MATHEMATICAL KNOWLEDGE FOR TEACHING PROOF: COMPARING SECONDARY TEACHERS, PRE-SERVICE TEACHERS AND UNDERGRADUATE STEM MAJORS

Orly Buchbinder
University of New Hampshire
orly.buchbinder@unh.edu

Sharon McCrone
University of New Hampshire
sharon.mccrone@unh.edu

Michelle Capozzoli
University of New Hampshire
Michelle.Capozzoli@unh.edu

Rebecca Butler
University of New Hampshire
rlb1053@wildcats.unh.edu

It has been suggested that integrating reasoning and proof in mathematics teaching requires a special type of teacher knowledge - Mathematical Knowledge for Teaching Proof (MKT-P). Yet, several important questions about the nature of MKT-P remain open, specifically, whether MKT-P is a type of knowledge specific to teachers, and whether MKT-P can be improved through intervention. We explored these questions by comparing performance on an MKT-P questionnaire of in-service secondary mathematics teachers, undergraduate STEM majors, and pre-service secondary mathematics teachers. The latter group completed the questionnaire twice- before and after participating in a capstone course, Mathematical Reasoning and Proving for Secondary Teachers. *Our data suggest that MKT-P is indeed a special kind of knowledge specific to teachers and it can be improved through interventions.*

Keywords: Mathematical Knowledge for Teaching, Reasoning and Proof, Preservice and In-service Secondary Teachers.

In recent years, there have been welcomed shifts in the research on teaching and learning of argumentation and proof towards increased focus on classroom-based interventions for supporting students' engagement with reasoning and proving (Stylianides & Stylianides, 2017). These studies have shown that students' opportunities to participate in proof-related practices such as generalizing, conjecturing, posing and critiquing arguments, are dependent on teachers' ability to design learning environments that foster such engagement and on teachers' ability to advance students' learning of reasoning and proof (Bieda, 2010; Cirillo, 2011; Martin, McCrone, Bower & Dindyal, 2005; Stylianides, Bieda & Morselli, 2016).

Given the critical role of the teacher in facilitating student engagement with reasoning and proving (Nardi & Knuth, 2017) and following Ball, Thames and Phelps' (2008) and Shulman's (1986) notion of Mathematical Knowledge for Teaching (MKT), several researchers have introduced the notion of Mathematical Knowledge for Teaching Proof (MKT-P). The latter has been posited as a special type of mathematical knowledge teachers need in order to carry out the work of teaching mathematics with an emphasis on reasoning and proving (e.g., Buchbinder & McCrone, 2020; Lin, et al, 2011; Lesseig, 2016; Stylianides 2011).

Although this line of research is fast growing, several key questions about the nature of MKT-P remain open. Specifically, it is unclear whether MKT-P is a type of knowledge that is specific to teachers of mathematics, or whether it should be viewed as general knowledge of mathematical content. If MKT-P can be shown to be distinctive to the act of teaching reasoning and proving, another important question is whether it is possible to facilitate the development of MKT-P through targeted interventions. Both questions have critical importance for preparation and professional development of mathematics teachers, yet, as far we know, the literature on this topic has been scant. Our study aims to provide some initial answers to both questions.

The study reported herein is part of a larger, NSF-funded 3-year design-based-research project (Edelson, 2002), which investigated how content and pedagogical knowledge of prospective secondary teachers (PSTs) developed as a result of their participation in a uniquely designed capstone course *Mathematical Reasoning and Proving for Secondary Teachers* (Buchbinder & McCrone, 2020). The design of the course and the MKT-P assessment instrument grew out of our conceptualization of MKT-P, described below. As we explored the growth of PSTs' knowledge in the course (our original research objective), we became intrigued by the specificity of the nature of MKT-P, which led to this current investigation. We administered the same MKT-P questionnaire to 17 in-service secondary mathematics teachers, 22 undergraduate STEM majors and 9 PSTs. These PSTs participated in the capstone course in Fall 2019. We hypothesized that the in-service teachers' performance would be quantitatively and qualitatively different from the other two groups. We also hypothesized growth in the PSTs' MKT-P as measured on the pre- and post-test questionnaires.

