The development of pre-service teachers' self-efficacy in teaching mathematics

ANNETTE HESSEN BJERKE

Teacher efficacy has received much attention in the general field of educational research, but applications in mathematics teacher education are few. In order to deepen the understanding of the nature and development of self-efficacy in teaching mathematics (SETM) during teacher education, the study presented here followed over a period of two years pre-service teachers (PSTs) preparing to teach primary school mathematics in Norway (grades 1–7, ages 6–13). Their developing SETM was investigated by means of an instrument designed to target the core activity of teaching mathematics: helping a generic child with mathematics tasks. A comparison of responses collected from 191 novice PSTs with those from the same cohort two years later ($n = 103$) shows a rise in SETM in the typical PST, and indicates the nature of the development of SETM during teacher education.

Back in 1990, Woolfolk and Hoy noted that "researchers have found few consistent relationships between the characteristics of teachers and the behavior or learning of students. Teachers' sense of efficacy [...] is an exception to this general rule" (Woolfolk & Hoy, 1990, p. 81). More recently, a synthesis of 40 years of research on the construct, found positive effects of high teacher efficacy, like teachers' well-being, classroom quality and student achievement (Zee & Koomen, 2016). These effects of teacher efficacy surely make the concept of great significance for teacher education, especially since teacher efficacy develops mainly during teacher education and tends to decline during the first year of teaching (Hoy & Spero, 2005).

While teacher efficacy has received considerable attention from researchers over recent years, studies directed specifically at mathematics teaching are relatively rare (Klassen, Tze, Betts & Gordon, 2011), and even

Annette Hessen Bjerke
Oslo and Akershus University College of Applied Sciences

Bjerke, A. H. (2017). The development of pre-service teachers' self-efficacy in teaching mathematics. *Nordic Studies in Mathematics Education*, 22 (3), 71–84.

fewer are directed at pre-service teachers (PSTs) (Charalambous, Philippou & Kyriakides, 2008). Except from investigations during fieldwork of PSTs' mathematics teacher efficacy beliefs (Charalambous et al., 2008), a recent review identifies another two studies examining the development of mathematics teacher efficacy in PSTs (Philippou & Pantziara, 2015): one found differences in the development of PSTs' beliefs and attitudes that relate to their mathematical background (Charalambous, Panaoura & Philippou, 2009), while another indicated no significant increase in teacher efficacy beliefs during a mathematics methods course (Evans, 2011). The research is still too sparse and inconclusive, and Philippou and Pantziara (2015) propose that further studies focusing on PSTs' mathematics teacher efficacy throughout educational programmes would help teacher educators to better understand the developmental process.

For PSTs, teacher education adds a new perspective to mathematics: it is no longer just about doing mathematics; it is also about helping others to do and preferably understand mathematics. It is crucial then that teacher educators understand the development of future mathematics teachers' confidence regarding this perspective that is captured in the construct of ”self-efficacy in teaching mathematics” (SETM). Studying the nature and development of SETM during teacher education fills a gap in the existing body of research (Klassen et al., 2011; Philippou & Pantziara, 2015; Wheatley, 2005). This paper addresses the research question: ”To what extent does PSTs' SETM develop during a mathematics methods course in primary teacher education, and what is the nature of this development?” The paper starts by defining the concept of SETM and presenting an instrument for its measurement, and continues with an analysis and discussion of data collected from a cohort of PSTs before and after a mathematics methods course in teacher education.

Self-efficacy in teaching mathematics

Self-efficacy is a person's judgment of his or her abilities to execute successfully a course of action (Bandura, 2006), a future-oriented belief about the level of competence one expects to show in a specific situation. Bandura's (1977) concept of self-efficacy has two components, i.e. a belief about action and outcome, an outcome expectancy; and a personal belief about one's own ability to cope with a task, a personal self-efficacy.

