Does the format matter? How the multiple-choice format might complicate the MKT items

JANNE FAUSKANGER, REIDAR MOSVOLD,
RAYMOND BJULAND AND ARNE JAKOBSEN

In order to design appropriate professional development programs for teachers, an instrument has been developed in the U.S. to measure teachers' mathematical knowledge for teaching. The process of translating and adapting these measures for use in other countries involves several challenges. This article focuses on issues related to the multiple-choice format of the items. Analyses of focus-group interviews reveal that the multiple-choice format may complicate the items. The teachers' reflections about the format in this Norwegian case contribute to the understanding of this important challenge.

Substantial progress has been made over the last two decades in understanding the knowledge that teachers need in their mathematics teaching (e.g. Sullivan & Wood, 2008). Researchers at the University of Michigan in the U.S. have contributed to this understanding with a concept they refer to as teachers' *mathematical knowledge for teaching* (MKT. Ball, Thames & Phelps, 2008). They claim that MKT, as assessed by their measures, made a difference to the mathematical quality of instruction (Hill, Blunk et al., 2008) and to students' achievement in mathematics (Hill, Rowan & Ball, 2005). The results from these researchers' efforts seem promising. Morris and colleagues (2009) even describe MKT as "the most promising current answer to the longstanding question of what kind of content knowledge is needed to teach mathematics well" (p. 492). Knowledge about the topics and tasks that teachers struggle with is useful when preparing professional development (PD) programs (Hill, 2010).

Janne Fauskanger, *University of Stavanger*
Reidar Mosvold, *University of Stavanger*
Raymond Bjuland, *University of Stavanger*
Arne Jakobsen, *University of Stavanger*

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.

Despite the promising results, these research efforts have also met criticism. Schoenfeld (2007) is one prominent critic. He has criticized the lack of a clear underlying framework and the use of multiple-choice (MC) format in the items. Schoenfeld argued that open-ended items would have been easier for teachers than MC items, and he claimed that the MC format might potentially complicate the items and thus make the MKT more difficult for the test-takers.

Care should always be taken when attempting to adapt measures for use in a new cultural context, and this is particularly important when it comes to the MKT measures (Fauskanger, Jakobsen, Mosvold & Bjuland, in press). The items in these measures relate to the practices of teaching mathematics in the U.S., and several researchers argue that teaching is a cultural activity (e.g. Stigler & Hiebert, 1999). An early work on translation (Mosvold, Fauskanger, Jakobsen & Melhus, 2009) pointed to a preliminary finding concerning teachers' perceived unfamiliarity with the MC format. This coincides with Tonheim and Torkildsen's (2010) findings that Norwegian students are seldom assessed with MC items in their mathematics teacher education. Different levels of experience with the item format might lead to differences in test-taking skills, and the item format is another important aspect of a discussion of the validity of an assessment instrument (see e.g. Haladyna, 2004).

Hambleton and Patsula (1998) argue that the choice of item format should be discussed when adapting measures for use in a country other than that for which the measures were originally intended. Based on previous experience, Schoenfeld (2007) argued that the format could complicate the items, and as a result of this the format might actually complicate the MKT being measured for the teachers. In this article, we present a further investigation of possible difficulties regarding the MC format of the MKT items. We address the following research question:

> What indicators are identified from teachers' reflections on how the multiple-choice format might complicate the content (MKT) being measured?

As an initial analysis of this question, we have decided to invite the test-takers (teachers) to reflect on the format. Through analyses of dialogues from focus-group interviews, we identify and discuss indicators of how the MC format can make the MKT being measured more complicated, as perceived by the teachers.

The MKT framework

The study of mathematics teachers' knowledge has been an active field of research for several decades (e.g. Sullivan & Wood, 2008) and various

methods have been used to assess different aspects of teachers' knowledge (e.g. Hill, Sleep, Lewis & Ball, 2007). Shulman's (1986) paper, focusing on knowledge unique to teaching, is frequently referred to (e.g. Graeber & Tirosh, 2008). His notions of subject matter knowledge (SMK) and pedagogical content knowledge (PCK) have subsequently been modified, criticized and expanded. One expansion is the empirically supported work carried out by Ball and colleagues in relation to MKT, which has been defined as "the mathematical knowledge used to carry out the work of teaching mathematics" (Hill et al., 2005, p. 373). The MKT construct was developed by studying several aspects of teaching (e.g. Hill, 2010).

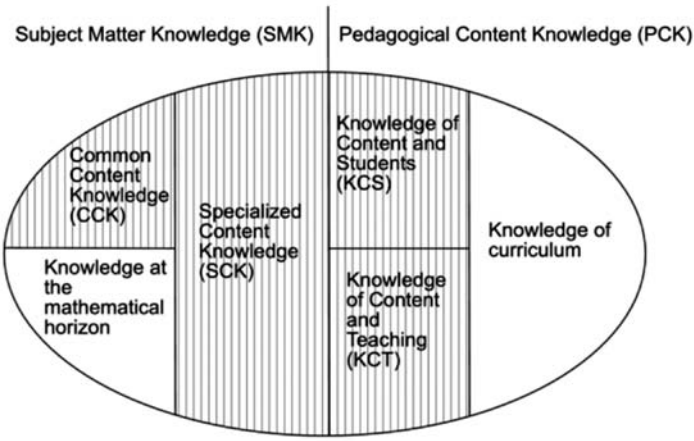

Figure 1. *Mathematical knowledge for teaching (Hill, Ball & Schilling, 2008, p. 377).*

