Why are mathematics teachers "not sure"?[1]

Janne Fauskanger
University of Stavanger

Reidar Mosvold
University of Stavanger

Received: 25 January 2016 Accepted: 11 July 2016

Researchers have widely adopted measures of teachers' mathematical knowledge for teaching (MKT). This paper investigates why teachers select "I'm not sure" as a suggested solution in MKT items. In this study, in-service teachers responded to multiple-choice MKT items, they submitted written responses to open-ended questions, and they discussed these items in group discussions. We analyse the different kinds of responses from three teachers in depth. A framework of cognitive types of teacher knowledge—distinguishing between Type 1, Type 2 and Type 3 knowledge—was applied in the analysis. We suggest that one teacher selected "I'm not sure" due to insufficient MKT, whereas the other two teachers appeared to have Type 1 and Type 2 knowledge. These findings provide proof that the knowledge teachers utilise in responses to open-ended questions and group discussions does not necessarily mirror the knowledge used when selecting a particular multiple-choice response.

Keywords · teachers' mathematical knowledge for teaching · measurement · multiple-choice items

Introduction

This paper builds upon the practice-based theory of mathematical knowledge for teaching (MKT), as it was developed by a group of researchers at the University of Michigan directed by Deborah Ball (e.g. Ball, Thames & Phelps, 2008). Associated with their development of the practice-based theory of MKT, the researchers in Michigan designed multiple-choice items to measure MKT (e.g. Hill, Sleep, Lewis & Ball, 2007). The item design was mainly carried out as part of the Learning Mathematics for Teaching (LMT) project, and the LMT project continues to administer and provide training in the use of these items—often referred to by researchers as the MKT items or MKT measures (e.g. Fauskanger, Jakobsen, Mosvold & Bjuland, 2012). Based on previous classroom studies, researchers in the LMT project developed items that focused on the mathematical knowledge needed for teaching (Ball et al., 2008). The MKT measures include items that reflect the mathematical tasks teachers face in classrooms. One example is assessing student work. According to Hill et al. (2007) assessments composed of these tasks of teaching can be used to measure the effectiveness of mathematics-focused professional development.

Developing assessment items that measure teachers' knowledge is complicated and expensive, and only a few standardised instruments for measuring the knowledge needed for

[1] This paper is developed from a paper that was originally presented at the Thematic Working Group (TWG) 20 on "Mathematics teacher knowledge, beliefs and identity" at CERME9 in Prague, Czech Republic, February 2015 (see Fauskanger & Mosvold, 2015c).

teaching mathematics internationally exist. In addition, the knowledge requirements of teaching are highly debated, and it is thus far from straightforward to talk about knowledge "needed" for teaching mathematics (e.g. Blömeke & Delaney, 2012). Among the instruments that aim at measuring such knowledge, however, the MKT measures are among the most widely adopted (Blömeke & Delaney, 2012). The MKT measures seem to function well psychometrically across countries (e.g. Blömeke & Delaney, 2012; Fauskanger et al., 2012). Like with any measures, there have also been challenges with the MKT measures. For instance, particular item stems and/or item distracters seem to be problematic (e.g. too difficult, difficult to understand) and display differences in item difficulty or negative point-biserial correlation; this has been observed in different countries (e.g. Ng, Mosvold & Fauskanger, 2012). It has also been observed that minor modifications of items might influence teacher performance (e.g., Kwon, Thames & Pang, 2012). In the context of writing multiple-choice items to measure teachers' specialised knowledge for teaching mathematics, Orrill et al. (2015) have identified five closely interrelated challenges. The purpose of this study is to dig deeper into one of these challenges: "writing clear stems and distracters" (p. 17).

A standard multiple-choice item includes a correct answer—referred to as the key—and a selection of distracters. The distracters are always coded as wrong answers. In the LMT project, some items were formatted with "yes" and "no" as alternative solutions (see Fig. 1). From a test-theoretical perspective, such items might invite test-takers to guess. To avoid this, "I'm not sure" was included as a suggested solution in such MKT items in 2001 (Hill, 2007). In 2002, item developers in the LMT project were required to include "I'm not sure" on all content knowledge items that were piloted in the area of number concepts and operations. To avoid variation in item parameters, "I'm not sure" was also included in the 2004 forms that were later translated and adapted for use in Norway (Fauskanger et al., 2012). "I'm not sure" is always coded as incorrect, and teachers who are experienced with multiple-choice items would immediately know that this suggested solution is a distracter (Osterlind, 1997). Some teachers avoid this suggested solution because they know it will be incorrect; other teachers might select it to avoid guessing or to avoid giving a wrong answer. In his handbook on development and validation of multiple-choice test items, Haladyna (2004) recommends that all distracters in multiple-choice items should be plausible, and it can thus be argued that the suggested solution "I'm not sure" should be avoided. Given that "I'm not sure" is used in many MKT items, further investigations of teachers' considerations when faced with such items are called for.

Previous studies indicate that teachers select this alternative solution for different reasons (Fauskanger & Mosvold, 2014; Haladyna, 2004), and a teacher might select a particular solution for reasons other than those anticipated (Orrill et al., 2015). In this study, we aim at digging deeper into this and investigate possible complexities of including "I'm not sure" as a distracter in MKT items. We analyse data from a collective case study in our endeavour to approach the following research question:

Why do mathematics teachers select "I'm not sure" when responding to MKT items?

If teachers select "I'm not sure" due to lack of knowledge or uncertainty, the inclusion of this suggested solution is unproblematic. If, however, teachers select this alternative solution in MKT multiple-choice items although their responses to other items indicate that they have strong knowledge, there might be a potential source of misreport in the measures. In answering this research question, two specific challenges that have been highlighted by the developers of MKT are addressed: the items might be answered using different types of knowledge such as procedural knowledge or conceptual knowledge (Ball et al., 2008) and a high MKT score might be obtained by utilising procedural knowledge only (Fauskanger, 2015; Hill, Umland & Litke, 2012).

