. Each expression is written out as a single line, using the division symbol ÷ and the multiplication symbol × rather than a fraction bar for division or a dot for multiplication. All four operations (addition, subtraction, multiplication, and division) are represented and parentheses are used, but there are no exponents. These details communicate to someone with knowledge about the teaching of this mathematics that the students are likely studying order of operations. Their use of the operations themselves is likely fluent at this point, but their ability to combine the operations correctly may not be. In the context of the assessment task, this is important because it makes some possible errors far less likely. For example, one could have assumed that Alexis misread the addition symbol or that she did not know how to perform the subtraction correctly, but this is an unlikely error for a student who is working with expressions of this type.

On the other hand, it is quite common for students at this level to make mistakes in the ordering of the operations. While the scenario does not state that this is an order of operations problem, the contextual clues embedded in the format of the content problems themselves make the work the test taker needs to do much easier by narrowing the field of all possible errors to a fairly small set of likely ones that need to be considered. This is a critical piece of information because it allows the test taker to rule out other competing, but unlikely theories. Again, one reason this set of assessment tasks was useful to study is that the prior interview work provides evidence to support such claims about the functioning of the context. And indeed in a prior study using this task, participants often referred explicitly to it being about order of operations, confirming this part of the design theory (Howell et al., 2013).

---

[1] PEMDAS is a mnemonic device commonly used in the U.S. to help students remember the order of operations. It stands for "parenthesis, exponents, multiplication, division, addition, subtraction," and is not strictly mathematically correct as written, although when used in instruction teachers generally qualify it by stating that the pairs "MD" and "AS" are performed in order, left to right, at the same time, not one before the other as the device implies.

The context also includes information about the student, Alexis, stating that she answered the first two problems incorrectly and the second two correctly. It is not strictly necessary to state which are correct and which incorrect, but providing the information up front may decrease the cognitive load on the test taker and encourage him to focus on the student's thinking rather than on whether the problems are correct. And pointing out that these are Alexis's answers also conveys a crucial piece of information about what the test taker needs to do by setting the condition to be met—the identified misconception must explain both Alexis's correct and incorrect work, and it must be a systematic error that the student makes consistently. A test taker who fails to attend to this aspect of the context may read through the problems assigning a unique diagnosis to each, or may cite difficulties students generally have with such problems without determining the specific difficulty Alexis is having. Both were patterns we observed in prior interview data and were associated with incorrect answers (Howell et al., 2013).

Finally, the assessment task presents an authentic scenario. Teachers frequently have to draw conclusions about student thinking from written work. The task of figuring out what Alexis is thinking seems not just plausible but worthwhile; teachers can't make informed decisions about next instructional steps without knowing first what their students understand and do not understand.

The summary above illustrates the type of descriptive account that was generated for each of the 10 tasks. These accounts provided rich descriptions of how the tasks functioned and more specifically the role that context played in these tasks. They also were used as the basis for generating provisional statements describing each context element. We then coded each identified context element inductively with short phrases describing the ways in which the context element functioned in the test taker's anticipated interaction with the task. We used a constant comparative method (Strauss & Corbin, 1990) to do this coding, which can be described in four steps: (1) independently analyzing a subset of tasks, (2) reconciling the coded elements and functions across tasks, (3) revising the list to reflect all elements and functions and testing the new categories by recoding the subset of tasks, and (4) expanding and iterating to a larger set of tasks until we had reached consensus on all codes for all context elements observed across all 10 tasks. Our goal in this work was not to achieve a particular level of coding reliability, but rather to generate a useful set of categories that captured the types of elements and functions we saw both in a given task and collectively across tasks. The short descriptors of the functions were then grouped together to form more general categories, and the entire set of tasks reviewed and recoded using these categories.

This process of task analysis generated three sets of categories that were relevant to describing the context and its function. Because we view these categories as an important outcome of this study, they are described in more detail below in the results section.

Results

The results are organized in two main sections. The first section presents the categories that were derived inductively from the analysis of the 10 MKT tasks. The second section focuses on the use of these categories to describe the context features and their function across these 10 tasks. While we present counts across the set of tasks to illustrate the frequency, distribution, and co-occurrences we observed, we remind the reader that for a study of this type the main results are the identification and description of the categories themselves.

