Why are Laura and Jane " not sure "?

Janne Fauskanger, Reidar Mosvold

Why are Laura and Jane "not sure"?

Janne Fauskanger and Reidar Mosvold

University of Stavanger, Department of Education and Sports Science, Stavanger, Norway, janne.fauskanger@uis.no.

The mathematical knowledge for teaching (MKT) measures have been widely adopted by researchers. This paper reports on a case study of two mathematics teachers. The aim of the study was to investigate challenges regarding extensive use of "I'm not sure" as a suggested solution in the multiple-choice items. The connection between teachers' responses to multiple-choice MKT items, their written responses to corresponding open-ended questions (long responses) and group discussions are analyzed. The findings indicate that teachers' responses to multiple-choice items do not always correspond with the understanding revealed in responses to open-ended questions and in group discussions.

Keywords: MKT, measuring teachers' knowledge, multiple-choice items.

INTRODUCTION

This paper has a focus on measuring teachers' mathematical knowledge for teaching (MKT), and it builds upon the practice-based theory of MKT that was developed by researchers in the US (Ball, Thames, & Phelps, 2008). In connection with the development of this theory, sets of multiple-choice items were developed to measure MKT (e.g., Hill, Sleep, Lewis, & Ball, 2007). These items—which are often referred to as MKT items—were developed in the Learning Mathematics for Teaching project (LMT).

A multiple-choice item includes a correct answer—referred to as the key—and several wrong answers or distractors. In order to reduce the possibility of guessing, "I'm not sure" was included as a suggested solution in MKT items formatted yes/no (see Figure 1) in 2001 (Hill, 2007). In 2002, the researchers who developed these items were required to include "I'm not sure" on all number and operations content knowledge items that were piloted. To avoid changes in item parameters, "I'm not sure" was also included in the forms from 2004—which were later translated and adapted for use in the Norwegian context (Fauskanger, Jakobsen, Mosvold, & Bjuland, 2012). "I'm not sure" is always coded as incorrect, and teachers who are experienced with such tests would therefore immediately know that this suggested solution is a distractor. Some teachers would thus avoid this suggested solution, whereas other teachers might select it to avoid giving a wrong answer or to avoid guessing. In his often-cited handbook of development and validation of multiple-choice test items, Haladyna (2004) recommends that all distractors in multiple-choice items should be plausible, and, as a consequence, the suggested solution "I'm not sure" should be avoided. Many issues regarding the MKT items have been investigated already, but possible challenges related to the use of "I'm not sure" have not received much attention.

Previous studies conclude that teachers seem to have different reasons for selecting this alternative solution; some select "I'm not sure" based on uncertainty, whereas other teachers reveals instrumental and even relational understanding of the content of the items—but still select this alternative (Fauskanger & Mosvold, 2014). In this paper, we analyze data from a collective case study in an attempt to approach the following research question:

> What differences can be found between teachers' arguments for choosing the suggested solution "I'm not sure" in multiple-choice items?

This question is important in order to reveal challenges of extensive use of "I'm not sure" as a suggested solution in multiple-choice items developed to measure MKT. We focus in particular on two 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 TEACHER KNOWLEDGE

Teacher knowledge is imperative to high-quality teaching (e.g., Davis & Simmt, 2006), and it is thus relevant to gain insight into methods used to access and assess different aspects of teachers' knowledge (e.g., Hill et al., 2007). At the University of Michigan, researchers have developed a practice-based theory of mathematical knowledge for teaching (MKT). Hill, Rowan and Ball (2005, p. 373) define MKT as "the mathematical knowledge used to carry out the work of teaching mathematics". In the LMT project, considerable resources were 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 and colleagues (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). Inspired by these promising results, researchers have adapted the measures for use both in and outside the US (Blömeke & Delaney, 2012).

Responses from multiple-choice items are expeditiously analyzed and they can be used at scale. Developing multiple-choice items that are intended for measuring something beyond procedural skills is, however, 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 result in trivialization of the complexities of teaching and thus threaten validity (Beswick, Callingham, & Watson, 2012; Haertel, 2004). In his critical discussion of the MKT measures, Schoenfeld (2007) argued that these measures might test something other than they are supposed to. He further suggested 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). An additional aspect of Schoenfeld's (2007) criticism was that the items actually measure a type of knowledge that is more procedural than intended, and more recent studies in Norway support this criticism (Fauskanger & Mosvold, 2015).