We analyse teachers' responses to a selection of multiple-choice MKT items and compare this with their written long responses concerning the content of the same items and their utterances in group discussions. Adding group discussions is based on recommendations from Adler and Patahuddin (2012) who report that the MKT items "have much potential in provoking teachers' talk and their mathematical reasoning in relation to practice-based scenarios; and exploring with teachers a range of connected knowledge related to the teaching of a particular concept or topic" (p. 17).

Directed content analysis is applied in our analysis of transcripts from group discussions as well as the teachers' long responses (Fauskanger & Mosvold, 2015a; Hsieh & Shannon, 2005). Tchoshanov's (2011) framework of cognitive types of teachers' content knowledge is used as an analytical lens. We focus in particular on three teachers' written long responses and discussions in group interviews regarding MKT items where they selected the alternative "I'm not sure". Cross case displays were constructed from teachers' long responses and their multiple-choice response supplemented with data from the interviews.

Measuring types of mathematics teachers' knowledge

It is relevant to gain insight into methods used to access and assess different aspects of teachers' knowledge (e.g. Hill et al., 2007). One reason why this is important is that teacher knowledge is allegedly imperative to high quality teaching (e.g. Davis & Simmt, 2006). Here, we concentrate in particular on accessing and assessing MKT, defined by Hill, Rowan and Ball (2005, p. 373) as, "the mathematical knowledge used to carry out the work of teaching mathematics". In the LMT project, substantial resources have been invested in developing and validating sets of multiple-choice items in order to assess and access teachers' MKT (e.g. Schilling, Blunk & Hill, 2007). Based on these efforts, Hill et al. (2004) suggest that the MKT measures can be used to measure growth in teachers' knowledge. The teacher knowledge that is measured by the MKT items further relates to the mathematical quality of instruction (Hill et al., 2008) and—to a certain extent—to student learning (Hill et al., 2005; Hill et al., 2012). Inspired by these promising results, researchers have used the measures extensively in the US, and MKT measures have also been adapted and used outside the US (Blömeke & Delaney, 2012). Researchers have discussed numerous concerns about adapting the measures; one concern relates to the challenge of writing clear stems and distracters (Orrill et al., 2015) and refers to what types of knowledge the measures access (cf. Fauskanger, 2015).

Multiple-choice items

There are some noticeable advantages of multiple-choice items. One advantage is that responses from multiple-choice items are expeditiously analysed, and they can be used at scale. Developing multiple-choice items that measure something beyond procedural skills, however, is both tedious and demanding (e.g. Haladyna, 2004; Osterlind, 1997). When discussing measurement of teacher knowledge, researchers have argued that use of multiple-choice items might lead to trivialisation of the complexities of teaching and thus threaten validity (Beswick, Callingham & Watson, 2012; Haertel, 2004). Even when the target construct is clear, items can measure an unintended construct, and a teacher can choose a correct or an incorrect answer for reasons other than those anticipated (Orrill et al., 2015). In his critical discussion of the MKT measures, Schoenfeld (2007) suggests that MKT items could be measuring something other than they were supposed to. He adds that the multiple-choice format might complicate the content for the test takers, and confirming evidence of this was found in a Norwegian context (Fauskanger, Mosvold, Bjuland &

Jakobsen, 2011). A supplementary aspect of Schoenfeld's (2007) criticism is that the items allegedly measure a type of knowledge that is more procedural than intended, and more recent studies in Norway support this criticism (Fauskanger & Mosvold, 2015b; Fauskanger, 2015).

A standard multiple-choice item consists of two parts. The problem is presented in the first part of the item, which is often referred to as the item stem. In MKT items, the stem typically presents a mathematical problem situated in a pedagogical context. In their analyses of the work of teaching mathematics, the Michigan researchers have sought to decompose the work of teaching into mathematical tasks of teaching. Hoover, Mosvold and Fauskanger (2014, p. 13) disclose that, in the MKT items, "mathematical tasks of teaching serve as a kind of backbone inextricably linking mathematical knowledge to the work of teaching." However, producing item stems linking mathematical knowledge to how the knowledge is used in the tasks of teaching mathematics is challenging. Carefully selected alternative solutions are presented in the second part of a multiple-choice item. The list of suggested solutions contains a key and one or more distracters (these are typically incorrect alternatives). In Figure 1, the stem of the question asked, "Which of these students would you judge to be using a method that could be used to multiply any two whole numbers?", but the distractors are Method A, Method B and Method C with 1, 2 and 3 for the different responses. Based on the stem of the question, initially, it might be challenging to work out what 1, 2 and 3 meant. Some MKT items differ from more standard multiple-choice items in that mathematically incorrect alternatives are not always included; the correct solution might then be: "all of the above" or "none of the above". The use of such items—where the correct response is to point to several or no correct solutions—are often discouraged (Haladyna, 2004). If used, they should at least be used with caution (Osterlind, 1997). Another way in which MKT items differ from more standard multiple-choice items is the extended use of the suggested solution "I'm not sure" (Fauskanger & Mosvold, 2014)—first included in items that were formatted yes/no and later for all number and operations content knowledge items (Hill, 2007).

3. Imagine that you are working with your class on multiplying large numbers. Among your students' papers, you notice that some have displayed their work in the following ways:

Student A	Student B	Student C
35	35	35
x 25	x25	x 25
125	175	25
+75	+700	150
875	875	100
		+600
		875

Which of these students would you judge to be using a method that could be used to multiply any two whole numbers?

	Method would work for all whole numbers	Method would NOT work for all whole numbers	I'm not sure
a) Method A	1	2	3
b) Method B	1	2	3
c) Method C	1	2	3

Figure 1. MKT testlet including "I'm not sure" as a suggested solution in all three multiple-choice items (Ball & Hill, 2008, p. 5).