Teaching Context and Function

Context focus. The various teaching contexts identified in the MKT tasks mapped onto three major components of instruction. These included features of *students* such as their history, learning needs, and actions; the *content* and how it is situated in the curriculum of school learning; and, the *setting*, which includes class size or grouping or mode of instruction such as lecture or discussion. Not only are these particular features central to instruction, but they have also recurred in many different heuristics and models used to characterize instruction (see, for example, Cohen, Raudenbush, & Ball, 2003; Hawkins, 1974; McDonald, 1992; Schwab, 1978). For each of the 10 MKT tasks, elements of the context could be identified as providing context for the *content*, *student*, or *setting* of instruction. These categories are useful for identifying the aspects of instruction that are the focus of the context features.

Context Function. The categories that were derived from the analysis describe the main function of the contexts identified in the 10 tasks. These categories are described in Table 1.

Table 1. *Context Functions.*

	Context function	Description
Critical functions	Narrows a set of possibilities	Context that functions by narrowing the set of possibilities that the test taker must consider – e.g., narrowing the possible answer choices or eliminating one or more options. This can be quite subtle, as in cases where the specified level of a student or class sets an expectation for the level of sophistication one might expect in an answer, which in turn serves to eliminate some set of possibilities. Sometimes it might be some other list that is narrowed rather than the answer choices. For example, in the task shared in Figure 2 the test taker needs to figure out what error the student has made before even considering the options, and the content context serves to narrow the possible errors.
	Sets condition for the answer	Context that functions to specify, explicitly or implicitly, what condition the answer needs to meet to be correct. For example, in the task shared in Figure 3, the setting context sets a condition for the answer – i.e., that selected problem needs to be one for which the student's answer will reveal the suspected misconception to the teacher.
Helpful functions	Direct the test taker's focus	Context elements that encourage the test taker to focus (or not to focus) on a particular aspect of the task. For example, in the task shared in Figure 2, the statement that the student answered two problems correctly and two incorrectly is intended in part to cue the test taker to pay attention to the correct work and not just the incorrect work.
	Provides additional information	Context that provides additional information that is useful but not critical. This might include defining a term that some test takers may not know. Or it may include context that reduces cognitive demand by stating up front that a student's work is incorrect so that the test taker knows that figuring this out is not part of the work he needs to do.
	Reinforces critical information	Context that reinforces a key idea. This can help ensure that a test taker is directed to pay attention to critical information and thus raise the likelihood that the test taker engages in the assessment task as intended.
Functions related to face validity	Authenticity	Context that helps support an authentic representation of the work of teaching. Perceived authenticity can be a key motivating factor and enhance validity.
	Plausibility	Context that specifically helps to add plausibility to an element of the task that would not otherwise seem reasonable. For example, in the task shown in panel C of Figure 1, the specification that the problem is a quick check at the end of class makes it feel reasonable that the teacher has a need to diagnose understanding on the basis of a single answer alone. Without this information, the test taker might wonder why the teacher does not simply ask the students to explain their work.
	Motivation	Context that creates a situation in which the test taker can better recognize the importance the task. For example, tasks that give specifics about a student and their learning needs can motivate because there seems to be a real and pressing need to help the student.

Context Relevance. Another pattern that emerged was one of relative levels of relevance or criticality. Some context functions, like those that set the condition the answer needs to meet to be correct, are essential for assessing the MKT content. Without that information the test taker would be unable to respond correctly and the particular type of MKT could not be assessed. Other functions were less essential in that it would still be possible to respond correctly absent that context. However, many of these context features were still quite helpful in directing the test taker and thus might serve to reduce cognitive load. For example, context that functions either to further define a key idea or direct the test taker to pay attention to something important falls into this second group. A third type of context functions to increase face validity, support the test taker's perception of authenticity, or to motivate in other ways that support completing an assessment task.

Teaching Context in MKT Assessment Tasks

To make these three sets of categories more concrete, an example task (Figure 3) is used to illustrate the process and the types of decisions that the coding and classification entailed.

Mr. Chamberlain is concerned that his students' use of the calculator has led them to view the equal sign as a signal to carry out an operation rather than as a symbol indicating equality. Of the following missing-number problems, which would best assess whether students understand the mathematically correct meaning of the equal sign?

