Studying teachers' knowledge by the use of multiple-choice items: the case of "I'm not sure"

JANNE FAUSKANGER AND REIDAR MOSVOLD

The mathematical knowledge for teaching (MKT) measures have been widely adopted by researchers in several countries. This article reports on a study on the connection between teachers' responses to multiple-choice MKT items, and in particular where they select the suggested solution "I'm not sure", and their written responses to corresponding open-ended questions (long responses). The findings from our analysis of 15 teachers' responses indicate that their long responses and their multiple-choice responses do not always correspond. Some teachers who selected "I'm not sure" showed uncertainty also in their long responses, whereas other teachers revealed instrumental and even relational understanding of the content.

High-quality teaching has several characteristics, and various components of teacher knowledge are included among them (e.g. Davis & Simmt, 2006; Tchoshanov, 2011). Over the years, a number of theories about teacher knowledge have emerged, and researchers have developed and used measures to learn more about the different components of teacher knowledge (e.g. Hill, Sleep, Lewis & Ball, 2007). We focus on one particular theory of mathematics teachers' knowledge, which is often referred to as Mathematical Knowledge for Teaching (MKT) (Ball, Thames & Phelps, 2008), and an instrument that was developed to measure teachers' MKT in the Learning Mathematics for Teaching (LMT) project at the University of Michigan (e.g. Hill et al., 2007). All the items in this instrument are situated in a classroom context. Based on their studies, Hill and colleagues (2004) claim that the LMT measures can be used to measure growth in teachers' knowledge, and they argue that teachers' scores on the measures can predict mathematical features of their instruction (Hill et al., 2008). The suggestion that teachers' knowledge relates to mathematical quality of instruction (e.g. Hill et al., 2008) and to student outcomes (Hill et al.,

Janne Fauskanger, *University of Stavanger*
Reidar Mosvold, *University of Stavanger*

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.

2005) has led to the adaptation of the measures by numerous researchers both in the United States and elsewhere (e.g. Blömeke & Delaney, 2012; Delaney, 2012).

In previous studies, the translation (Mosvold, Fauskanger, Jakobsen & Melhus, 2009), adaptation and use (Drageset, 2009; Fauskanger, Jakobsen, Mosvold & Bjuland, 2012) of LMT items have been studied in a Norwegian context. One issue that has been investigated in these previous studies relates to the use of multiple-choice format in the LMT measures. Fauskanger and colleagues (2011) argued that the multiple-choice format might complicate the items for Norwegian teachers, and they suggested that attempts should be made to investigate ways of "opening up" the LMT items (cf. Schoenfeld, 2007). Following up on these results, a study was conducted in which open-ended questions were added to a selection of multiple-choice items from the LMT project. Fauskanger and Mosvold (2012) reported on a potential mismatch between the 30 participating teachers' responses to a randomly chosen multiple-choice item and their written responses to the associated open-ended questions (hereafter referred to as "long responses"). Their results also indicated that it might be interesting to look closer into the teachers' reasoning when they selected the solution "I'm not sure" to multiple-choice items. In this article, we build on these previous results when we focus our attention on the connection between teachers' responses to a selection of 28 LMT multiple-choice items and the associated long responses. The items were selected because they represent important themes in the Norwegian curriculum as well as in the in-service education the teachers participated in. In particular, we focus on the multiple-choice items where teachers selected the suggested solution "I'm not sure", and we investigate the connection between these responses and the corresponding long responses. We address the following research question:

> Which (if any) relationship(s) exist between teachers' content knowledge as expressed through teachers' responses to a selection of multiple-choice LMT items, including the suggested solution "I'm not sure", and their written responses to open-ended questions concerning the content of the same items?

We investigate the relationships between teachers' multiple-choice responses to LMT items and their corresponding long responses from a cognitive perspective, and we use qualitative content analysis (Hsieh & Shannon, 2005) as an analytical approach. Skemp's (1976) distinction between instrumental and relational understanding serves as a starting point.