The MKT measures are widely adopted (Blömeke & Delaney, 2012). Still, the process of translating and adapting items is complex (e.g. Fauskanger et al., 2012), and some particular item distracters seem to be problematic across countries (Ng et al., 2012). Orrill et al. (2015) have identified five closely interrelated challenges:

1. creating items with appropriate difficulty levels;
2. creating items for the target constructs;
3. using precise language;
4. incorporating pedagogical concerns appropriately; and
5. writing clear stems and distracters (p. 17).

In this article, the fifth challenge is addressed in order to dig deeper into the vagaries of analysing responses to the MKT multiple-choice items.

Types of teacher knowledge

Different approaches have been made to distinguish between categories of teachers' knowledge and understanding of the mathematical content; Skemp's (1976) categorisation of instrumental and relational understanding is a classic example. Rote memorisation of algorithms for two-digit multiplication (Fig. 1) is an example of instrumental understanding, whereas relational understanding encompasses a deep, conceptual understanding of novel strategies for multi-digit multiplication. Skemp's work—along with the work done by the researcher at the University of

Michigan (e.g. Ball et al., 2008)—was one among several sources of influence when Tchoshanov (2011) developed his framework of cognitive types of teacher content knowledge. Tchoshanov differentiated between knowledge of facts and procedures (termed Type 1 knowledge), knowledge of concepts and connections (termed Type 2 knowledge), and knowledge of models and generalisations (termed Type 3 knowledge). A person with only procedural knowledge can multiply 35 by 25 (Fig. 1); this person would multiply the numbers by using an algorithm without necessarily knowing why it works. A person may, however, also identify the key (i.e., the correct suggested solution) by employing additional knowledge of concepts and connections (Type 2 knowledge). "Knowing why" is closely related to Tchoshanov's (2011) Type 2 knowledge. This type of knowledge is necessary to respond to an item on whether or not a selection of algorithms for two-digit multiplication could be used to multiply any two whole numbers (as in the testlet in Figure 1). Type 3 knowledge would relate to the generalised algorithm for multi-digit multiplication. In his studies, Tchoshanov (2011) did not find strong support for a hypothesis about the effect of Type 3 knowledge on the quality of instruction and student achievement. The MKT items included in our study (table 1) do not focus on models or generalisations, and we have therefore not focused on Type 3 knowledge in the present study.

Skemp (1976) argued that students cannot develop relational understanding from instrumental teaching. Based on empirical evidence, Tchoshanov (2011) concludes that Type 2 knowledge could be a good predictor of teaching that will have a positive impact on students' achievement. This conclusion highlights the importance of investigating whether or not measures of teacher knowledge access Type 2 knowledge (Fauskanger, 2015). For this reason, the importance of exploring cognitive types of teacher knowledge, accessed by measures of teachers' MKT as well as possible challenges regarding the use of multiple-choice items to investigate something as complex as teachers' MKT, becomes evident. It is relevant to carefully investigate these challenges (Orrill et al., 2015), and in particular the challenges related to writing stems and distracters revealed from extensive use of "I'm not sure" as a suggested solution.

Methodology

According to the official coding manuals from the LMT project, the suggested solution "I'm not sure" should be coded as incorrect. An underlying hypothesis is that teachers who select this response do not have the MKT necessary to identify the key; they select "I'm not sure" to avoid giving the wrong answer based on guesswork (Hill, 2007). In order to learn more about why teachers still select "I'm not sure" in multiple-choice items (e.g. Haladyna, 2004), we compare each teacher's responses to the multiple-choice items with the types of knowledge (Tchoshanov, 2011) that can be identified from their written (long) and oral responses to items where they select the answer "I'm not sure" in the multiple-choice item. The intention is not to compare different teachers' MKT, but to compare and contrast the different types of responses each teacher gives to items. The underlying hypothesis is that teachers' MKT should be the same regardless of whether it is accessed by multiple-choice items, written reflections or teachers' individual voices in group discussions.

Participants

We applied a case-oriented approach in this study (Silverman, 2006). We considered the case to be an entity and first looked for configurations and characteristics within the case before we searched for similarities and patterns across cases. Based on a previous study (Fauskanger & Mosvold, 2014), and for the purpose of this paper, three in-service teachers were selected as cases—from a larger sample of 30 teachers (Fauskanger, 2015)—to investigate the topic under investigation. The three teachers, who have been assigned the pseudonyms Laura, Ola and Jane, were participants in a professional development course. They were all experienced mathematics teachers with more than ten years of experience. The course had 38 participants altogether—all in-service teachers—and 30 teachers submitted multiple-choice responses to 28 MKT items (including the three items in the testlet in Fig. 1). Number concepts and operations was the common theme for all 28 items, and 18 of the items included "I'm not sure" as a suggested solution (see e.g. Fauskanger & Mosvold, 2014). Fifteen teachers selected this alternative solution in one, two or three items each (see table 1).

Table 1. *Teachers responding "I'm not sure" (Adjusted table from Fauskanger & Mosvold, 2014, p. 47).*

Item number	1b	1c	1d	5	6a*	6b*	6c*	7d	9d
Number of teachers	2	2	3	1	2	1	5	1	5
Names (pseudo-nyms)	Are and Laura	Are and Laura	Are, Laura and Mons	Harald	Jane and Ragna	Jane	Jane, Ragna, Nina, Ola and Jan	Frøya	Ada, Sara, Pia, Erna and Inge

*6a, 6b and 6c are the multiple-choice items in the testlet presented in Figure 1.