A standard multiple-choice item consists of two parts: a problem (often referred to as stem) and a list of suggested solutions. In the MKT items, the stem is typically situated in the context of the work of teaching mathematics. The list of suggested solutions contains a key and one or more distractors (often incorrect alternatives). 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 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

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

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).

Different approaches have been made to distinguish between different categories of teachers' knowledge and understanding of the mathematical content; Skemp's (1976) distinction between instrumental and relational understanding is an archetype. Rote memorization of algorithms for two-digit multiplication (Figure 1) is an example of instrumental understanding, whereas relational understanding encompasses a deep, conceptual understanding. Skemp argued that students cannot develop relational understanding from instrumental teaching. Recent research has investigated this connection empirically and concludes that teachers' knowledge of facts and procedures have less positive effect on the quality of instruction and students' achievement relative to knowledge of concepts and connections (Tchoshanov, 2011). For this reason, the importance of exploring teacher knowledge accessed by measures as well as possible challenges regarding the use of multiple-choice items to investigate something as complex as teacher knowledge, becomes evident. It is relevant to carefully investigate these challenges—in particular the challenges revealed from extensive use of "I'm not sure" as a suggested solution.

METHODOLOGY

This study is part of a collective case study (Silverman, 2006), and two in-service teachers were chosen as cases to investigate the topic under investigation. When adopting a case-oriented approach, we considered the case as an entity and first looked for configurations and characteristics within the case before we searched for similarities and patterns across cases. The two teachers, who have been assigned the pseudonyms Laura and Jane, were participants in a professional development course. A total of 38 in-service teachers participated in the course, and 30 of these teachers agreed to submit multiple-choice responses to 28 MKT items (including the testlet in Figure 1). All of these items had a focus on number concepts and operations, and 18 of the items included "I'm not sure" as a suggested solution. 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 questions prompting long responses were developed to tap into teachers' instrumental and relational understanding (Skemp, 1976) and varied across the 28 items. The group discussions were based on the same 28 items.

In a previous publication, we have reported on results from analysis of multiple-choice responses and long responses for all the teachers who selected "I'm not sure" as a response to a particular item (Fauskanger & Mosvold, 2014). Those analyses revealed that three groups of teachers could be distinguished between (see Table 1).

Laura and Jane both responded "I'm not sure", but their long responses indicated that Jane did so due to insecurity (group 1), whereas Laura indicated relational understanding in her long response (group 3). In our attempt to investigate different arguments for selecting "I'm not sure" as a response to MKT items, we therefore selected these two teachers as contrasting cases.

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; they select "I'm not sure" to avoid guessing. In order to learn more about what challenges are revealed from extensive use of "I'm not sure" in multiple-choice items, we focus on what types of understanding could be found in teachers' long and oral responses when they select the answer "I'm not sure" on a multiple-choice item.

The unit of analysis is the individual teachers' multiple-choice responses, their long responses and their

Group	Group 1: not sure	Group 2: instrumental understanding	Group 3: relational understanding
Name of teacher	Erna, Frøya, *Jane*, Jan, Ada and Nina	Pia, Mons, Harald, Ola and Are	Sara, Inge, Ragna and *Laura*

Table 1: Teachers grouped according to their long responses

individual voices as revealed 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).

We used Skemp's (1976) categories for our coding of the textual data. Excerpts from teachers' written and oral 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 were coded as relational. A third code, low/no MKT, 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 low level of MKT. 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.

THE CAse of Jane

Jane's long responses indicated insecurity related to the content in focus, and she responded "I'm not sure" 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 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." 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 Jane, and that was the intention of including this suggested solution in the MKT items (Hill, 2007).