Theoretical background

Shulman's (1986, 1987) work on teacher knowledge has been decisive for the development of quite a few successive frameworks (e.g. Graeber & Tirosh, 2008), and his categorization of teacher knowledge has engendered subsequent categorizations. Shulman suggested the following domains of teachers' content knowledge: *subject matter knowledge*, *pedagogical content knowledge* and *curricular knowledge*. Subject matter knowledge involves "the amount and organization of knowledge per se in the mind of the teachers" (Shulman, 1986, p. 9); this is content knowledge generally shared by all well-educated people in some domain (e.g. mathematics). Pedagogical content knowledge refers to instruction and consolidation of teachers' knowledge of content with their pedagogical knowledge, and – according to Shulman – this is the only special content knowledge teachers needs to possess.

In their efforts to develop a practice-based theory of mathematical content knowledge for teaching, Ball and colleagues (2008) focused on the "work of teaching"; according to them, the work of teaching entails everything a teacher does that is related to teaching (mathematics). This work led to identification of specific mathematical tasks of teaching. Based on analyses of these tasks, the theory of MKT was developed to consist of Shulman's (1986, 1987) two broad categories: subject matter knowledge and pedagogical content knowledge. Each of these broad categories has been further subdivided into three domains. Pedagogical content knowledge has been divided into: (1) *knowledge of content and students*, (2) *knowledge of content and teaching* and (3) *knowledge of content and curriculum*. Subject matter knowledge, on the other hand, has been divided into: (1) *common content knowledge*, (2) *specialized content knowledge* and (3) *horizon content knowledge*. In the MKT framework, pedagogical content knowledge is not the only knowledge domain that is special to teachers; aspects of mathematics content knowledge – like specialized content knowledge – are also particular to the profession. The development of measures has played an important part in the process of developing the practice-based theory of MKT (Ball et al., 2008). An example from one of the released testlets (composed by more than one multiple-choice question) is given in figure 1.

Being able to multiply 35 and 25 relates to common content knowledge, which is the mathematical knowledge that any educated person would know (Ball et al., 2008). Common content knowledge thus includes – but is not limited to – knowledge of number facts and algorithms; it could also include knowledge of concepts and connections. Within the frame of common content knowledge, a person could solve $35 \cdot 25$ instrumentally (Skemp, 1976) and get the correct answer – without knowing why the

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 x 25 125 +75 875	35 x25 175 +700 875	35 x 25 25 150 100 +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. *LMT testlet developed to measure teachers' subject matter knowledge (Ball & Hill, 2008, p. 5)*

algorithm works. Strong specialized content knowledge does, however, require understanding a variety of algorithms for two-digit multiplication; it also indicates ability to assess whether or not the algorithms could be used to multiply any two whole numbers (as in figure 1). Specialized content knowledge "allows teachers to engage in particular teaching tasks, including how to accurately represent mathematical ideas, provide mathematical explanations for common rules and procedures, and examine and understand unusual methods to problems" (Hill, Ball & Schilling, 2008, p. 378). Specialized content knowledge is thus closely related to Skemp's (1976) relational understanding, but there is more to MKT than relational understanding.

Skemp (1976) divided teachers' mathematical understanding into two divergent categories: (1) *Instrumental understanding* – a less robust version of understanding – for example rote memorization of algorithms for two-digit multiplication and (2) *Relational understanding*, which encompasses a deep, conceptual understanding. When relating his framework to teaching, Skemp (1976) suggested that a teacher who teaches from an instrumental paradigm cannot produce students who learn mathematics relationally. Recent studies confirm that conceptual and connected mathematical knowledge is a premise for conceptual teaching (e.g. Tchoshanov, 2011). This confirmation highlights the importance of studying possible relationship(s) between teachers' content knowledge as expressed through teachers' responses to multiple-choice LMT items and their long responses.

Multiple-choice items

An advantage of multiple-choice items is that they can be used at scale, and responses are quickly analyzed. However, multiple-choice items are challenging to develop (e.g. Haladyna, 2004; Osterlind, 1997), in particular if the items should not measure skills such as recall and procedures in mathematics (Haertel, 2004; Haladyna, 2004). Several studies have compared the use of multiple-choice and open-ended items (e.g. Hollingworth, Beard & Proctor, 2007), and there are indications that these different formats actually measure different types of knowledge. Hollingworth and colleagues (2007) argued that the two formats are equally effective and conclude that both open-ended and multiple-choice items are related to a common factor. Concerning the LMT measures in particular, Schoenfeld (2007) argued that these measures might test something else than they are supposed to. Others claim that validity might be threatened because the use of multiple-choice items can lead to trivialization of the complexities of teaching (Haertel, 2004) because "various facets of teacher knowledge develop together" (Beswick, Callingham & Watson, 2012, p. 131) and cannot be measured separately by the use of multiple-choice items. The format of the LMT items might also complicate the content for the test takers (Fauskanger et al., 2011; Schoenfeld, 2007). There are, then, challenges regarding the use of multiple-choice items to investigate something as complex as teacher knowledge, and it is relevant to carefully investigate these challenges.