There is a wealth of different ways of both defining and measuring teacher efficacy, and numerous researchers have raised concern about conceptual problems in self-efficacy measurement (Klassen et al., 2011). Some studies include both components (”personal self-efficacy” and ”outcome expectancy”), but as suggested by Ross (1994) after analysing

88 studies of teacher efficacy, studies should examine the distinct components separately. This study builds on Ross' (1994) understanding of the "personal self-efficacy" component of Bandura's theory: "the respondent's expectation that he or she will be able to bring about student learning" (p. 4), as opposed to outcome expectancy (the belief that certain actions will lead to learning).

However, teacher efficacy is dependent upon subject-matter, context and the population on which it is measured (Riggs & Enochs, 1990). Seen in the subject-specific situation of teaching mathematics, and in line with Ross' (1994) understanding, in this paper, SETM relates to the component of teacher efficacy corresponding to Bandura's concept of personal self-efficacy. Despite the fact that SETM is already a more focused construct than teacher efficacy (in that it narrows down from generalist teaching to teaching mathematics), it is still too comprehensive to be measured well in novice PSTs – the sixteen recurrent tasks of teaching mathematics (Ball, Thames & Phelps, 2008), are arguably not all appropriate for novices (Bjerke & Eriksen, 2016).

Self-efficacy in tutoring children in primary mathematics (SETcPM) is proposed to be a central part of SETM possible to be measured in the intended population (Bjerke & Eriksen, 2016). Tutoring is an important subset of teaching, and tutoring is argued to be "socially dynamic, but not reflecting the full-blown complexity of the mathematics classroom" (Schoenfeld, 2010, p. 6). Additionally, SETcPM holds in it those of the sixteen tasks of teaching mathematics possible to be measured in novices, such as "responding to students' 'why' questions", "finding an example to make a specific mathematical point", and "linking representations to underlying ideas and to other representations" (Ball et al., 2008, p. 400). In this paper, I focus on the particular context of SETcPM as a central part of SETM in PSTs preparing to teach mathematics in primary school.

Method

In response to the literature on teacher efficacy, a 20-item SETcPM-instrument was developed with the intention of contributing to a better understanding of novice PSTs' initial SETM, and to allow mapping the development of SETM during teacher education (Bjerke & Eriksen, 2016). The instrument targets SETM in light of the core activity of teaching mathematics: helping a generic child with mathematics tasks.

In the SETcPM-instrument, PSTs are asked to indicate on a four-point Likert scale (with response categories *Not confident*, *Somewhat confident*, *Confident* and *Very confident*) their level of confidence in helping a generic child with 20 mathematics tasks. Ten items focus on rules and algorithmic procedures in mathematics (e.g. "How confident are you that you can help a child with the task "Calculate 4.14 + 3.190"?") which simply ask for instrumental calculations. The other ten items are based on what Skemp refers to as relational understanding, requiring "knowing both what to do and why" (Skemp, 1976, p. 20). These items focus on reasoning and explanations (e.g. "How confident are you that you can help a child with the task "Explain why 0.3 is ten times larger than 0.03"?"). These two categories of items (listed in figure 2) are intended to capture what Skemp labels the two "kinds of mathematics" (Skemp, 1976, p. 26).

The participants are a cohort of PSTs enrolled in a four-year programme for primary school teachers (grades 1–7, ages 6–13) at a university college in Norway. The programme includes a 30 credits compulsory course in mathematics pedagogy spanning the first two years. PSTs undergo a total of 30 days of school placement during each of the two first years, and the first data collection point (the pre-test) took place at the start of the first semester, before the first period of placement, while the second data collection point were conducted about 18 months later (the post-test). The 191 novice PSTs that undertook the pre-test had a mean age of 22.5 years, and ware 30 % men. At the end of this course, 103 2nd year PSTs (a subgroup of the original 191 novice PSTs) completed the post-test. Comparisons of the results from the pre- and post-test implementations enable investigations into the nature and the extent to which PSTs' SETM develops during this particular mathematics methods course in primary teacher education.

In line with the approach taken when validating and reporting the SETcPM-instrument (Bjerke & Eriksen, 2016), the *Rasch rating scale model* (RSM) was applied when analysing the data. The rationale for choosing RSM in the first place, was the RSM's emphasis on identifying and studying anomalies in the data disclosed by the Rasch model, instead of choosing a model that best characterizes the given data. In this way, RSM helped to identify which items best measure the underlying constructs, and moreover, it revealed which items are interpreted in the same way when the same test is implemented at different occasions (like is the case here).