- ○ __ + __ = 18
- ○ 7 + 5 = __ + 6
- ○ __ = 17 + 9 + 5
- ○ 23 + 4 = __ = 4 + 23

Figure 3. The Chamberlain task.

The context for content in this task is given directly and indirectly. The scenario indicates that Mr. Chamberlain's concern is focused on the meaning of the equals sign. The format of the missing number equation problems communicates the level of the students as early elementary and signals that the use of the equal sign is likely new to them. This bolsters the authenticity and appropriateness of Mr. Chamberlain's concern as represented in the problem, as students often misunderstand the equals sign to be a command to perform an operation. It both makes sense that students working at this level would have this confusion and it conveys that the confusion is important for a teacher to attend to. Thus, in this case, the content context provides authenticity and contributes to the face validity of the task.

Unlike the task in Figure 2, in which information about the student was given directly, the student context in this problem is given indirectly in the form of the teacher's concern. What we know about the students is that they have used a calculator, and further that the teacher believes they may hold a particular misconception (that the equals sign is a command to perform an operation). Knowing that that the suspected misconception is connected to calculator use in this way provides key information to the test taker by defining, if indirectly, the operational view of the equals sign. Understanding the difference between the operational and equality views of the equal sign is key to answering correctly, and this piece of context reduces the cognitive load for a test taker unfamiliar with the misconception or with the terminology used to describe it. It also provides a plausible basis for the students to have that misconception, as calculator use is common and can lead to exactly this type of misunderstanding. The student context here serves dual functions. It supports the plausibility of the scenario, contributing additionally to face validity, and it also provides helpful but non-critical information to the test taker by defining a key idea.

We point out here an ambiguity in the coding and classification. One could argue that the teacher's concern is a part of the setting context, and not really information about the students. We acknowledge this, and use this example to draw attention to a necessary imprecision in the categories we have proposed. In many cases the distinctions are subtle and a piece of context might well fall into multiple categories. In fact, in this case we listed the teacher's concern about the operational view as setting context as well as coding the student's use of the calculator as student context. As a feature of the setting, the teacher's concern motivates the task by providing a plausible reason to care which problem is selected, further supporting face validity. More importantly, it sets the condition the answer needs to meet in order to be correct; the correct answer must be a problem that will reveal the given misconception to the teacher. This function of context (setting the condition the answer needs to meet to be correct) is at the highest level of relevance because it is critical that it be included in order for the task to function as designed. That the context is difficult to assign to the categories of setting or student is less important than the critical function it serves in orienting the test taker's thinking. We draw the reader's attention to the ambiguity here to illustrate clearly that our goal is not to create strict divisions between context types so much as to name categories that are useful for systematically analyzing or generating MKT tasks.

We also draw a distinction between the context that is situating the test taker and the actual knowledge or ability that the test taker must have in order to respond to the task correctly. This last piece of context sets the condition the answer needs to meet, and the test taker must distill this understanding from the context in order to answer correctly. But the test taker still needs to know which problem will meet that condition. While the context clues situate the test taker so that she is answering the right question, they do not answer the question for her. In this case, the test taker still needs to know or be able to anticipate that a student with the given misconception will likely write 12 in the blank on the second problem, having interpreted the equal sign as a command to add 5 and 7. For this option,12 is incorrect because 5 + 7 is not equal to 12 + 6. While the student might think about the equal sign incorrectly in each of the other options, the answer the student gives would be the same as the correct answer and would not reveal the error to the teacher. This is the only problem that makes the misconception visible.

Table 2 gives an overview of the context features coded for each of the three MKT tasks that have been discussed in depth so far in the paper (Figures 1, 2, & 3). It is worth noting that while we made efforts to reach consensus in the coding, we do not propose that the context elements for which we coded are fixed or that there is always a clear classification. Rather, we find these elements useful in providing conceptual tools that help to identify and understand the function of context. Specifically, this makes these context elements more visible and provides a language that can be used to evaluate and critique the design of assessment tasks. The examples in Table 2 also illustrate that not every type of context element or function appears in every task. This is typical of what was represented across the set of analyzed tasks and suggests as a cautionary note that while the proposed categories are analytically useful, they are not strictly necessary. They do not form a template for assessment task construction.