A standard multiple-choice item consists of two parts: a problem (also called a stem), and a list of suggested solutions. This list normally contains one correct alternative (referred to as the key) and one or more incorrect alternatives (distractors). Some LMT items differ from more standard multiple-choice items. As an example, some items do not include any incorrect alternatives; the correct solution is the alternative "all of the above". Osterlind (1997) recommends that such items should be used with caution, whereas others suggest that they should be avoided altogether (Haladyna, 2004). If a test taker knows that two of the three alternative solutions are correct, he or she can use this information to correctly choose "all of the above". Another difference in the LMT items is the widespread use of the suggested solution "I'm not sure" (see figure 1) – which is always coded as incorrect. Teachers therefore immediately know that one of the suggested solutions will be a distractor in these items.

This might invite teachers to avoid this suggested solution; other teachers might choose this solution to avoid giving a wrong answer. Haladyna (2004) recommends that all distractors should be plausible. It is thus interesting to study the plausibility of "I'm not sure".

Methodology

Participants and instrument

The participants in this study, 30 in-service teachers, were participants in a professional development course. They taught different grade levels: 17 in grades 1–4, 8 in grades 5–7 and 5 in grades 8–10. Their formal education in mathematics/mathematics education also varied, and their teaching experience varied from less than 5 years to more than 20 years. Our focus in the analysis is on relationship(s) between teachers' content knowledge as expressed through multiple-choice questions and long responses – not on differences between gender, grade level, formal education or experience. To inform the reader, however, we have presented these data – as background information – in table 2

The teachers agreed to submit multiple-choice responses to 28 LMT items, which had been translated into Norwegian (Fauskanger et al., 2012; Mosvold et al., 2009), as well as long responses related to each item. All of these items had a focus on number concepts and operations. The teachers worked on the LMT items and long responses at home. Although this has also been done in other studies – both inside and outside the U.S. – it can be seen as a limitation. For this study, however, our focus was to investigate the connection between teachers' different kinds of responses rather than providing a measure of their MKT as such, and we therefore found the approach acceptable. The questions prompting long responses were developed to tap into teachers' instrumental and relational understanding (Skemp, 1976) and varied across the 28 items. We analyzed long responses in the cases where teachers selected the suggested solution "I'm not sure"; 18 out of the 28 items included this suggested solution.

Analysis

The unit of analysis is individual teachers' multiple-choice responses to LMT items and their long responses. We used qualitative content analysis to investigate the relationships between these two kinds of responses. This approach is regarded as a flexible way of analyzing textual data, and a systematic approach to classify and identify themes or patterns (Hsieh & Shannon, 2005).

We used Skemp's (1976) categories as a starting point for our coding in order to learn more about what types of understanding could be found in teachers' long responses when they select the answer "I'm not sure" on a multiple-choice item. Excerpts from teachers' long responses reflecting memorization of facts or rules, procedural computations or other aspects related to instrumental understanding were coded as *instrumental*, whereas excerpts reflecting understanding of concepts and connection between them multiple solutions to non-routine problems or other aspects related to relational understanding – like in the excerpt presented in the results section – were coded as *relational*. A third code was *low/no MKT* used to code excerpts where teachers' explicitly wrote that they did not know the content of the item(s) or excerpts revealing low level of MKT.

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 would then be that teachers who select this response do not have the proper level of MKT to identify the key.

In order to increase the reliability of the coding, the two authors coded the data independently and reconciled. In the few instances where there was a mismatch between our initial coding, we discussed and reached agreement.

Results – the case of "I'm not sure"

Out of the 28 multiple-choice items, 18 items included the suggested solution "I'm not sure" (see e.g. figure 1). Fifteen teachers selected this alternative solution in one, two or three items each (see table 1).