The Course: Mathematical Reasoning and Proving for Secondary Teachers

Our prior interest in proof and reasoning (Buchbinder, 2010; McCrone and Martin, 2009) and the current work with preservice mathematics teachers has culminated in our design-based research project in which we designed a capstone course *Mathematical Reasoning and Proving for Secondary Teachers* and studied the development of PSTs' knowledge in it (Buchbinder & McCrone, 2020). The course comprised four modules focused on the following proof themes: (1) direct reasoning and argument evaluation, (2) conditional statements, (3) quantification and the role of examples in proving, and (4) indirect reasoning. These topics are known to be particularly difficult to learn and to teach (e.g. Antonini & Mariotti, 2006; Stylianides & Stylianides, 2018).

Each module includes activities aimed to crystalize, connect and apply the PSTs' knowledge of proof and reasoning across a range of secondary mathematics topics. The *crystalize* activities aimed to help PSTs refresh their memory of a particular proof theme. The *connect* activities provided opportunities to connect PSTs' mathematical knowledge with knowledge of students' proof related conceptions and misconceptions. PSTs were then required to *apply* their knowledge in actual secondary classrooms by developing lessons related to a specific proof theme and teaching those lessons to small groups of middle school and high school students. Collectively these activities aimed to enhance PSTs' MKT-P.

Mathematical Knowledge for Teaching Proof Framework

Our conceptualization of MKT-P draws on Schulman's original framework (1986), with the broad categories of subject matter and pedagogical knowledge. Within these categories, we distill those elements that have particular relevance to teaching of reasoning and proving. We also drew inspiration from the existing MKT-P literature (e.g. Corleis et al., 2008; Lin, et al., 2011; Lesseig, 2016; Stylianides 2011), but ultimately developed our own comprehensive MKT-P framework. The framework distinguishes between three interrelated facets: Knowledge of Logical Aspects of Proof (KLAP), Knowledge of Content and Students specific to proving (KCS-P) and Knowledge of Content and Teaching specific to proving (KCT-P). KLAP describes elements of subject matter knowledge specific to proof, such as knowledge of valid and invalid modes of reasoning, knowledge of logical forms of proof, such as direct proof or proof by contradiction, knowledge of a range of accepted definitions, theorems and their proofs, knowledge of logical connections and relations, such as converse, inverse, bi-conditional, etc. The pedagogical content knowledge specific to proof is represented by two types: KCS-P and KCT-P. KCS-P includes knowledge of students' proof related conceptions and misconceptions such as a tendency to rely on inductive reasoning when attempting to prove general statements, or view counterexamples as mere exceptions. KCT-P describes knowledge of pedagogical strategies for supporting student learning of reasoning and proving, such as designing and enacting proof-related tasks, questioning techniques and providing instructional feedback on students' arguments.

The three facets of MKT-P are interrelated. For example, designing proof-oriented tasks (KCT-P) must take into account students' conceptions (KCS-P); and assessing the validity of students' arguments (KCT-P) requires robust knowledge of logical aspects of proof (KLAP). Distinguishing between the knowledge facets was useful for designing both the MKT-P questionnaire and the capstone course targeting MKT-P development.

Methods

Participants and Data Collection

The participants in the study were nine PSTs who participated in the capstone course in Fall of 2019, 17 in-service secondary mathematics teachers and 22 undergraduate STEM majors. All three groups completed the same MKT-P questionnaire, described below. The PSTs were seniors who had successfully completed most of their mathematical coursework, including a *Mathematical Proof course*, and at least one methods course, but had no prior classroom teaching experience. The PSTs completed the MKT-P questionnaire twice, at the beginning and the end of the course.