In addition to submitting their multiple-choice responses to the items, the teachers also agreed to submit long responses related to each item and to discuss the items in groups afterwards. The teachers were grouped by the grade level they taught. In this article the focus is not on differences between grade level, formal education or experience; an investigation of such differences is reported elsewhere (Fauskanger, 2013). The questions prompting long responses were developed to tap into teachers' Type 1 and Type 2 knowledge (Tchoshanov, 2011) and varied across the 28 items (see Fauskanger, 2015). These associated open-ended questions were developed by our group of fellow researchers based on extensive analysis of data from research on accessing and assessing aspects of mathematics teachers' knowledge. In MKT, knowledge is defined "in broad terms" (Ball et al., 2008, p. 399). Consequently, questions were also developed to tap into the type of knowledge teachers emphasise in connection with their teaching. The questions were associated with the content of each MKT item specifically and therefore differed from item to item. The open-ended questions first invited teachers to reflect upon the knowledge demonstrated by students (e.g. when using a particular algorithm, see Fig. 1). Second, the questions invited teachers to reflect upon what the students might need to learn as a possible next step. Third, the teachers were invited to connect the content of the MKT items to their own teaching practice and to discuss whether or not they would encourage their students to use, for example, a particular algorithm. Finally, the teachers were asked to provide in-depth reflections about what knowledge they believe is important for their teaching of mathematics, and whether or not the 28 MKT items represented important knowledge for them as teachers. The teachers were asked to justify all their responses.

Efforts have earlier been made to validate the open-ended questions for use in this context (e.g. Fauskanger & Mosvold, 2012, 2014). In a previous publication, we have reported from analysis of multiple-choice responses and long responses for all the teachers in general (Fauskanger, 2015), and for those who selected "I'm not sure" as a response to one or several particular items in particular (Fauskanger & Mosvold, 2014). The results from this latter study indicated that three groups of teachers could be distinguished between: 1) teachers who were not sure about the content in a particular MKT item, 2) teachers whose long responses indicated Type 1 knowledge, and 3) teachers whose long responses indicated Type 2 knowledge (see table 2, first row). Laura, Ola and Jane all selected "I'm not sure" in one or more items, but their long responses indicated that their understanding differed. We consider these three teachers to be representative of the teachers in their respective groups, and in our attempt to further investigate the rationale for selecting "I'm not sure" as a response to MKT items, we therefore selected these three teachers as contrasting cases when analysing group discussions.

Data collection and analyses

Based on analyses of teachers' written responses (Fauskanger, 2015), group discussions were implemented. Adding group discussions is based on recommendations from Adler and Patahuddin (2012) to use MKT items to provoke teachers' talk and their mathematical reasoning. Furthermore, Orrill et al. (2015) recommend interviews as a critical approach to meeting the challenges of multiple-choice item development in general, and to meeting the challenge of writing clear stems and distracters in particular. The group discussions were based on the same 28 MKT items (Adler & Patahuddin, 2012). The questions forming the point of departure for the discussions were developed to facilitate more in-depth discussions of the content of 28 MKT items. Only discussions related to the 18 items including "I'm not sure" as a suggested solution are analysed for the purpose of this paper.

The unit of analysis is the individual teachers' multiple-choice responses, their individual long responses and their individual utterances in group discussions. The group discussions were recorded and transcribed. We have applied an iterative strategy weaving back and forth between the empirical material and theories (Alvesson & Karreman, 2011). For the analyses we used Tchoshanov's (2011) cognitive types of teacher knowledge (Type 1 and Type 2 knowledge) as analytic framework. We considered Tchoshanov's framework to be relevant for this study, since it builds upon Shulman's (1986) categorisation of teacher knowledge and also relates to the MKT framework (Fauskanger, 2015). Tchoshanov's types were used as categories, and we developed related codes that were used for coding of teachers' written long responses and transcripts of their individual utterances from group discussions. Excerpts from teachers' written and oral responses reflecting memorisation of facts or rules, procedural computations or other aspects related to Type 1 knowledge were coded as "Type 1". Excerpts reflecting understanding of concepts and connection between them, multiple solutions to non-routine problems or other aspects related to Type 2 knowledge were coded as "Type 2". A third code, low/no knowledge, was used to code excerpts where teachers explicitly wrote or said that they did not know the content of the item(s), or excerpts revealing insufficient knowledge. In order to increase the reliability of the coding, two researchers coded the complete data set independently. In the few instances where there was a mismatch between our initial coding, we discussed and reached agreement. Since this was the case only in very few instances, we did not find it necessary to use any inter-rater reliability method other than individual coding followed by discussion.

Results and discussion

When we analysed the connection between teachers' multiple-choice responses and their long responses according to Tchoshanov's (2011) framework, three groups of teachers emerged (see table 2, first row). When analysing the teachers' responses from group discussions, the same three groups of teachers could still be distinguished between (see table 2, second row). With one exception (Jan), the teachers remained in the same groups across written responses and group discussions.

Table 2. *Teachers grouped according to their individual long and oral responses.*

Group	Group 1: not sure	Group 2: Type 1 knowledge	Group 3: Type 2 knowledge
Long responses	Erna*, Frøya, **Jane**, Jan, Ada and Nina	Pia, Mons, Harald, **Ola** and Are	Sara, Inge, Ragna and **Laura**
Oral discussion	Frøya, **Jane**, Ada and Nina	Jan, Pia, Mons, Harald, **Ola** and Are	Sara, Inge, Ragna and **Laura**

*Not present in the group discussions.

When analysing data from the group discussions, we found it difficult to distinguish between all the teachers in groups 1 and 2. This becomes apparent from our discussions of the three selected cases below. In the following, we present results from our analysis of data—multiple-choice responses, written long responses as well as oral responses in group discussions—from three teachers: Jane, Ola and Laura.

The case of Jane

Jane's long responses indicated insecurity related to the content in focus, and she selected "I'm not sure" as her response to all three items in the testlet in Figure 1 (6a, 6b and 6c in our form). When asked how she would approach students who use methods like A, B and C, Jane wrote: "It is difficult to know when you do not understand the methods [the students have] used." This long response—along with the other long responses written by Jane—thus seems to support the hypothesis that the selection of "I'm not sure" implies lack of knowledge or insecurity. Coding her responses to the multiple-choice items in this testlet as incorrect thus seems reasonable. It can also be argued that the inclusion of "I'm not sure" has reduced the possibility of guessing with teachers like Jane, and that was the original intention of including this suggested solution in the MKT items (Hill, 2007).

