



Preparing prospective secondary teachers to teach mathematical reasoning and proof: the case of the role of examples in proving

Orly Buchbinder¹  · Sharon McCrone¹

Accepted: 25 April 2023
© The Author(s) 2023

Abstract

Mathematics teacher education programs in the United States are charged with preparing prospective secondary teachers (PSTs) to teach reasoning and proving across grade levels and mathematical topics. Although most programs require a course on proof, PSTs often perceive it as disconnected from their future classroom practice. Our design research project developed a capstone course *Mathematical Reasoning and Proving for Secondary Teachers* and systematically studied its effect on PSTs' content and pedagogical knowledge specific to proof. This paper focuses on one course module—*Quantification and the Role of Examples in Proving*, a topic which poses persistent difficulties to students and teachers alike. The analysis suggests that after the course, PSTs' content and pedagogical knowledge of the role of examples in proving increased. We provide evidence from multiple data sources: pre-and post-questionnaires, PSTs' responses to the in-class activities, their lesson plans, reflections on lesson enactment, and self-report. We discuss design principles that supported PSTs' learning and their applicability beyond the study context.

Keywords Reasoning and proving · Mathematical Knowledge for Teaching Proof · Prospective secondary teachers

1 Introduction

Recognizing the critical role of proof in mathematics, policy and curricular documents worldwide emphasize teaching proof across grade levels and mathematical topics (e.g., ACARA, 2022; MINEDUC, 2019; NCTM, 2009; NGA & CCSSI, 2010). Teacher education programs are expected to prepare their graduates to implement this vision of teaching. Many secondary teacher preparation programs provide strong content preparation, including courses on proof taught in mathematics departments by mathematicians, to both mathematics and mathematics education majors (Blömeke et al. 2014; Tattro, 2013). Studies have shown that prospective secondary teachers (PSTs) experience difficulties with proof at the university level, struggling with topics such as understanding the relationship between empirical and deductive reasoning, including the role of examples in proving (e.g., Weber, 2010); reasoning with conditional statements (e.g., Durand-Guerrier, 2003); proof production

and comprehension (e.g., Hodds, et al., 2014) and proof by contradiction (Antonini & Mariotti, 2008).

With few opportunities for making connections between university proof courses and their future classroom practice, PSTs often develop views of proof as formal exercises in university mathematics, disconnected from secondary classrooms (Stylianides et al., 2017). For example, Schwarz et al. (2008) examined future teachers' professional knowledge of argumentation and proof in Germany, Hong Kong, and Australia. Across all three countries, the authors concluded that "possessing a tertiary mathematical background as required for teaching and having a high affinity with proving in mathematics teaching at the lower secondary level are not sufficient preparation for teaching proof" (p. 808).

Felix Klein (1932) alluded to the problem of *double discontinuity*—the feeling of disconnect that future teachers experience when first encountering university-level mathematics, and then again, at the start of classroom teaching. While the feelings of disconnect apply to university mathematics in general, (e.g., Goulding et al., 2003; Winsløw & Grønbæk, 2014), the situation with proof seems to be unique due to the explicit expectation by policy documents and national curricula that teachers will integrate proof in their classroom teaching. This suggests that universities need

✉ Orly Buchbinder
orly.buchbinder@unh.edu

¹ University of New Hampshire, Durham, USA

to develop more effective approaches for supporting PSTs developing proof-specific knowledge and practices.

Toward this end, we conducted a three-year design research study (Sandoval, 2014) in which we developed a capstone course¹ *Mathematical Reasoning and Proving for Secondary Teachers*, as a culminating experience of a mathematics education program. The course was offered at the mathematics department and taught by mathematics education faculty, the first author of this paper. The course consists of four modules, each addressing a topic identified in the literature as posing persistent difficulties to students and teachers alike: (a) direct proof and argument evaluation, (b) conditional statements, (c) quantification and the role of examples in proving, and (d) indirect reasoning. The course activities intended to help PSTs to crystalize their mathematical knowledge, connect it to secondary curriculum, learn about students' conceptions, and apply this knowledge by designing lessons that integrate proof with secondary mathematical topics and enacting them in local schools (Buchbinder & McCrone, 2020).

During the three-year design research project, we systematically studied the impact of the course on PSTs' content and pedagogical knowledge of proof. This paper provides evidence that the course design and activities were conducive to PSTs' learning and points to the design principles supporting that learning. Due to space constraints, we focus on one course module: *Quantification and the Role of Examples in Proving* (QRE) and formulate our research questions with respect to it:

1. How do PSTs' knowledge and practices related to quantification and the role of examples in proving change due to participation in the course?
2. How did the course activities contribute to the observed changes in PSTs' knowledge and practices?

The broader project examined similar questions with respect to other modules and the whole course; presentation of all results is beyond the scope of this paper.

2 Theoretical perspectives

2.1 Reasoning and proving at the secondary level: the focus on QRE

One of the main challenges to conceptualizing reasoning and proving in schools has been to differentiate school-level proofs from formal, university-level proofs while preserving the integrity of proof as a hallmark of mathematical validity (Harel & Sowder, 2007). To address these challenges, we adopt a definition of proof as "a mathematical argument for or against a mathematical claim that is both mathematically sound and conceptually accessible to the members of the local community where the argument is offered" (Stylianides & Stylianides, 2017, p. 121). This definition can equally apply to a community of mathematicians and of school students. Importantly, it suggests that the validity of student arguments should be grounded in deductive reasoning rather than authority of a teacher, textbook or empirical evidence, and does not require proof to be formal or have a certain format.

Recent reviews of international literature (e.g., Mariotti, et al., 2018; Stylianides et al., 2017) show that proof has been a challenging topic to learn and to teach. Students at all levels, as well as prospective and even practicing teachers experience challenges with understanding the relationship between empirical arguments and deductive reasoning. A recurring finding around the world is that students rely on supportive examples as proof, without realizing the limitations of such reasoning (Stylianides & Stylianides, 2017). PSTs also, tend to consider empirical arguments as more convincing than deductive proofs (Ko, 2010), and both pre- and in-service teachers have been shown to struggle distinguishing between valid and invalid arguments (Harel & Sowder, 2007; Ko, 2010). Thus, teachers may miss opportunities to address the limitations of empirical arguments with their students, or unintentionally reinforce this problematic conception.

Another well-documented finding is that students and teachers treat counterexamples as exceptions, rather than disproof, or tend to prefer multiple counterexamples (e.g., Lee, 2016; Weber, 2010). Tabach et al. (2010) identified challenges related to existential (there exist) statements. Teachers in that study struggled to accept correct students' proofs of existential statements when those relied on a supportive example, or to reject incorrect "disproof" by non-supportive examples.

As teachers are charged with helping students develop mathematically accepted notions of proof and the roles of examples in proving/disproving quantified statements, it is important that teachers themselves have a strong mathematical and pedagogical knowledge base.