Table 1. *Teachers responding "I'm not sure" (names are fictitious)*

Item number	No. of teachers	Teachers
1b	2	Are and Laura
1c	2	Are and Laura
1d	3	Are, Laura and Mons
5	1	Harald
6a	2	Jane and Ragna
6b	1	Jane
6c	5	Jane, Ragna, Nina, Ola and Jan
7d	1	Frøya
9d	5	Ada, Sara, Pia, Erna and Inge

When analyzed through the lens of Skemp's (1976) categories, we would expect that most teachers in their long responses might indicate instrumental understanding of the content. In our analysis of these 15 teachers' long responses, however, the teachers were spread across all three categories. As can be seen in table 2, four of the teachers even showed relational understanding in their long responses – although they had selected the distractor "I'm not sure".

Table 2. *Nature of teachers' corresponding long responses*

Groups	Name of teacher	Grade level taught	ECTS in mathematics (education)	Years of teaching experience
1: not sure	Erna	5–7	1–15	2–5
	Frøya	5–7	No	11–20
	Jane	1–4	1–15	21+
	Jan	5–7	1–15	11–20
	Ada	1–4	1–15	11–20
	Nina	8–10	No	11–20
2: instrumental understanding	Pia	1–4	1–15	11–20
	Mons	8–10	1–15	2–5
	Harald	5–7	No	2–5
	Ola	8–10	31–60	6–10
	Are	5–7	No	2–5
3: relational understanding	Sara	1–4	1–15	11–20
	Inge	5–7	No	11–20
	Ragna	1–4	1–15	11–20
	Laura	1–4	1–15	6–10

Six teachers indicated uncertainty in their long responses. This group of teachers could be split into two sub-groups. The first four teachers in group 1 – Erna, Frøya, Jane and Jan – explicitly wrote in their long responses that they did not understand the content of the item(s). In relation to a testlet focusing on rules of thumb (testlet 7 in our form, see table 1), one of these four teachers, Frøya, wrote: "I would not have used [the rule of thumb presented in] d [in my class] because I did not understand it [the rule] myself." A second example is Jane who responded "I'm not sure" to all three items in the testlet in figure 1 (6a, 6b and 6c in our form, see table 1). When asked how she would approach students who used methods like A, B and C, Jane wrote: "It is difficult to know when you do not understand the methods [the students have] used."

The remaining two teachers in group 1 – Ada and Nina – selected "I'm not sure", but they did not provide any additional reflections regarding this choice in their long responses. These two teachers were therefore placed in group 1, although their answers to other items might indicate a higher level of MKT. When analyzing Nina's long responses to item c in the testlet presented in figure 1 (item 6c in our form) – where she had responded "I'm not sure" – it seems like she gave this multiple-choice response based on uncertainty related to whether or not to use this method in her teaching. Although she did not provide any explicit explanation for why she selected "I'm not sure", she did not seem to be uncertain about whether or not method c could be used to multiply any two whole numbers; her long response was incomplete. The long responses of Erna, Frøya, Jane and Jan thus seemed to support the hypothesis that the selection of "I'm not sure" implies lack of knowledge, but the long responses of Ada and Nina did not include so much supporting information.

The teachers who were placed in group 2 in our analysis indicated an instrumental understanding of the content in their long responses. Typically, a teacher's response would be placed in this category if (s)he referred to a standard algorithm or textbook definition when trying to elaborate on why (s)he chose this particular alternative solution in the multiple-choice item. An example of this is Ola's long response to the item in figure 1. He wrote that the students need to learn more about "the standard algorithm for multiplication. [It] is faster for them to work with." Harald's long response related to an item focusing on possible definitions of a prime number is another example: "I have to admit that I find it difficult to give an answer to the question if it is important to know the definition of a prime number and definitions in general." Harald continued to write that definitions are easy to look up in books or on the Internet and that this item number 5 in our form (see table 1) had invited him to think that if definitions are used in tasks and on exams in higher grades he might focus more on definitions in his future teaching. Harald's long responses indicate that he might have given his multiple-choice response based on uncertainty related to the importance of knowing definitions, rather than uncertainty related to the particular definition of a prime number given in this item.