Figure 1 shows the correspondence between Shulman's (1986) categories and the current map of MKT. At present, items have been developed to measure teachers' knowledge in four of the MKT domains (striped in figure 1, Hill, 2010). The left side of the oval is related to Shulman's SMK. Common content knowledge is knowledge that is used in the work of teaching, in ways that correspond with how it is used in settings other than teaching. Specialized content knowledge is the mathematical knowledge "that 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 & Shilling, 2008, p. 378). Both are forms of mathematical knowledge. The right side of the oval contains knowledge related to Shulman's PCK. The researchers have to a lesser extent succeeded in developing items to measure the PCK domains so far (Hill, 2010). Hill, Ball and Schilling

(2008) describe an effort to conceptualize and develop measures of teachers' knowledge of content and students. This work suggests that a possible direction for future item development in this domain is to invest in open-ended items. We focus on the left side of the oval in this context. Figures 2 and 4 are examples of items developed to measure teachers' specialized content knowledge, while in figure 3 teachers' common content knowledge is in focus.

Hill, Sleep and colleagues (2007) claim that different assessments constitute different theories about teachers' knowledge. The differences between e.g. constructivism and socio-cultural theories are evident in the attribution of individual knowledge structures on the one hand and knowledge being the internalization or appropriation of social practices on the other (Goodchild, 2001). Situated perspectives turn attention away from individual knowledge (Boaler, 2000), and it is considered inadequate to focus on knowledge alone, outside of the practices of its production and use. The MKT measures are not grounded in these overarching and more generic theories (Hill, Ball & Schilling, 2008) and are criticized on the basis of the claim that MKT is a personal construct (e.g. Stylianides & Delaney, 2011). Measuring teachers' MKT is related to a more positivistic perspective in which teachers are seen as having inert or "in the head" knowledge (e.g. Williams, 2011) that it is possible to measure. This does not however exclude the fact that teachers' situated or enacted knowledge is equally important.

While the measurement of student teachers' knowledge is a widely accepted practice, that of practicing teachers' knowledge is not (e.g. Hill, Sleep et al., 2007), at least not in Norway (Lysne, 2006). In order to consider how teachers' knowledge might be responsibly assessed, the goal of Hill, Sleep and their colleagues (2007) is to move the debate concerning assessment of teachers "from one of argument and opinion to one of professional responsibility and evidence" (ibid., p. 112). To make advances in developing tools to study teachers' knowledge, as well as to understand the MKT, a set of agreed-upon, reliable and valid methods for assessing teachers' MKT is required (Hill, Sleep et al., 2007). These authors argue that assessing teachers' knowledge:

> [...] can be done in ways that honor and define the work of teaching, ratify teachers' expertise, and help to ensure that every child has a qualified teacher. Doing so requires carefully constructed instruments that take seriously the work of teaching and that can be used at scale. (ibid., p. 150)

Hill and colleagues see further development of the MKT measures as one attempt to attain this goal. A close consideration of the format will be an important contribution to this.

The format of the MKT items

An advantage of using the MC format in the MKT items is that they can be used at scale and are less time consuming to analyze than open-ended items would be (Hill, Sleep et al., 2007). According to Burton and his colleagues (1991), a standard MC item consists of two parts: a problem (also called stem), and a list of suggested solutions. This list normally contains one correct alternative, which is referred to as the key, and a number of incorrect alternatives, termed distractors.

In some of the MKT items the key is the "incorrect" answer to the mathematical problem presented. An example is given in figure 2. In this particular item, alternative C) is the key, although that alternative in itself is mathematically incorrect, whereas the other alternatives are distractors that can all be used to represent this particular multiplication of fractions.

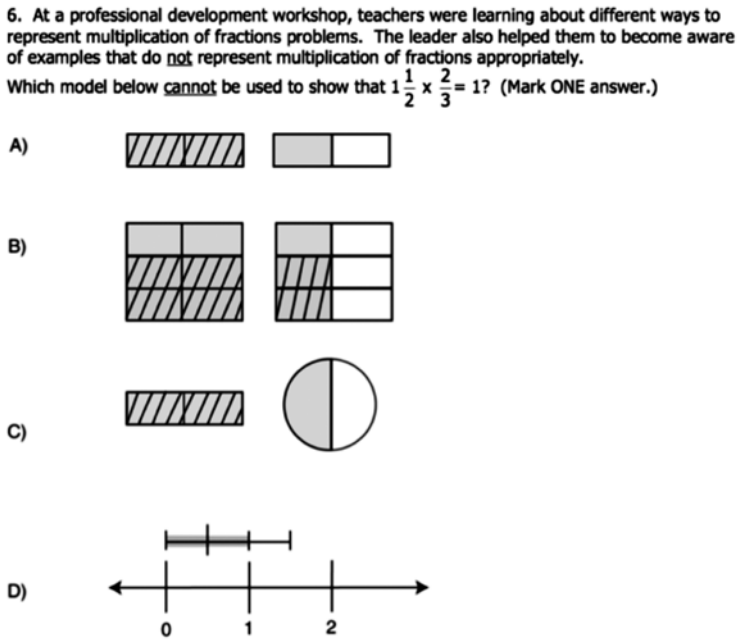

Figure 2. *Item developed to measure teachers' specialized content knowledge. Item number 6 from the set of released items (Ball & Hill, 2008, p.7). (Items used in our adapted measures are not released for publication).*

In addition to such standard MC items, there are MKT items that differ in at least two ways. First, some items may not include any incorrect alternatives. The key is then a suggested solution such as "they are making all

11. Students in Mr. Hayes' class have been working on putting decimals in order. Three students — Andy, Clara, and Keisha — presented 1.1, 12, 48, 102, 31.3, .676 as decimals ordered from least to greatest. What error are these students making? (Mark ONE answer.)

a) They are ignoring place value.

b) They are ignoring the decimal point.

c) They are guessing.

d) They have forgotten their numbers between 0 and 1.

e) They are making all of the above errors.

