Mathematical Knowledge for Teaching Proof: Comparing Secondary Teachers, Pre-Service Secondary Teachers, and Undergraduate Majors

Orly Buchbinder[1] · Sharon McCrone[1] · Michelle Capozzoli[1] · Rebecca Butler[1]

Abstract

The calls for reasoning and proof to have a prominent place in secondary mathematics classrooms raise the question of knowledge that teachers need to support student learning of reasoning and proof. Several conceptualizations of Mathematical Knowledge for Teaching Proof (MKT-P) have been proposed over the years, yet some key questions about the nature of MKT-P remain unclear, such as whether MKT-P is a special kind of knowledge specific to teaching or whether it is just common mathematics knowledge. Another question is whether MKT-P can be improved through targeted intervention. Our study attempts to respond to both questions. An MKT-P questionnaire was administered to in-service secondary teachers, undergraduate mathematics and computer science (M&CS) majors, and pre-service secondary teachers before and after taking a capstone course Mathematical Reasoning and Proving for Secondary Teachers. *The results suggest that MKT-P is indeed a specialized type of knowledge, which can be improved through targeted intervention.*

Keywords Reasoning and proof · Pre-service and In-service secondary Teachers · Mathematical Knowledge for Teaching · Proof conceptions · Undergraduate students

Introduction: Background and Motivation for the Study

Student engagement with mathematical reasoning, justification, argumentation and proof across grade levels and mathematical topics is widely recognized as essential for meaningful learning of mathematics and for preparing students for STEM careers

✉ Orly Buchbinder
  Orly.Buchbinder@unh.edu

1 University of New Hampshire, Durham, United States

(NGA & CCSSO, 2010; NCTM, 2009, 2014, 2018; Stylianou et al., 2009). In school mathematics, the term *reasoning and proof* is often construed broadly and includes a variety of proving *processes* such as exploring, observing patterns, conjecturing, generalizing, justifying, constructing arguments, proving, and critiquing arguments of others (Ellis et al., 2012; Hanna & deVillers, 2012; Jeannotte & Kieran, 2017; Stylianides, 2008).

The teaching and learning of reasoning and proof at the secondary level has changed little over recent years and leaves much to be desired (Nardi & Knuth, 2017; Stylianides et al., 2017). This can be attributed partially to curriculum materials that offer limited opportunities for teachers to engage students in proving (Otten et al., 2014; Thompson et al., 2012), partially to teachers' own narrow conceptions of proof (Knuth, 2002; Ko, 2010), and partially to teachers' limited knowledge and skill in supporting student engagement with proof in classrooms (e.g., Martin & McCrone 2003; Otten et al., 2017).

The quality of student engagement with reasoning and proof depends on teachers' ability to create learning environments that provide appropriate opportunities for such engagement (Nardi & Knuth, 2017). Researchers have conjectured that this requires a special type of teacher knowledge - Mathematical Knowledge for Teaching Proof (MKT-P), and have proposed various conceptualizations of MKT-P, delineating its components and designing instruments for capturing MKT-P (e.g., Lesseig 2016; Stylianides, 2011, Steele & Rogers, 2012). Within this body of research, there are some key questions that remain unanswered about the nature of MKT-P. Specifically, attempts to conceptualize and capture MKT-P rely on the assumption that MKT-P is a special type of knowledge, specific to teaching, which differs from other forms of mathematical knowledge. However, to the best of our knowledge, this assumption has not been tested empirically.

The MKT-P literature relies heavily on the larger body of research on general Mathematical Knowledge for Teaching - MKT (e.g., Ball et al., 2008) inspired by the work of Shulman (1986). Over the last few decades, research conducted across many countries accumulated evidence on the nature of teacher professional knowledge, its connection to the quality of classroom teaching and student learning outcomes (Baumert et al., 2010; Blömeke & Kaiser, 2014; Depaepe et al., 2020; Hill et al., 2005, 2008; Krauss et al., 2008). Some studies focused on the ability to improve MKT through interventions. For example, Ko et al., (2017) showed that prospective secondary teachers (PSTs) demonstrated improved MKT-Geometry as a result of their coursework.

To validate the claim that MKT is a special type of knowledge for teaching, some studies compared how different groups, with hypothesized different MKT, perform on the same MKT instrument. Validity is a continuous and iterative process of "constructing and evaluating arguments for and against the intended interpretation of test scores and their relevance to the proposed use" (AERA et al., 2014, p. 11). The validation of the MKT construct by "comparison of contrasting groups" (Krauss et al., 2008) rests on the assumptions that (a) groups with similar mathematical but different pedagogical backgrounds would perform differently on the same MKT test, and (b) more pedagogically experienced participants would show better MKT performance. Indeed, Phelps et al., (2020) found that practicing elementary teachers significantly

outperformed prospective elementary teachers, when controlling for participants' backgrounds. To examine whether knowledge for teaching is separate from content knowledge, Hill et al., (2007) compared 27 elementary-level teachers, 18 non-teachers and 18 mathematicians' performance on items measuring content knowledge and items assessing knowledge of content and students (KCS). The researchers found that many teachers relied on their pedagogical knowledge to respond to KCS items, while other groups were able to respond successfully to these items by relying solely on their content knowledge.

At the secondary level, studies comparing MKT performance across populations are rare. A notable exception is a study by Krauss et al., (2008) of the COACTIV project, who compared subject matter knowledge and pedagogical content knowledge of six groups: 85 mathematics teachers in academically oriented schools, 113 mathematics teachers in non-academically oriented schools, 16 science teachers, 137 university mathematics majors, 30 advanced high-school students, and 90 mathematics PSTs. The study found that mathematics teachers in academic track had stronger content and pedagogical knowledge than PSTs, who in turn, outperformed non-academic track mathematics teachers, school students and science teachers. The undergraduate mathematics majors performed similar to in-service mathematics teachers of academic track on content knowledge and, interestingly, also on the pedagogical content knowledge. The only significant difference between mathematics teachers' and mathematics majors' performance was found on a subset of questions directly related to classroom instruction.

With respect to MKT-P, we are not aware of any study that made such comparisons, nor examined growth of MKT-P over time. Our study attempts to address this gap. The study reported herein is part of a larger, NSF-funded research project, whose goals were to design and study an undergraduate capstone course *Mathematical Reasoning and Proving for Secondary Teachers*. The course activities had PSTs refresh and strengthen their own knowledge of proof, explore student conceptions of proof, then develop and teach lessons in local schools that integrate reasoning and proving within a regular school curriculum. The research goals of the larger project were to study how MKT-P and dispositions towards proof of the PSTs evolve due to their participation in the course (Buchbinder & McCrone, 2020).

In the third year of the project, we conducted an additional study, which is the focus of this paper. In this study, we compared the MKT-P performance of in-service secondary mathematics teachers, PSTs, both before and after taking our capstone course, and undergraduate M&CS majors taking the course *Mathematical Proof*. Methodologically, we would prefer the latter group to comprise only mathematics majors, however, at the institution in which the study was conducted, the *Mathematical Proof* course it taken by a mix of pure mathematics and computer science majors. Also important to note is that at the university of the study, mathematics education and pure mathematics majors take identical mathematics coursework for the first two years, and some overlapping courses (e.g. Geometry and Abstract Algebra) during their second two years. The teacher preparation courses start in year three. Thus, we expected M&CS majors and PSTs to have comparable mathematical background but differ in terms of their pedagogical knowledge. By examining how MKT-P performance compares across these groups, we sought to examine whether MKT-P, as mea-

sured by our instrument, requires a special type of professional knowledge beyond mathematical knowledge of proof and reasoning. Further, we wanted to investigate whether PSTs' MKT-P performance improves following their participation in the specialized capstone course.

In the following, we outline our MKT-P framework and compare it to other existing ones. Then, we describe the MKT-P instrument and present data comparing in-service teachers, M&CS majors, and PSTs before and after the course (hereafter denoted by PSTs-Pre and PSTs-Post, respectively).

Theoretical Background

Prior Research on Mathematical Knowledge for Teaching Proof: MKT-P

Mathematical Knowledge for Teaching (MKT) is a special type of professional knowledge that distinguishes the mathematical work of teaching from that of other professionals (Ball et al., 2008). It has been suggested that MKT is content specific. For example, Phelps et al., (2020) found differences in teacher performance on Number and Operations, Algebra, and Geometry versions of an MKT test even when controlled for years of teaching, socioeconomic status, gender and holding a math major. McCrory et al., (2012) developed a theory of MKT in algebra, while Herbst & Kosko (2014) proposed a conceptualization and assessments of MKT in geometry.

Although proof is a concept that spans mathematical subjects, researchers have suggested that teaching mathematics with an emphasis on reasoning and proof requires a special type of knowledge: mathematical knowledge for teaching proof (MKT-P). Several researchers (e.g., Buchbinder & McCrone 2020; Corleis et al., 2008; Lesseig, 2016; Steele & Rogers, 2012; Stylianides, 2011) proposed frameworks for MKT-P. These frameworks seem to vary by the extent to which they align with either Shulman's (1986) or Ball et al.'s (2008) general MKT frameworks, and by the extent of inclusion (or not) of meta-mathematical domains, such as knowledge of the roles of proof in mathematics, or beliefs about proof.

Some MKT-P frameworks, such as Lesseig's (2016), follow Ball's et al. (2008) division of MKT-P into four[1] domains. Two domains of subject matter knowledge: common content knowledge, with elements such as ability to construct a valid proof; and specialized content knowledge, which attends to multiple roles of proof in mathematics. The two pedagogical content knowledge domains are: knowledge of content and students, which includes elements like knowledge of students' proof schemes (Harel & Sowder, 2007); and knowledge of content and teaching, which intertwines knowledge of proof with pedagogical strategies. Speer et al., (2015) raise a critical concern that at the secondary or post-secondary levels the common content knowledge and specialized content knowledge are too close to each other to be reliably distinguishable. We agree with this assumption, specifically with respect to proof-related knowledge. Thus, instead of following Lesseig (2016) or Ball et al.'s (2008)

---

[1] Excluding Horizon Content Knowldge and Knowledge of Content and Curriculum.