¹ A capstone course is "a study unit which is located towards the end of an academic study program, with the aim of concluding or 'crowning' the experience, and to link academic competence and training with the needs of a professional occupation" (Winsløw & Grønbæk, 2014, p. 4).

2.2 Mathematical Knowledge for Teaching Proof: MKT-P with the focus on QRE

Researchers have conjectured that teaching mathematical reasoning and proving requires a special type of teacher knowledge: Mathematical Knowledge for Teaching Proof (MKT-P). Several frameworks delimiting this concept have been proposed over the years (e.g., Corleis et al., 2008; Harel, 2008; Lesseig, 2016; Lin et al., 2011; Stylianides, 2011). Following Stylianides' (2011) and Harel's (2008) approaches, we conceptualize MKT-P as comprised of three facets: Knowledge of the Logical Aspects of Proof (content knowledge of proof), Knowledge of Content and Students specific to proof (KCS-P) and Knowledge of Content and Teaching specific to proof (KCT-P) (Buchbinder & McCrone, 2020).

Knowledge of the Logical Aspects of Proof includes knowledge of different types of proofs, valid and invalid modes of reasoning, logical relations, a range of definitions and theorems and the roles of examples in proving. The latter includes the types of inferences that can be drawn from examples with respect to two types of quantified statements: *universal* statements—“all objects in a domain D, satisfy some property P(x),” and *existential* statements—“there is an object in D that satisfies P(x).” Examples that are in the domain and satisfy the property, support but do not prove a universal statement, while one such example proves the existential statement is true. Examples that are in the domain but do not satisfy the property are counter-examples disproving a universal statement but are merely non-confirming for an existential statement; insufficient to disprove it. Examples that are not in the domain are irrelevant to proving or disproving either type of statement [cf., Buchbinder and Zaslavsky's (2019) Role of Examples in Proving framework]. In classrooms, teacher knowledge of logical aspects of proof is manifested in their use of clear and accurate mathematical language, notation, and ability to identify and correct students' logical mistakes.

Knowledge of Content and Students specific to proof involves knowledge of students' proof-related conceptions, misconceptions, and common mistakes. With respect to the role of examples in proving, KCS-P involves recognizing student challenges with this topic, such as confusion between universal and existential quantifiers, the difficulty to discern between examples, counterexamples and irrelevant examples for a given statement, and making inferences that consider both the type of quantifier and the type of example (Buchbinder & Zaslavsky, 2019; Durand-Guerrier, 2003). Related classroom practices involve the teacher's ability to identify and anticipate students' proof-related misconceptions, facilitate discussions and explain proof concepts.

Stylianides (2011) notes that while classroom discussions may offer opportunities to discuss the differences

between empirical examples and proof, helping students to overcome their proof-related misconceptions requires carefully planned classroom interventions. Thus, Knowledge of Content and Teaching specific to proof (KCT-P) involves pedagogical strategies like identifying curriculum opportunities for reasoning and proving, designing, and enacting proof-related tasks. Specific to the roles of examples in proving, it involves designing tasks with opportunities for students to learn about mathematically acceptable ways to prove or disprove quantified statements and the roles examples in these processes.

The three facets of MKT-P are closely intertwined. Drawing distinctions between them serves the operational purpose of capturing and assessing MKT-P. The conceptualization of the classroom manifestations of MKT-P draws on the literature connecting teacher knowledge to quality of instruction (e.g., Charalambous, 2020; Kunter et al., 2013). The theoretical perspective on the multidimensionality of MKT-P as comprised of content knowledge, pedagogical knowledge, beliefs, and practices aligns with other general frameworks on teacher competence (Blömeke, et al., 2015; Kunter et al., 2013), mathematical knowledge for teaching (Ball et al., 2008; Shulman, 1986), and frameworks utilized in international studies like MT21 (Schmidt 2013), and TEDS-M (Tatto, 2013). Our MKT-P framework distills elements specific to reasoning and proof from this research base and MKT-P literature.

3 Method

3.1 Design research methodology

Design research methodology intertwines instructional design and educational research (Gravemeijer & Prediger, 2019) through iterative stages of design, implementation, analysis, reflection, and refinement of learning environments. Researchers concurrently design environments and study phenomena emerging in them to develop local—domain or topic-specific—learning theories that connect instructional design with learning outcomes (Fig. 1).

The final stage of design research is a retrospective analysis of all data and partial theories from earlier cycles to produce contextually sensitive design principles (Sandoval, 2014). We report on the results of this reflective analysis following three iterations and provide evidence for how the finalized design principles contributed to the development of PSTs' knowledge and practices specific to the Quantification and the Role of Examples module.

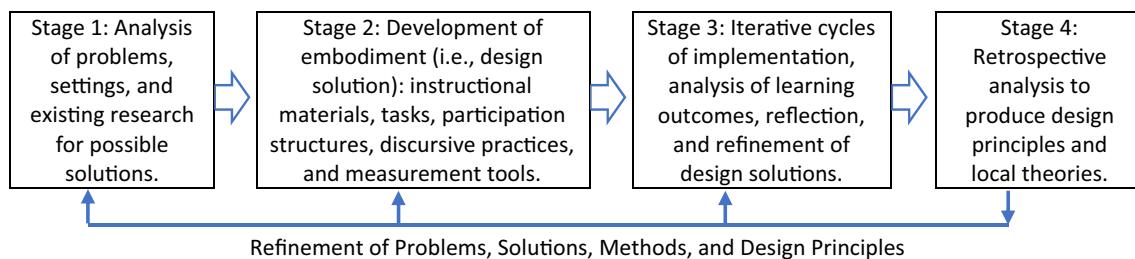


Fig. 1 Four phases of design-based research

3.2 The capstone course design

The initial design principles for the course *Mathematical Reasoning and Proving for Secondary Teachers* came from the analysis of literature on students' and teachers' conception of proof and difficulties with proving (e.g., Ko, 2010; Schwarz et al., 2008). Through this analysis, we identified four proof themes for the course modules: direct proof and argument evaluation, conditional statements, quantification and the role of examples in proving, and indirect reasoning.

Teacher knowledge and practices outlined in our MKT-P framework, represent the desired learning outcomes and the course objectives. To identify pedagogical strategies for supporting these objectives, we consulted literature on PSTs' learning in undergraduate programs (e.g., Grossman et al., 2009; Kunter et al., 2013; Wasserman et al., 2019). The key elements emerging from the literature inspired the three types of course activities: crystallize, connect, and apply. Later, after iteratively testing the course design and conducting retrospective analysis, these types of activities became formulated as design principles (Sandoval, 2014). *Crystallize* activities provide PSTs opportunities to refresh and enhance their content knowledge of the four proof themes. The *connect* activities focus on connecting university-level knowledge of proof to secondary mathematics, and on increasing PSTs' awareness of students' difficulties with proving. *Apply* activities provide opportunities to enact content and pedagogical knowledge of proof. These types of activities integrate various aspects of MKT-P as they support proof-specific content and pedagogical knowledge and practices. The crystallize-connect-apply structure was reproduced in each module with specific activities (Buchbinder & McCrone, 2020), including QRE module, described below.