The in-service teachers were recruited through in-person presentations at local schools and professional development workshops. Of the 17 participants, five teachers were from the same school; the rest were from different schools or districts. Their teaching experience ranged from two to 25 years ($\bar{x}$ = 12.18, SD = 8.00). The teachers completed the Qualtrics Research Suite online version of the questionnaire and received $35 honorarium.

The 22 undergraduate STEM majors were recruited through in-person presentations in three sections of a *Mathematical Proof* course at the same university in which the capstone course was given. The group comprised 11 computer science majors, 9 mathematics majors, 1 mathematics education major and 1 philosophy major. Twelve participants were sophomores, eight juniors and two seniors. The questionnaire was administered in a paper and pencil version during the final weeks of the *Mathematical Proof* course and students received extra credit for this.

MKT-P Questionnaire

We developed a 29-item MKT-P questionnaire, with some questions having a common stem. Ten items were in the area of KLAP, 11 in KCS-P and 8 in KCT-P. The items spanned four proof themes: (1) direct proof and argument evaluation, (2) conditional statements and logical equivalence, (3) quantification and role of examples in proving, and (4) indirect proof, matching the four proof themes of the capstone course (Buchbinder & McCrone, 2020). The mathematical content was middle- to high-school level algebra, geometry and functions.

The KLAP questions were multiple-choice items with a box for justification. The questions called for detecting correct assumptions for a proof by contradiction, determining logically equivalent statements, recognizing circular reasoning steps in given proofs, and identifying counterexamples. The KCS-P questions were grounded in pedagogical context (Baldinger & Lai, 2019), describing classroom situations where students presented arguments for or against a particular conjecture. The participants were to interpret the students' arguments, assess their correctness on a 4-point scale and describe any errors or potential misconceptions (if any) they notice. The KCT-P items had a similar setup as KCS-P items, but instead of numeric assessment, the participants were asked to provide feedback to the hypothetical student, highlighting strengths and weaknesses of their arguments (see Figure 1 for a sample KCT-P item).

Mr. Briggs asked his students to prove the following statement: ***The sum of any two rational numbers is a rational number.***

Molly's solution:

Suppose r and s are rational numbers. By definition of a rational number, let $r = \frac{a}{b}$ where a and b are integers, and $b \neq 0$. Similarly, s is rational so let $s = \frac{a}{b}$ where a and b are integers, and $b \neq 0$.

Then $r + s = \frac{a}{b} + \frac{a}{b} = \frac{2a}{b}$.

Let $p = 2a$. Then p is an integer because it is the product of two integers.

Hence $r + s = \frac{p}{b}$, where p and b are integers and $b \neq 0$.

Thus, $r + s$ is a rational number by definition of a rational number. ■

i) Identify errors (if any) in the student's argument. If none, write "no errors".
ii) Provide feedback to the student, highlighting strengths and weaknesses of their argument.

Figure 1: Sample KCT-P item

Large-scale validation of the instrument was beyond the scope of our original research. Thus, we used expert validation with three mathematicians and one mathematics education expert, and tested the instrument for two years. Cronbach alpha for the entire MKT-P questionnaire was 0.892, with 0.81 for KLAP, 0.71 for KCS-P and 0.76 for KCT-P.

Data Analysis

For the quantitative data analysis, each KLAP item was scored out of 3 points: 1 point for correct choice and 2 points for correct explanation, or 1 point for partially correct explanation. The KCS-P and KCT-P items were scored on a 0-4 point rubric, with 0 points given to a mathematically incorrect response and 4 points to a correct answer that showed deep engagement with the student's argument (exceeding expectations). The research team developed the scoring rubric jointly, by analyzing about 20% of the data. Next, two researchers scored the rest of the data individually and met regularly with the rest of the team to reconcile any discrepancies. The Kappa scores for inter-rater reliability were 0.78 for KCS-P and 0.8 for KCT-P.