MKT framework, we gravitated towards a more general model - such as Stylianides' (2011) "comprehensive knowledge package for teaching proof."

Stylianides' model (2011) contains three dimensions: mathematical knowledge about proof, knowledge about students' conceptions of proof, and pedagogical knowledge for teaching proof which describes practical knowledge teachers need to successfully implement proof-related activities in their classrooms. Since Stylianides' framework is oriented towards elementary teachers, we extended it to include aspects of knowledge that are appropriate and common in secondary education. Building on the literature on teachers' conceptions of proof (e.g. Knuth 2002; Ko, 2010; Lin et al., 2011), we included in our MKT-P framework topics that prospective and even practicing secondary teachers find challenging. For example, understanding the relationship between empirical and deductive reasoning (e.g., Weber 2010); reasoning with conditional statements and logical connectors (e.g., Dawkins & Cook 2016; Durand-Guerrier, 2003); analyzing and evaluating mathematical arguments (Hodds et al., 2014; Selden & Selden, 2003); and indirect proof (Antonini & Mariotti, 2008; Thompson, 1996).

Another relevant framework was developed by the *Mathematics Teaching in the 21st Century* (MT21) project (Schmidt, 2013). Their framework for teacher professional knowledge contained four facets: mathematical content knowledge, pedagogical content knowledge, general pedagogical knowledge (e.g., classroom management) and beliefs about mathematics and its teaching. Although originally designed as a general MKT framework, several studies have used it specifically to examine teacher knowledge of argumentation and proof (e.g., Corleis et al., 2008; Schwarz et al., 2008). For example, Corleis et al., (2008) found that prospective teachers in Hong Kong had stronger content knowledge of argumentation and proof, while prospective teachers in Germany performed better on items assessing pedagogical, and pedagogical content knowledge related to proof. Since we view beliefs and dispositions towards proof as an affective domain of teacher expertise, we refrained from adopting a framework that counted beliefs within MKT-P. Rather, we conceptualize the affective domain, dispositions towards proof, and cognitive domain, MKT-P, as separate but mutually impacting dimensions of expertise for teaching reasoning and proof.

Steele and Rogers' (2012) MKT-P framework focuses on the subject matter knowledge of proof and roles of proof in mathematics, and pedagogical content knowledge (PCK). Although Steele and Rogers do not elaborate on the components of proof-specific PCK, they capture it by examining teachers' proof-related classroom practices.

Depaepe et al., (2020) advocate for the need to develop frameworks that integrate cognitive, affective and practice domains of teacher expertise. Following this suggestion, our resulting framework for *expertise for teaching reasoning and proof* comprises a triad of interrelated domains: MKT-P, proof-related classroom practices and dispositions towards proof. It is beyond the scope and the aims of this paper to present the complete framework, and we report on this elsewhere (Buchbinder & McCrone, 2020, 2021). Here, we focus on the cognitive dimension of this expertise - the MKT-P, which we view as a declarative knowledge that can be captured in a written assessment. Narrowing the focus this way provides a common ground for comparing across

different groups of participants: practicing teachers, PSTs, and M&CS majors, who do not teach in classrooms. Below, we describe our MKT-P framework.

Our MKT-P Framework

Our MKT-P framework is comprised of three interrelated facets. One is the subject matter knowledge specific to proof, which we term Knowledge of the Logical Aspects of Proof (KLAP) and two facets of Pedagogical Content Knowledge specific to proof: Knowledge of Content and Students (KCS-P), and Knowledge of Content and Teaching (KCT-P). It is not our intention to describe the development of the framework here, but rather to provide descriptions of framework components that were used to structure our research.

The KLAP includes knowledge of different types of proofs (e.g., proof induction, by contradiction, direct proof), valid and invalid modes of reasoning, the roles of examples in proving; logical relations (e.g., implication, biconditional, converse), knowledge of a range of definitions of mathematical objects (e.g., various definitions of a trapezoid), and the range of theorems (e.g., various proofs of the Pythagorean Theorem).

Of the two pedagogical domains, Knowledge of Content and Students (KCS-P) refers to knowledge of common student conceptions and misconceptions related to proof, such as student tendency to rely on empirical evidence, or to treat a counterexample as an exception (e.g., Buchbinder 2010; Reid, 2002). Knowledge of Content and Teaching proof (KCT-P) relates to pedagogical practices for supporting students' engagement with proof. This includes knowledge of instructional activities, questioning techniques, providing feedback on student work and teaching strategies that help to advance students' naïve conceptions towards conventional mathematical knowledge of proof. This domain of MKT-P is challenging to capture in a written format, due to its intricate connection to classroom teaching. In our MKT-P questionnaire, we operationalized the KCT-P facet through a single aspect of providing instructional feedback on student mathematical work.

The three domains of MKT-P are interrelated even when considering declarative knowledge. For example, providing instructional feedback on student's work (KCT-P) requires identifying misconceptions underlying that student's work (KCS-P), which in turn, requires strong mathematical knowledge (KLAP). Nevertheless, the separation of the three domains was useful in the development of the MKT-P questionnaire and in designing the course *Mathematical Reasoning and Proving for Secondary Teachers* (Buchbinder & McCrone, 2020).

The Course *Mathematical Reasoning and Proving for Secondary Teachers*

As mentioned above, this study grew out of a larger, 3-year NSF-funded project, during which we developed and studied the course *Mathematical Reasoning and Proving for Secondary Teachers* (Buchbinder & McCrone, 2018, 2021). The course

comprised four instructional modules corresponding to teachers' difficulties with reasoning and proving identified in the literature (e.g., Antonini & Mariotti 2008; Dawkins & Cook, 2016; Durand-Guerrier, 2003; Hodds et al., 2014; Selden & Selden, 2003; Thompson, 1996; Weber, 2010). The four modules are: (1) direct reasoning and argument evaluation, (2) conditional statements, (3) quantification and the role of examples in proving, and (4) indirect reasoning. Although these are not the only areas of difficulties reported in the literature, we felt that they are most relevant to the secondary curriculum and resonate with our own teaching experience at the secondary and tertiary levels.

The key course objective was to bridge between university experiences of mathematics education majors and the realm of classroom teaching (c.f, Wasserman et al., 2019), with a specific focus on reasoning and proving. Thus, the course activities aimed to help PSTs enhance their content and pedagogical knowledge of reasoning and proving, connect it to the secondary school curriculum and apply this knowledge by designing and teaching four proof-oriented lessons in local schools. These activities were inspired by our framework for expertise in teaching reasoning and proving. They also reflect recommendations for teacher preparation, such as deepening both content and pedagogical knowledge, active learning, and making direct connections to teachers' classroom practices (Association of Mathematics Teacher Educators, 2017; Conference Board of the Mathematical Sciences, 2012).

The MKT-P questionnaire was developed as an assessment for the main study and used to capture changes in the PSTs' written performance before and after the course. Due to the design-based and exploratory nature of our study, formal instrument validation was not part of our research objectives. Nevertheless, after using and gradually improving the MKT-P questionnaire over two years, we were curious whether it indeed captured knowledge that is special to mathematics teaching. Hence, we devised the current study to examine the following research questions:

1. How do practicing teachers, M&CS majors and prospective teachers compare on their performance on our MKT-P questionnaire?
2. How does the prospective teachers' MKT-P performance change (if at all) following their participation in the course *Mathematical Reasoning and Proving for Secondary Teachers?*

Methods

Participants and Procedures

The study involved three groups of participants. Group one included nine PSTs taking the course *Mathematical Reasoning and Proof for Secondary Teachers*, taught by the first author of this paper. Of the nine PSTs, five were pursuing high-school certification, and four middle-school certification; eight females and one male. All PSTs except one were in the final year of their program. All had successfully completed a course on *Mathematical Proof* and had taken a proof-intensive *Geometry* course, either prior to or concurrently with our course. The PSTs had also taken two educa-

tional courses: an introduction to mathematics education, and methods for teaching mathematics in grades 6–12. The PSTs had no classroom teaching experience prior to this course.

The PSTs completed the same MKT-P questionnaire twice. It was first administered during the first week of the course – this questionnaire was not graded and not returned to the PSTs. The second time, the questionnaire was given as a final exam, one week after the course completion. The questionnaire took between 1.5 and 2 hours to complete in a paper and pencil format. The differences in stakes inevitably affected the data and should be taken into consideration in interpreting the results.

The second group of participants were 17 practicing secondary teachers, 11 female and 6 males, whose teaching experience ranged from two to 25 years ($\bar{x}$ = 12.18, SD = 8.26). The teachers were recruited through in-person presentations at local schools or professional development conferences. Five of the 17 teachers were from the same school, but the rest came from different schools and districts around the region. The teachers taught a variety of courses, 40 in total, such as Algebra 1, Algebra 2, Financial Mathematics, Statistics, Geometry, Pre-calculus and Calculus. Five out of 17 teachers taught Geometry. The MKT-P questionnaire was administered online, via Qualtrics. The teachers were given a two-week window to complete the questionnaire and received a small honorarium.