3.3 The QRE module

3.3.1 What Can You Infer from This Example?

The module included two *What Can You Infer from This Example?* activities, (Buchbinder et al., 2017), completed by PSTs individually, online. In Activity 1 PSTs were given

a false universal statement: "A quadrilateral whose diagonals are congruent and perpendicular to each other is a kite," and asked to determine its logical structure and truth-value, and to justify their response. Next, PSTs examined five fictitious students' examples and determined what can be inferred from each example about the statement: (1) a square and (2) non-convex kite with congruent and perpendicular diagonals only support the statement, (3) an isosceles trapezoid with perpendicular diagonals and (4) a general quadrilateral with congruent and perpendicular diagonals disprove the statement, and (5) a general convex kite—irrelevant to the statement, since its diagonals are not congruent (Fig. 2). Next, the PSTs could revise their original assessment of the statements' truth-value. Then, they watched a cartoon-based classroom scenario of students confused about which quadrilateral is "the best" counterexample: an isosceles trapezoid that has congruent and perpendicular diagonals but is not a kite, or a general kite, which does not have congruent diagonals. The PSTs wrote a scenario (Zazkis et al., 2013) describing how they, as teachers, would lead the discussion to resolve the confusion.

In Activity 2 the PSTs analyzed a true existential statement: "There exist three consecutive even numbers whose sum is divisible by four." They examined six fictitious students' examples; determining, for each example, whether it proves the statement, disproves it, neither proves nor disproves, or cannot be used to evaluate the statement. For instance, (8, 10, 12) are consecutive even numbers whose sum is not divisible by 4; they neither prove nor disprove the statement, since an existential statement cannot be disproved by non-supportive examples. A triplet (4, 6, 10) has non-consecutive even numbers, making it irrelevant, even though the sum is divisible by 4. These distinctions were not clarified in advance but were left for the PSTs to deduce from the activity. Next, the PSTs could revise their original assessment of the truth-value of the statement.

Although these activities are embedded in a school context, they intend to *crystallize* PSTs' content knowledge of QRE, since identifying what can be inferred from a given example does not require any pedagogical knowledge. Analyzing the classroom scenario and writing its continuation supports

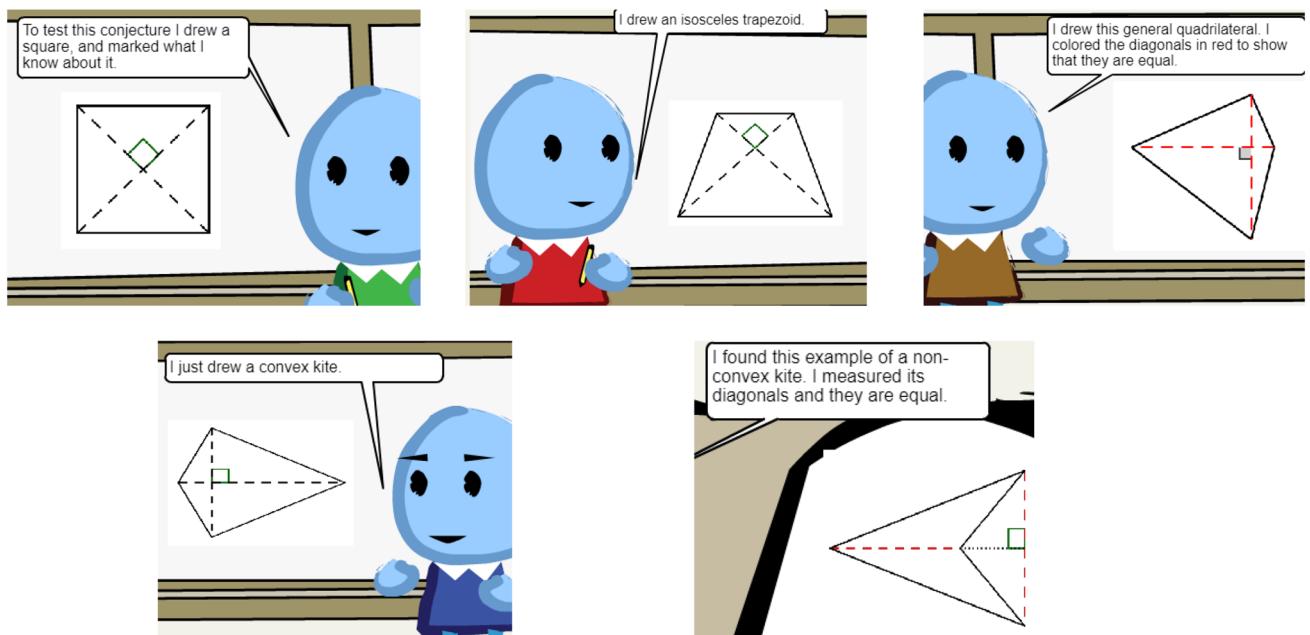


Fig. 2 What Can You Infer from This Example? (Graphics are © 2021, The Regents of the University of Michigan, used with permission)

a. True-False task example
For each statement below decide whether it is true or false and justify your answer.

1. Every three numbers a, b, c satisfy the equation: $\frac{a}{b+c} = \frac{a}{b} + c$.

b. Always-Sometimes-Never task example
For each equation, indicate whether it is true for all values of the parameters, some values of the parameters, or no values of the parameters.

1. $\frac{a}{b+c} = \frac{a}{b} + c$

Fig. 3 a, b Items from the *True-False* and *Always-Sometimes-Never* tasks

pedagogical knowledge of QRE, embodying the *connect* design principle.

3.3.2 Task analysis: true-false vs. always-sometimes-never

In this activity the PSTs solved two tasks with the same mathematical content but different requirements. The *True-False* task required sorting of six statements into non-mutually exclusive categories: universal, existential, require a general proof, require a disproof by a counterexample, require a proof by supportive example, have only supportive examples, have both supportive and counterexamples, and have no supportive examples. The *Always-Sometimes-Never* task contained the equations from the *True-False* task *not* embedded in quantified statements (Fig. 3). For each equation, the PSTs determined if it is true for all values of parameters, some values, or no values of the parameters. This intended to emphasize that while a predicate $P(x)$ can be true or false, depending on the variable (x), a quantified

statement must have a single truth-value. Next, for all six equations, the PSTs wrote their own, true, universal or existential statements.

The PSTs completed these tasks in groups, each group with one version of the task, and then compared their work. This activity is mainly focused on *crystallizing* content knowledge of QRE. The *connect* aspect is manifested in the use of secondary school mathematical content, and in the follow-up discussion where PSTs contemplated learning opportunities for reasoning and proving afforded by these tasks.