For each group: Teachers, STEM majors and PSTs, we calculated the mean total scores for the overall MKT-P. Since the number of items in each subdomain: KLAP, KCS-P and KCT-P, was different, we calculated the mean average score per subdomain per group. Using JMP® Pro statistical software version 15.0.0 we performed one-way ANOVA to determine whether the three groups differed statistically. In addition, we used Welch's Test for the presence of non-constant variance and Tukey-Kramer's Honestly Significant Difference Test for multiple comparisons. Since PSTs' pre and post-course scores are dependent on each other, we performed two separate analyses, once comparing the performance of teachers, STEM majors and PSTs' pre-course scores, and once comparing teachers, STEM majors and PSTs' post-course scores. In the rest of the paper, we use PSTs-pre and PSTs-post to denote this distinction. We also used matched pairs t-tests to compare PSTs-pre to PSTs-post performance.

To capture qualitative differences among the groups we used open coding and thematic analysis (Miles, Huberman, & Saldana, 2018; Yin, 2011). In particular, we coded for the use of first-person language in providing feedback to hypothetical students, and for the types of negative and positive appraisals of student arguments given by the study participants.

Results

The overall MKT-P performance of the three groups is shown in Table 1. The maximum possible score on the test was 86, suggesting that teachers' mean total was about 60%, STEM majors scored around 50% and PSTs went from 41% on the pre- to 70% total score on the post.

The one-way ANOVA and the Welch's Test showed that the three groups: teachers, STEM majors and PST-pre are statistically different from each other ($p = 0.0219$). The difference was due to teachers scoring significantly higher than PSTs-pre ($p = 0.0379$). The differences between STEM majors and PSTs-pre or between teachers and STEM majors were not statistically significant. However, it is notable that while the maximum total score in the STEM majors' group was 66, four teachers had a total score above 75, meaning that the lack of significance can be due to the high variability of performance in the teachers' group.

Table 1: Overall MKT-P performance of the three groups

Group	No of participants	Mean Total Score	SD
Teachers	17	51.4	21.11
STEM majors	22	42.5	13.12
PSTs **pre**	9	35.4	7.59
PSTs **post**	9	59.7	11.23

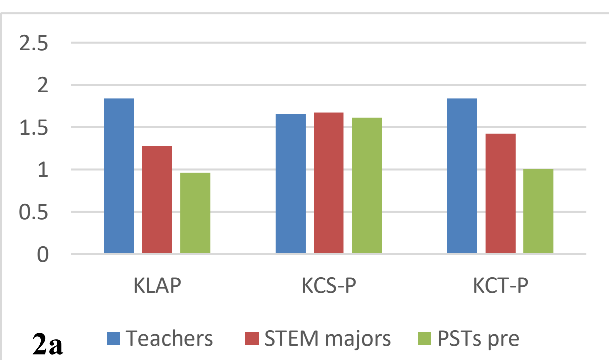
When comparing the mean scores of teachers, STEM majors and PSTs-post, the differences were still significant ($p = 0.0053$), but in this case, the difference was due to PSTs-post scoring higher than the other two groups. In particular, PSTs-post significantly outperformed STEM majors ($p = 0.0274$), but not the teachers ($p = 0.4319$).

The pairwise *t*-test comparing PSTs' pre and post-course performance revealed significant growth in overall MKT-P ($p < 0.0001$); the 95% confidence interval showing the average increase between 16 and 29 points. This outcome supports our assumption that MKT-P can be improved by targeted intervention, such as our capstone course, so much so that PST-post outperformed both the teachers and STEM majors.

Table 2 shows the results of the analysis broken down by MKT-P subdomains: KLAP, KCS-P and KCT-P. In this table, we calculated the mean scores for each domain, rather than total points, since the number of items (and points) in each domain was different.