Table 2. *Sample Coding Classifications.*

Task Description	Content context and its function	Student context and its function	Setting context and its function
Ms. Hupman wants to select a brief assessment problem to ascertain whether her students understand how to evaluate exponential expressions. (Figure 1, panel C)	That the lesson is introductory **narrows the likely errors students would make**[1] to those above the level of arithmetic (students probably know how to multiply).	No student context is given.	That the problem is a quick proficiency check provides **plausibility**[3] for why the answer alone needs to convey information and also **sets the condition for the answer**[1] – that it must reveal to the teacher whether or not the student is proficient. The focus on the <u>least</u> useful problem **decreases authenticity**[3] as a teacher would generally look for the most useful, not the least.
One of Ms. Santiago's students has answered two order of operations problems correctly and two incorrectly, and the test taker must figure out what she has done wrong and predict which additional problems she will answer incorrectly. (Figure 2)	The types of mathematical symbols used ("x" for multiplication, for example) coupled with the specification that this is prior knowledge for the students **narrows the possible error types**[1] to exclude arithmetic errors and include errors related to ordering of steps.	The specification that the student answered two problems correctly and two incorrectly **encourages the test taker to attend to both**[2] the correct and the incorrect work, **suggests a systematic error**[2], and **sets the condition**[1] the selected error needs to meet – it needs to explain the given work.	That the student work shown was in response to a worksheet suggests that the teacher is looking at the work after the fact, with time to reflect, making the work needed to analyze the errors more **plausible**[3].

Mr. Chamberlain is concerned that his students may have a specific misconception about the equal sign and must choose an assessment problem that would reveal to him whether or not they have that misconception. (Figure 3)	The format of the missing number equation problems communicates the level of the students as early elementary and hints that the use of the equal sign symbol is likely new to them, supporting the **authenticity**[3] of the teacher's concern that they might not understand it.	The teacher expresses concern that students use of the calculator may have caused them to have an operational view of the equals sign, providing a **plausible**[3] explanation for why they would misunderstand, and **defines**[2] for the test taker what is meant by the operational view.	The teacher's concern that the students may have an operational view of the equals sign provides **motivation**[3] for the task of teaching, as well as **setting the condition**[1] to evaluate the answers as those that reveal that incorrect operational view.

Note: Bold text indicates the function of a context element: 1) critical context function, 2) useful context function, 3) face validity context function. A full version of this table and all tasks analyzed is available from the corresponding author upon request.

Table 3 provides an overview of how often each function type for each context category appeared across the 10 tasks analyzed. For many of these tasks, various context functions and features could appear multiple times. For example, the task presented in Figure 3 was coded as having content context that supports authenticity, student context that makes the situation more plausible and defines a key term, and setting context that motivates the situation as well as setting the condition the answer needs to meet. This particular task contributes one count to the content context category and two counts each to student and setting categories.

Table 3. *Context Type and Function for Ten MKT Tasks.*

Type of function	Type of context			Total occurrences over 10 tasks
	Content	Student	Setting	Total
Critical context functions:				
Narrows a set of possibilities	5	2	0	7
Sets the condition for the answer	2	3	2	7
Helpful context functions:				
Directs the test takers focus	1	4	1	6
Provides additional information	4	4	0	8
Reinforces critical information	1	1	0	2
Face Validity				
Authenticity	1	2	2	5
Plausibility	0	1	8	9
Motivation	0	2	1	3

Our summary suggests that for the 10 tasks analyzed the context functions are relatively equally represented across all the major coding categories. While critical functions seem to occur slightly less often in the "setting" column than elsewhere, and face validity functions noticeably more, the overall distribution suggests that all three types of context elements can serve all functions and can also vary in their criticality. This suggests that teaching context, at least as it appears in these particular MKT tasks, can play a variety of functions across a number of major features of instruction.

Discussion

Mathematical knowledge for teaching includes the full range of mathematics used in teaching the subject. This is a form of applied knowledge that teachers draw on and use as they engage in and carry out the many practices that make up the moment-to-moment and day-to-day work of mathematics teaching (Ball & Bass, 2002; Ball et al., 2008). In this study, we conducted an analysis of 10 tasks designed to assess MKT. These tasks assess types of MKT that go beyond the common content knowledge used in doing the work of the student curriculum (e.g., the first example in Figure 1), with the goal of measuring specialized types of MKT used in practices only encountered in mathematics teaching (Phelps et al., 2014). Because these tasks focus on types of MKT applied in teaching practice, they all include teaching context. We found that across these tasks the context served a number of different functions. In fact, for many tasks, the context served multiple functions. We coded almost 50 instances of context serving an identified function across just 10 tasks.