As we have seen already, there are indications that the connection between teachers' responses to the multiple-choice questions and their long responses is not necessarily straightforward. The long responses of the teachers in group 3 are perhaps the most interesting. These four teachers – Sara, Inge, Ragna and Laura – also responded "I'm not sure" to some of the multiple-choice items. When analyzing the additional long responses given by these four teachers, however, we concluded that they seem to have a relational understanding of the content. The long responses from this group of teachers indicate that their reasons for responding ''I'm not sure'' relate to aspects such as the wording or context of the items and thus to interpretation of the items rather than to their level of MKT. When analyzing the teachers' long responses, it is evident that the wording is an issue for three out of the teachers responding ''I'm not sure'' related to item 9d – an item part of a testlet focusing on whether or not word problems correctly represent 3 ÷ ½. Sara indicated a relational understanding in her long response, and she seemed to have responded ''I'm not sure'' due to the wording of item 9d. This brings forth issues related to item development and translation (cf. Fauskanger et al., 2012; Mosvold et al., 2009).

Laura is another example of a teacher who displayed deep conceptual knowledge (Skemp, 1976) related to place value (items 1a–d in our form) even though her multiple-choice responses were ''I'm not sure''. In this particular testlet, the stem presents a context dealing with groups of students who have decomposed a three-digit number (e.g. 456) into hundreds, tens, ones and tenths in different ways. In the first item (1a), the students have decomposed the three-digit number incorrectly (e.g. 456 decomposed into 4 hundreds, 50 tens and 6 ones). The remaining three items all represent correct decompositions including hundreds, tens and ones (1b), hundreds, tens and tenths (1c) and tens and ones (1d). Laura argued in her long response that the stem could be interpreted in different ways and that the choice of key for each item would depend on this. The following is an excerpt from Laura's long response: ''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.'' By ''closed problem'' Laura seemed to have in mind the standard decomposition, and by ''open problem'' she meant ''open'' to non-standard ways of decomposing three-digit numbers. 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 tenths 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.

Laura is one of the teachers whose long response – by relating the decomposition of numbers to understanding of ''the four arithmetical operations'' and "the development of flexible strategies" – indicates relational understanding (Skemp, 1976) of the content. In her long response, Laura related multiple decompositions to arithmetic, and multiple decompositions seem to be just as important for her as standard decompositions (c.f. Jones et al., 1996). Her incorrect multiple-choice responses are thus inconsistent with her long response, and the reason why she responds "I'm not sure" relates to her high level of MKT.

Conclusion

From studying the long responses related to 15 teachers' multiple-choice response(s) "I'm not sure", our findings indicate that this multiple-choice response seems to be given by teachers with relational understanding (as the teachers in group 3, table 2) as well as by teachers who explicitly indicated that they could not identify the key due to their low level of *local* MKT (as the teachers in group 1, table 2). The long responses from the 15 teachers can be grouped into three – as presented in table 2. Included in group 1 are the teachers who wrote that they responded "I'm not sure" because they did not know the content the item was developed to measure. The long responses from the teachers in group 2 indicated an instrumental understanding. In the validation studies of the LMT measures (e.g. Delaney, 2012), instrumental understanding was coded as MKT. Instrumental understanding does not, however, relate to teaching that promotes conceptual understanding (Tchoshanov, 2011), and the coding of "I'm not sure" for this group of teachers should be discussed. The third group includes the teachers who responded "I'm not sure" but whose long responses indicated a high level of MKT.

Our results indicate that the knowledge teachers utilize does not necessarily mirror the multiple-choice response given. Teachers might draw on deep conceptual or relational knowledge, procedural or instrumental knowledge (Skemp, 1976), or their lacking knowledge when responding "I'm not sure". The assumption that the multiple-choice response "I'm not sure" is correctly coded as *incorrect* should thus be subject to further scrutiny.

Research related to multiple-choice item development (Haladyna, 2004; Osterlind, 1997) recommends test developers to avoid suggested solutions including such as "all of the above" and to avoid negative words such as *not* in the item stem as well as in the suggested solutions. One way of explaining the results in this study might then be that some of the multiple-choice items were ambiguous and in need of revision; some items could even have been replaced by better items in the first place. Another possible explanation is that the teachers who seemed to understand the content but still answered "I'm not sure" had poor test-taking strategies – and this might be culturally related. Norwegian teachers are not so familiar with multiple-choice measures, and the format itself could have made it more complicated for the teachers to pick the correct alternative solution from the multiple-choice item (cf. Fauskanger et al., 2011). Finally, the apparent mismatch between several teachers' responses to the multiple-choice items and their responses to the open-ended questions might indicate that these two formats measure different aspects of teacher knowledge (e.g. Hollingworth et al., 2007) or that teachers' MKT is too complex to be measured by multiple-choice items (e.g. Haertel, 2004; Beswick et al., 2012).