The third group of participants were 22 undergraduates: mathematics and computer science majors taking a mandatory course *Mathematical Proof.* This is an introduction to proof course, usually taken by mathematics and computer science students in the second year of their studies, after the completion of Calculus I and II courses. The course had three sections, each taught by an experienced mathematics professor. The professors agreed to provide students with an extra credit for participating in the study. Twenty-two students volunteered, 4 females, 18 males. Twelve participants were in their second year of studies, eight in the third and two in the last, fourth year. The breakout of the group by major was: 11 computer science, nine mathematics, one mathematics education[2] and one philosophy major. The participants completed a paper and pencil version of the MKT-P questionnaire outside of their course class time; taking about 1.5 hours to finish. The authors and graduate assistants, who were not familiar with the participants prior to the study, administered the questionnaire. The data collection occurred in the final weeks of the *Mathematical Proof* course. Before administering the questionnaire, the authors met with the course instructors to confirm that the instructors had taught all the proof topics included in it, so students should be familiar with the content. Due to ethical considerations, we could not collect any additional data about this group, such as, what motivated them to volunteer for our study. Some students may pursue an extra credit opportunity to boost a low grade; others may wish to protect and solidify an existing good grade. The results of our study should be interpreted considering both possibilities. In addition, we note that due to this self-selection bias, we are not claiming that the outcomes of our study are representative of all the M&CS majors in the proof course.

---

[2] At the time of the study, this participant has not taken any educational coursework but only mathematics courses together with other pure mathematics majors. Hence, this participant's data were left in the sample.

Table 1 Distribution of number of items across MKT-P facets and proof themes

	KLAP: Knowledge of Logical Aspects of Proof	KCS-P: Knowledge of Content and Students	KCT-P: Knowledge of Content and Teaching	Total
Direct Proof	2	5	6	**13**
Conditional Statements	5	0	0	**5**
Role of Examples	1	6	2	**9**
Indirect Proof	2	0	0	**2**
Total	**10**	**11**	**8**	**29**

MKT-P Questionnaire Design

As a part of the larger study, we conducted a thorough review of literature on existing tools for measuring secondary teachers' conceptions of proof. In particular, we looked for instruments that focus on content and pedagogy specific to proof, rather than surveys of beliefs about proof (e.g., Kotelawala 2016) or general secondary MKT instruments (e.g., Howell et al., 2016). We found very few instruments devoted specifically to MKT for proof (e.g., Lesseig et al., 2019) and none that addressed all four proof themes central to our study. Hence, we developed our own questionnaire using a combination of original items and items adapted from the literature (e.g., Lesseig et al., 2019; Riley, 2003).

The resulting MKT-P questionnaire contained 29 items, some having a common stem (Table 1). The items were distributed among the three facets of MKT-P: KLAP, KCS-P and KCT-P, as well as across the four proof themes, which are the foci of the four modules of the course: (1) direct proof and argument evaluation, (2) conditional statements, (3) quantification and the role of examples in proving, and (4) indirect reasoning. The distribution of items was relatively close between the MKT-P facets, but to a lesser degree among the four proof themes; most items fell under *direct proof* and the *role of examples* categories. The items spanned mathematical topics such as: number and operations (10 items), geometry (7), algebra and functions (7), and five general items (see an example in Fig. 1). We also took into consideration levels of difficulty of the questions and the overall length of the questionnaire, to minimize the effect of test taking fatigue on the quality of data.

The resulting MKT-P questionnaire was validated by three mathematics professors with expertise in teaching proof courses and one mathematics education professor whose area of research is in undergraduate education and proof. We piloted the questionnaire in the course *Mathematical Reasoning and Proof for Secondary Teachers* for two years, and gradually improved it by fine tuning the wording of the questions and replacing items that were too easy or too difficult. The MKT-P questionnaire used in the study and described in this paper is the third version of the instrument. The internal consistency of the entire questionnaire measured by Cronbach alpha is 0.8918; the Cronbach alphas for each domain are: 0.8070 for KLAP, 0.7098 for KCS-P and 0.7591 for KCT-P[3]. Below, we illustrate the types of items in each MKT-P facet and describe the scoring procedures.

---

[3] Cronbach alpha values above 0.7 are considered acceptable, above 0.8 good, and above 0.9 excellent.

1. A theorem in geometry has been proved. Which of the following statements is correct?

   a) The converse of the theorem must be true and does <u>not</u> need further proof.

   b) The converse of theorem must be true, but it needs further proof.

   c) The converse of the theorem must be false and does <u>not</u> need further disproof.

   d) The converse of theorem must be false, but this needs a disproof.

   e) None of the above.

Explanation:

Fig. 1 Sample KLAP item

Data Analysis Methods

Scoring Procedures

Teacher data was downloaded from Qualtrics in the form of an Excel spreadsheet; PSTs' and M&CS majors' hand-written answers were digitized and added to the same spreadsheet. The first three authors of this paper individually scored about 20% of the questionnaires across all groups of participants and met regularly to discuss and fine-tune the scoring system. The responses were assessed for mathematical correctness and quality of explanations (see examples below).

Figure 1 shows a sample KLAP item. The item assesses knowledge of the relationship between truth-values of a statement and its converse, corresponding to the Conditional Statements proof theme.

The KLAP items were scored on a 3-point scale: 0–1 points for choosing the answer, and 0–2 points for the explanation. For example, a participant teacher (T18) chose a distractor (b) "The converse of theorem must be true but it needs further proof" and justified their answer by writing "If p then q has the converse if q then p. The original If p then q is not a biconditional, so the converse must be proven in order to be shown as true." This response got 0 points for the answer choice but was awarded 1 point for correctly defining a converse in logical notation and knowing that it "requires a separate proof." The following response was scored as 3 out of 3 points: The correct answer is (e) "None of the above," justification: "The converse of a true statement may be either true or false and would have to be proved separately."

The KCS-P and KCT-P items intended to assess proof-related pedagogical knowledge. Hence, the items positioned participants in the role of a teacher in a hypothetical classroom situation (Baldinger & Lai, 2019) and targeted aspects of pedagogical knowledge such as interpreting student mathematical work, identifying proof-related misconceptions (KCS-P) and providing instructional feedback on student work (KCT-P). These items intertwine pedagogical and subject matter knowledge, and were designed to require pedagogical knowledge, such as an ability to analyze students' mathematical work and provide instructional feedback.

The KCS-P items, except one, had a common structure. Given a mathematical statement and a sample student's argument showing the statement is true or false, the participants were asked to evaluate the correctness of the argument on a scale from

9. Mrs. Martinez asked her students whether the following statement is true or false:

| *Any quadrilateral with congruent diagonals has a right angle or two congruent sides.* |

Below are responses of some students.

a) **Angela**: There are not so many quadrilaterals with congruent diagonals, only a rectangle and an isosceles trapezoid. An isosceles trapezoid has one pair of opposite congruent sides, and a rectangle has two such pairs, making this statement true.

i) Assign a score from 1 (lowest) to 4 (highest) to the student's argument.

ii) If you assigned a score of 4, explain why. If you assigned a score less than 4 explain what errors or potential misconceptions, you noticed in the student's solution.

Fig. 2 Sample KCS-P item

one to four, and describe any kinds of misconceptions they noticed in the student's work. Successful completion of this item required interpreting student mathematical thinking and knowledge of student proof-related misconceptions.

For example, in item 9 (Fig. 2), Angela arrived at an incorrect conclusion by limiting the domain of quadrilaterals that satisfy the statement to special quadrilaterals while excluding general quadrilaterals. This misconception corresponds to misunderstanding the domain of a universally quantified statement and the role of counterexamples in disproving such statements.

The items in the KCS-P category were scored on a 3-point scale. One point was given for a correct assessment of student work in part (i). Although the hypothetical student argument in the item is either mathematically correct or incorrect, we allowed the participants to evaluate its correctness on the four-point scale. Prior research (Baldinger & Lai, 2019; Buchbinder, 2018) showed that PSTs draw on a combination of mathematical and pedagogical resources when evaluating student work and shy away from using a dichotomous scale. For the analysis, we collapsed the four options into two to create a dichotomous score. Thus, if the participant correctly identified a logical flaw in the student argument, and assessed it as either 1 or 2, they received a full score on part (i). However, if the participant gave high marks (3 or 4) to a faulty argument, they received a score of zero for their answer. On part (ii) of the item, participants could score 0–2 points. Two points were awarded for a complete correct explanation of the misconceptions in the student's argument; a partial score of one point was given for incomplete or a partially correct explanation. Table 2 shows examples of varied responses to the "Angela" item with respective total scores.

A general KCT-P item contained a mathematical statement and a sample student argument. For example, Molly's argument in Fig. 3 uses the same algebraic notation for two different variables, limiting the generality to a special case. The participants had to identify any mistakes in the argument and provide feedback to the fictitious student highlighting strengths and weaknesses of the solution.

The KCT-P items were scored on a three-point scale: 1 point for identifying student errors (if any) in part (i), and 0–2 points for providing instructional feedback in part (ii). Zero points were given to feedback that did not explain what was problematic in the student answer, or simply offered a different solution path; partial credit of one point was given for feedback that did not sufficiently engage with student think-

14. Mr. Briggs asked his students to prove the following statement:

> ***The sum of any two rational numbers is a rational number.***

Below are solutions made by three students in Mr. Briggs' class.

a) **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.

Fig. 3 Sample KCT-P item

ing, and two points were given for any actionable feedback that highlighted strengths and weaknesses of the student's argument. Table 2 shows examples of responses to the "Molly" item, with the respective total scores.

As we coded the data, we were surprised to encounter a small set of 40 responses, which showed an exceptional depth of engagement with student thinking, and strong pedagogical orientation. For example, consider the following response to KCT-P item "Molly":

> Molly assumed that r and s where the same rational numbers. Very clear proof. You show your strong understanding of what a rational number is and how to use variables to generalize a situation. Your reasoning of why p is an integer is strong. Would your proof hold true if r and s were equal to different fractions? What if $s=x/y$? Could you still show that your argument is true for all values of s and r?