Figure 3. *Item developed to measure teachers' common content knowledge. Item number 11 from the set of released items (Ball & Hill, 2008, p.10).*

of the above errors". The set of released items does not contain any items with this alternative as a key (figure 3), but several of the non-released items are of this kind.

In this item b) is the key since the students' ordering of the decimals would have been correct if the decimal point had been removed. Alternatives a), c), d) and e) are distractors. As an example it can be observed that alternative a) is incorrect because the students do not ignore place value, as is c) since the ordering of the numbers seems to follow some kind of pattern.

A second difference is that some of the MKT items have one stem and multiple MC questions related to this stem (figure 4). These items are called testlets (Hill, 2010). Alternative 1 is the key for all the MC questions in this particular testlet, because the three methods presented could all be used to multiply any two whole numbers.

The use of MC items is not unproblematic, and every format has advantages as well as disadvantages. Several studies have compared the use of open-ended vs. MC items (e.g. Hollingworth, Beard & Proctor, 2007), and there are indications that these different formats might actually measure different types of knowledge. Others, however, suggest that there are no evident differences between the use of open-ended items and MC items (van den Bergh, 1990). When investigating the hypothesis that open-ended items measure something other than MC items, Hollingworth and colleagues (2007) conclude that both open-ended and MC items are related to a common factor. They also argue the two formats are equally effective.

The use of MC items to assess teachers' knowledge has been subject to criticism. Schoenfeld (2007) acknowledges that measures based on MC

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 x25 125 +75 875	35 x25 175 +700 875	35 x25 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 4. *Testlet developed to measure teachers' specialized content knowledge. Testlet number 3 from the set of released items (Ball & Hill, 2008, p. 5).*

items can serve certain functions, but he argues that they test something other than what they are intended to. He recommends open-ended questions to reflect the teachers' "desired competencies" (ibid., p. 204). Others claim that validity might be threatened because the use of MC items can lead to trivialization of the complexities of teaching (Haertel, 2004). MC items may also be limited in their cognitive range (Boodoo, 1993) and the interpretation of scores could measure test-taking strategies rather than MKT (Martinez, 1999). The format may also unwittingly involve "greater implications than intended by the developers" (Hill, Sleep et al., 2007, p. 150). The MC format may solidify the misconception that mathematical competence is demonstrated by quick solutions (Schoenfeld, 1992) among some teachers, and teachers who do not think of mathematics as quick solutions to routine problems may feel marginalized by the format. Even if Hill, Dean and Goffney (2007) conclude that their work on validation corrects for common problems of MC items, the aspects mentioned above should be taken into consideration when translating and adapting measures. Challenges related to MC format are important to investigate further when the MKT items are used in different cultural settings.

Cultural aspects of MKT

The use of MC format in the MKT items has provided some promising research results so far (Hill et al., 2005; Hill, Blunk et al., 2008). However, the development as well as the validation of the measures was originally done in a U.S. context only (e.g. Hill, Dean & Goffney, 2007) and more recently in Ireland (Delaney, 2008), Ghana (Cole, 2011) and Indonesia (Ng, 2012). Since the knowledge required for teaching may be more culturally-based than pertaining simply to mathematical knowledge (Stylianides & Delaney, 2011), a new debate concerning the cultural aspects of MKT has emerged (see e.g. Ng, Mosvold & Fauskanger, 2012).

Attempts to adapt and use the MKT measures in a different cultural context should include careful analyses of the challenges involved on different levels. As an example, prior research on U.S. teachers' subject-matter knowledge found that many teachers hold procedural understandings of algorithms, which stands in contrast to that of teachers in China (Ma, 2010). When adapting an item focusing on algorithms (see figure 4), this is an important issue to take into consideration. In his efforts to adapt MKT items into an Irish context, Delaney (2008) discussed aspects related to cultural differences extensively. He referred to this as a challenge of establishing equivalence, and he particularly used Singh's (1995) model for establishing construct equivalence. Building on the attempts and experiences of Delaney and colleagues (2008) in translating and adapting MKT items for use in Ireland, several researchers have followed up with similar attempts. Mosvold and his colleagues (2009) used a similar framework in their attempt to translate and adapt MKT items for use in a Norwegian context, and they had a particular focus on the challenges of translation. Other researchers have used MKT items in countries like Indonesia (Ng, 2012), South-Korea (Kwon, 2009) and Ghana (Cole, 2011). Most of these studies build on the experiences of Delaney, but only Cole (2011) discusses the format of the items, and the discussions about the test-takers' reflections are provided only to a limited extent. There have been some efforts to study the challenges of adapting the items into a different cultural context (Fauskanger et al., in press), comparing the challenges of translation and adaptation in two different cultures (Ng et al., 2012) and the performance of the items in the Norwegian context (Jakobsen, Fauskanger, Mosvold & Bjuland, 2011). In spite of such attempts, however, much work is still needed in order to learn more about the cultural issues related to the translation, adaptation and use of MKT items in different cultural contexts. Such studies are also needed to investigate the possible cultural aspects of the MKT framework itself. The present article is an attempt to approach one of these cultural issues by examining the complexities of using the MC format of the MKT items in a different cultural context.