Table 2: Performance of the groups by MKT-P subdomain

		KLAP		KCS-P		KCT-P	
Group	No	Mean Score	SD	Mean Score	SD	Mean Score	SD
Teachers	17	1.84	1.26	1.66	1.09	1.84	1.29
STEM majors	22	1.28	1.29	1.67	1.09	1.42	1.23
PSTs **pre**	9	0.96	1.17	1.61	1.01	1.01	1.12
PSTs **post**	9	2.15	1.15	2.04	1.03	1.97	1.22



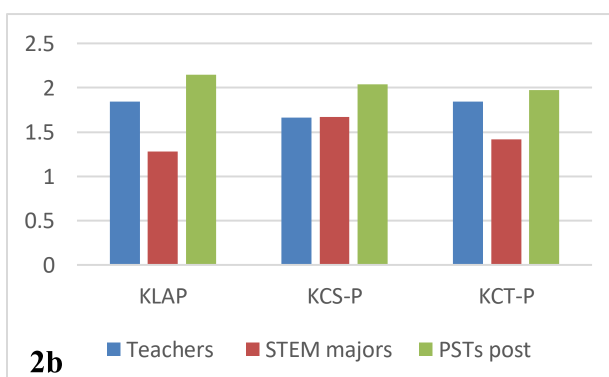


Figure 2 a & b: Performance of the groups by MKT-P subdomain

Figure 2 shows the same information as Table 1, but in a graphic format: Figure 2a (left) compares performance of teachers, STEM majors and PSTs-pre. Figure 2b (right) compares teachers, STEM majors and PSTs-post.

When comparing teachers, STEM majors and PSTs-pre, the analysis showed that there three groups differed significantly on KLAP ($p < 0.0001$) and KCT-P ($p < 0.0001$), but not on KCS-P ($p = 0.8843$). The KCS-P portion of the questionnaire intended to assess participants' ability to identify proof-related misconceptions. All three groups performed very similarly. We do not have an explanation for that, except that our KCS-P items probably measure mostly mathematical knowledge, despite their pedagogical framing. The PSTs performance on the KCS-P items improved significantly from pre to post ($p = 0.0013$), and was significantly higher than of teachers ($p = 0.0162$) and of STEM majors ($p = 0.0126$).

Considering the KLAP portion of the questionnaire, the teachers significantly outperformed both STEM majors ($p < 0.0001$) and PSTs-pre ($p < 0.0001$). This result is interesting since KLAP items measure pure mathematical knowledge. A closer analysis revealed that teachers were better than other groups at identifying logical forms such as converse and contrapositive and tended to use proper mathematical vocabulary. The PSTs' KLAP performance improved significantly on the post-questionnaire ($p < 0.0001$). The PSTs-post scored significantly higher than STEM majors ($p < 0.0001$) but not significantly higher than teachers ($p = 0.1343$).

A similar tendency was observed with respect to KCT-P portion of the test – items that called for identifying logical errors in student arguments and providing instructional feedback to the students. Not surprisingly, teachers significantly outperformed STEM majors ($p =0.0090$) and PSTs-pre ($p < 0.0001$). But when compared to PSTs-post, the PSTs closed the gap and scored very similar to the teachers, and significantly higher than STEM majors ($p = 0.0050$).

Qualitative Differences Between the Groups

The differences between the groups also had a qualitative nature, as revealed in the analysis of written feedback to hypothetical students' arguments on the KCT-P items. STEM majors tended to use third person language talking about the student work rather than addressing the student directly (contrary to the task requirements). STEM majors tended to compliment student work for brevity or clarity, focusing more on the presentation rather than on the content of the argument, e.g. "clear and appropriate assumptions, well ordered." Positive appraisals often merely reiterated the student's approach, e.g. "Anthony was smart in using variable *a* and *b* to help prove the conjecture." Despite praising the student, this comment shows neither analysis of nor engagement with the student's proof strategy.