This response is arguably better than the sample response in Table 2, coded as 3 (i.e., a complete correct response) exceeding our own highest expectations. To capture this phenomenon, we assigned such exceptional responses one additional bonus point above 3. There were only 40 responses, which is less than 4% of all KCS-P and KCT-P data points.

Once the codebook was finalized, team members coded the data individually. Each week, we coded a portion of data from all participant groups and met to compare and discuss the scores, to ensure consistent application of the scoring system. In total, we coded 1653 responses. The initial agreement between the coders was 91% on KLAP items, 82% on KCS-P items and 84% on KCT-P items. The interrater reliability measured by Cohen's Kappa were 0.89 for KLAP; 0.78 for KSC-P and 0.80 for KCT-P. All discrepancies between the coders were discussed until fully reconciled.

Table 2 Coding rubric for KSC-P item "Angela" (Fig. 2) and KCT-P item "Molly" (Fig. 3)

KSC-P item "Angela"	KCT-P item "Molly"
Score 0: Response absent or mathematically incorrect Example: Student is correct (4 points) and explains reasoning well.	Score 0: Response absent or mathematically incorrect Example: No errors. Well formulated. Good definitions. Clear and concise.
Score 1: A participant gave 3 or 4 points for student work, and either did not notice the misconception, or gave incorrect alternative. Example: There are other quadrilaterals that fit the statement like parallelogram. However, a rectangle and isosceles trapezoid do fit the criteria.	Score 1: A participant noticed the error but did not communicate this clearly in the feedback to the student. Example: No errors. Great job but only proves for $r = s$.
Score 2: A participant gave 1 or 2 points for student work. Possibly noticed the misconception but explained it vaguely. Example: Using examples to prove general statements.	Score 2: A participants notices the error. Does not explain why student work is wrong but tells the student how to fix it. Example: Should not use a/b as both fractions. Use c/d for second fraction.
Score 3: A participant gave 1 or 2 points for student work. Noticed and correctly explained the misconception. Example: Angela only thought of the two most common examples of convex quadrilaterals with congruent diagonals. She did not try all examples or construct a proof of some kind.	Score 3: A participant noticed the error, explained it and provided actionable feedback. Example: You didn't prove it for all rational numbers because you had r equal to s. This simply meant adding a fraction to itself is rational. r and s are not necessarily equal, so $r = a/b, s = c/d$

Next, we imported the scores into JMP® Pro statistical software version 15.0.0 for analysis. We used one-way ANOVA to explore whether there are differences between the groups of participants on the overall MKT-P performance, and on each MKT-P domain: KLAP, KCS-P and KCT-P. One-way ANOVA assumes equal variances among the groups; however, this assumption was often violated in our data, mostly due to the high variability in the teachers' group. To compensate, we conducted the Welch's test. Since the PSTs completed the questionnaire twice, before and after taking the course *Mathematical Reasoning and Proof for Secondary Teachers*, their pre- and post-scores were dependent on each other. Thus, we conducted the analysis in two rounds: once comparing M&CS majors, Teachers and PSTs-pre, and for a second time comparing M&CS majors, Teachers, and PSTs-post. In addition, we used Tukey-Kramer's Honestly Significant Difference Test for pairwise comparisons between groups. Finally, we conducted a matched pairs *t*-test analysis comparing PSTs' performance between pre- and post- questionnaires. These two types of analyses correspond to our two research questions.

Results

Overall MKT-P

Table 3 shows overall MKT-P performance across all groups: M&CS majors, Teachers, PSTs-Pre, and PSTs-Post. Of the 29 items on the MKT-P questionnaire, there were 28 items worth 3 points and one item worth 2, making the maximum possible score of 86, and maximum average item score of 2.97. The last two columns of Table 3 represent three-way comparisons: Teachers, M&CS majors, and PSTs-Pre (one before last column) and Teachers, M&CS majors, and PSTs-Post (last column). The shaded boxes in the last two columns indicate a group not included in the comparison.

Table 3 shows that all groups scored relatively low overall. M&CS majors scored at about 50%, teachers scored at about 60%, and PSTs' scores went from 41% on the pre-test to 69% on the post-test. One-way ANOVA supplemented by the Welch's test, indicated a significant difference at a 5% significance level among the three groups: M&CS majors, Teachers, and PSTs-Pre ($p = 0.0219$). The difference remained significant when comparing M&CS majors, Teachers, and PSTs-Post ($p = 0.0053$). Further, pairwise comparisons using the Tukey-Kramer Honest Significant Difference test showed that the statistical significance among the groups is due to certain pairwise differences, and that these differences shifted between PSTs' pre- and post- scores. Figures 4a and b show box and whisker plots for each group comparing M&CS majors and Teachers compared to PSTs-Pre (4a, left) and to PSTs-Post (4b, right).

The pairwise comparisons of M&CS majors, Teachers and PSTs-Pre showed that the Teachers significantly outperformed the PSTs-Pre ($p = 0.0181$). This is not surprising and is consistent with general MKT literature (e.g., Phelps et al., 2020). The Teachers also scored higher than M&CS majors, although the difference was not statistically significant.

The PSTs-Pre scored lower, although not significantly than M&CS majors. The relationship between the scores of M&CS majors and PSTs-Pre is not easy to interpret, as it is affected by multiple factors. On the one hand, the M&CS majors were currently enrolled in a proof course so the proof-specific subject matter was possibly fresher in their minds than that of the PSTs-Pre, who completed the same proof course at least a year or two prior to taking the capstone course. On the other hand, since completing the proof course, the PSTs had a chance to apply their proof knowledge in courses like Geometry or Abstract Algebra.

The matched pairs t-test, corrected for multiple comparisons, indicated that there is a significant difference between the PSTs' pre and post course performance

Group	No' of participants	Mean	SD	Teachers, M&CS majors and PSTs-Pre	Teachers, M&CS majors and PSTs-Post
PSTs-**Pre**	9	35.38	7.59		
M&CS majors	22	42.54	13.12	*p*-value: 0.0219*	*p*-value: 0.0053*
Teachers	17	51.35	21.11		
PSTs-**Post**	9	59.67	11.24		

Table 3 Overall MKT-P performance across all groups
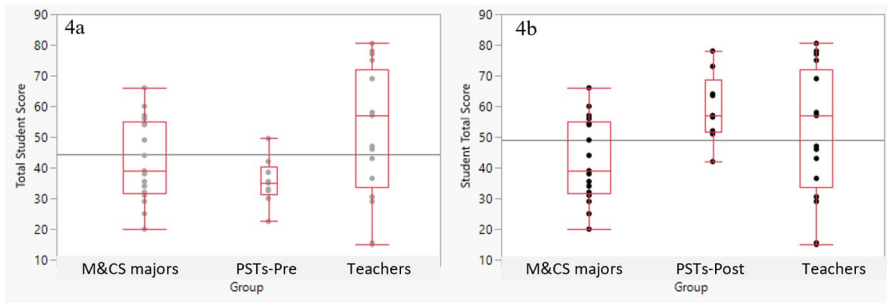
Figs. 4a & 4b Overall MKT-P scores box and whisker plots for M&CS majors and teachers compared to PSTs-Pre (4a, left) and to PSTs-Post (4b, right)

($p < 0.0001$). PSTs-Post significantly outperformed the M&CS majors ($p = 0.0274$), and even scored higher than practicing teachers, although not significantly ($p = 0.4319$) (Table 3; Fig. 4b).

In the following sections, we present the results of comparing participant performance along the three facets of MKT-P: knowledge of the logical aspects of proof (KLAP), knowledge of content and students specific to proof (KCS-P) and knowledge of content and teaching specific to proof (KCT-P).

Knowledge of the Logical Aspects of Proof: KLAP

There were 10 KLAP items, each worth 3 points, making the maximum mean score of 3. Table 4 shows mean and standard deviation of scores for Teachers, M&CS majors, PSTs-Pre and PSTs-Post. The number of data points is the product of the number of items and the group size. Figures 5a and b show box and whisker plots for M&CS majors and Teachers compared to PSTs-Pre (5a, left) and to PSTs-Post (5b, right).

One way ANOVA showed statistically significant differences in performance on KLAP questions between teachers, M&CS majors, and PSTs both on the pre-questionnaire ($p < 0.0001$) and on the post questionnaire ($p < 0.0001$). According to the Tukey-Kramer Honestly Significant test, all groups of participants were statistically different from each other. Teachers significantly outperformed both the M&CS majors ($p < 0.0001$) and the PSTs-Pre ($p < 0.0001$). Teachers tended to score on average 0.88 points higher than PSTs-Pre and 0.56 points higher than M&CS majors. The M&CS majors scored on average 0.32 points higher than PSTs-Pre, which is statistically significant ($p = 0.0409$).

This outcome is interesting because it is reasonable to expect that M&CS majors currently enrolled in a proof course would outperform both the teachers and the PSTs on KLAP, which is a pure mathematical section of the MKT-P instrument. Yet, it was secondary teachers who seemed to have the strongest mathematical knowledge specific to proof among the three groups.

Matched pairs analysis of the difference between pre and post PSTs' KLAP scores revealed statistically significant improvement ($p < 0.0001$). The improvement was apparent in both the rate of correct answers and in the quality of PSTs' explanations.