3.3.3 QRE-integrated lesson cycle

For the *apply* activity, each PST reached out to their cooperating school teacher to find out the mathematical topic to be taught. The PSTs designed a lesson plan integrating that topic with some key ideas of QRE. To support the PSTs, the instructor devoted class time to lesson planning, and consultation with peers and the course instructor. Next,

Table 1 Summary of evaluation methods by data source and knowledge type

Data Source	Type of knowledge/practice	Type of knowledge/practice		
		Proof-specific content knowledge	Proof-specific Knowledge of content and students (KCS-P)	Proof-specific knowledge of content and teaching (KCT-P)
MKT-P questionnaire	Quantitative analysis of changes in aggregated pre-post MKT-P scores on QRE items. ANOVA and paired <i>t</i> test			
In-class activities: <i>What Can You Infer from This Example?</i>	Qualitative analyses of shifts in PSTs' responses before and after examining student work			
Lesson plans				Quantitative assessment of Proof and QRE integration in lesson plans
Reflections on enacted lessons			Qualitative analysis of reflective noticing of pedagogical practices and student thinking	
Course essays	Qualitative open-coding of PSTs' self-reported evidence of learning from the course (types of knowledge / practices) with connections to specific course activities or design principles			

PSTs taught small groups of students in a local school, video-recorded the lesson and wrote a reflection report. PSTs' lesson plans were graded for scope and richness of proof-related tasks and correctness of mathematical explanations; the classroom teaching was *not* assessed to reduce performance pressure; the post-lesson reflections were graded for completion. For research purposes, we re-analyzed the entire data corpus, as described below.

3.4 Participants

The participants were 34 PSTs (22 females, 12 males) who took the capstone course in the three years of the project. The PSTs were in their final, fourth year of secondary mathematics education program, having completed most of their mathematical courses alongside mathematics majors, including proof-intensive courses like Mathematical Proof, Geometry, and Abstract Algebra (some PSTs took it concurrently with the capstone course). Also, all PSTs had completed one or two mathematics education courses, which had no classroom practicum component.

3.5 Data sources and analysis

We used a combination of descriptive statistics and qualitative analyses (Table 1).

To assess changes in PSTs' knowledge we compared their pre-post course performance on the MKT-P questionnaire (Cronbach's alpha = 0.892; moderate to good internal consistency). The number of QRE items varied between 9 to 11 across versions of the questionnaire, which changed slightly over the years. The items were distributed among the three

facets of MKT-P and mathematical topics: algebra, functions, and geometry; but we report on them in aggregate, since there were not enough data points in each set. Each item was scored on a scale from 0 to 3. Three points were given for a correct answer supported by a correct explanation; intermediate scores of 1 or 2 were given for partially correct responses. We used ANOVA and paired *t* test to compare pre-to-post-course means, with Bonferroni correction for multiple comparisons. To compensate for unequal variances, we used Tukey's Honest Significant Difference test.

Questionnaire data were triangulated using multiple sources: PSTs' responses to the course activities, lesson plans, and written artifacts. For each data source, we developed an analytic rubric based on our MKT-P framework. To examine the development of PSTs' content knowledge, we analyzed shifts in their responses to *What Can You Infer from This Example?* tasks before and after the PSTs examining students' examples, coding for the mathematical correctness and types of justifications.

To examine the PSTs' proof-specific knowledge of content and teaching, we analyzed the extent to which they integrated QRE in their lesson plans. Lesson planning is one of the core tasks of teaching, often used to assess pedagogical knowledge. Of the common criteria for evaluating lesson plans, we focused on objectives, formative assessment, explanations- adaptation of mathematical content to facilitate learning, and task design (Blömeke et al., 2008; Silver et al., 2009). Each lesson plan was scored on four parameters: (1) the ratio of QRE-related lesson objectives; (2) the ratio of QRE-specific discussion prompts in lesson's summary; (3) a dichotomous score of 1 or 0, based on inclusion, or not, of an explanation of the roles of examples in

Janet, a student in Mr. Brown's class, notices that ***increasing a numerator and denominator of a fraction by the same number results in a larger fraction.***

Mr. Brown asked the class what they think about Janet's conjecture. Consider students' responses.

Kacey: Let's say we take a random fraction $\frac{5}{7}$, we increase the numerator and the denominator by another random number, say 3. We get $\frac{5+3}{7+3} = \frac{8}{10}$. We can compare the two fractions by calculating a common denominator. We get: $\frac{5}{7} = \frac{50}{70}$ and $\frac{8}{10} = \frac{56}{70}$. Obviously, the new fraction is larger. Since we used random numbers in all our calculations, it is safe to say that any other numbers will produce the same result. Which means that Janet's conjecture is 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. 4 A KCS-P item *Kacey*

proving; (4) the ratio of tasks containing opportunities for students to engage with QRE. The unit of analysis was the smallest unit in which students were asked to do something, e.g., calculate, justify, generalize, prove. Adding the four ratios resulted in a numeric score ranging from 0 for lessons with no proof content, to 4 for lessons devoted to QRE entirely.

Following lesson enactment, the PSTs watched the video of their lesson, wrote a reflection on what they noticed (about 8–9 comments per 50-min video), and responded to prompts like: What QRE-related ideas were included in your lesson? How do you know if students understood them? We analyzed PSTs' pedagogical learning from these reflections using the concept of reflective noticing (Buchbinder et al., 2021), which combines teacher noticing (Dindyal et al., 2021) and reflection (Moore-Russo & Wilsey, 2014). We also analyzed PSTs' summative, post-course reflections, using open coding (Patton, 2002), to identify recurring themes of PSTs' perceived challenges with, and learning from the course activities.

4 Results

4.1 Enhancement of PSTs' knowledge of QRE

4.1.1 Evidence from the MKT-P questionnaire

We present evidence for qualitative and quantitative improvements in PSTs' pre-to-post course performance. Figure 4 shows a KCS-P item in which a student *Kacey*, attempted to prove a general conjecture with random examples. Successful completion of the item entails recognizing the flaw in Kacey's argument, rating it low and explaining the misconception.

To illustrate improvement in PSTs' performance, consider Eva's² response. On the pre-test Eva rated Kacey's argument high (3 out of 4) since "it is not a reasoning that can be used with only one example." The high rating and focusing on a single example may suggest Eva's own fragile understanding. On the post-test Eva rated Kacey's work low (2 out of 4) and explained:

Kacey shows a supporting example of the conjecture. Her example supports the statement. However, this one example is not enough information to prove the statement is true in general, she needs a general proof.

Here Eva used precise vocabulary and provided clear indication of understanding the limitation of supportive examples for proving universal statements, which require a general proof.

These types of changes were typical of many PSTs; reflecting stronger content knowledge and improved ability to identify students' misconceptions. The mean post-score on Kacey item in years two and three increased by a full point; in year one the mean score increased by 0.4 points due to a high mean pre-score of 2.57 out of 3 (on a scale from 0 to 3, as described in Sect. 3.5).