In their critiques, STEM majors tended to point out that a student's argument did not constitute a mathematical proof but without clarifying the concern, e.g., "not proven enough," "isn't concrete enough to prove the statement." More substantive critiques referred to incorrect assumptions, e.g., "the student assumed their conclusion by saying the sum of two fractions is a fraction," and lack of generality, e.g. "there is no generality, it is only examples." Overall, STEM majors tended to focus feedback on the mathematical validity and form of a student's argument.

Alternatively, participants in the teacher group tended to speak directly to hypothetical students and focus their comments on students' conceptual understanding of the given problem. For example, "you show your strong understanding of what a rational number is and how to use variables to generalize a situation." Teachers' critiques of student arguments tended to focus less on the form and more on the mathematical validity of the arguments. Moreover, the critiques were often phrased as open questions, e.g. "can we look at this algebraically?" or "would your proof hold true if r and s were equal to different fractions?" or "is there a way to show this is true for all real numbers?" This rhetorical style of feedback shows teachers' concern for student understanding and engaging students in revisionary work.

The PSTs comments fell between the student-oriented feedback of the teachers and the mathematics-oriented feedback of the STEM majors. Some PSTs worded their feedback in the question format e.g., "How can you say that only numbers that satisfy Sam's conjecture are 2 and 0?" But the majority of PSTs used third person language and made mathematics-oriented comments, e.g., "Anthony made a valid argument by turning the numbers into a general expression." There were shifts towards more frequent use of first person language and question-posing feedback from pre- to post-questionnaire.

Discussion

The objectives of our study were to examine whether the Mathematical Knowledge for Teaching Proof, as measured by our MKT-P questionnaire, differs from pure knowledge of mathematical content. We hypothesized that if this knowledge is special to mathematics teachers, it would show as better performance on the MKT-P questionnaire when compared to STEM majors or PSTs. We also conjectured that it would be possible to facilitate MKT-P growth through a targeted intervention such as our capstone course *Mathematical Reasoning and Proving for Secondary Teachers* (Buchbinder & McCrone, 2020), which would be visible in improved PSTs' performance on the MKT-P questionnaire.

The data presented above supports both of our assumptions. The teachers outperformed STEM majors and PSTs-pre on the overall MKT-P, and on two MKT-P subdomains: knowledge of the logical aspects of proof (KLAP) and knowledge of content and students (KCT-P). The three groups performed similarly on the KCS-P portion of the questionnaire - items intended to assess knowledge of students' proof-related (mis)conceptions. This may be reflective of a limitation of our instrument, which did not discriminate between the different groups.

The fact that teachers outperformed PSTs-pre on almost every measure is not surprising; it is consistent with the general MKT literature (e.g., Phelps, Howell, & Liu, 2020). Our study adds to this literature by showing that the differences between prospective and practicing teachers appear also in MKT-P. The teachers also scored higher than STEM majors, whose knowledge of proof was fresh in their minds due to their enrollment in a proof course at the time of the study. This outcome may support our assumption that MKT-P is a special kind of knowledge, beyond mathematical content knowledge. Alternatively, this difference can be due to self-selection bias of the participants in the two groups of STEM majors and in-service teachers. Our study design does not allow distinguishing between these alternatives. Future studies should explore this issue.

Another support for our hypothesis about the special nature of MKT-P comes from the qualitative analysis of the feedback provided by the participants on sample student arguments. Particularly striking were the differences between teachers and STEM majors, while PSTs were somewhere between those two groups. The teachers' comments were characterized by a tendency to use first person language addressing the student directly, deeper engagement with a student's argument, attempts to gauge and advance student understanding through guiding questions and suggestions for revisions. On the contrary, the STEM majors' comments were characterized by the tendency to use third person language, focus on the form of the argument rather than its logical structure, critiquing student work for the lack of mathematical rigor but without explaining insufficiencies in student work. Thus, teachers' MKT-P is evident in their ability to provide feedback of higher potential for educative impact than STEM majors (Hattie, & Timperley, 2007).