Traditionally responses on the Likert-scale were assigned numbers (e.g. 0 for *Not confident* to 3 for *Very confident*) and these numbers were added together to provide a raw score on the test. However, there are some problems when simply adding up a PST's responses on a Likert-scale and use this raw score to address this PST's level of SETcPM. Likert-scale are ordinal data that cannot be assumed to be linear (Boone, Staver & Yale, 2014). The Rasch measurement, which is based on an equation developed by George Rasch, converts these ordinal data to linear mea-sures. In this way, the strength of the Rasch model is that it supports the construction of a genuine interval estimate for the underlying construct, and both items and persons are measured on the same scale in unit logits, the logarithm of the odds of success. This allows comparisons between items, between persons, and between items and persons, in the form of establishing a person's probable answer on an item.

Analysis

When the data satisfy the conditions of the Rach model, the model provides both items and persons with measures, where the higher the person measure, the more evidence of the presence of SETcPM, and the higher an item measure, the more SETcPM is needed to endorse it.

All the 20 items in the SETcPM-instrument met the requirements from the Rasch model for the 191 novice PSTs, and were therefore considered as contributing to the measurement of novice PSTs' SETcPM (Bjerke & Eriksen, 2016). Since validation is a continuous process, I reiterated the process for the mixed sample discussed in this paper, which includes 103 post-test responses (2nd year PSTs) in addition to the original 191 pre-test responses (novice PSTs). A potential limitation to the usefulness of the instrument for measurement of *developing* SETcPM is that items might be interpreted in significantly different ways by the respondents undertaking the instrument at different stages in the course of the programme (Bond, 2004; Wolfe & Smith Jr, 2007b). In order to investigate this closer, I next present the steps taken to reduce potential skewing of the results due to different interpretations by different groups (for details see Boone et al., 2014).

A Rasch analysis was completed based on the 20 items for the mixed sample ($n = 293$) of novice PSTs and 2nd year PSTs. The unidimensionality condition of the Rasch model held sufficiently well for the data, with mean square fit statistics (MNSQ) showing fit values within acceptable limits for all 20 items (between 0.83 and 1.11), with the exception of item 11 (outfit MNSQ = 0.58) possibly overfitting the model. The Rasch reliability estimates (which in general underestimate reliability) were 0.89 for persons and 0.99 for items, indicating a reproducible measure.

In order to investigate the invariance of item difficulties (i.e. whether item difficulties are the same for the two groups – or whether the novice PSTs and 2nd year PSTs interpreted any of the items differently), a differential item functioning (DIF) analysis was conducted, revealing that five items exhibited DIF (2, 7, 8, 9, and 15). Those items are treated as distinct items (renamed as 2', 7', 8', 9' and 15' for 2nd year PSTs) for the two groups in the following analyses, whereas the remaining 15 items were used to anchor the two tests. This anchoring enables an accurate comparison between the pre- and post-test results.

A new analysis of the anchored pre- and post-test was conducted. Table 1 shows that the unidimensionality condition of the Rasch model holds sufficiently well for the data, with MNSQ fit values within acceptable limits for all 20 items (with item 11 possibly overfitting the model). Further DIF-analysis of high versus low scoring PSTs revealed no problematic items, except item 11 (contrast of 1.03, over the limit of 0.64 suggested by Linacre (2014)). The lack of additional background data makes it hard to ascertain why, but it is important to keep this in mind when interpreting the results of the analysis.