Group	No' of data points	Mean	SD	Teachers, M&CS majors and PSTs-Pre	Teachers, M&CS majors and PSTs-Post
PSTs-**Pre**	90	0.96	1.17		
M&CS majors	220	1.28	1.29	p-value <.0001*	p-value <.0001*
Teachers	170	1.84	1.26		
PSTs-**Post**	90	2.15	1.16		

Table 4 KLAP performance across all groups
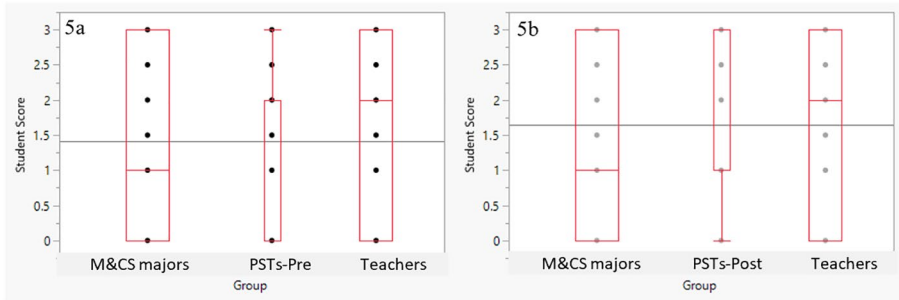
Figs. 5a & 5b KLAP portion of MKT. Box and whisker plots for M&CS majors and teachers compared to PSTs-pre (5a, left) and to PSTs-post (5b, right)

For example, in response to the KLAP item in Fig. 1, one PST on the pre-questionnaire chose incorrect distractor (c) explaining that "The converse is the opposite, so when the theorem is proved you know the opposite can't be true". However, on the post-questionnaire, this PST chose the correct answer (e) and explained that "options a-d state that the converse must be either always true or always false and there are cases where the converse can be true and also cases where it can be false".

When comparing Teachers, M&CS majors and PSTs-Post, the equal variances assumption for ANOVA was violated; there appears to be a lot of variability in Teachers' scores but less variability in the PSTs' scores. To compensate, we conducted Welch's test, which showed that Teachers and PSTs-Post scored similarly ($p=0.1343$) on KLAP items and significantly higher than M&CS majors ($p<0.0001$). One possible interpretation of this result is that applying proof-related subject matter knowledge to teaching, which was the case with practicing teachers and PSTs-Post, strengthens that subject matter knowledge (Leikin & Zazkis, 2010).

To illustrate some of the differences between the participant groups, consider the item in Fig. 1. In the Teacher group, the modal response (12 out of 17) was the correct distractor (e) followed by a correct explanation that the truth value of the converse is independent of the truth-value of the original implication. Five teachers even accompanied their response by examples illustrating this point. The quote below shows a particularly detailed response to this item provided by teacher T6. We bring the quote verbatim including the original text in the parentheses.

The converse of the theorem may be true or false. The converse must either be proven true or proven false using a counterexample. EX: Theorem: If two

angles are vertical angles, then they are congruent. (True) Converse: If two angles are congruent, then they are vertical angles. (False - counterexample could be two 40 degree angles that do not share a vertex.) EX: Theorem: If a triangle is equilateral, then it is equiangular. (True) Converse: If a triangle is equiangular, then it is equilateral. (True - can be proven.)

Although such a detailed response was exceptional, it shows that T6 had a rich knowledge of examples and counterexamples of theorems illustrating the relationship between a statement and a converse, which could be used in an instructional explanation. Other correct explanations included contrasting a converse with a contrapositive, which is equivalent to the original statement, or contrasting a converse with a biconditional, which requires proof of two directions.

The M&CS majors performed poorly on this item, only 5 out of 22 participants responded to it correctly, using justifications that referenced the contrapositive or the biconditional, or simply stating that a converse "does not need to be true or false". One student used an example "a square is a rectangle, but a rectangle is not a square".

The PSTs performance on this item improved quite dramatically. While there was only one correct response on the pre-questionnaire, seven out of nine participants responded correctly on the post-questionnaire. The positive changes were apparent both in the rate of correct choices and in the quality of explanations. For example, on the pre-questionnaire Gemma chose the answer (b) The converse of the theorem must be true but it needs further proof. She justified her answer by essentially restating the distractor. She wrote: "Although the theorem has been proved, the converse of the theorem must also be proved. It is not given that it is true, so it must be shown." On the post-questionnaire, Gemma chose the correct distractor (e) none of the above, and explained it:

> If a theorem in the form P→Q has been proven, it does not mean that the converse of the theorem (Q→P) is true. It can be true if it is a biconditional statement, meaning P⇔Q, however that is not always the case. If it is true, it must be proven. If it is false, a counterexample can be used to show it is false.

These examples illustrate the differences among the groups as well as the growth in PSTs' KLAP.

Knowledge of Content and Students: KCS-P

The KCS-P section had 11 items, 10 worth 3 points, and one worth 2 points, making the maximum mean score of 2.9 across all 11 items. Table 5 shows means and standard deviations of Teachers, M&CS majors, PSTs-Pre, and PSTs-Post. The number of data points is the product of the number of questions and the group size. Figures 6a and b show box and whisker plots for all groups.

When comparing Teachers, M&CS majors and PSTs-Pre, one-way ANOVA showed that the groups were not statistically different ($p=0.8843$). However, considering the PSTs-Post, all three groups are significantly different ($p=0.0095$), with

Group	No' of data points	Mean	SD	Teachers, M&CS majors and PSTs-Pre	Teachers, M&CS majors and PSTs-Post
PSTs-**Pre**	99	1.61	1.09		
M&CS majors	242	1.67	1.09	*p*-value: 0.8843	*p*-value: 0.0095*
Teachers	187	1.66	1.09		
PSTs-**Post**	99	2.04	1.03		

Table 5 KCS-P performance across all groups
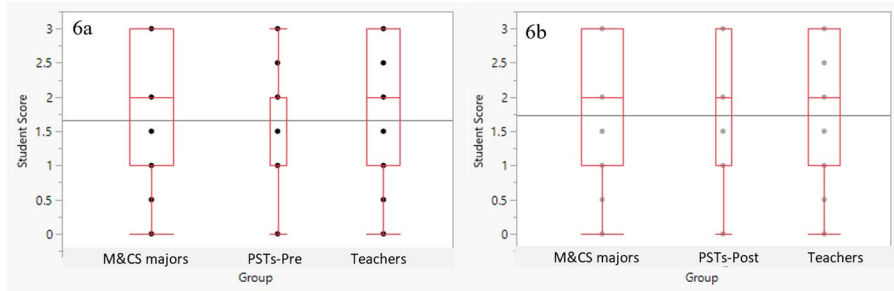
Figs. 6a & 6b KCS-P portion of MKT. Box and whisker plots for M&CS majors, and teachers compared to PSTs-pre (6a, left) and to PSTs-post (6b, right)

the PSTs-Post scoring significantly higher than both M&CS majors ($p=0.0126$) and Teachers ($p=0.0162$). This is because the PSTs performance improved significantly between pre and post ($p=0.0013$), according to the matched pairs *t*-test. We find it encouraging that the PSTs' performance significantly improved on the post-questionnaire and was significantly higher even than that of practicing teachers.

We were surprised to see that Teachers and M&CS majors scored similarly on the KCS-P portion of the questionnaire, since KCS-P items were intended to assess participants' knowledge of students' proof-related conceptions and misconceptions. We considered this as pedagogical knowledge, which would be characteristic of practicing teachers. It is possible that our items did not require special pedagogical knowledge beyond pure mathematical knowledge, despite our best intentions. At the same time, this outcome concurs with the findings of Krauss et al., (2008), who found that undergraduate mathematics majors performed similar to in-service teachers on all PCK items, except ones directly related to classroom teaching.

We illustrate these data using the example of the Angela item in Fig. 2. This item appeared to be rather difficult for all groups of participants. Only four out of 17 Teachers and five out of 22 M&CS majors noticed and correctly explained the nature of Angela's misconception: limiting the domain of the statement to only special quadrilaterals, which happen to satisfy the statement. Other M&CS majors, who received partial credit of 1 or 2 points for this item, criticized Angela's argument for its reliance on specific examples, e.g., "The student used two specific examples to prove the general statement." Although it is true that a general statement requires a general proof, if it can be shown that the domain of the statement is finite, or has a finite number of cases, then it is possible to prove the statement by exhaustion of

cases. Angela's argument attempted to follow this route but failed due to her restriction of the domain of the statement. Yet, her mistake does not delegitimize her proof approach, hence the critique of "relying on examples" is misplaced.

This type of critique also appeared in the Teachers' group and in the PSTs-Pre group. In addition, the participants in these groups suggested that Angela left out a square, which also has congruent diagonals. For example, one teacher wrote: "A square could also have congruent diagonals and her statement says there are only two convex quadrilaterals with those characteristics." This is another illegitimate critique of Angela's argument, since a square would have been accounted for by Angela's mentioning of rectangle.

Four out of nine PSTs correctly responded to this item on the post-questionnaire, as opposed to none on the pre-test. In addition, the way the PSTs worded their responses on the post-questionnaire suggests that their analyses went beyond identifying a mathematical mistake, towards conjecturing about student thinking and understanding, as the next quote by Gavin shows:

> The student rationalizes what she knows well and if those two were the only quadrilaterals with congruent diagonals she would be right, but she mostly generalizes all quadrilaterals and forgets about a kite, which also fits the parameters but just shows she is speaking in generalizations.

Some of the nuances of participants' responses, across all the groups, were captured by our coding scheme, although they are less visible in the aggregated averages.

Knowledge of Content and Teaching: KCT-P

Similar to KCS-P items, the KCT-P items were grounded in teaching context of interpreting student work, but KCT-P items also asked participants to provide feedback to the student, highlighting strength and weakness of their argument. There were eight items, each worth 3 points, making the maximum possible mean of 3. Table 6 shows the performance of all groups.

One-way ANOVA combined with the Tukey-Kramer Honest Significant test showed that all three groups were statistically different when comparing Teachers, M&CS majors and PSTs-Pre ($p < 0.0001$). Teachers significantly outperformed both groups; they tended to score on average 0.83 points higher than PSTs-Pre ($p < 0.0001$) and 0.42 points higher than M&CS majors ($p = 0.0090$). PSTs-Pre performed rather poorly on KCT-P items, scoring significantly lower even than M&CS majors ($p = 0.0498$). However, on the post-questionnaire, the PSTs significantly improved their performance ($p < 0.0001$). The PSTs-Post scored similar to teachers ($p = 0.7433$) and significantly outperformed M&CS majors ($p = 0.0050$).