Our second research question was whether MKT-P can be enhanced through intervention. Note that exploring *how* PSTs' MKT-P evolves throughout the capstone course, connecting the learning processes to the design features of the course and examining factors that promote or inhibit MKT-P development were the core objectives of our three-year long study. Presentation of these findings is beyond the scope of this paper. The significance of this paper is in comparing PSTs' pre- and post-course performance with other groups who may have similar characteristics to our PSTs. We do not see STEM majors or teachers as control groups in any sense. Comparing the MKT-P performance across all groups allows putting the observed changes in the PSTs' MKT-P into broader perspective, adding methodological strength to the simple pre-post design.

The data presented above show that STEM majors performed slightly better than PSTs-pre, although the differences were not statistically significant on any measure. Despite the fact that all PSTs had successfully passed the *Mathematical Proof* course in the second or third year of their program, prior to taking the capstone course, the proof-specific mathematical content was fresher in the minds of the STEM majors than of the PSTs. The course *Mathematical Reasoning and Proving for Secondary Teachers* provided the PSTs with opportunities to refresh and strengthen their proof-related content knowledge. More importantly, the course activities challenged the PSTs to connect this knowledge to teaching secondary mathematics by analyzing sample student arguments, providing feedback on hypothetical student work, planning proof-oriented tasks, enacting them in real classrooms and reflecting on their teaching. These types of activities help to bridge the gap between university-level mathematics preparation and the practice of teaching secondary mathematics (Grossman et al., 2009; Wasserman et al., 2018).

Our study concurs with that literature. After participating in the capstone course, the PSTs' MKT-P improved significantly both overall and in each subdomain. The PSTs closed the gap with in-service teachers on the overall MKT-P, KLAP and KCT-P, and scored significantly higher than the teachers did on KCS-P. The PSTs-post also performed significantly higher than STEM majors did on the overall MKT-P and on each of the MKT-P subdomains. Overall, these results support our hypothesis that MKT-P can be enhanced through intervention.

Our study is exploratory, small scaled and localized. In our data analysis, we utilized statistical techniques that are robust to small numbers of participants (see methods section). However, we make no claims to generality and the results should be interpreted as preliminary. Nevertheless, our study makes several notable contributions to the existing body of knowledge. We proposed an MKT-P framework and a questionnaire for assessing MKT-P at the secondary level, which spans four proof themes - key areas of difficulty with reasoning and proof, according to the research literature. The comparison of the MKT-P performance of in-service teachers, STEM majors and PSTs suggests that MKT-P, as measured by our instrument, is indeed a type of knowledge that is special to teachers, as opposed to other groups with presumably similar mathematical content knowledge. Finally, our study has shown that MKT-P can be enhanced by targeted intervention, such as our capstone course. This was evident in the significant improvement of PSTs' performance from pre- to post, and in comparison of PSTs-post scores with STEM majors and teachers. It would be important to replicate this study on a larger scale and with other, more diverse, populations.

Acknowledgments

This research was supported by the National Science Foundation, Award No. 1711163. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Antonini, S., & Mariotti, M. A. (2006). Reasoning in an absurd world: difficulties with proof by contradiction. In *Proceedings of the 30th PME Conference, Prague, Czech Republic* (Vol. 2, pp. 65-72).

Baldinger, E. E., & Lai, Y. (2019). Pedagogical context and proof validation: The role of positioning as a teacher or student. *The Journal of Mathematical Behavior*, *55*, 100698.

Bieda, K. N. (2010). Enacting proof-related tasks in middle school mathematics: Challenges and opportunities. *Journal for Research in Mathematics Education*, *41*(4), 351-382.