Table 1. *Fit statistics for SETcPM in mixed, anchored sample. Item 2, 7, 8, 9 and 15 measured only on novices, item 2', 7', 8', 9' and 15' only on 2nd year PSTs*

Item entry	SETM measure	Outfit MNSQ	Outfit ZSTD	Item entry	SETM measure	Outfit MNSQ	Outfit ZSTD
1	.77	1.05	.6	14	.48	1.21	2.5
2	.20	1.08	.8	15	.19	1.17	1.6
3	.31	1.12	1.5	16	.17	.84	-2.1
4	-.58	.89	-1.1	17	1.47	.94	-.7
5	-1.26	.94	-.4	18	.93	1.11	1.3
6	-1.61	.94	-.4	19	-1.16	.88	-1.0
7	1.65	.88	-1.1	20	1.12	1.05	.6
8	-1.59	.81	-1.2	2'	-.56	.89	-.6
9	.53	1.00	.0	7'	2.40	1.02	.2
10	-.10	.97	-.3	8'	-.87	.82	-.8
11	-2.08	.58	-2.7	9'	-.26	.88	-.7
12	-.39	1.01	.1	15'	-.54	.99	.0
13	.78	1.03	.4				

The analysis revealed that each answer category is most probable for some combination of item measure and person measure. For the mixed sample, all answer categories are used, the observed average measure for respondents endorsing the four answer categories increases as expected with the categories, and, overall, there is a higher usage of the categories showing high level of SETcPM. Additionally, there is no disordering of the average person abilities by category for the individual items, except for items 2' and 8'. For item 2' the observed mean measure for respondents endorsing *Not confident* was .15, slightly higher than that for the *Somewhat confident* (-.08), and for item 8' the observed mean measure for respondents endorsing *Somewhat confident* was .68, slightly higher than that for the C*onfident* (.60). Given the small difference and the low count for respondents endorsing *Not confident* on item 2' (only 4) and *Somewhat confident* on item 8' (only 5), I did not act on this, but will monitor the items in the future.

Following the steps taken when analysing the pre-test (Bjerke & Eriksen, 2016), a principal component analysis of residuals was conducted to look for potential multidimensionality in the post-test. A secondary dimension with a strength (eigenvalue) of 2.8 was identified, with the content of the items at the two ends of the contrast exhibiting substantive differences, with calculation items at the top and explanation items at the bottom. Additional analyses were conducted based on data simulated to fit the Rasch model, but no extra dimension was found, and I conclude that SETcPM as measured by this instrument in the post-test is sufficiently unidimensional.

The collected results of these analyses enable me to compare the two tests and to express the responses of the novices and the 2nd year PSTs on the same linear metric. This allows the investigations and interpretations on how the SETcPM has developed in this particular cohort of PSTs, and also how SETcPM has developed in individuals.

Results

The overall rationale for developing the instrument was to measure development of self-efficacy in teaching (not just tutoring) mathematics in primary school. In Bjerke and Eriksen (2016) we have argued for the central role of SETcPM in SETM. Building on that, I argue here that the development of SETcPM is worthwhile investigating, as it encompasses issues at the core of the development of SETM.

Five of the 20 items were perceived differently by novice PSTs and 2nd year PSTs and were for that reason excluded from the anchor. Upon closer inspection, the DIF-analysis revealed that, on two items (items 7 and 8, see figure 2 for wording) 2nd year PSTs were less confident at the end of the course than expected based on their perception at the start of the course. A possible explanation for this is that teacher education might have made the 2nd year PSTs aware of the challenges of being able to help children with knowing both what to do and why, as in Skemp's relational understanding. Item 7 is algebraic in nature, and a novice perhaps would not see the challenges in the nature of the task. The same is the case on item 8, a novice might not perceive the difficulties associated with negative numbers.

In figure 1, the proportion of responses in each category (from *Not confident* at the bottom to *Very confident* at the top of each column) for each item provides a clear picture of the pattern of endorsement of the survey items for the mixed sample. Here the items are deliberately displayed in two groups, according to the two types of tasks ("rules" and "reasoning") embedded in the instrument. To the left in figure 1 are the items demanding explanations, and to the right the items focusing on calculations. There is a clear trend in the data from the mixed group: calculate-items received a higher proportion of *Very confident* answers (the black portion of the bars). This stands in contrast to the explain-items, where the categories *Not confident* and *Somewhat confident* are noticeably more frequent. This tells us that the trend is that the items focusing on reasoning and explanations are harder to endorse for the typical PST.