Methods

In this study, we decided to solicit the opinions of the test-takers (the teachers) in order to learn more about the challenges raised by the MC format in the MKT measures. Seven semi-structured focus-group interviews (FGIs) were organized, and a total of fifteen teachers participated. Teachers from different schools, grade levels and with different levels of experience were selected for participation. The first group consisted of two experienced teachers, whereas the second group consisted of three inexperienced teachers. The participants in these two groups were selected on the basis of their level of experience and special interest in mathematics education, and were all from different schools. The other five groups were randomly selected from schools that were connected to our university as practice schools, and they consisted of teachers from different schools and different levels. All the participants had a special interest in mathematics and mathematics teacher education. For these five FGIs, pairs of teachers were selected in association with their respective headmasters. The first two interviews were held at the university, whereas the other five were held at the teachers' respective schools.

The participants worked individually with a set of MKT items before the interviews. The FGIs were designed with the following structure, eliciting questions with a focus on: a) background information of the teachers, b) general considerations of the MKT measures, c) particular considerations in relation to the MC format (e.g. "Do you have any comments in connection with the multiple-choice format of the measures' items?"), d) comments on the mathematical topic, structure and difficulty item by item, and finally e) comments and reflections that supplement the other issues discussed in the interviews. We focus on the teachers' reflections expressed in the dialogues, trying to capture the main challenges associated with the MC format, if such exist.

The FGIs were recorded and transcribed, and these transcriptions were analyzed through content analysis (e.g. Törner, Rolka, Rösken & Sriraman, 2010), which aims "to obtain descriptive information about a topic" (Fraenkel & Wallen, 2006, p. 485). One approach to content analysis is to start with previously determined categories (Fraenkel & Wallen, 2006); we began by using the two categories that appeared from Schoenfeld's (2007) criticism as an analytical framework. He argued that there were 1) more general challenges related to the MC format, and 2) that the MC format might complicate the content being measured and thus make the MKT items more difficult than if they were open-ended. After having organized the data material in these two categories through content analysis, we used a more grounded approach (Bryman, 2004) to uncover subcategories of the two main categories. For a subcategory to be established, the aspect in focus had to be discussed by the teachers

in at least two FGIs. In this article, we have focused on the second main category. Some of the transcripts have been slightly adapted to avoid too many gap fillers and repetitions.[1]

Results and discussion

In order to approach the question of what indicators the teachers introduce as to how the MC format might complicate the MKT items, we present and discuss results from a study of teachers' reflections concerning the MC format of the MKT items in FGIs. In his criticism of the MC format of these items, Schoenfeld (2007) distinguished between two main issues. First, he claimed that more general problems with MC were involved, and, second, he argued that the MC format could complicate the items and make the MKT being measured more difficult for the teachers than if it was measured by open-ended items. Our main focus here is on how the item format could make the MKT being measured more difficult for the participating teachers, but we start by providing a brief report on more general issues that were indicated by the teachers in our FGIs.

Indicators of general challenges with the MC format

In addition to these more MKT-specific challenges with the MC format, the teachers' reflections also indicated several issues concerning the item format that are not specific to the MKT items. Some of the more experienced teachers' reflections on the format indicate an anticipation that mathematical competence is demonstrated by quick solutions, as reported by Schoenfeld (1992). The teachers also raised some issues related to more general test-taking strategies in their reflections. On the one hand, they expected that they should be able to find the correct answer (the key) by eliminating the implausible answers (the distractors). Such a response elimination strategy is a common test-taking strategy when MC format is used (Martinez, 1999).

Another issue related to test-taking strategies is that of guessing, and teachers in our interviews suggested that this might be another issue to consider (see also Martinez, 1999 for a discussion on such issues). One teacher argued that the format has a weakness since it does not provide any information about students' mistakes and their thinking and could measure test-taking strategies rather than MKT (see also Hill, Dean & Goffney, 2007 for a similar discussion). This suggests that the teacher was not aware that distractors are often chosen to reflect incorrect strategies.

In their work on validity, Hill, Dean and Goffney (2007) conclude that their work rules out common problems and critiques of MC items.

As an example they claim that test-taking strategies are not widely used. The teachers in our interviews offered some reflections concerning more general problems related to MC format, and these issues are important to investigate further in different cultural settings. In this connection, however, we focus more on the teachers' reflections on how the format could make the MKT being measured more difficult for them.

Indicators of the MC format complicating the MKT being measured

In five of the seven FGIs, teachers indicated that it was challenging to be assessed by MC items. In the FGIs, the teachers discussed three different but closely related indicators regarding how the MC format could complicate the MKT being measured. First, they argued that the suggested solutions (the distractors and the key) made the items more complicated than open-ended items would have been. Second, the set of suggested solutions that were given in the items were experienced by the teachers as something that forced them into a particular way of thinking. Third, the teachers argued that some items lacked important and correct solutions, and this made it all the more difficult for them. These reflections by the teachers form the categories in which we present the results below.

The suggested solutions complicate the MKT items

The first transcript example, taken from the interview with three inexperienced teachers (TU1A, 1B and 1C), indicates that the suggested solutions may complicate the items, meaning that an item without suggested solutions would be easier. This aspect was brought up in four of the FGIs, three with inexperienced teachers and one with one experienced and one inexperienced teacher. When asked about what it was like to work on a measure made up of MC items, one of the teachers said[1]:

5. TU1A: I have never done this before, so I thought it was (...). Well, I did it [a MC test] when I took my [theoretical] driver's license test (laughter). But I think it's a difficult way to be assessed.