In the group interview, Jane explains that she wants to teach an algorithm that she is more comfortable with herself:

94. Jane: There is a point to explaining that your own way [of calculating it] is all right, in a way. I have used it for calculating for years, so it is natural to me. But, then again, we are different. Some like this and some like that. And it is similar to subtraction, when you borrow 10, if you say for instance 15 minus, or if you say three minus and then add the five. We have shown both ways, and then it is up to them [the students] what they… Then we have said that it relates to what they (...) some like this and some like that, and≈ [sign used to indicate that one interviewee is interrupted by another]

95. Ragna: ≈Open for all [to choose]≈

96. Inga:	≈And then, many [students] have the parents show it to them in a different way than the one they have been taught. Yes, that is difficult.
97. Interviewer:	Do you experience that the parents have a common [algorithm]?
98. Jane:	No. I haven't asked about that, but I know that many [parents] help their children setting it up, and with subtraction, it is quite similar with borrowing (...)
99. Interviewer:	Anything else you want to say in relation to the items that were concerning different algorithms?
100. Jane:	No, and we have been taught that there isn't only one standard algorithm (laughter).

As displayed in this excerpt, the reason why Jane selected "I'm not sure" seems to be that she is insecure about the mathematical content, and she wants to teach the students an algorithm with which she is familiar herself (94). The excerpt thus indicates that Jane holds Type 1 knowledge (Tchoshanov, 2011) related to multi-digit multiplication, at least related to the algorithms present in this testlet (Fig. 1). When seen in relation to Jane's long responses, the excerpts from the group discussion in which Jane participated could also be interpreted as "not sure" based on her writing that she does not understand the methods used by the students (see Fig. 1). Jane reveals, however, that the professional development course has made her aware that multiple algorithms exist, and she seems to agree that different algorithms might be useful for different students (100).

Based on analyses of Jane's written and oral responses, her MKT multi-digit multiplication seems to be somewhere between low and Type 1 (see table 2). It is difficult to distinguish between these, however, indicated by the arrow between the two cells in table 2. This is also true for other teachers who were placed in Group 1 (table 2) based on analyses of their written responses. Jan is one example, and he explicitly said in the group discussion that his view on novel algorithms for multi-digit multiplication had changed.

The case of Ola

Ola is one of five teachers who selected "I'm not sure" in one of the 18 multiple-choice items that included this suggested solution. His long responses indicate that he selected "I'm not sure" based on procedural understanding of the content of the particular item (see table 2). In his long response related to the testlet presented in Figure 1, Ola writes that his students mainly "learn only one method by rote without looking for connections." When faced with students who use the methods A, B and C in Figure 1, he writes that he would teach them "the standard algorithm for multiplication." He will not encourage his students to use the methods A, B and C, because it will be "too much work." Ola teaches in the lower secondary grades, and when writing about algorithms for multi-digit subtraction and multiplication he has not made use of more than one algorithm because "most students have learned the same algorithm I learned myself, 40 years ago, and I don't know if I have experienced anyone not using this when subtracting. There has been one or two examples where students have used a method of multiplication that was unknown to me." Ola also explicitly writes that his knowledge related to algorithms is procedural. At the end of the first semester, he writes "After 3 months in this course, I consider it [knowing more than one algorithm for multi-digit arithmetic] more useful than previously thought. I have previously considered the algorithms for the four mathematical operations only as learned, automatic procedures. I have thought of our place value system mainly as 'innate' knowledge." This excerpt indicates that, during the course, he might have developed a type of knowledge that exceeds Type 1 knowledge (Tchoshanov, 2011), but these written reflections alone are not sufficient to confirm his increased knowledge.

When discussing the MKT items in a group of five teachers, another teacher (Lars) argues that one of the methods used by the students who solved a subtraction problem by addition is "ingenious." Ola questions Lars's argument:

61. Ola: But would you stop, would you think that "but they should learn to do this in another way. They should learn subtraction with borrowing." The common [algorithm], isn't that what you are aiming for?

Ola highlights a standard algorithm related to all items including novel algorithms for multi-digit arithmetic. Later in the discussion, Ola continues to say that "I must admit that I have never cared about why I borrow, even as a grown up, I hardly understand why I carry 9 instead of 10, and then (.) borrow 1 more. (...) This is how we were taught" (line 71). He contends that the professional development course has opened his eyes. However, he would still like his students to use the algorithm that he is familiar with himself: "But we still want them to do it like that, I am still thinking about it, that someone could have had a good and sensible way of doing this in Norway" (line 71). When challenged by the interviewer who points out that the teachers taking part in the professional development course use at least four different algorithms for multi-digit multiplication, Ola says "most [teachers] mean that we should have a standard, that you get there. You have to follow the statistics or the majority" (line 73). From the discussion, Ola displays familiarity with different algorithms of multi-digit arithmetic—or at least he knows that several algorithms exist—but he still argues that the most commonly used algorithm is the one he would prefer in his own teaching (line 73).

Later in the group discussion, the interviewer asks how they would meet students who use the algorithm of student C in Figure 1. This question prompts the following discussion:

80. Ola: We don't want them if it is like that≈

81. Kjell: ≈That we have regarded them as back-up strategies≈

82. Ola: ≈Yes.

83. Kjell: It is good as a temporary strategy, but if it is supposed to be used, then it takes time (...) in the end. So (...) in a starting phase, when they do either like A or B or whatever, then I hope that they manage to take it further eventually, so they manage to get to a higher level, really. That they don't continue to draw helping lines and calculations.

84. Interviewer: I would say that C is more effective. It is faster than the one with borrowing that you now (...) What about those who use one [algorithm] that is faster than (...) "the standard"?