We illustrate some of the differences between the groups' KCT-P performance on the item Molly, shown in Fig. 3. Fourteen out of 17 Teachers and 14 out of 22 M&CS majors noticed Molly's error – using the same variable to represent two different fractions, and by this limiting the proof to two *same* rational numbers. Below are examples of two such responses:

Group	No' of data points	Mean	SD	Teachers, M&CS majors and PSTs-Pre	Teachers, M&CS majors and PSTs-Post
PSTs-**Pre**	72	1.01	1.337	*p*-value:	
M&CS majors	176	1.42	1.23	<.0001*	*p*-value: 0.0012*
Teachers	136	1.84	1.29		
PSTs-**Post**	72	1.97	1.22		

Table 6 KCT-P performance across all groups
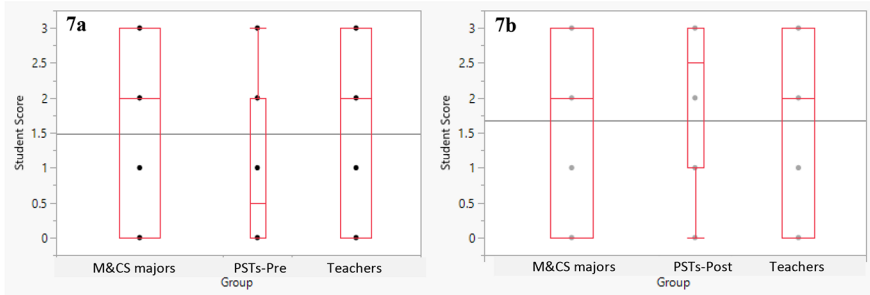
Figs. 7a & 7b KCT-P portion of MKT. Box and whisker plot for M&CS majors and teachers compared to PSTs-pre (7a, left) and to PSTs-post (7b, right)

> Assumed *r* and *s* are equal rational numbers - > not accounting for all possibilities. Strength = using variables / algebra to proof. Weakness = should have made *r* and *s* using different variables. i.e., $r = a/b$ where *b* not equal 0, $s = c/d$ where *d* not equal 0. [LL, M&CS major]
>
> Molly's definition of "*s*" assumes it is equal to "*r*". Her proof relies on *r* and *s* being fully general. The overall outline of your proof is sound. However, your definition of *s* makes it identical to *r*. You want for *r* and *s* to be *any* arbitrary rational numbers - not necessarily identical ones. So you should think of a more general way to define *s*, not reusing "*a*" and "*b*". [T4]

Both responses are completely correct, and the participants provided actionable feedback to Molly, that is, feedback that contained enough information for Molly to improve her work. Therefore, both responses received the high score of 3 according to our codebook (Table 2). Our coding scheme captured mathematical correctness and richness of pedagogical considerations, however, the analysis revealed that there were some qualitative differences between the groups, which our scheme was not capturing. For example, we noticed that Teachers tended to write longer and more involved feedback than M&CS majors and were more likely to address the hypothetical student in the first person, as illustrated in the quotes above. Teachers also often framed their feedback using guiding questions, and oriented students towards revising their arguments instead of simply pointing out errors.

On the item Molly (Fig. 3), the PSTs improved only slightly; from three correct responses on the pre-questionnaire to four correct responses on the post. The number of PSTs who did not notice Molly's error remained the same (four PSTs) suggesting

that completing the questionnaire 14 weeks earlier was not the sole reason for the pre-to-post change, and there are reasons to assume that the improved PSTs' performance is due to their learning. We again noted qualitative differences in the PST-Post's feedback, which was closer to that of the Teachers. To capture these and other qualitative differences between the participant groups, we developed a completely new coding scheme and re-coded all responses to KCT-P items. The description of this analysis and its results are beyond the scope of this paper, and we report on them elsewhere (Butler et al., 2022).

Discussion

As we turn to summarize and discuss the outcomes of our study, we remind the reader of the context of our study and address its methodological limitations. Next, we frame our findings within the broader literature, explicate its contributions and outline some directions for further research.

Study Context and Limitations

This study is a part of the larger design-based research project which aimed to enhance prospective secondary teachers' expertise for teaching reasoning and proof as they participate in a specially designed capstone course *Mathematical Reasoning and Proving for Secondary Teachers.* The design of the course and the assessment instruments were inspired by our conceptualization of MKT-P outlined above. Given the local and design-oriented nature of our main study, situated in a single university course, it was beyond the scope and goals of our project to conduct a validation study of our MKT-P instrument using a traditional factor analysis approach. But as we used our MKT-P questionnaire across the three years of the study, we used the annual data to refine the items, improving the overall clarity and performance of the questionnaire. Thus, we feel that the current version of our MKT-P instrument was sufficiently supported by theory and prior empirical evidence for conducting this comparison study.

There are several other limitations to our methods. The number of participants overall and in each group was quite small. Nevertheless, we assert that examining the nature of MKT-P, even on a small scale, can be advantageous and informative for the field of mathematics education.

We also note the self-selection bias in the groups of M&CS majors and Teachers, which showed in the very high variance in the Teacher group performance. The M&CS majors' group does not have as high variation as the Teachers' group, but their overall performance on the MKT-P instrument was quite low. This may be due to the self-selection bias or to other factors. We have no data on the M&CS majors' performance in the *Mathematical Proof* course, nor how representative our study participants were in relation to other M&CS majors taking the proof course. In any case, we make no claims about the generalizability of our results.

Another limitation of our study design is the difference in stakes of taking the MKT-P questionnaire by the PSTs before and after the capstone course. The post-questionnaire was the final course exam for the PSTs. Although, it was only 10% of the course grade (the rest being distributed between lesson planning, ongoing assignments and course portfolio), it was the course summative assessment, and the PSTs most likely tried to put their best effort into it.

Despite these limitations, to the best of our knowledge, our study is the first one attempting to examine the nature of MKT-P by comparing and contrasting performance of groups with varied subject matter and pedagogical backgrounds on the same MKT-P questionnaire. Our study is also unique in our attempt to document and characterize changes in PSTs' MKT-P over time.

Summary and Discussion of the Results

The two key findings of the analysis presented above can be summarized as follows. In general, the teachers outperformed M&CS majors who, in turn, outperformed PSTs-Pre. Secondly, the PSTs' performance improved significantly after participating in the capstone course. In fact, the PSTs-Post scored similar to teachers and outperformed M&CS majors. We now discuss these findings in greater detail.

With respect to our first research question (How do practicing teachers, M&CS majors and prospective teachers compare on their performance on our MKT-P questionnaire?) our results showed that there were statistically significant differences between the groups' performance on the overall MKT-P, and on two facets: KLAP and KCT-P. On the KCS-P items, the differences between the groups only showed for PSTs-Post compared to M&CS majors and teachers. These outcomes concur with the *Learning through Teaching* paradigm and the numerous studies showing that mathematics teachers' expertise develops through their classroom teaching experiences (e.g., Chazan 2000; Leikin & Zazakis, 2010; Mason 2008). Closer examination of the groups' performances across MKT-P facets revealed interesting patterns.

On the KLAP and KCT-P facets, the Teachers significantly outperformed M&CS majors. Teachers' success on the pedagogical items assessing ability to provide instructional feedback (KCT-P) was less surprising than their success on items that rely solely on proof-related mathematical knowledge (KLAP). Despite the high variance in the Teachers' group, most teachers were fluent with such concepts as the converse, contrapositive, and indirect reasoning and were able to identify logically equivalent and non-equivalent statements. We find it interesting, since, as mentioned above, only five out of 17 teachers taught Geometry, i.e., taught content directly related to proof. Thus, it is even more intriguing that Teachers' performance on KLAP items was significantly higher than that of M&CS majors, who, at the time of the study were enrolled in a proof course and had proof-related concepts fresh in their mind.

The M&CS majors scored significantly higher than PSTs-Pre on both KLAP and KCT-P facets. The outcome relating to KLAP is interesting when considering that all PSTs have successfully completed the *Mathematics Proof* course in the past and at least one proof-oriented course, such as Geometry or Abstract Algebra, in which they had ample opportunities to apply their knowledge of proof and proving and therefore

were expected to perform well. Moreover, we were surprised that M&CS majors outperformed PSTs-Pre on KCT-P items. We initially assumed that since PSTs are in the last year of their program and had taken two education courses, they would perform better than M&CS majors did on this pedagogical knowledge portion of the questionnaire. The results did not confirm our expectation. One possible explanation for the overall low performance of the PSTs-Pre is that their pre-questionnaire was low-stakes, non-graded class assignment; hence, the PSTs might not have given it their best effort. M&CS majors, on the other hand, were trying to earn an extra credit and might have invested more into completing the MKT-P questionnaire.

With respect to our second research question (How does the prospective teachers' MKT-P performance change (if at all) following their participation in the capstone course *Mathematical Reasoning and Proving for Secondary Teachers?)* the analysis shows that the PSTs improved dramatically on the post-questionnaire, both on the overall MKT-P and on each subdomain. The PSTs-Post significantly outperformed M&CS majors on all subdomains of MKT-P and scored similar to Teachers on KLAP and KCT-P, and even higher than Teachers on KCS-P. These improvements in PSTs'-Post performance are partially due to the change in the test-taking stakes for the PSTs, and partially due to PSTs completing the same questionnaire twice - before and after the course. However, we doubt that these are the only reasons for such significant improvements. The PSTs'-Post responses had higher rates of correct answers, richer explanations, and utilized proof-specific vocabulary; these responses were also qualitatively closer to those of practicing teachers. Importantly, the observed changes are reflective of the content of the course modules and of the pedagogical issues addressed during class discussions.