6. [...]

7. TU1A: Because it [the suggested solutions] makes you doubt, because everything is in a way similar, to some degree correct. And then you have to select an alternative, then I think it's easier when you get to come up with your own answer. Instead of being forced to select among alternatives that someone else has produced. It takes time because there are many similar suggested solutions. (. . .) I wanted to calculate myself in a way.

[UiS FGI1, October 7, 2008]

Later in the interview, when asked to comment on the first items of the MKT measures, TU1A and TU1B indicated that the suggested solutions related to whether or not 1 is a prime number confused them.

66. TU1A: And there you have that multiple-choice makes me think, yes everything is right.

67. TU1B: Yes, that's what you immediately think.

In this item one of the four proposed definitions (including and excluding 1) is correct. We observe that TU1A thinks that a measure built from MC items is more challenging than one made from open-ended items would be (7), as Schoenfeld (2007) claims. One of the reasons given by these inexperienced teachers is that all the suggested solutions are perceived to be similar and thus seem correct (7, 66 and 67). This aspect may be illustrated by the released item in figure 3, where the suggested solutions are quite similar and several might seem correct. The MKT items in which all the suggested solutions are correct and the key is "all of the above" (figure 3) may thus complicate these items even more (Burton et al., 1991). The teachers' confusion may also be due to an expectation that basic skills are in focus in MC items and that MC items cannot be written to elicit complex cognition (Boodoo, 1993), such as the MKT measures (Hill, Sleep et al., 2007). The item presented in figure 2 illustrates this, and the teacher has to figure out what use other teachers might make of the representations presented in relation to each of the four models. Issues related to translation may also have complicated the items (e.g. Mosvold, et al., 2009). Not being able to choose the correct definition of a prime number among four suggested definitions may also be due to the teachers' level of MKT in general or, in particular, to what Ball and colleagues (2008) call "common content knowledge". It could also be due to cultural issues related to which definitions are used and how they are used in different countries (Ng, 2012).

The next transcript example selected from an interview with two inexperienced teachers at a junior high school (grades 8-10) indicates that the format's suggested solutions makes even basic knowledge appear more complicated:

17. Int.: Yes, when you worked on these measures, did you react to the way the questions were posed? (. . .)

18. T6A: No, in my case I thought this was a fairly standard multiple-choice test. You have basic knowledge presented in a bit complicated way, (. . .) and we are given alternatives that are fairly similar so that you, (. . .) the things that once were basic [knowledge] suddenly become more complicated.

[FGI, School 6, March 5, 2009]

T6A's response might be taken to indicate that the similarity of suggested solutions complicates what would otherwise be relatively straightforward (18). According to this teacher, results from teachers' responses to the MKT measures may show that teachers hold less MKT than they really do. This is in line with what Schoenfeld (2007) points out, but contradicts Haertel's (2004) position, suggesting that MC items can trivialize the complexities of teaching. Nevertheless, the MKT items are constructed to differentiate between teachers (e.g. Hill, 2010). T6A may expect the knowledge measured by the MKT items to be basic content (Boodoo, 1993). Another interpretation of this statement might be that the teachers do not have any problem with the format, but rather that the presentation of the alternatives in the item actually makes the teacher aware of the deep knowledge that is required to deal with the problem presented.

The issue of suggested solutions complicating the items, and therefore the MKT being measured, was brought up mostly by inexperienced teachers. This may indicate that the experienced teachers hold more MKT and therefore find the items easier, or that they are more experienced when it comes to the MC format. However, in the interview at School 2, the teachers indicate the opposite. This is in line with Sirne's (2005) assertion that MC makes items easier than an open-ended format would do.

The MC format forces one way of thinking on the teachers

The next transcript example illustrates issues raised in two FGIs with inexperienced teachers. In both FGIs this aspect was brought up in relation to discussion of the MKT measures as a whole. The example indicates that these inexperienced teachers do not like the fact that the MC format and the suggested solutions lead them into one way of thinking. They do not like being forced to choose among alternatives already produced but would rather do their own calculations to find the correct answer.

9. TU1B: I don't have anything against multiple-choice, not when it comes to these kind of items or measures [given in order to learn more about teachers' MKT to be able to build professional development on what teachers already know], but if I were supposed to have produced something on an exam I'm not very enthusiastic about this. I feel I filter the alternatives I can give an answer to, but I would have thought differently, or done things in a different way (. . .).

(. . .)

12. TU1A: You're in a way forced into another person's way of thinking. You're not able to use your own [thinking] in the same way.

[UiS FGI1, October 7, 2008]

TU1A finds the MC format challenging because the suggested solutions make it difficult to think independently (12). The inexperienced teachers prefer to think independently and TU1B argues that she thinks differently from the alternatives given (9). The teachers' arguments are in line with Schoenfeld's (2007) claim that there is a danger that teachers' answers may reveal neither anything about the process used to find the answers nor whether the teachers are actually answering the questions posed. Using figure 2 as an example, the teachers may be able to use area models or other representations to illustrate multiplication of fractions appropriately, but the suggested solutions may still appear to disturb their independent thinking. Schoenfeld's statement is strong, but corresponds rather closely to statements made by some of the teachers in our study. Since this was only brought up by the inexperienced teachers, it may indicate a lack of experience with the MC format in pre-service teacher education (Tonheim & Torkildsen, 2010).