85. Ola: Yes, if they. Yes.

86. Mons: Yes, but I think it helps them. Using some of these methods demand structure from the students. That you write down the numbers in an orderly way, so that you don't mix up the numbers in the end. (...) And then my experience is that if you write more numbers, there is a larger possibility of errors and lack of structure. So, I am a supporter of cutting the amount of numbers you write and write as little as possible. That is a good system. (...) Those who were lazy and found an easier way of getting at the right [answer]. There was a student who received appraisal for that, because laziness should kind of be rewarded from time to time. And then these smart little calculation methods proved efficient. So, I am not a big opponent of all these columns and things like that, because I know it leads to errors. We know that students kind of miss it there≈

87. Ola: ≈But the fast-division [e.g., $7^{1}4^{2}1 : 3 = 247$] was quite ingenious, since it included some mental calculation too. But I thought we could rather do it

like that, because you can use it instead of spending time on making all of those columns, so they make mistakes down there≈

88. Mons: ≈That is why girls win and boys lose, because girls want to make it pretty and boys don't understand their own numbers.

89. Lars: But it demands a lot of practice on mental calculation then.

90. Ola: Yes it does.

When discussing students' use of different algorithms, Ola argues that he does not want students with different algorithms in lower secondary school (80). In their group discussion, Ola and his colleagues show familiarity with different algorithms, but they are concerned that alternative algorithms might be too time consuming (83). They believe that other algorithms might be useful to experience along the way (83), but they are afraid that the students get confused about what carries where (86). Ola states that the "fast division" is an algorithm he wants to use, that is dividing e.g. 741 by three as follows: $7^14^21 : 3 = 247$. However, he becomes hesitant when Lars argues that this algorithm demands a lot of mental calculation (89–90). Ola concludes—and he repeats this several times in the group discussion—that algorithms are not so important for lower secondary school students after all, since they use a calculator.

The case of Laura

Laura also selected "I'm not sure" as her response to several multiple-choice items. In contrast with Jane and Ola, however, Laura's long responses indicate that this was not due to insecurity or Type 1 knowledge. On the contrary, Laura's long responses indicate deep conceptual knowledge, or Type 2 knowledge (Tchoshanov, 2011). In her long response related to a particular testlet (items 1a-d in our form), Laura argued that the stem could be interpreted in different ways and that the choice of key for each item would depend on this interpretation. In this testlet, the stem presents students working in groups decomposing three-digit numbers (e.g., 574) into hundreds, tens, ones and tenths in different ways. The four multiple-choice items invite teachers to evaluate which of the decompositions they, as the teacher, should accept as correct. In the first item (1a), students have decomposed the three-digit number incorrectly: $574=5\times100+70\times10+4$. The remaining three items all represent correct decompositions including: item 1b) $574=4\times100+17\times10+4$, item 1c) $574=4\times100+7\times10+40\times10^{-1}$ and item 1d) $574=57\times10+4$.

The following is an excerpt from what Laura wrote: "Item a) is wrong by all means. Items b), c) and d) are wrong if it [the problem presented in the stem] is a closed problem, but they are correct if it is an open problem." When highlighting testlet 1 as mirroring knowledge important for her as a teacher, Laura wrote:

> To be able to do arithmetic one has to think flexibly when it comes to decomposing a number. 574 is not only 500+70+4. It could also be 400+170+4. 500 is 5 hundreds, 50 tens or 500 ones, etc. The students need to be familiar with this [non-standard ways of decomposing numbers] in order to be able to understand the four arithmetical operations [addition, subtraction, multiplication and division] and in order to develop flexible strategies for multi-digit arithmetic.

Based on these statements, we suggest that when Laura talks about "closed problem", she seems to have the standard decomposition in mind, i.e. $574=5\times100+7\times10+4\times1$. By "open problem", she seems to mean "open" to non-standard ways of decomposing three-digit numbers.

Laura is one of the teachers whose long responses—by relating the decomposition of numbers to understanding of "the four arithmetical operations" and "the development of flexible strategies"—indicate Type 2 knowledge (Tchoshanov, 2011) of the content. In her long responses, Laura relates multiple decompositions to arithmetic, and multiple decompositions seem to be just as important for her as standard decompositions (cf. Jones et al., 1996). Her incorrect multiple-choice responses are thus inconsistent with her long response; she selects “I’m not sure” as a response despite strong MKT. It appears that an explanation for Laura’s choice of “I’m not sure” as a response can be found in the wording of the items included in this testlet. This brings forth issues related to item development and translation (cf. Fauskanger et al., 2012).

Analysis of Laura’s utterances from the interviews supports the hypothesis that she has Type 2 knowledge, in terms of both mathematical content knowledge and pedagogical content knowledge. The interview data thus seems to support the long responses from Laura.

63. Laura: I want to argue that these items are very much about understanding. (...) You have these [MKT items], which relate to understanding what the students actually do. And being able to identify it. And it is what they were able to divide into, I mean, normally we divide into all the hundreds, all the tens and all the ones. That is how to do it. And then they don’t remember that they can change into ones and tens, and then they are stuck there. So this is very relevant.

(…)

137. Laura: It relates to what grade level it is in. So you could say, now I work in third grade, and we have approached the bigger numbers. The first point then is that they know that the numbers have different value if they are in the one, ten or hundred place. Then they know this, and this is the first thing they have to know. And then they know the place value system. But if they are not able to calculate, for instance 200 minus four, because that doesn’t work since there are no ones there [to subtract from in 200]. Then there is something they don’t know after all, about knowing that they have different value. That is the first point. And then it is concerning the flexibility that this item implies. To see if they have [this]. This also has to become natural eventually. But I think it is important to know that there and there and there [points to the digits in the three-digit number] the values are different.