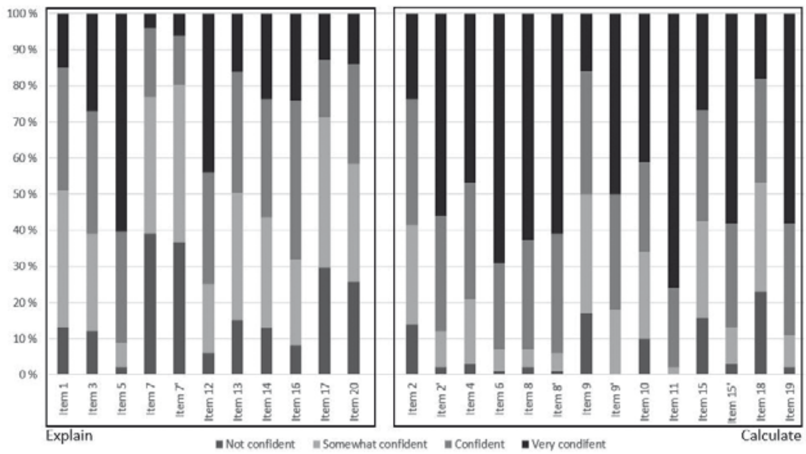

Figure 1. *Proportion of responses in each response category for the mixed sample*

Figure 2 shows an ordinal map where the frequency of responses from the mixed sample on each item is shown a "bar", with changes in colour positioned at the Andrich thresholds, the points of equal probability of adjacent categories (Bond & Fox, 2007). For example, for Item 14, the threshold between *Somewhat confident* and *Confident* is located at about the same place as the mean for novices (.55), which means that a PST with an estimate at this position on the scale will have a 50 % chance to choose either of the two categories *Somewhat confident* and *Confident*. Since the thresholds are ordered, a person with an estimate above .55 (but below the next threshold) will most probably endorse category *Confident* on item 14.

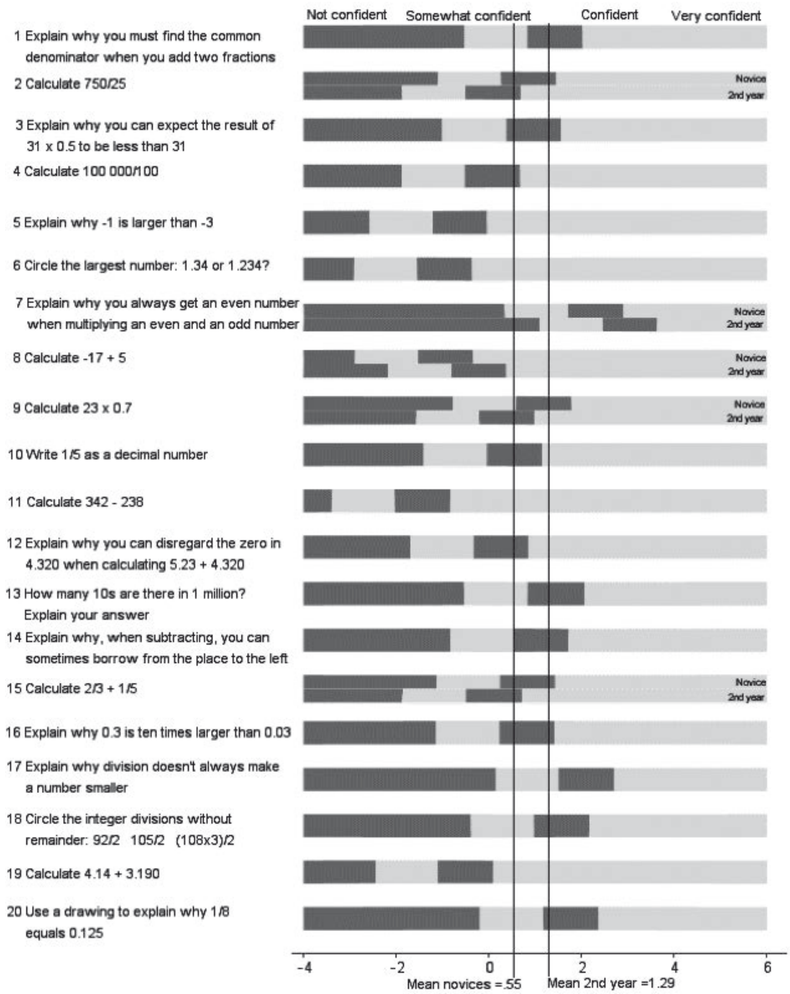

Figure 2. *The mixed sample (n = 294) reported on an* Ordinal Map. *Not anchored items are shown with two bars, for responses from novices, and 2nd year PSTs respectively*