The suggested solutions lack important alternatives

The two last transcript examples in this section aim to illustrate the discussion in three of the FGIs involving experienced and inexperienced teachers in both primary and junior high school. The first example is taken from an interview with two experienced teachers (TU2A, TU2B) and the second from an interview with two inexperienced teachers (T6A, T6B). The examples suggest that MC-based measures could lack other correct alternatives than the key presented. According to the teachers in our study, this complicates the items (81).

81. TU2A: At first I became a bit confused. Because you go into the role of a student, in a way. And then you first search for what's correct [the key]. And then you don't find the one you expected to find (. . .) But, none [of the suggested solutions] were exactly correct.

[UiS FGI2, October 28, 2008]

TU2A is commenting on an item focusing on place value. The item deals with students working to decompose a three-digit number into hundreds, tens, ones and tenths. The students have arrived at different answers and the teachers are asked in the item to evaluate which of four different ways to represent the three digit number they will accept as correct. TU2A did not find the solution she expected to be the correct one (e.g. 456 divided into 4 hundreds, 5 tens and 6 ones) and thus found none of the suggested solutions to be the key (81). This could illustrate that MC items do not always measure what they are supposed to measure (Schoenfeld, 2007), because this teacher would have been able to give at least one correct answer (e.g. 456 equals 4 hundreds, 5 tens and 6 ones) if

this item was not given in MC format. On the other hand, the items are made to differentiate among teachers' knowledge (Hill, 2010) and when TU2A does not find the expected key (81), it may be due to the teacher's level of MKT or lack of what Ball and colleagues (2008) call specialized content knowledge. Not being able to evaluate students' different ways to decompose a three-digit number when three out of four suggested solutions are mathematically correct (e.g. 456 divided into 3 hundreds, 15 tens and 6 ones), as is the case in this particular item, leads us to discuss the level of MKT.

According to the teachers in our data, this could also be a cultural issue. The experienced teachers in this particular interview argue that it is unusual in a Norwegian school context to divide the three digit number into hundreds, tens and ones, rather than following the positions. TU2A says that she was unable to see "the most usual (...) that people will look for (...) the correct one" (line 71). This issue is commented on by teachers in four of the other interviews as well. In this item it appears evident that the cultural issues are related to the format of the item, since an open-ended item would have prompted the teachers to provide a solution (and possibly a supporting argumentation), and the cultural issues that became evident through the suggested solutions in the MC item would not have appeared in an open-ended version.

A second example illustrates the issue of missing but correct alternatives. T6B in the transcript example below is commenting on an item in which the teachers are invited to figure out what students performing calculations in a specific way (written two-digit subtraction) were probably doing (see testlet in figure 4). Four solutions are suggested, and only one of them is correct and matches the algorithm presented.

91. T6B: There you have a wonderful example of [an item] where I would have asked [the students]: What have you done, could you show me what you have done?

92. Int.: Yes.

93. T6B: Instead of me using 20 minutes to try and figure out what on earth they have done (. . .)

94. Int.: You would have asked? (. . .)

95. T6B: Yes, I missed that solution.

96. Int.: Yes.

97. T6B: But I have to reach an answer by guess work.

[FGI, School 6, March 5, 2009]

T6B says that she had to guess (97) because she was unable to see the solution by focusing on what she as a teacher would have been doing if she

had seen this specific way of calculating in her classroom (95). This can be seen as relating to a disadvantage to the MC format emphasized by Clauser and Margolis (2006): The MC items assess what e.g. the teacher knows, not what the teacher can do. This could also be seen in relation to Schoenfeld's (2007) critique and may indicate that this item does not measure what it is supposed to measure and that the results would not tell anything about teachers' MKT. T6B would have asked the student about what she was doing and about the thinking behind this specific way of calculating two-digit subtraction (91 and 93). When this alternative was not present, the teacher guessed the way to the key (97). Guessing in this situation may, on the other hand, be due to not being able to analyze and understand different students' written algorithms related to two-digit subtraction, and thus lack of what Ball and colleagues (2008) call specialized content knowledge. Prior research has found that many teachers (in the U.S.) hold narrow, procedural understandings of algorithms (e.g. Ma, 2010) and the discussions related to algorithms in the seven FGIs indicate that this might be an issue among several of the fifteen Norwegian teachers as well. The MKT items are made to measure teachers' MKT, and the suggested solution of "ask the student" would not be related to mathematics at all, even if it is a good idea to let students explain their thinking.

The algorithm presented in this item is not one that the teachers in our study consider standard in the Norwegian school context. The unfamiliarity with this particular algorithm was discussed in several FGIs. T6B not being able to find the key may therefore be related to cultural issues because it is easier to identify what is done in relation to well-known algorithms. T6B's wishing to ask the student about her thinking rather than analyzing her written work may also be a cultural issue: it might be more common in Norwegian classrooms to ask students to explain what they are doing, and it might also be due to the cultural issue of not being used to reflecting on "unusual" algorithms (Ma, 2010).

Hill and her colleagues (Hill, Ball, Blunk, Goffney & Rowan, 2007) claim that MC assessments validly represent the knowledge involved in an actual teaching practice in the U.S. The fifteen teachers' reflections on the MKT measures suggest that the MC format of the MKT items can be a critical issue when measuring teachers' MKT. Schilling, Blunk and Hill (2007) seek to improve the items and say that, in order to be better able to measure teachers' knowledge of content and students (figure 1), they may reorient their measurement strategy away from the MC format. At school 13, the teachers underlined the importance of developing the MKT items' MC format further by adding commentary-boxes to the items.