During the course, the PSTs had numerous opportunities to strengthen their knowledge of the logical aspects of proof, connect this knowledge to the secondary school curriculum, learn about students' proof-related conceptions and misconceptions, and apply their knowledge to teaching (Buchbinder & McCrone, 2020). We assert that these rich course experiences contributed to PSTs' development of proof-specific content and pedagogical knowledge and showed in their improved MKT-P performance. These outcomes are consistent with MKT literature showing that prospective teachers' content and pedagogical knowledge increase during university training (e.g., Blömeke et al., 2008), and with Ko's et al. (2017) study, which found a strong positive correlation between PSTs' field experiences in K-12 classrooms and their overall performance on an MKT-Geometry questionnaire. Our study, therefore, provides evidence that an integrated content and pedagogy course with an embedded practice component provides a rich learning environment fostering the development of prospective teachers' MKT-P.

It is important to note that while we observed significant improvement in PSTs' performance on MKT-P post-questionnaire, we do not claim that this can be used as sole evidence of their increased expertise in teaching mathematics with the emphasis on reasoning and proving. Although the literature suggests that there is cumulative evidence for a positive relationship between teacher knowledge as measured by paper and pencil instruments and the quality of their instructional practices (Hill et al., 2005, 2008), we agree with the scholars who view teacher knowledge as contextualized, situated and mediated by beliefs and experiences (Depaepe et al., 2020;

Mason, 2008; Phelps et al., 2020). In the larger study, we developed a suite of instruments for capturing PSTs' dispositions, lesson planning and classroom performance to assess how their expertise for teaching reasoning and proving develops (Buchbinder & McCrone, 2021). The discussion of these instruments and report on their use is beyond the scope of this paper. In the study reported herein, we use the data on improved PSTs' MKT-P performance to substantiate the claim supporting MKT-P as a construct by showing that MKT-P grows and develops over time as a result of targeted intervention in the course.

The MKT-P Questionnaire and the Nature of MKT-P

To summarize the contributions of our study, we proposed a conceptualization of MKT-P as comprised of three facets: KLAP, KCS-P and KCT-P. This framework was useful in designing the course *Mathematical Reasoning and Proving for Secondary Teachers* and the MKT-P questionnaire. Our MKT-P questionnaire is not perfect: achieving a balanced representation of items along the three MKT-P facets, the four proof themes, and three mathematical areas was challenging and not always possible. However, we are not aware of an alternative MKT-P instrument that achieves such goals in a better way. Importantly, the results show that our MKT-P questionnaire can reliably distinguish between different groups of participants – Teachers, M&CS majors, and PSTs - and can detect improvement in MKT-P performance over time.

These two findings provide validation of the MKT-P constructs, similar to the methodological approach of the COACTIV project, i.e., "construct validation by reference to contrast populations" (Krauss et al., 2008 p. 880). Krauss et al., (2008) found that in-service teachers and mathematics majors outperformed prospective teachers on both content knowledge (CK) and PCK items. This result concurs with our study outcome that M&CS majors and teachers scored higher than PSTs-Pre on KLAP and on pedagogical domains like KCS-P and KCT-P. A surprising finding in the Krauss et al.'s study was that mathematics majors and teachers scored almost identically on both CK and PCK items. The differences between the mathematics majors and teachers only became apparent on one subset of PCK items focused on classroom instruction, such as constructing explanations and choosing representations, but *not* on the items that tested knowledge of student misconceptions. Similarly, in our study, there were no differences between M&CS majors' and Teachers' performances on KCS-P items, which aimed to assess knowledge of proof-related misconceptions.

One possible explanation for this outcome is that our KCS-P items can be solved by applying purely mathematical considerations rather than pedagogical ones, which is a limitation of our MKT-P questionnaire. However, given that the larger-scale, COACTIV study (Krauss et al., 2008) did not detect differences between teachers and mathematics majors on this type of items, the issue may lie with the particular genre of items. Note that similar findings were observed with KCS items in Hill's et al., (2007) study at the elementary level, where success on KCS items could be achieved *either* through utilization of pedagogical consideration or with pure mathematical knowledge. This implies the need for further research and development of KCS items that isolate the pedagogical and mathematical components. However, it

remains an open question whether such isolation is feasible at the secondary level (cf., Speer et al., 2015). It is possible that at the secondary level, identifying proof-related misconceptions requires mathematical rather than pedagogical knowledge.

Alternatively, the issue may run deeper. Recent large-scale MKT studies (e.g., Blömeke et al., 2008; Copur-Gencturk et al., 2019) and theoretical analyses of MKT research (Speer et al., 2015) suggest that at the secondary and tertiary levels, content and pedagogical knowledge are maybe too closely intertwined to be empirically distinguishable. Hence, these researchers argue for a unidimensional model of MKT. If this was the case for MKT-P as well, it would explain the lack of difference between the groups on the KCS-P portion of our questionnaire.

At the same time, a hypothesis of unidimensional MKT-P would leave unanswered the differences between the groups on other MKT-P facets, such as KLAP and KCT-P, observed in our study. Some of these differences, e.g., the nature of instructional feedback in the KCT-P items, were qualitative in nature (Butler et al., 2022) and, therefore, required different analytic techniques to capture. It seems important to develop a coding scheme that can take these types of qualitative differences into consideration within the quantitative data analysis. Such a unified coding scheme could further illuminate the differences between various groups of participants and shed light on the nature of MKT-P.

It is important to note that this study was situated in the United States context, within the specifics of a particular university teacher preparation program, which may differ significantly even among other U.S. institutions. Our results should be interpreted within the context in which they were obtained. At the same time, we do not think that our outcomes are completely unique. As shown in the discussion section, many of our outcomes concur with prior research in the U.S. and internationally, specifically with Krauss et al.'s (2008) COACTIV study. This may suggest broader implications of our findings, which could be explored in additional contexts and educational settings. The findings of our study and the alternative explanations for the outcomes suggest the need for further empirical research into the nature of MKT-P and its dimensionality, and the need for designing items and instruments that can capture the variation in MKT-P performance at scale.

Implications for Education

Our study also has implications for undergraduate mathematics education of both mathematics and mathematics education majors. Our findings point to the situated nature of knowledge (Depaepe et al., 2020), not just pedagogical but also the purely mathematical one, such as the knowledge of the logical aspects of proof. In other words, the abstract nature of KLAP, which students are expected to develop in a course on mathematical proof, does not seem to transfer easily to teaching of mathematics, nor to other mathematical disciplines. Our study shows that intentionally designed activities, which aim to connect between advanced mathematical concepts specific to proof and their applications to mathematics teaching can be beneficial for future teachers. Although not every university program can add a whole course like we did, the modular structure of our course design allows applying modules

or activities on their own. Such activities can be integrated in proof courses and can potentially benefit not only future teachers but also other undergraduate majors, such as mathematics or computer science majors. Such benefits have been documented in undergraduate courses such as abstract algebra, real analysis, geometry, statistics, and others (Álvarez et al., 2020; Lai et al., 2018; Wasserman et al., 2019). Future studies can examine the effectiveness of stand-alone proof-oriented activities on undergraduates' learning.

List of Abbreviations

KLAP	Knowledge of the logical aspects of proof
KCS-P	Knowledge of Content and Students specific to proof
KCT-P	Knowledge of Content and Teaching specific to proof
MKT-P	Mathematical Knowledge for Teaching Proof
M&CS	Mathematics, and computer science majors
PST	Prospective secondary teachers
PST-Pre	Prospective secondary teachers prior to taking the course
PSTs-Post	Prospective secondary teachers after taking the course

Statements and Declarations We have no conflicts of interest to disclose.

References

American Educational Research Association, American Psychological Association, and National Council for Measurement in Education (AERA (2014). APA, and NCME). *Standards for Educational and Psychological Testing.* https://www.testingstandards.net/open-access-files.html

Álvarez, J. A. M., Arnold, E. G., Burroughs, E. A., Fulton, E. W., & Kercher, A. (2020). The design of tasks that address applications to teaching secondary mathematics for use in undergraduate mathematics courses. *Journal of Mathematical Behavior*, *60*, 100814

Antonini, S., & Mariotti, M. A. (2008). Indirect proof: what is specific of this mode of proving? *ZDM-Mathematics Education, 40*(3), 401–412

Association of Mathematics Teacher Educators. (2017). *Standards for Preparing Teachers of Mathematics*. amte.net/standards

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

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

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180

Blömeke, S., & Kaiser, G. (2014). Theoretical framework, study design and main results of TEDS-M. In S. Blömeke, F. J. Hsieh, G. Kaiser, & W. Schmidt (Eds.), *International perspectives on teacher knowledge, beliefs, and opportunities to learn* (pp. 19–47). Springer

Blömeke, S., Kaiser, G., Lehmann, R., & Schmidt, W. H. (2008). Introduction to the issue on empirical research on mathematics teachers and their education. *ZDM-Mathematics Education*, *40*(5), 715–717

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

Buchbinder, O. (2018). Supporting prospective secondary mathematics teachers in creating instructional explanations through video-based experience. *Journal of Technology and Teacher Education 26(1), 33–56.*

Buchbinder, O., & McCrone, S. (2018). Mathematical Reasoning and Proving for Prospective Secondary Teachers. In A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, and S. Brown, (Eds.) *Proceedings of the 21st Annual Conference of the Research in Undergraduate Mathematics Education*, San Diego, CA. (pp. 115–128)

Buchbinder, O., & McCrone, S. (2020). Preservice Teachers Learning to Teach Proof through Classroom Implementation: Successes and Challenges. *Journal of Mathematical Behavior*, *58*, 100779. https://doi.org/10.1016/j.jmathb.2020.100779