As shown in figure 2, the means for novice PSTs and 2nd year PSTs are 0.55 logits ($SD = 1.16$) and 1.29 logits ($SD = 1.10$) respectively, and an unpaired t-test shows that the difference is statistically significant ($t(102) = 5.387$, $p < 0.01$). The result confirms the theory-based expectation that SETM increases during teacher education. I will now elaborate in two ways on the development identified by comparing the pre- and post-test at cohort and at person level.

To establish the nature of pre- to post-test development, Boone et al. (2014) propose deriving a more concrete understanding of the differences between groups by interpreting the gap between the means on the scale, in terms of the probable answer on each item for the "typical" member of each group. Figure 2 reveals that on seven of the 15 comparable items the means of the two groups correspond on the scale to different response categories, while the remaining eight items correspond to the same response category. The typical PST's responses to four items (1, 13, 14, 20) change from *Somewhat confident* at the beginning of the course to *Confident* by the end of the course, where three of the items ask about confidence in helping a generic child explain why. The typical PST's responses to three items (4, 10, 12) change from *Confident* to *Very confident*, with only one explain-item. This difference between the mean for novice PSTs and that for 2nd year PSTs suggests that, even for the typical 2nd year PST, it is easier to tick *Very confident* on teaching tasks concerned with rules and calculations than on tasks asking to explain why. Additionally, figure 2 reveals that a typical novice PST (mean = .55) is *Very confident* on one explain-item and four calculate-items, while a typical 2nd year PST (mean = 1.29) is *Very confident* on two explain-items and nine calculate-items. The analysis indicates that the typical 2nd year PST does not reach, during teacher education, the highest level of SETM on most of the items involving explain-tasks that require him to address relational understanding.

While the expectation that SETcPM should increase during teacher education was confirmed for the cohort as a group, it is also of interest to see the individual trajectories, which give a more detailed picture of the complexity in the group. 78 PSTs can be identified in both tests, and for this group there is an overall moderate correlation between the two sets of SETcPM measures (Spearman's rho coefficient 0.48, significant at the 0.01 level), however there are some switches in the rankings, as illustrated when comparing the positions of individual PSTs in figures 3a and 3b. There are some individual movements from pre-test to post-test. In figure 3a and figure 3b fifteen of the PSTs identifiable on both pre- and post-tests are marked out: the five with lowest SETcPM-scores on the pre-test are marked with a's, the five with pre-test scores closest to the mean for novices are marked with b's, and the five with highest SETcPM-scores on the pre-test are marked with c's. Despite the fact that the measures in the pre- and post-test seem to give approximately the same distributions, these comparisons reveal movement in the group. Those traceable PSTs with lowest SETcPM-scores in the pre-test (marked with a's) are no longer amongst those with lowest scores in the post-test, while those closest to the mean in the pre-test seem to have spread more in the post-test, but more in the positive direction. Those with highest SETcPM-scores on the pre-test are still among those with highest scores in the post-test. However, caution is needed when discussing those PSTs at the top of the scale because their measures are higher than that of the most difficult item to endorse, meaning that they are not well measured by the instrument, but it is nevertheless noticeable that they have not dropped on the scale.