Figure 5 shows a KCT-P item in which a student *Sam* conjectured that 0 and 2 are the only numbers whose product is equal to their sum. This non-existence statement is equivalent to a universal statement: all numbers except 0 and 2, do *not* have this property. A student *Calvin* finds only pairs of numbers whose sum is not equal to their product and concludes that the conjecture is true. But the inability to find supportive examples is insufficient to prove nonexistence. In fact, there are infinite number pairs whose product is equal to

² All names are pseudonyms.

Sam noticed that $2 + 2 = 2 \cdot 2$, and $0 + 0 = 0 \cdot 0$. *Sam thinks that except for 0 and 2, there do not exist two numbers whose sum is equal to their product.*

The teacher, Mrs. Chang decided to have the class explore Sam's conjecture, and invited students to share their ideas. Below are their arguments.

Calvin: I tested multiple numbers: same and different, positive and negative, and even fractions and zero. I could not find any other numbers except for 2 and 0 that have this feature. $3 + 3 \neq 3 \cdot 3$, $4 + 5 \neq 4 \cdot 5$, $(-2) + 7 \neq (-2) \cdot 7$, $0 + 1 \neq 0 \cdot 1$, and even fractions: $\frac{1}{2} + \frac{1}{4} \neq \frac{1}{2} \cdot \frac{1}{4}$. I think this is something unique only to 2 and 0.

- 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. 5 A KCT-P item *Calvin*

Table 2 Change in performance on QRE portion of MKT-P questionnaire

	Year	No. of PSTs	No. of data points	Mean Pre	Mean Post	Effect size: Cohen's d ^a	p value	Confidence interval
	Year 1	14	126	2.135	2.603	0.4	<0.0001	(0.264, 0.672)
	Year 2	11	121	1.612	2.202	0.5	<0.0001	(0.402, 0.779)
	Year 3	9	90	1.011	1.772	0.7	<0.0001	(0.520, 1.002)

^aCohen's d values less than 0.5 is a small effect size; 0.5–0.8 medium effect size

their sum, e.g., 3 and $3/2$, 4 and $4/3$. Regardless of whether the PSTs noticed that, they were expected to identify a flaw in Calvin's argument and provide feedback on his work.

On the pre-test Bella complimented Calvin on trying multiple types of numbers but criticized the use of examples as proof, writing: "errors and weaknesses are that you can't prove by example with just a handful." This response makes one suspect whether more examples would be acceptable to Bella. On the post-test she wrote:

He used proof by examples to show no more exist, and that is a flaw in thinking for universal statement. Calvin tried more examples than Sam but that is not enough to prove that 2, 0 are only ones. His thinking of finding a counterexample is correct, but his counterexample must be a number that works to show 0, 2 aren't only it. Showing proof by example here isn't good enough.

This answer illustrates typical pre-post changes in PSTs' responses: the increased use of correct vocabulary and correct explanation of inapplicability of a proof by example. It also specifies what type of object constitutes a counterexample for the conjecture. The change in the mean score on this item was negligible in year one; but in years two and three it increased by 0.77 and 1.22 points respectively.

Table 2 shows the aggregated mean change in pre-post-course performance on all QRE items. The number of data

points is the product of the number of participants and the number of QRE items.

The effect size in year 1 was small, due to this cohort having relatively high pre-mean, leaving limited space for growth. The effect size in years two and three was medium. In each year, the pre-to-post performance improved significantly. These results may be partially explained by repeated exposure to proof content, but not entirely, since despite taking several proof-intensive courses, the pre-course means in all years were quite low (on 0–3 scale) (Table 2). Thus, PSTs' experiences in the capstone course likely contributed to the observed growth. PSTs' self-report data supports this assumption, as this quote shows:

As a student who was previously a pure mathematics major and have taken many proof-based classes before this course, this was the only course that allowed students to truly explore these types of proof and apply the knowledge gained from these ways of proving in an alternative setting.

This quote, and similar ones, suggest that PSTs perceived their engagement with proof in the capstone course as qualitatively different from their experiences in other mathematics courses. The PSTs expressed appreciation for the opportunities to contextualize their knowledge of proof and apply it in situations approximating teaching practice.

Table 3 Justifications of the truth-value of the statement: *a quadrilateral whose diagonals are congruent and perpendicular to each other is a kite* ($N=34$)

Truth value	Justification type/example	Before examining student work	After examining student work
True	“A square is a type of kite”; “non-convex kite satisfies the statement”	3	1
False	Correct or unspecified counterexample “Other quadrilaterals have these properties”	6	17
	Irrelevant counterexample “Square/Rhombus”	8	1
	Incorrect explanation: “Kites don’t have congruent diagonals”	17	7
	Both correct and irrelevant Counterexamples	0	8

The universal quantifier is implicit. The statement is equivalent to a conditional “If a quadrilateral has congruent and perpendicular diagonals, then it is a kite.” (For detailed analysis of this statement see Buchbinder et al., 2017)

Table 4 Justifications for the truth-value of the statement: *there exist three consecutive even numbers whose sum is divisible by four* ($N=34$)

Truth value	Justification type	Before examining student work	After examining student work
False	Non-confirming example used as a counterexample	1	1
True	Gave supportive example or said such examples exist	12	8
	Correct explanation that one supportive example proves existential statement	4	11
	Correct explanation and an example	17	14
	Attempt at general proof	1	1

4.1.2 Evidence from the What Can You Infer from This Example? activities

Tables 3 and 4 show the shifts in PSTs' justifications for the truth-value of the two statements, after examining student work.

Almost all PSTs correctly identified the truth-value of both universal and existential statements. However, only 6 out of 34 PSTs correctly justified the falsehood of the universal statement referring to the existence of a counterexample—a quadrilateral with congruent and perpendicular diagonals, which is not a kite. Meaning that most of PSTs' initial justifications were incorrect (Table 3). Eight PSTs considered a square or a rhombus as counterexamples, which is incorrect, since both are kites. Half of the PSTs wrote that “kites do not have congruent diagonals.” This is geometrically imprecise and logically inapplicable to disprove this statement.

After examining students' examples, the number of PSTs' incorrect justifications and irrelevant counterexamples decreased, while the number of correct responses more than doubled. These shifts occurred even before the whole class discussion, solely due to the PSTs' interaction with the activity. The increase from 0 to 8 in the number of

justifications that use *both* correct and irrelevant counterexamples is problematic. It suggests that the PSTs maintained their initial incorrect idea while also accepting correct counterexamples as legitimate.

For existential statements (Table 4), almost all PSTs initially proved it with a correct supportive example or explained that such an example exists, or both. After exploring student work there was an increase in the number of correct general explanations that a single supportive example proves an existential statement. We interpret this as PSTs' increased ability to verbalize the mathematical warrant behind existential proof.