These context features focus on different aspects of instruction. We grouped these under the larger categories of content, student, and setting. These categories provide a useful set of lenses for considering which core aspects of instruction are represented in the context. It also seems likely that different types of MKT tasks might require context that focuses on aspects of instruction that did not come up in our analysis. For example, tasks like the Chamberlain task provide background information about the teacher's concern. This suggests that *teacher* might be an additional useful category that simply did not appear often in the set of tasks we examined. This category might include information such as teachers' pedagogical motivations, purposes, or constraints. Other MKT assessment tasks might call for a more fine-grained list of the major components or aspects of instruction, such as separate categories for curriculum materials and content.

We also identified across the context features a variety of functions (Table 1). Again, it is important to emphasize that these functions are almost certainly not exhaustive. Additional functions might be identified for a different set of MKT tasks. Although the list of context functions is likely incomplete, we think it nonetheless provides a useful start and important insights into the assessment of MKT. One insight that emerged from this analysis was that these functions could be placed into three larger groups describing the degree to which the context was critical to assessing the MKT construct. We discuss each group of functions briefly below.

We described one group as "functions related to face validity" (Table 1) because their only role was to support the test taker's perception of the situation as authentic, to make the work seem plausible, or to motivate. This group of context functions is arguably the least critical for supporting the test taker in providing an answer. In fact, in some situations, these context features may not be needed at all. If the test taker, for example, is familiar with the content and accepts that it is important and used in teaching, the context may do little more than add to reading load and may even introduce construct irrelevant variance. On the other hand, context that adds face validity can support the test taker in important ways. Michael Kane (2006), in his seminal chapter on validity, argues that tests that lack face validity can introduce construct-irrelevant variance since the test may in part measure a test taker's disengagement with the tasks rather than the construct of interest. Context associated with this group of context functions should be examined with special care to make sure that it plays a sufficiently important role to be included in the assessment task.

We described a second group of context functions as "helpful functions" (Table 1) because they served to support the test taker in providing an answer (e.g., directing the test taker's focus toward a particular aspect of the work or reinforcing critical information). While this type of context was not critical to answering the task, it played an important role, often reducing burden for the test taker. As was the case for the face validity functions, the context associated with this second group is not critical to answering. However, it is not obvious that the context is construct irrelevant, since it appears to support the test taker in productively and efficiently engaging the task.

We described the final group of functions as "critical functions" (Table 1) because the test taker needs to consider the associated context in order to provide an answer. This included cases where the context information narrows the answer possibilities or sets a condition the answer needs to meet. If the context were removed entirely for these instances then the MKT that was the focus of the task simply could not be assessed. In these cases the context is not only critical, but arguably an integral part of the construct itself (Phelps, Howell, & Kirui, 2015). Removing context from these tasks would fundamentally change the MKT assessed and would likely lead to tests that suffered from construct under-representation (Messick, 1989).

Our analysis also revealed that because context can simultaneously serve multiple functions of varying criticality, it cannot easily be labeled as strictly construct relevant or irrelevant. A passage that increases reading load may support the test taker's work in other ways. We also note that identifying context in an assessment task requires more than a surface analysis of its presence or absence. Tasks with a very limited instructional scenario may very well be rich in context, and others, like the task shown in Figure 1, Panel B, may have an instructional scenario that contributes little to the knowledge that the task assesses.

We recognize that both the specific ways that context functions and also their occurrences could vary for a different sets of MKT tasks. This analysis represents only a snapshot of possible context types, the ways in which they are hypothesized to function, and variation of each type. While we have no evidence that particular patterns or lack of patterns would generalize to other measurement situations, we do have evidence from a related study that the patterns are similar when looking at comparable measures in other subjects (Phelps, Howell, & Kirui, 2015).

In conclusion, we think that the approach to describing teaching context in this paper is likely to be useful in better understanding and evaluating MKT task design and as a basis for designing studies that systematically vary the use of context to further explore how those designs function. The analysis illuminates the relation between the types of knowledge a test taker uses in answering a task, the design of assessment tasks including relevant features of context, and the MKT domain assessed by the task. Explicit attention to the role that context plays in the design of MKT assessments offers the potential to better understand not only the content knowledge that is assessed in particular tasks, but also to begin to develop a theory of how teaching context itself may serve to define this knowledge.