As displayed in these example excerpts, and in particular the last one (137), Laura selected “I’m not sure” despite being secure about the mathematical content. This is the case for all of Laura’s individual contributions to the group discussion. The voice of Laura in the group discussion thus seems to support findings from her long responses, and the reason why she—and the other teachers in this group (group 3, table 2)—responded “I’m not sure” in the multiple-choice items does not relate to insufficient MKT, but rather to deep, conceptual, or Type 2 knowledge (Tchoshanov, 2011) of the content. This might relate to the fact the in Norwegian textbooks the standard decomposition (i.e. $574=5\times100+7\times10+4\times1$) is the focus of attention.

Concluding discussion

In a previous study, we analysed 15 teachers’ long responses related to a set of multiple-choice items from the LMT project (Fauskanger & Mosvold, 2014). Based on those results, three teachers were chosen as contrasting cases in the present study. We analysed the teachers’ responses to multiple-choice MKT items, written long responses as well as discussion of the items in group interviews in order to learn more about what differences can be found between teachers’ arguments for choosing the suggested solution “I’m not sure” in MKT multiple-choice items. Each of the three teachers was chosen because they are representative for their group (table 2).

The distracter “I’m not sure” is always coded as incorrect. However, the results from our analyses indicate that interpretation of teachers’ selection of “I’m not sure” as a multiple-choice response in the MKT items is far from straightforward (cf. Haladyna, 2004). We found that some teachers who selected "I'm not sure" explicitly indicated that they could not identify the key due to their insufficient local MKT (as Jane, see table 2). This seems to be in line with the intention of introducing "I'm not sure" to the items, and the suggested solution reduces the possibility of guessing for these teachers (Hill, 2007). Preferably, all teachers who select "I'm not sure" should have been in this category (Haladyna, 2004), but in our study they were not. Other groups of teachers selected "I'm not sure" as a response to multiple-choice items although their long responses as well as their discussions in group interviews indicated that they had Type 2 knowledge (e.g. Laura, see table 2) or Type 1 knowledge (e.g. Ola, see table 2). Teachers might draw on deep conceptual, relational Type 2 knowledge (Laura), procedural, instrumental Type 1 knowledge (Ola), or their lacking knowledge or insecurity (Jane) when responding, "I'm not sure". Our results thus indicate that the type of knowledge (Tchoshanov, 2011) teachers utilise in long responses or discussions does not necessarily mirror the type of knowledge (that seems to be) used when selecting a certain multiple-choice response. The assumption that the multiple-choice response "I'm not sure" is correctly coded as incorrect should therefore be subject to further scrutiny (Haladyna, 2004; Osterlind, 1997), and the inclusion of this alternative response in MKT items should also be critically discussed.

Our analysis indicates that none of the 30 participating teachers were guessing—thus conforming to the intention of including "I'm not sure"—but the inclusion of "I'm not sure" is still problematic. The choice of this suggested solution, it appears, could hinge on anything from lack of knowledge to deep conceptual knowledge, or Type 2 knowledge (Tchoshanov, 2011). One might advocate removal of "I'm not sure" as an option in the items, but this would change item parameters and should not be done hastily. Instead, we suggest two possible solutions. On the one hand, we argue that researchers who apply MKT measures need to be careful when many teachers select "I'm not sure" as response. Our study indicates that this alternative response does not always imply weak MKT, and a large number of such responses represents an uncertainty that needs to be addressed when analysing the results. On the other hand, we suggest that the presence of this alternative solution in a number of MKT items could be used as a starting point for cross-cultural studies that investigate teachers' motivation for selecting "I'm not sure" when responding to MKT items. We suspect that there might be cultural differences in teachers' selection of responses in MKT items, and comparison studies would be relevant in order to learn more about the challenges of "I'm not sure" (Haladyna, 2004; Osterlind, 1997) as a suggested solution in the MKT multiple-choice items across cultural contexts.

A limitation of our study is the relatively small sample size of 30 teachers, and the results from this study cannot be used to make claims about general response patterns. On the other hand, our study provides an existence proof that different reasons for selecting "I'm not sure" are possible. This indicates that a potential source of misreport seems to be embedded in the MKT measures, and this should be subject to further scrutiny — in particular related to the challenge in writing clear stems and distracters (Orrill et al., 2015). Larger studies would reveal if these tendencies can be found at scale, and cross-cultural studies could show if there are different tendencies across cultures. Cross-cultural studies would also have the potential of revealing why teachers select a particular multiple-choice response (e.g. "I'm not sure"), how this differs across countries, and whether and how this relates to the work of teaching mathematics and the tasks of teaching embedded in this work (Hoover et al., 2014; Kazima, Jakobsen & Kasoka, 2016).

References

Adler, J., & Patahuddin, S. M. (2012). Recontexualising items that measure mathematical knowledge for teaching into scenario based interviews: An investigation. *Journal of Education*, *56*, 17–43.

Alvesson, M., & Karreman, D. (2011). *Qualitative research and theory development: Mystery as method*. London: Sage Publications.

Ball, D. L., & Hill, H. C. (2008). Mathematical knowledge for teaching (MKT) measures. Mathematics released items 2008. Retrieved from http://www.umich.edu/~lmtweb/files/lmt_sample_items.pdf

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.

Beswick, K., Callingham, R., & Watson, J. (2012). The nature and development of middle school mathematics teachers' knowledge. *Journal of Mathematics Teacher Education, 15*(2), 131–157.

Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM – The International Journal on Mathematics Education, 44*(3), 223–247.

Davis, B., & Simmt, E. (2006). Mathematics-for-teaching: An ongoing investigation of the mathematics that teachers (need to) know. *Educational Studies in Mathematics, 61*(3), 293–319.

Fauskanger, J. (2015). Challenges in measuring teachers' knowledge. *Educational Studies in Mathematics, 90*(1), 57–73.

Fauskanger, J. (2013). Teachers' epistemic beliefs about mathematical knowledge for teaching two-digit multiplication. In M. S. Hannula, P. Portaankorva-Koivisto, A. Laine & L. Näveri (Eds.), *Current state of research on mathematical beliefs XVIII: Proceedings of the MAVI-18 conference, September 12-15, 2012* (pp. 271–284). Helsinki: University of Helsinki.