510. Int.: Finally, what do you think about this way of finding out more about teachers' knowledge before planning professional development for teachers?

511. T13A: I think (. . .) if one manages to find out more about the competence that is needed [among teachers] by carrying out such a measure, then I think this is really smart, so one can hit [what the teachers' need] in a proper way, that's my thoughts.

512. T13B: Yes, (...) If I'm to observe students who solve tasks in this way, it could be wise to insert a commentary box [beside the item] where they are allowed to do calculations, because you can then identify what they don't understand, where do they miss? What is the difficulty?

513. Int.: What you say is that we get more information if we let the teacher [have the possibility] to write something more than a circle or a cross?

514. T13B: Yes

[FGI, School 13, March 13, 2009]

This idea might be worth following up in future research, as the researchers in Michigan do (e.g. Hill, Dean et al., 2007).

Conclusion

In our analyses of seven FGIs, we have identified three indicators given by the teachers as to how the format may complicate the items and MKT being measured. First, some teachers expressed the view that the MC items were more complicated than comparable open-ended items, and this corresponds with what Schoenfeld (2007) argues. For example, the teachers find it hard to identify the key from the distractors because all the suggested solutions were perceived to be similar (and correct) for some items. The teachers perceive an open-ended item to be easier. The measures are developed to differentiate among teachers (Hill, 2010). Teachers being unable to identify the key in all the items can be taken as an indication that the items function as intended. Second, some teachers argued that the MC format forced them into a particular way of thinking. These teachers do not like being forced to choose among pre-made alternatives, and they would rather calculate and reflect on the items in order to find the correct answer. A third indicator is that the teachers experience that important and correct alternatives are not included among the suggested solutions.

In this article, we have let the test-takers themselves reflect on the challenges they perceive in relation to the format of the MKT items. Through their reflections, we have suggested some extensions of Schoenfeld's (2007) criticism in relation to how the MC format might actually make the MKT being measured more difficult for the teachers. It

is important to stress, however, that there are several possible explanations for these indicators. The teachers' conception of the MC format as a difficult way to be assessed may be related to their level of MKT. It may also be related to cultural differences, both when it comes to MKT and to test-taking strategies (in relation to the MKT measures). These perceived difficulties may also be due to a lack of experience with this kind of measure and format. Further studies are needed in order to learn more about these difficulties and their possible impact in different cultural settings. If such difficulties seem to have an impact on the results, the problematic item(s) might have to go through a new round of adaptation (Fauskanger et al., in press). One option would be to investigate different ways of making the items more open. This could be done by rewriting the items and making them open-ended, by keeping the MC format and adding commentary boxes, or by designing new items to fit the particular cultural context in focus.

Most of the research concerning the use of MC format is related to students, whereas little has been done in relation to teachers and none in relation to Norwegian teachers. In this respect, our article makes a significant contribution to the field. First and foremost, however, the article is a contribution when it comes to making other researchers more aware of possible difficulties related to the item format that are specific to the adaptation of MKT measures for use in other countries.

Acknowledgements

Our research project is supported by The Norwegian Oil Industry Association, and would not have been possible without the participating teachers and our colleagues' help with the collection of data. We would also like to thank the three reviewers and the editor for their recommendations, which helped us improve this article.

References

Ball, D. L. & Hill, H. C. (2008). *Mathematical knowledge for teaching (MKT) measures*. Mathematics released items 2008. Retrieved September 3, 2011, 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.

Bergh, H. van den (1990). On the construct validity of multiple-choice items for reading comprehension. *Applied Psychological Measurement*, 14 (1), 1–12.

Boaler, J. (2000). Introduction: intricacies of knowledge, practice, and theory. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching & learning* (pp. 1–17). Westport: Greenwood Publishing Group.

Boodoo, G. M. (1993). Performance assessments or multiple choice? *Educational Horizons*, 72 (1), 55–56.

Bryman, A. (2004). *Social research methods* (2nd ed.). New York: Oxford University Press Inc.

Burton, S. J., Sudweeks, R. R., Merrill, P. F. & Wood, B. (1991). *How to prepare better multiple-choice test items: guidelines for university faculty*. Brigham Young University Testing Services & The Department of Instructional Science. Retrieved February 13, 2012, from https://www2.kumc.edu/comptraining/documents/WrittingbetterMCtestitems1991BrighamYoung.pdf

Clauser, B. E. & Margolis, M. J. (2006). Testing for licensure and certification in the professions. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 701–731). Santa Barbara: Greenwood Publishing Group.

Cole, Y. A. (2011). *Mathematical knowledge for teaching: exploring its transferability and measurement in Ghana* (Unpublished doctoral dissertation). Ann Arbor: University of Michigan.

Delaney, S. F. (2008). *Adapting and using U.S. measures to study Irish teachers' mathematical knowledge for teaching* (Unpublished doctoral dissertation). Ann Arbor: University of Michigan.

Delaney, S., Ball, D., Hill, H., Schilling, S. & Zopf, D. (2008). "Mathematical knowledge for teaching": adapting U.S. measures for use in Ireland. *Journal of Mathematics Teacher Education*, 11 (3), 171–197.

Fauskanger, J. & Mosvold, R. (2010). Undervisningskunnskap i matematikk: Tilpasning av en amerikansk undersøkelse til norsk, og læreres opplevelse av undersøkelsen. *Norsk Pedagogisk Tidsskrift*, 94 (2), 112–123.