The follow-up whole-class discussion was centered on crystallizing content and pedagogical knowledge of proof through collective analysis of the hypothetical student work and of the PSTs' own anonymized responses. The PSTs found this process beneficial as this comment shows:

Sometimes, a student's reasoning seemed valid and correct to me, but I later learned how their reasoning could be more developed, or why it was invalid. Over time, I improved upon recognizing invalid proofs and techniques.

The PSTs attributed their improvement to the situated nature of the activities and the opportunity to

Number of congruent elements in two triangles	Options of congruent elements in two triangles	True or False: “If two triangles have <u> </u> corresponding <u> </u> that is/are congruent, then the triangles are congruent”	True or False: “There exist two triangles that have <u> </u> corresponding <u> </u> that is/are congruent such that the triangles are congruent.”
One congruent element	1 angle	“If two triangles have 1 corresponding angle that is congruent, then the triangles are congruent.”	“There exist two triangles that have 1 corresponding angle that is congruent such that the triangles are congruent.”
	1 side		
Two congruent elements	2 angles	If two triangles have 2 corresponding congruent angles, then the triangles are congruent.	There exist two triangles that have 2 corresponding congruent angles, such that the triangles are congruent.
	1 side and 1 angle		
Three congruent elements	3 sides		
	3 angles		
	2 sides and 1 angle		
Four congruent elements	3 sides and 1 angle		

Fig. 6 Silvia’s worksheet integrating QRE in a lesson on congruent triangles

analyze student mathematical work both individually and collectively.

4.2 Enhancement of PSTs’ QRE-related practices

4.2.1 Lesson planning

The *apply* activities involved lesson planning, enactment, and reflection. The PSTs designed QRE-integrated lessons in a variety of mathematical topics and grade levels such as: equations, proportional reasoning, functions, congruent triangles, parallel lines, standard deviation, and matrix operations.

The success of QRE integration in the lesson plans varied. There were 16 lesson plans (47%) with none or low QRE integration scores of 0–1.3 (see the Methods section). For example, Jane’s high school-level lesson on matrices had a QRE integration score of 0.4 out of 4. The plan contained no explanation of quantified statements or the role of examples in proving, and no relevant summary questions, as Jane expected students to “learn about quantified statements indirectly.” Of the eight tasks in the lesson plan, two were *True–False* questions about matrix multiplication, one universal and one existential.

There were 12 lessons (35%) with the medium QRE integration scores of 1.3–2.6. These lessons mentioned QRE in either the objectives, explanations, or summary, and/or had a relatively high ratio of QRE-related tasks. For example, Emily’s lesson on quadratic functions scored 2.2 on QRE

integration. Of the six lesson objectives, one was “students will learn about quantification and the role of examples in proving”. Of the 17 tasks, 12 were devoted to QRE. Emily had students examine six *Always–Sometimes–Never* questions about quadratic functions, e.g., “knowing the vertex and the y-intercept we can graph the parabola,” and then consider how many examples would be sufficient to prove/disprove each statement depending on if it is always, sometimes, or never true. Emily used these questions to discuss the role of examples in proving/disproving quantified statements. One out of three lesson summary questions assessed student understanding of this topic.

Six lesson plans (18%) had high QRE-integration scores of 2.6–4. Such lessons contained explicit explanations of the role of examples in proving/disproving quantified statements, QRE-related objectives and summative questions, and at least half of the tasks directly dealing with QRE. For example, in Silvia’s lesson on congruent triangles (QRE integration score of 3.15) students had to prove or disprove eight pairs of quantified statements, which differed only by the type of quantifier (Fig. 6). Moreover, students had to formulate the statements themselves, from the given information about the pairs of triangles. The lesson contained an exposition on proving/disproving quantified statements; three out of four lesson objectives and two out of five summative questions specifically addressed QRE.

A prevalent theme in the PSTs’ post-course reflections was that lesson planning was “the most challenging” but also “most worthwhile” aspect of the course. Angie wrote:

Table 5 Categories indicating PSTs' learning from reflection

Category	Example PST reflection response
Reflecting on one's teaching in relation to student learning	A student asks for help on the first problem. Eventually we come to an answer together. Overall, I think going over this problem step by step really helped him
Reflecting on student understanding	There appears to be only one student, who admittedly does not understand the jump from solving the linear equation to deciding that the entire statement is false. I am trying to explain this to her. I'm not sure if she actually understands
Critically assessing pedagogical choices	I like the way I asked students why we can say those four sides are all congruent [...] but I don't like the way I chose to explain why we can't assume the third side is congruent
Contemplating alternative teaching move	I should have thought about how this is a new concept to them and should have done a better job of explaining and encouraging questions
Connecting to a general pedagogical principle	These kinds of discussions are very important in a math classroom because it allows students to see their peers' thinking
Reflecting on the efficiency of an in-the-moment decision	Since I was hearing various answers , I called on each student to tell me what they think. It was essential to intervene and assist students because this led to them being able to prove the conjecture

While incorporating the proof themes into our lessons was challenging, it was also very eye-opening into the multitude of ways that higher-level mathematics topics can be brought into lower-level subjects.

In summary, 53% of the lesson plans had medium or high QRE integration scores, suggesting that despite the challenges, about half of the PSTs succeeded in designing lesson plans that creatively integrated QRE with regular mathematics topics. Notably, 60% of the lessons used *True–False* or *Always–Sometimes–Never* task formats. The post-course reflections provided supportive evidence that PSTs drew inspiration from the course activities in their lesson plans.

4.2.2 PSTs' learning from reflection on enacted lessons

We outline the main findings of the analysis of PSTs' reflections on enacted lessons to illustrate their effect on PSTs' pedagogical growth (for more details see Buchbinder et al., 2021). Effective reflection entails teachers noticing multiple aspects of classroom environment, critically analyzing them, and making connections to past experiences, theoretical principles, and future actions (Moore-Russo & Wilsey, 2014). The analysis revealed four broad categories of PSTs' noticing: mathematical content, teaching, students, and interactions. The two modal categories, accounting for 45% of codes, were PSTs' noticing their own teaching of mathematics and their interactions with students. This is typical of novice teachers, who tend to focus on aspects of the classroom situation directly involving them (Dindyal et al., 2021).

Next, we used literature-based categories to identify evidence of PSTs' learning from reflections (shown in Table 5 in the descending order of frequency).

Given the foci of PSTs' noticing, it is not surprising that the modal category (33% of codes) was *reflecting on one's teaching in relation to student learning*. Post-course reflections support this observation, as this comment shows:

It was a learning experience to teach the lesson, but a great deal of learning occurred watching the videos because I could analyze my lesson in depth. I was able to go back in time to my lesson and see how what I said or what I did affected the discourse that took place.

The added value of this analysis is illustrating that PSTs' reflections bear characteristics of effective reflection, which may have contributed to PSTs' learning.