Buchbinder, O., & McCrone, S. (2021). Characterizing mathematics teachers' proof-specific knowledge, dispositions and classroom practices. Paper presented at the *ICME 14th International Congress on Mathematical Education*, Shanghai, 2020

Butler, B., Buchbinder, O., & McCrone, S. (2022). Comparing STEM Majors, Practicing and Prospective Secondary Teachers' Feedback on Mathematical Arguments: Towards Validating MKT-Proof. In Karunakaran, S. S., & Higgins, A. (Eds.). *Proceedings of the 24th Annual Conference on Research in Undergraduate Mathematics Education*, (pp. 91–99). Boston, MA

Chazan, D. (2000). *Beyond formulas in mathematics and teaching: Dynamics of the high school algebra classroom*. Teachers College Press

Conference Board of the Mathematical Sciences (CBMS) (2012). *The Mathematical Education of Teachers II*. American Mathematical Society. http://www.cbmsweb.org/MET2/

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-Mathematics Education*, *40*(5), 813–832. https://doi.org/10.1007/s11858-008-0149-1

Copur-Gencturk, Y., Tolar, T., Jacobson, E., & Fan, W. (2019). An empirical study of the dimensionality of the mathematical knowledge for teaching construct. *Journal of Teacher Education*, *70*(5), 485–497

Dawkins, P. C., & Cook, J. P. (2016). Guiding reinvention of conventional tools of mathematical logic: students' reasoning about mathematical disjunctions. *Educational Studies in Mathematics*.http://link.springer.com/article/10.1007/s10649-016-9722-7

Depaepe, F., Verschaffel, L., & Star, J. (2020). Expertise in developing students' expertise in mathematics: Bridging teachers' professional knowledge and instructional quality. *ZDM-Mathematics Education*, *52*(2), 1–14

Durand-Guerrier, V. (2003). Which notion of implication is the right one? From logical considerations to a didactic perspective. *Educational Studies in Mathematics*, *53*(1), 5–34

Ellis, A. B., Bieda, K., & Knuth, E. (2012). *Developing Essential Understanding of Proof and Proving for Teaching Mathematics in Grades 9–12*. National Council of Teachers of Mathematics

Hanna, G., & deVillers, M. D. (2012). *Proof and Proving in Mathematics Education: The 19th ICMI Study*. Springer

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 805–842). Information Age Publishing, Inc

Herbst, P., & Kosko, K. (2014). Mathematical knowledge for teaching and its specificity to high school geometry instruction. In J. J. Lo, K. R. Leatham, & Van L. R. Zoest (Eds.), *Research Trends in Mathematics Teacher Education* (pp. 23–45). Springer International Publishing

Hill, H. C., Rowan, B., & Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406. https://doi.org/10.3102/00028 31204 20023 71

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., et al. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*(4), 430–511. https://doi.org/10.1080/07370-00080-21772-35

Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. *Measurement*, *5*(2–3), 81–92

Hodds, M., Alcock, L., & Inglis, M. (2014). Self-Explanation training improves proof comprehension. *Journal for Research in Mathematics Education*, *45*(1), 62–101

Howell, H., Lai, Y., Phelps, G., & Croft, A. (2016). Assessing Mathematical Knowledge for Teaching Beyond Conventional Mathematical Knowledge: Do Elementary Models Extend? https://doi.org/10.13140/RG.2.2.14058.31680

Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in Mathematics*, *96*(1), 1–16

Knuth, E. J. (2002). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of Mathematics Teacher Education*, *5*(1), 61–88

Ko, Y. Y. (2010). Mathematics teachers' conceptions of proof: Implications for educational research. *International Journal of Science and Mathematics Education*, *8*(6), 1109–1129 https://doi.org/10.1007/s10763-010-9235-2

Ko, I., Milewski, A., & Herbst, P. (2017). *How are pre-service teachers' educational experiences related to their mathematical knowledge for teaching geometry*. https://www.researchgate.Net/publication/311727654

Kotelawala, U. (2016). The status of proving among US secondary mathematics teachers. *International Journal of Science and Mathematics Education*, *14*(6), 1113–1131

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM-Mathematics Education*, *40*(5), 873–892

Lai, Y., Strayer, J. F., & Lischka, A. E. (2018). Analyzing the development of MKT in content courses. In T.E. Hodges, G. J. Roy, & A. M. Tyminski (Eds.). *Proceedings of the 40th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 496–499). University of South Carolina & Clemson University

Leikin, R., & Zazkis, R. (2010). Teachers' opportunities to learn mathematics through teaching. In Learning through teaching mathematics. In R. Leikin, & R. Zazkis (Eds.), *Learning through Teaching Mathematics: Developing teachers' knowledge and expertise in practice* (pp. 3–22). Springer

Lesseig, K. (2016). Investigating mathematical knowledge for teaching proof in professional development. *International Journal of Research in Education and Science*, *2*(2), 253–270

Lesseig, K., Hine, G., Na, G. S., & Boardman, K. (2019). Perceptions on proof and the teaching of proof: a comparison across preservice secondary teachers in Australia, USA and Korea. *Mathematics Education Research Journal*, *31*(4), 393–418

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

Martin, T. S., & McCrone, S. S. (2003). Classroom factors related to geometric proof construction ability. *The Mathematics Educator*, *7*(1), 18–31

Mason, J. (2008). PCK and beyond. In P. Sullivan, & T. Wood (Eds.), *International handbook of mathematics teacher education knowledge and beliefs in mathematics teaching and teaching development* (1 vol., pp. 301–322). Sense Publishers

McCrory, R., Floden, R., Ferrini-Mundy, J., Reckase, M. D., & Senk, S. L. (2012). Knowledge of algebra for teaching: A framework of knowledge and practices. *Journal for Research in Mathematics Education*, *43*(5), 584–615

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. https://doi.org/10.1007/s10649-017-9785-0

National Council of Teachers of Mathematics (2009). *Focus in High School Mathematics: Reasoning and Sense Making*. NCTM

National Council of Teachers of Mathematics. (2014). *Principles to Actions: Ensuring Mathematical Success for All*. NCTM

National Council of Teachers of Mathematics. (2018). *Catalyzing Change in High School Mathematics: Initiating Critical Conversations*. NCTM

National Governors' Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards: Mathematics*. Council of Chief State School Officers

Otten, S., Bleiler-Baxter, S. K., & Engledowl, C. (2017). Authority and whole-class proving in high school geometry: The case of Ms. Finley. *The Journal of Mathematical Behavior*, *46*, 112–127

Otten, S., Gilbertson, N. J., Males, L. M., & Clark, D. L. (2014). The mathematical nature of reasoning-and-proving opportunities in geometry textbooks. *Mathematical Thinking and Learning*, *16*(1), 51–79

Phelps, G., Howell, H., & Liu, S. (2020). Exploring differences in mathematical knowledge for teaching for prospective and practicing teachers. *ZDM-Mathematics Education*, *52*(2), 255–268. https://doi.org/10.1007/s11858-019-01097-x

Reid, D. A. (2002). Conjectures and refutations in grade 5 mathematics. *Journal for Research in Mathematics Education*, *33*(1), 5–29

Riley, K. J. (2003). *An Investigation of Prospective Secondary Mathematics Teachers' Conceptions of Proof and Refutations.* Unpublished Doctoral dissertation, Montana State University, Bozeman, Montana.   https://scholarworks.montana.edu/xmlui/bitstream/handle/1/8385/31762103856330.pdf?sequence=1

Schmidt, W. (2013). *Mathematics Teaching in the 21st Century.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/ICPSR34430.v1

Schwarz, B., Leung, I. K., Buchholtz, N., Kaiser, G., Stillman, G., Brown, J., & Vale, C. (2008). Future teachers' professional knowledge on argumentation and proof: a case study from universities in three countries. *ZDM-Mathematics Education*, *40*(5), 791–811

Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, *34*(1), 4–36

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

Speer, N. M., King, K. D., & Howell, H. (2015). Definitions of mathematical knowledge for teaching: Using these constructs in research on secondary and college mathematics teachers. *Journal of Mathematics Teacher Education*, *18*(2), 105–122

Steele, M. D., & Rogers, K. C. (2012). Relationships between mathematical knowledge for teaching and teaching practice: The case of proof. *Journal of Mathematics Teacher Education*, *15*(2), 159–180. https://doi.org/10.1007/s10857-012-9204-5

Stylianides, G. J. (2008). An analytic framework of reasoning-and-proving. *For the Learning of Mathematics*, *28*(1), 9–16

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), 10

Stylianides, G. J., Stylianides, A. J., & Weber, K. (2017). Research on the teaching and learning of proof: Taking stock and moving forward. In J. Cai (Ed.), *Compendium for Research in Mathematics Education* (pp. 237–266). NCTM

Stylianou, D. A., Blanton, M. L., & Knuth, E. J. (2009). *Teaching and Learning Proof across the Grades: A K-16 Perspective*. NCTM/ Routledge. Studies in Mathematical Thinking and Learning Series

Thompson, D. R. (1996). Connecting Research to Teaching: Learning and Teaching Indirect Proof. *The Mathematics Teacher*, *89*(6), 474–482

Thompson, D. R., Senk, S. L., & Johnson, G. J. (2012). Opportunities to learn reasoning and proof in high school mathematics textbooks. *Journal for Research in Mathematics Education*, *43*(3), 253–295

Wasserman, N., Weber, K., Fukawa-Connelly, T., & McGuffey, W. (2019). Designing advanced mathematics courses to influence secondary teaching: Fostering mathematics teachers' 'attention to scope'. *Journal of Mathematics Teacher Education*, *22*(4), 379–406. https://doi.org/10.1007/s10857-019-09431-6

Weber, K. (2010). Mathematics majors' perceptions of conviction, validity, and proof. *Mathematical Thinking and Learning*, *12*(4), 306–336