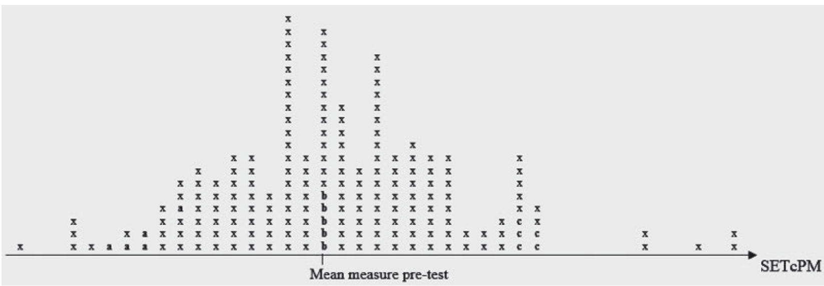

Figure 3a. *The distribution of 191 novice PSTs' responses on the pre-test*

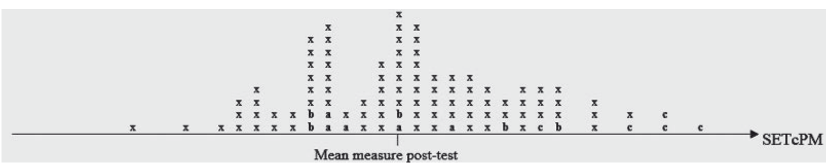

Figure 3b. *The distribution of 103 2nd year PSTs' responses on the post-test*

Having identified movements in the group, the question is whether it is possible to learn more about who benefits (in terms of SETcPM) from this particular mathematics methods course and what can be said about the increase in SETcPM for the PSTs that entered the programme with low SETcPM. When comparing scores on pre- and post-test for the 78 traceable PSTs, 16 exhibit almost no development of SETcPM (change in SETcPM < |.2|), 53 have positive SETcPM-development, and nine negative. Those with negative development are especially interesting, as they go against the trend. Seven of them had very high measures as novices (above the mean for 2nd year PSTs (>1.29)) – it could be that teacher education has been a revelation for them, perhaps the programme has enabled them to judge more accurately their confidence, in contrast with the initial over-confidence. Teacher education might have tuned them in by addressing some of the challenges embedded in the 20 items, which in turn can be interpreted as a positive outcome.

Discussion and conclusion

This study adds to the body of research showing that teacher efficacy, and SETM in particular, develops during teacher education. Previous research in the domain of mathematics teacher efficacy lacks a focus on PSTs (Charalambous et al., 2008) and a focus on relevance to practice (Klassen et al., 2011). There is a need to understand how teacher efficacy is fostered by teacher education programmes, which in turn requires a clearer understanding of how efficacy beliefs change over time (Klassen et al., 2011). The two samples in the pre- and post-test are relatively small, with only 78 PSTs identifiable in the intersection; nevertheless, comparisons of their responses can inform teacher educators on the nature of the development of SETcPM as a core component of SETM.

In particular, these data have shown that PSTs tend to be less confident when it comes to tasks requiring them to explain (rather than calculate). These findings are closely connected to a point initially made at the very start of this paper: mathematics is no longer just about doing mathematics following predetermined rules; it is also about helping others to do and preferably understand mathematics (as needed in order to being able to *explain why*). These analyses show that this layer, the need to be able to explain why and how things work in mathematics, is something that teacher education needs to focus on even more. It is visible in the comparison of the typical novice with the typical 2nd year PST: the typical novice PST was *Very confident* on four calculation-tasks, compared to nine for 2nd year, while the typical novice PST was *Very confident* on only one explaining-task, raised to two for the typical 2nd year PSTs.

One could expect that all PSTs have lower SETcPM as novices than they do two years later, however, this did not turn out to be the case. Teacher education has different effects on different PSTs. One needs, for instance, to understand more on what makes some PSTs lose confidence while others gain confidence. In addition to the positive trends in the group as a whole, this study shows some of the advantages of looking at individual trajectories, as they give more details of the complexity in the group. Despite the fact that the PSTs in the cohort discussed in this paper attended the same mathematics methods course with the same length of school placements, this research reminds us not to expect PSTs to develop SETcPM in similar manners.