5 Discussion

5.1 Summary and limitations

We presented evidence of PSTs' learning from the Quantification and the Role of Examples in Proving (QRE) module, and from the course overall. The evidence for strengthened PSTs' content knowledge is shown in significant improvement in PSTs' performance on the QRE portion of the MKT-P questionnaire and an increased number of correct justifications on *What Can You Infer from This Example?* activities, even prior to the whole class discussion. The evidence for enhanced pedagogical knowledge and practices came from PSTs' lesson plans and reflections on enacted lessons. PSTs' written comments provide additional support, connecting course activities to their learning, although self-reported data should be treated with caution.

These results must be interpreted within methodological limitations of the study. First, the number of the participants is small, not allowing for generalizing the outcomes. Second, the feasibility of assessing MKT with questionnaires has been contested due to the situated nature of teacher knowledge (Charalambous, 2020). Our study followed the established methodological tradition of using questionnaires for assessing *cognitive* aspects of MKT-P (cf., Krauss et al., 2008; Tatto, 2013) while *practices* were assessed using scenario-based instruments (e.g., Zazkis et al., 2013) and a lesson enactment cycle. Third, since the course design and measurements are aligned, as typical of design research (Sandoval, 2014), the observed improvement in MKT-P scores may be, at least partially, attributed to this alignment. To mitigate these challenges we used multiple data sources to triangulate evidence of PSTs' enhanced knowledge of QRE. Another limitation related to the nature of MKT is that if teacher knowledge or competence are special to the teaching profession and grow with experience (Krauss et al., 2008; Shulman, 1986), it is unclear how much change can occur in a single semester. We addressed this challenge by treating lesson enactment as a non-assessed learning experience, and only considering lesson planning and reflective noticing as indicators of growth of MKT-P (cf., Blömeke et al., 2008; Dindyal et al., 2021).

Finally, the absence of a control group, although typical of design research (Gravemeijer & Prediger, 2019), makes it difficult to attribute the outcomes to the course. In a related study (Buchbinder et al., 2022) we partially address this issue by comparing PSTs' MKT-P performance to that of other mathematically knowledgeable groups, like undergraduate mathematics majors and inservice secondary teachers. That study showed that PSTs' post-course performance is closer to that of inservice teachers, who outperformed all other groups. However, such comparison is not a substitute for a controlled study, which can be conducted in the future.

5.2 Design principles: *crystallize*–*connect*–*apply*

University mathematics programs seek to bridge the double discontinuity (Klein, 1932) and to address the needs of future mathematics teachers in various ways (Tatto, 2013). Instructors may explicate connections between university and school mathematics by introducing tasks grounded in secondary context or adopt curricula emphasizing such connections (e.g., Wasserman et al., 2023). To the best of our knowledge, these efforts are less common with respect to the topic of proof. Our study addressed this gap by designing a novel capstone course, bridging between university-level proof and secondary teaching. Although the course targets both content and pedagogical MKT-P, its focus is inherently mathematical, organized around four proof themes. This approach asserts the primacy of subject matter knowledge

in teachers' knowledge (Harel, 2008), and positions teaching as a form of applied mathematics (Stylianides & Stylianides, 2010).

Since adding a whole course to an existing program may not be feasible for most universities, it is important to clarify the contributions of our study. The outcomes of design research are not intended to be generalized in the same way as those of experimental research, but the design principles generated in the retrospective analysis may apply beyond the research context (Gravemeijer & Prediger, 2019). The motivation for the three design principles: *crystallize*–*connect*–*apply*, came from extensive analysis of research literature (see Method section), which typically treats each principle separately. Our study distills these design principles from various strands of literature and illustrates how they can be integrated in a holistic design, embodied in specific activities, which correspond to three types of learning opportunities afforded by the course.

In the QRE module, the *crystallize* design principle: *Use activities that strengthen (crystallize) PSTs' content knowledge specific to proof*, was embodied in the activities *What Can You Infer from This Example?*, *True–False*, and *Always–Sometimes–Never*. These activities aimed to crystallize PSTs' knowledge of the roles of examples in proving/disproving quantified statements relying solely on content knowledge. The use of secondary mathematics content in the statements intended to focus PSTs' attention on the logical aspects of QRE rather than on the complexity of university-level content (Dawkins, 2017). Concurrently, the use of secondary mathematics content in these tasks supported the second design principle: *Connect university-level knowledge of proof with secondary school mathematics*. In our study, this principle entails two types of connections (1) the mathematical connections between university-level proof and secondary curriculum, and (2) pedagogical connections to student mathematical conceptions. In the QRE module these connections were closely intertwined and linked with the first design principle. That is, the opportunities to strengthen content knowledge of QRE were situated in secondary classroom contexts and linked with pedagogical opportunities to examine students' conceptions. By presenting mathematical arguments as products of student work, the activities support positioning of PSTs as future teachers, affecting their interactions with the tasks (e.g., Baldinger & Lai, 2019).

The *crystallize* and *connect* design principles may be implemented in a variety of courses. Tasks requiring analysis of fictitious students' arguments have been successfully used in university courses like Calculus, Real Analysis, Abstract Algebra and capstone courses to strengthen connections between university and school mathematics (Álvarez, et al., 2022; Wasserman et al., 2019; Winsløw & Grønbæk, 2014). Analyzing written arguments, identifying mistakes, and responding to incorrect solutions provide students in

university mathematics courses with opportunities to reflect on their own mathematical understanding, become aware of their own possible misconceptions and engage in self-explanation and justification (Hodds et al., 2014). Thus, tasks embedding the *crystallize* and *connect* design principles could benefit PSTs, and other majors, in a variety of university mathematics courses. Future studies may examine applications of our tasks, or modifications thereof, in proof-intensive university courses.

The *apply* design principle: *Provide opportunities to apply (enact) content and pedagogical knowledge specific to proof in environments approximating school teaching* was realized through the plan-enact-reflect lesson cycle. Originally, we envisioned this as an assessment component but early on realized that this is a learning opportunity for the PSTs to develop practical skills for integrating proof in teaching mathematics. The benefits of including this design principle in our study are evident in the relative success of PSTs in designing QRE-oriented lessons, the improved performance on pedagogical items of the MKT-P questionnaire, and PSTs' self-report. The *apply* design principle emphasizes the importance of engaging future teachers with approximations of practice (Grossman et al., 2009) and the value of reflecting on one's practice (Moore-Russo & Wilsey, 2014).

Implementing the *apply* design principle in our study required conditions that may be challenging to replicate: an instructor specializing in mathematics education, the course being offered to PSTs exclusively, and coordination with local schools. However, some alternatives to lesson planning and enacting can be utilized in other university courses. For example, scripting tasks and lesson plays, where students explain a concept or a procedure in the form of a written scenario and visual images have been used in courses like Abstract Algebra, problem-solving for secondary teachers, and others (Zazkis et al., 2013; Zazkis & Herbst., 2017).

Our study serves as a proof of existence for the possibility of supporting PSTs' learning to teach reasoning and proving at the secondary level via a university mathematics course. Future studies may explore the effects of individual design principles, modules, or activities in other settings.