Our results can be criticized regarding the size of the sample, and we cannot argue that our results can be generalized to a larger population. The relatively small sample size did, however, provide us with the opportunity to make more in-depth qualitative analyses of the data material, and the results from this study would be relevant to follow up in future studies. More research is necessary in order to investigate whether or not the same tendencies can be found in a larger population of teachers and whether the same pattern can be found for all sets of LMT items. The results from our study indicate that the "opening up" of the multiple-choice items might provide us with more information about teachers' knowledge, and we argue that this might enable researchers to get a more complete view of teachers' MKT. The downside, however, is that the increased amount of information makes the data material much more difficult to analyze, and the question of whether or not questions of different format actually measure the same thing should be taken into consideration. We do, however, argue that it is still relevant to open up the items like this, and we believe this can be done both as part of the process of ensuring the quality of the measures as well as in order to learn more about teachers' MKT. The results from our study also indicate that the use of "I'm not sure" as a suggested solution in MKT items can be problematic, and future research is called for in order to investigate whether or not the tendencies observed in our data material can also be found in a larger sample of teachers. It would also be relevant to investigate whether or not this suggested solution is particularly problematic in a context where teachers are not so familiar with the multiple-choice format.

References

Ball, D. L. & Hill, H. C. (2008). *Mathematical knowledge for teaching (MKT) measures. Mathematics released items 2008*. Retrieved from http://sitemaker.umich.edu/lmt/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*, 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.

Delaney, S. (2012). A validation study of the use of mathematical knowledge for teaching measures in Ireland. *ZDM*, 44 (3), 427–441.

Drageset, O. G. (2009). Exploring mathematical knowledge for teaching. In M. Tzekaki, M. Kaldrimidou & C. Sakonidis (Eds.), *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 487–480). Thessaloniki: PME.

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*, 44 (2), 387–399.

Fauskanger, J. & Mosvold, R. (2012). "Wrong, but still right". Teachers reflecting on MKT items. In L. R. van Zoest, J.-J. Lo & 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: 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.

Graeber, A. & Tirosh, D. (2008). Pedagogical content knowledge. Useful concept or elusive notion. In P. Sullivan & T. Wood (Eds.), *Knowledge and beliefs in mathematics teaching and teaching development* (pp. 117–132). Rotterdam: Sense Publishers.

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

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

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.

Hill, H. C., Blunk, M., Charalambous, C. Y., Lewis, J. M., Phelps, G. et al. (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: Information Age Publishing.

Hollingworth, L., Beard, J. J. & Proctor, T. P. (2007). An investigation of item type in a standards-based assessment. *Practical Assessment Research & Evaluation*, 12 (18), 1–13.

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. et al. (1996). Multidigit number sense: a framework for instruction and assessment. *Journal for Research in Mathematics Education*, 27 (3), 310–336.

Mosvold, R., Fauskanger, J., Jakobsen, A. & Melhus, K. (2009). Translating test items into Norwegian – without getting lost in translation? *Nordic Studies in Mathematics Education*, 14 (4), 101–123.

Osterlind, S. J. (1997). *Constructing test items: multiple-choice, constructed-response, performance and other formats* (2 ed.). Hingham: Kluwer Academic Publishers.

Schoenfeld, A. H. (2007). Commentary: the complexities of assessing teacher knowledge. *Measurement: Interdisciplinary Research and Perspectives*, 5 (2), 198–204.

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

Shulman, L.S. (1987). Knowledge and teaching: foundations of the New reform. *Harvard Educational Review*, 57 (1), 1–22.

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.

Janne Fauskanger

Janne Fauskanger is Assistant Professor in mathematics education at the University of Stavanger, Norway. Her main interest is related to primary school teachers' mathematical knowledge for teaching and their practice.

janne.fauskanger@uis.no

Reidar Mosvold

Reidar Mosvold is Associate Professor in mathematics education at the University of Stavanger, Norway. His interests are related to teachers' beliefs and knowledge of mathematics, and their influence on practice.

reidar.mosvold@uis.no