An important question concerning those losing confidence might be to ask whether this is always undesirable. This particular research suggest that it is not, as it might mean that the programme enables the PSTs to gauge their confidence more accurately as it prepares PSTs on what teaching mathematics demands of them when it comes to their own subject matter knowledge, as in their actual *knowledge of mathematics*. Following this thought, in future research, there is a need to bring together two main bodies of research – teacher self-efficacy and subject matter knowledge – in order to understand more about how PSTs recognise and relate to their subject matter knowledge when developing SETcPM and hence SETM.

Limitations

The study has certain limitations that need to be addressed. As pointed out in Bjerke and Eriksen (2016), items harder to endorse need to be added to the instrument for future implementations of pre/post-test designs, especially for the post-test administration. This would allow a more accurate measurement of PSTs with high SETcPM on the pre-test. A recommendation would be to add items that cover more of Ball et al.'s (2008) sixteen recurrent tasks of teaching mathematics.

References

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.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84 (2), 191.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich: IAP.

Bjerke, A. H. & Eriksen, E. (2016). Measuring pre-service teachers' self-efficacy in tutoring children in primary mathematics: an instrument. *Research in Mathematics Education*, 18 (1), 61–79.

Bond, T. (2004). Validity and assessment: a Rasch measurement perspective. *Metodologia de las Ciencias del Comportamiento*, 5, 179–194.

Bond, T. & Fox, C. (2007). *Applying the Rasch model: fundamental measurement in the human sciences*. Mahwah: L. Erlbaum.

Boone, W. J., Staver, J. R. & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer.

Charalambous, C. Y., Panaoura, A. & Philippou, G. (2009). Using the history of mathematics to induce changes in preservice teachers' beliefs and attitudes: insights from evaluating a teacher education program. *Educational Studies in Mathematics*, 71 (2), 161–180.

Charalambous, C. Y., Philippou, G. N. & Kyriakides, L. (2008). Tracing the development of preservice teachers' efficacy beliefs in teaching mathematics during fieldwork. *Educational Studies in Mathematics*, 67 (2), 125–142.

Evans, B. R. (2011). Content knowledge, attitudes, and self-efficacy in the mathematics New York City Teaching Fellows (NYCTF) program. *School Science and Mathematics*, 111 (5), 225–235.

Hoy, A. W. & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: a comparison of four measures. *Teaching and teacher education*, 21 (4), 343–356.

Klassen, R. M., Tze, V. M., Betts, S. M. & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: signs of progress or unfulfilled promise? *Educational Psychology Review*, 23 (1), 21–43.

Linacre, J. M. (2014). *Winsteps® Rasch measurement computer program user's guide*. Retrieved from www.winsteps.com

Philippou, G. N. & Pantziara, M. (2015). Developments in mathematics teachers' efficacy beliefs. In B. Pepin & B. Roesken-Winter (Eds.), *From beliefs to dynamic affect systems in mathematics education* (pp. 95–117). Cham: Springer International Publisher.

Riggs, I. M. & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, 74 (6), 625–637.

Ross, J. A. (1994, June). *Beliefs that make a difference: the origins and impacts of teacher efficacy*. Paper presented at the annual meeting of the Canadian Association for Curriculum Studies, Calgary.

Schoenfeld, A. H. (2010). *How we think: a theory of goal-oriented decision making and its educational applications*. New York: Routledge.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 20–26.

Wheatley, K. F. (2005). The case for reconceptualizing teacher efficacy research. *Teaching and teacher education*, 21 (7), 747–766.

Wolfe, E. W. & Smith Jr, E. V. (2007b). Instrument development tools and activities for measure validation using Rasch models: part II -- validation activities. *Journal of Applied Measuremen*t, 8 (2), 204–234.

Woolfolk, A. E. & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology*, 82 (1), 81–91.

Zee, M. & Koomen, H. M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: a synthesis of 40 years of research. *Review of Educational Research*, 86 (4), 981–1015. doi: 10.3102/003465431562680

Annette Hessen Bjerke

Annette Hessen Bjerke got her PhD degree in September 2017 and this article is a part of her thesis. She has worked as a teacher educator in mathematics at Oslo and Akershus University College since 2004, and is a textbook author in elementary school mathematics. Her research interest concerns how teacher education fosters future mathematics teachers.

annette.hessen@hioa.no