Fauskanger, J., Jakobsen, A., Mosvold, R. & Bjuland, R. (in press). Analysis of psychometric properties as part of an ongoing adaptation process of MKT items for use in other countries. *ZDM The International Journal on Mathematics Education*, 44 (2).

Fraenkel, J. R. & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill.

Goodchild, S. (2001). *Students' goals: a case study of activity in a mathematics classroom*. Bergen: Caspar Forlag.

Graeber, A. & Tirosh, D. (2008). Pedagogical content knowledge. Useful concept or elusive notion. In P. Sullivan & T. Wood (Eds.), *The international handbook of mathematics teacher education* (Vol. 1., pp. 117–132). Rotterdam/Taipei: 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* (3rd ed.). New Jersey: Lawrence Earlbaum Associates.

Hambleton, R. K. & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45 (1–3), 153–171.

Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. *Journal for Research in Mathematics Education*, 41 (5), 513–545.

Hill, H. C., Ball, D. L., Blunk, M., Goffney, I. M. & Rowan, B. (2007). Validating the ecological assumption: the relationship of measure scores to classroom teaching and student learning. *Measurement: Interdisciplinary Research and Perspectives*, 5 (2&3), 107–118.

Hill, H. C., Ball, D. L. & Schilling, S. (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. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L. 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., Dean, C. & Goffney, I. M. (2007). Assessing elemental and structural validity: data from teachers, non-teachers, and mathematicians. *Measurement: Interdisciplinary Research and Perspectives*, 5 (2), 81–92.

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., 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). Retrieved April 3, 2011, from http://pareonline.net/pdf/v12n18.pdf

Jakobsen, A., Fauskanger, J., Mosvold, R. & Bjuland, R. (2011). Comparison of item performance in a Norwegian study using US developed mathematical knowledge for teaching measures. In M. Pytlak, T. Rowland & E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education* (pp. 1802–1811). University of Rzeszów, Poland.

Kwon, M. (2009). *Validating the adapted mathematical knowledge for teaching (MKT) measures in Korea*. Paper presented at the AERA 2009 Annual Meeting.

Lysne, A. (2006). Assessment theory and practice of students' outcomes in Nordic countries. *Scandinavian Journal of Educational Research*, 50 (3), 327–359.

Ma, L. (2010). *Knowing and teaching elementary mathematics: teachers' understanding of fundamental mathematics in China and the United States*. Anniversary edition. New York: Routledge.

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34 (4), 207–218.

Morris, A. K., Hiebert, J. & Spitzer, S. M. (2009). Mathematical knowledge for teaching in planning and evaluating instruction: What can preservice teachers learn? *Journal for Research in Mathematics Education*, 40 (5), 491–529.

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), 9–31.

Ng, D. (2012). Using the MKT measures to reveal Indonesian teachers' mathematical knowledge: challenges and potentials. *ZDM The International Journal on Mathematics Education*, 44 (2). (doi: 10.1007/s11858-011-0375-9)

Ng, D., Mosvold, R. & Fauskanger, J. (2012). Translating and adapting the mathematical knowledge for teaching (MKT) measures: the cases of Indonesia and Norway. *The Mathematics Enthusiast*, 9 (1&2), 149–178.

Schilling, S., 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. (1992). Learning to think mathematically: problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). New York: Macmillan.

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.

Singh, J. (1995). Measurement issues in cross-national research. *Journal of International Business Studies*, 26 (3), 597–619.

Sirnes, S. M. (2005). *Flervalgsoppgaver – konstruksjon og analyse*. Bergen: Fagbokforlaget.

Stigler, J. W. & Hiebert, J. (1999). *The teaching gap: best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.

Stylianides, A. J. & Delaney, S. (2011). The cultural dimension of teachers' mathematical knowledge. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 179–191). London: Springer.

Sullivan, P. & Wood, B. (Eds.). (2008). *Knowledge and beliefs in mathematics teaching and teaching development* (Vol. 1). Rotterdam: Sense Publishers.

Tonheim, O. H. M. & Torkildsen, O. E. (2010). Matematikk 1 i lærerutdanninga – kvalifiserande? In P. Haug (Ed.), *Kvalifisering til læreryrket* (pp. 209–226). Oslo: Abstrakt forlag.

Törner, G., Rolka, K., Rösken, B. & Sriraman, B. (2010). Understanding a teacher's actions in the classroom by applying Schoenfeld's theory Teaching-In-Context: reflecting on goals and beliefs. In B. Sriraman & L. English (Eds.), *Theories of mathematics education. Seeking new frontiers* (pp. 401–420). Heidelberg: Springer.

Williams, J. (2011). Audit and evaluation of pedagogy: towards a cultural-historical perspective. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 161–178). London: Springer.

Notes

1. Conventions used in the transcriptions:

   [...] indicates a short comment from the interviewer or the interviewee(s) difficult to hear.

   (...) indicates a short break.

   (. . .) indicates that a part of the interviewers or the interviewee's statement is left out.

   (. . . .) indicates that a part of the focus-group interview is left out.

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

Raymond Bjuland

Raymond Bjuland is Associate Professor in mathematics education at the University of Stavanger, Norway. His interests are related to students' collaborative problem solving in small groups, the use of gestures in teacher-student dialogues, and mathematical knowledge for teaching.

raymond.bjuland@uis.no

Arne Jakobsen

Arne Jakobsen is Associate Professor in mathematics at the University of Stavanger, Norway. His interests are mathematics, mathematical knowledge for teaching, and quantitative studies in mathematics education.

arne.jakobsen@uis.no