Buchbinder, O. (2010). The role of examples in establishing the validity of universal and existential mathematical statements. Unpublished doctoral dissertation (in Hebrew). Haifa: Technion

Buchbinder, O., & McCrone, S. (2020). Preservice teachers learning to teach proof through classroom implementation: Successes and challenges. *The Journal of Mathematical Behavior*, *58*, 100779.

Cirillo, M. (2011). "I'm like the Sherpa guide": On learning on teach proof in school mathematics. *Proceedings of PME 35*, 2, 241–248.

Corleis, A., Schwarz, B., Kaiser, G., & Leung, I. K. (2008). Content and pedagogical content knowledge in argumentation and proof of future teachers: A comparative case study in Germany and Hong Kong. *ZDM - The International Journal on Mathematics Education, 40*(5), 813–832.

Edelson, D. C. (2002). Design research: What we learn when we engage in design. Journal of the Learning Sciences, 11(1), 105-121.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers college record*, *111*(9), 2055-2100.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

JMP®, Version 15.0.0. SAS Institute Inc., Cary, NC, 1989-2019.

Lesseig, K. (2016). Investigating Mathematical Knowledge for Teaching Proof in Professional Development. International Journal of Research in Education and Science, 2(2), 253-270.

Lin, F. L., Yang, K. L., Lo, J. J., Tsamir, P., Tirosh, D., & Stylianides, G. (2011). Teachers' professional learning of teaching proof and proving. In *Proof and proving in mathematics education* (pp. 327-346). Springer, Dordrecht.

Martin, T. S., McCrone, S. M. S., Bower, M. L. W., & Dindyal, J. (2005). The interplay of teacher and student actions in the teaching and learning of geometric proof. *Educational Studies in Mathematics*, *60*(1), 95-124.

McCrone, S. M. S**.** & Martin, T. S. (2009). Formal proof in high school geometry: Student perceptions of structure, validity and purpose. In D. Stylianou, M. Blanton & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspective* (pp. 204-221). New York, NY: Routledge.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). *Qualitative data analysis: A methods sourcebook*. Sage publications.

Nardi, E., & Knuth, E. (2017). Changing classroom culture, curricula, and instruction for proof and proving: how amenable to scaling up, practicable for curricular integration, and capable of producing long-lasting effects are current interventions?. *Educational Studies in Mathematics*, *96*(2), 267-274.

Phelps, G., Howell, H., & Liu, S. (2020). Exploring differences in mathematical knowledge for teaching for prospective and practicing teachers. *ZDM - Mathematics Education*, 52, 255–268.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, *15*(2), 4-14.

Stylianides, A. J. (2011). Towards a comprehensive knowledge package for teaching proof: A focus on the misconception that empirical arguments are proofs. *Pythagoras*, *32*(1), 1-10.

Stylianides, A. J., Bieda, K. N., & Morselli, F. (2016). Proof and argumentation in mathematics education research. In A. Gutierez, G. Leder, & P. Boero (Eds.), 2nd *handbook of research on the psychology of mathematics education* (pp. 315-351). Rotterdam: Sense Publications.

Stylianides, A. J., & Stylianides, G. J. (2018). Addressing key and persistent problems of students' learning: The case of proof. In *Advances in mathematics education research on proof and proving* (pp. 99-113). Springer, Cham.

Stylianides, G.J., Stylianides, A.J. (2010). Mathematics for teaching: A form of applied mathematics. *Teaching and Teacher Education*. *26*(2) 161-172.

Stylianides, G. J., & Stylianides, A. J. (2017). Research-based interventions in the area of proof: the past, the present, and the future. *Educational Studies in Mathematics*, *96*(2), 119-127.

Wasserman, N., Weber, K., Villanueva, M., & Mejia-Ramos, J.P. (2018). Mathematics teachers' views of the limited utility of real analysis: A transport model hypothesis. *The Journal of Mathematical Behavior*, *50*, 74-89.

Yin, R. K. (2015). *Qualitative research from start to finish*. Guilford publications.