References

Ball, D. L., & Bass, H. (2002, May). Toward a practice-based theory of mathematical knowledge for teaching. In *Proceedings of the 2002 annual meeting of the Canadian Mathematics Education Study Group* (pp. 3–14).

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. doi: 10.1177/0022487108324554

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119–142. doi: 10.3102/01623737025002119

Gitomer, D., Phelps, G., Weren, B., Howell, H., & Croft, A. (2014). Evidence on the validity of content knowledge for teaching assessments. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project* (pp. 493–528). San Francisco: Jossey-Bass.

Hawkins, D. (1974). I, thou, and it. In D. Hawkins (Ed.), *The informed vision*: *Essays on learning and human nature* (pp. 48–62). *New York*: *Agathon* Press.

Herbst, P., & Kosko, K. (2014). Mathematical knowledge for teaching and its specificity to high school geometry instruction. In J. Lo, K. Leatham, & L. Van Zoest (Eds.), *Research trends in mathematics teacher education* (pp. 23–46). New York, NY: Springer.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400. Retrieved from http://www.jstor.org/stable/40539304

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching*. The Elementary School Journal, 10*5(1), 11–30. doi: 10.1086/428763

Hill, H. C., Sleep, L., Lewis, J., & Ball, D. L. (2007). Assessing teachers' mathematical knowledge: What knowledge matters. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learnin*g (pp. 111–156). Charlotte, NC: Information Age Publishing.

Howell, H., Phelps, G., Croft, A., Kirui, D., & Gitomer, D. (2013). *Cognitive interviews as a tool for investigating the validity of content knowledge for teaching assessments* (ETS Research Report 13–19). Princeton, NJ: Educational Testing Service.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: Praeger.

Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement, 68*(5), 845–861. doi: 10.1177/0013164407313369

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM Mathematics Education, 40*(5), 873–892. doi: 10.1007/s11858-008-0141-9

McCrory, R., Floden, R., Ferrini-Mundy, J., Reckase, M. D., & Senk, S. L. (2012). Knowledge of algebra for teaching: A framework of knowledge and practices. *Journal for Research in Mathematics Education, 43*(5), 584–615. doi: 10.5951/jresematheduc.43.5.0584

McDonald, J. (1992). *Teaching: Making sense of an uncertain craft*. New York: Teachers College Press.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education/Macmillan.

Mislevy, R., & Haertel, G. (2006). Implications of Evidence-Centered Design for Educational Testing. *Educational Testing: Issues and Practice, 25*(4), 6–20. doi: 10.1111/j.1745-3992.2006.00075.x

Phelps, G., Howell, H., & Kirui, D. (2015). *Situating the assessment of content knowledge for teaching in the context of teaching practice* (ETS Research Report). Manuscript submitted for publication.

Phelps, G., Howell, H., Schilling, S. G., & Liu, S. (2015). *Exploring differences in mathematical knowledge for teaching for prospective and practicing teachers.* Manuscript in preparation. *Journal of Mathematics Teacher Education*.

Phelps, G., Weren, B., Croft, A., & Gitomer, D. (2014*). Developing content knowledge for teaching assessments for the Measures of Effective Teaching Study* (ETS Research Report 14-33). Princeton, NJ: Educational Testing Service. doi: 10.1002/ets2.12031

Schwab, J. J. (1978). The practical: A language for curriculum. In I. Westbury & N. Wilkof (Eds.), *Science, curriculum, and liberal education: Selected essays* (pp. 287–321). Chicago: University of Chicago Press.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage Publications.

Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M): Conceptual framewor*k. Lansing, MI: Teacher Education and Development International Study Center.

Thompson, P. W. (2015). Researching mathematical meanings for teaching. In English, L., & Kirshner, D. (Eds.), *Third Handbook of International Research in Mathematics Education* (pp. 435–461). London: Taylor and Francis.

Turner, F. and Rowland, T. (2008) *The knowledge quartet: a means of developing and deepening mathematical knowledge in teaching*. Mathematics Knowledge in Teaching Seminar Series: Developing and Deepening Mathematical Knowledge in Teaching (Seminar 5), Loughborough University, Loughborough.