Fauskanger, J., & Mosvold, R. (2015a). En metodisk studie av innholdsanalyse – med analyser av matematikklæreres undervisningskunnskap som eksempel. [English: A methodolodical study of content analysis – with analyses of teachers' mathematical knowledge for teaching as an example]. *Nordic Studies in Mathematics Education, 20*(2), 79–96.

Fauskanger, J., & Mosvold, R. (2015b). The difficulty of measuring types of mathematics teachers' knowledge. In Silfverberg H., Kärki T. and Hannula M. S. (Eds.), *Nordic research in mathematics education – Proceedings of NORMA14, Turku, June 3-6, 2014. Studies in Subject Didactics 10* (pp. 71–80). Turku: The Finnish Research. Association for Subject Didactics.

Fauskanger, J., & Mosvold, R. (2015c). Why are Laura and Jane "not sure"? In K. Krainer & N. Vondrova (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (pp. 3192–3198). European Society for Research in Mathematics Education.

Fauskanger, J., & Mosvold, R. (2014). Studying teachers' knowledge by the use of multiple-choice items. The case of "I'm not sure". *Nordic Studies in Mathematics Education, 19*(3-4), 41–55.

Fauskanger, J., & Mosvold, R. (2012). "Wrong, but still right". Teachers reflecting on MKT items. In L. R. Van Zoest, J.-J. Lo and J. L. Kratky (Eds.), *Proceedings of the 34th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 423–429). Kalamazoo, MI: Western Michigan University.

Fauskanger, J., Mosvold, R., Bjuland, R., & Jakobsen, A. (2011). Does the format matter? How the multiple-choice format might complicate the MKT items. *Nordic Studies in Mathematics Education, 16*(4), 45–67.

Fauskanger, J., Jakobsen, A., Mosvold, R., & Bjuland, R. (2012). Analysis of psychometric properties as part of an iterative adaptation process of MKT items for use in other countries. *ZDM – The International Journal on Mathematics Education, 44*(3), 387–399.

Haertel, E. (2004). Interpretive argument and validity argument for certification testing: Can we escape the need for psycological theory? *Measurement: Interdisciplinary Research and Perspectives, 2*(3), 175–178.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). New Jersey, NJ: Lawrence Earlbaum Associates Inc.

Hill, H. C. (2007). *Introduction to MKT scales. Mathematical Knowledge for Teaching (MKT) measures*. Ann Arbor, MI: University of Michigan.

Hill, H. C., Blunk, M., Charalambous, C. Y., Lewis, J. M., Phelps, G., Sleep, L., & Al., E. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.
Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematical knowledge for teaching. *The Elementary School Journal, 105*(1), 11–30.
Hill, H. C., Sleep, L., Lewis, J. M., & Ball, D. L. (2007). Assessing teachers' mathematical knowledge. What knowledge matters and what evidence counts? In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 111–156). Charlotte, NC: Information Age Publishing.
Hill, H. C., Umland, K., Litke, E., & Kapitula, L. R. (2012). Teacher quality and quality teaching: Examining the relationship of a teacher assessment practice. *American Journal of Education, 118*(4), 489–519.
Hoover, M., Mosvold, R., & Fauskanger, J. (2014). Common tasks of teaching as a resource for measuring professional content knowledge internationally. *Nordic Studies in Mathematics Education, 19*(3–4), 7–20.
Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288.
Jones, G. A., Thornton, C. A., Putt, I. J., Hill, K. M., Mogill, T. A., Rich, B. S., & Van Zoest, L. R. (1996). Multidigit number sense: A framework for instruction and assessment. *Journal for Research in Mathematics Education, 27*, 117–143.
Kazima, M., Jakobsen, A., & Kasoka, D. N. (2016). Use of Mathematical Tasks of Teaching and the Corresponding LMT Measures in the Malawi Context. *The Mathematics Enthusiast, 13*(1–2), 172–187.
Kwon, M., Thames, M. H., & Pang, J. (2012). To change or not to change: adapting mathematical knowledge for teaching (MKT) measures for use in Korea. *ZDM – The International Journal on Mathematics Education, 44*(3), 371–385.
Ng, D., Mosvold, R., & Fauskanger, J. (2012). Translating and adapting the mathematical knowledge for teaching (MKT) measures: The cases of Indonesia and Norway. *The Montana Mathematics Enthusiast, 9*(1–2), 149–17.
Orrill, C. H., Ok-Kyeong, K., Peters, S. A., Lischka, A. E., Jong, C., Sanchez, W. B., & Eli, J. A. (2015). Challenges and Strategies for Assessing Specialised Knowledge for Teaching. *Mathematics Teacher Education and Development*, *17*(1), 12–29.
Osterlind, S. J. (1997). *Constructing test items: Multiple-choice, constructed-response, performance and other formats* (2nd ed.). Hingham, MA: Kluwer Academic Publishers.
Schilling, S. G., Blunk, M., & Hill, H. C. (2007). Test validation and the MKT measures: Generalizations and conclusions. *Measurement: Interdisciplinary Research and Perspectives, 5*(2-3), 118–128.
Schoenfeld, A. H. (2007). Commentary: The complexities of assessing teacher knowledge. *Measurement: Interdisciplinary Research and Perspectives, 5*(2), 198–204.
Silverman, D. (2006). *Interpreting qualitative data: Methods for analysing talk, text and interaction* (3rd ed.). Thousand Oaks, CA: Sage Publications Inc.
Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*, 20–26.
Tchoshanov, M. A. (2011). Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics. *Educational Studies in Mathematics, 76*(2), 141–164.

Authors

Janne Fauskanger
Department of Education and Sports Science
University of Stavanger
4036 Stavanger
Norway
email: janne.fauskanger@uis.no

Reidar Mosvold
Department of Education and Sports Science
University of Stavanger
4036 Stavanger
Norway
email: reidar.mosvold@uis.no