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

> 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 5. 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≈
>
> Ragna: ≈Open for all [to choose]≈
>
> 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.
>
> Interviewer: But do you experience that the parents have a common [algorithm]?
>
> 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 (...)
>
> Interviewer: Anything else you want to say in relation to the items that were concerning different algorithms?
>
> 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. The excerpt thus indicates that Jane's knowledge related to multi-digit multiplication is instrumental (Skemp, 1976). She reveals, however, that the professional development course has made her aware of the existence of multiple algorithms, and she seems to agree that different algorithms might be useful for different students.

THE CASE OF LAURA

Laura also selected "I'm not sure" as her response to several multiple-choice items. In contrast with Jane, however, Laura's long responses indicates that this was not due to low level of MKT or insecurity. On the contrary, Laura's long responses indicate deep conceptual knowledge (Skemp, 1976). In her long response related to a particular testlet — with a content focus related to place value and non-standard ways of decomposing three-digit numbers (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. 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." 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 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.

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 relational understanding (Skemp, 1976) 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, and she responds "I'm not sure" despite of a high level of MKT. Laura seems to have responded "I'm not sure" due to the wording of the items included in this testlet. This brings forth issues related to item development and translation (cf., Fauskanger et al., 2012).

When analyzing Laura's utterances from the interviews, it also appears that she has a deep understanding—both in terms of mathematical content knowledge and pedagogical content knowledge. The interview data thus, seems to support the long responses from Laura.

> 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.
>
> (...)
>
> 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 4, 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 excerpts, Laura selected "I'm not sure" despite being secure about the mathematical content. The voice of Laura in the group discussion thus seems to support findings from her long responses and the reason why she responded "I'm not sure" in the multiple-choice items does not relate to her low level of MKT but rather her relational understanding (Skemp, 1976) of the content.

CONCLUDING DISCUSSION

In a previous study, we analyzed 15 teachers' long responses related to a set of multiple-choice items from the LMT project (Fauskanger & Mosvold, 2014). Based on those results, two teachers were chosen as contrasting cases in the present study. We analyzed

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 the challenges related to extensive use of "I'm not sure" as a suggested solution in the MKT multiple-choice items.

The results from our analysis indicate that we need to be careful about how we interpret it when teachers select "I'm not sure" as a multiple-choice response. "I'm not sure" is always coded as incorrect, and our analysis of teachers' long responses as well as their interview discussions reveals that this alternative was indeed selected by some teachers who explicitly indicated that they could not identify the key due to their low level of local MKT (as Jane, see Table 1). 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). Other teachers, however, 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 relational understanding (e.g., Laura). Teachers might draw on deep conceptual or relational knowledge (Laura), procedural or instrumental knowledge (Schoenfeld, 2007), or their lacking knowledge (Jane) when responding "I'm not sure". Our results thus indicate that the knowledge teachers utilize in long responses or discussions does not necessarily mirror the 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, and the inclusion of this alternative response in MKT items should also be critically discussed.

Our analysis of teachers' long responses as well as their discussions in group interviews indicate that none of the 30 participating teachers were guessing—thus conforming to the intention of including "I'm not sure". This suggested solution is still problematic, however, since one cannot conclude whether the choice of this suggested solution is based on lack of knowledge as one extreme point or deep conceptual knowledge as the other extreme. One might advocate removal of the "I'm not sure" option from the items, but this would change item parameters and should not be done hastily. Instead, we call for cross-cultural studies that investigate teachers' motivation for selecting a suggested solution like "I'm not sure" when responding to MKT items. Recontextualizing the MKT items into scenarios for use in qualitative studies related aspects of teachers' MKT as done by Adler and Patahuddin (2012) might be a fruitful approach. These researchers argue that the carefully constructed MKT items provokes teachers' mathematical reasoning in relation to practice-based scenarios.

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, UK: Sage.

Ball, D. L., & Hill, H. C. (2008). *Mathematical knowledge for teaching (MKT) measures. Mathematics released items 2008*. Available at 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.

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

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.

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, NJ: Lawrence Earlbaum.

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. 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. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 111–156). Charlotte, NC: Information Age Publishing.

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*(3), 310–336.

Osterlind, S. J. (1997). *Constructing test items: Multiple-choice, constructed-response, performance and other formats* (2 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* (3 ed.). Thousand Oaks, CA: Sage.

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.