Acknowledgements 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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not

permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Álvarez, J. A., Kercher, A., Turner, K., Arnold, E. G., Burroughs, E. A., & Fulton, E. W. (2022). Including school mathematics teaching applications in an undergraduate abstract algebra course. *Primus*, 32(6), 685–703.
- Antonini, S., & Mariotti, M. A. (2008). Indirect proof: what is specific of this mode of proving? *ZDM - the International Journal on Mathematics Education*, 40(3), 401–412.
- Australian Curriculum, Assessment and Reporting Authority (ACARA). (2022). Critical and creative thinking (version 8.4). <https://www.australiancurriculum.edu.au/f-10-curriculum/general-capabilities/critical-and-creative-thinking/>
- 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.
- Blömeke, S., Hsieh, F.-J., Kaiser, G., & Schmidt, W. (Eds.). (2014). *International perspectives on teacher knowledge, beliefs and opportunities to learn*. Springer.
- Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (Eds.). (2015). *Assessment of competencies in higher education*. Hogrefe.
- Blömeke, S., Paine, L., Houang, R. T., Hsieh, F. J., Schmidt, W. H., Tatto, M. T., et al. (2008). Future teachers' competence to plan a lesson: First results of a six-country study on the efficiency of teacher education. *ZDM - the International Journal on Mathematics Education*, 40(5), 749–762.
- Buchbinder, O., Brisard, S., Butler, R., & McCrone, S. (2021). Pre-service secondary mathematics teachers' reflective noticing from 360-degree video recordings of their own teaching. *Journal of Technology and Teacher Education*, 29(3), 279–308.
- Buchbinder, O., & McCrone, S. (2020). Preservice teachers learning to teach proof through classroom implementation: Successes and challenges. *Journal of Mathematical Behavior*, 58, 100779.
- Buchbinder, O., McCrone, S., Butler, R., & Capozzoli, M. (2022). Mathematical knowledge for teaching proof: Comparing secondary teachers, pre-service secondary teachers, and undergraduate majors. *International Journal for Research in Mathematics Education*. <https://doi.org/10.1007/s40753-022-00187-8>
- Buchbinder, O., Ron, G., Zodik, I., & Cook, A. (2017). What can you infer from this example? Applications of on-line, rich-media task for enhancing pre-service teachers' knowledge of the roles of examples in proving. In A. Leung & J. Bolite-Frant (Eds.), *Digital technologies in designing mathematics education tasks – Potential and pitfalls* (pp. 215–235). Springer.
- Buchbinder, O., & Zaslavsky, O. (2019). Students' understanding of the role of examples in proving: Strengths and inconsistencies. *Journal of Mathematical Behavior*, 53, 129–147.
- Charalambous, C. Y. (2020). Reflecting on the troubling relationship between teacher knowledge and instructional quality and making a case for using an animated teaching simulation to disentangle this relationship. *ZDM*, 52(2), 219–240.
- 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.
- Dawkins, P. C. (2017). On the importance of set-based meanings for categories and connectives in mathematical logic. *International Journal of Research in Undergraduate Mathematics Education*, 3, 496–522.
- Dindyal, J., Schack, E. O., Choy, B. H., & Sherin, M. G. (2021). Exploring the terrains of mathematics teacher noticing. *ZDM*, 53, 1–16.
- 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.
- Goulding, M., Hatch, G., & Rodd, M. (2003). Undergraduate mathematics experience: Its significance in secondary mathematics teacher preparation. *Journal of Mathematics Teacher Education*, 6, 361–393.
- Gravemeijer, K., & Prediger, S. (2019). Topic-specific design research: An introduction. In G. Kaiser and N. Presmeg (Eds.), *Compendium for early career researchers in mathematics education*, ICME-13 Monograph 33–57.
- Grossman, P., Compton, C., Igla, D., Ronfeldt, M., Shah, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, 111(9), 2065–2100.
- Harel, G. (2008). A DNR perspective on mathematics curriculum and instruction. Part II: with reference to teacher's knowledge base. *ZDM - the International Journal on Mathematics Education*, 40, 893–907.
- 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). NCTM.
- Hodds, M., Alcock, L., & Inglis, M. (2014). Self-explanation training improves proof comprehension. *Journal for Research in Mathematics Education*, 45(1), 62–101.
- Klein, F. (1932). *Elementary mathematics from an advanced standpoint: Volume I: Arithmetic, algebra, analysis*. Springer.
- Ko, Y. Y. (2010). Mathematics teachers' conceptions of proof: Implications for educational research. *International Journal of Science and Mathematics Education*, 8, 1109–1129.
- Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM - the International Journal on Mathematics Education*, 40(5), 873–892.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820.
- Lee, K. (2016). Students' proof schemes for mathematical proving and disproving of propositions. *The Journal of Mathematical Behavior*, 41, 26–44.
- 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 G. Hanna & M. de Villiers (Eds.), *Proof and Proving in Mathematics Education*. The 19th ICMI Study (pp. 327–346). Springer.
- Mariotti, M. A., Durand-Guerrier, V., & Stylianides, G. J. (2018). Argumentation and proof. In T. Dreyfus, M. Artigue, D. Potari, S. Prediger, & K. Ruthven (Eds.), *Developing research in mathematics education* (pp. 75–89). Routledge.
- MINEDUC. (2019). *Bases curriculares IIIº y IVº Medio*. MINEDUC.
- Moore-Russo, D. A., & Wilsey, J. N. (2014). Delving into the meaning of productive reflection: A study of future teachers' reflections on representations of teaching. *Teaching and Teacher Education*, 37, 76–90.
- National Council of Teachers of Mathematics. (2009). *Focus in high school mathematics: Reasoning and sense making*. 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.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods*. Sage.
- Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences*, 23(1), 18–36.
- 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 - the International Journal on Mathematics Education*, 40(5), 791–811.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Silver, E. A., Mesa, V. M., Morris, K. A., Star, J. R., & Benken, B. M. (2009). Teaching mathematics for understanding: An analysis of lessons submitted by teachers seeking NBPTS certification. *American Educational Research Journal*, 46(2), 501–531.
- 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. (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, 119–127.
- 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.
- Tabach, M., Levenson, E., Barkai, R., Tirosh, D., Tsamir, P., & Dreyfus, T. (2010). Secondary school teachers' awareness of numerical examples as proof. *Research in Mathematics Education*, 12(2), 117–131.
- Tatto, M. T. (2013). *The Teacher Education and Development Study in Mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries*. Technical Report. International Association for the Evaluation of Educational Achievement.
- 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.
- Wasserman, N., Buchbinder, O., & Buchholtz, N. (2023). Making university mathematics matter for secondary teacher preparation. *ZDM Mathematics Education*. <https://doi.org/10.1007/s11858-023-01484-5>
- Weber, K. (2010). Mathematics majors' perceptions of conviction, validity, and proof. *Mathematical Thinking and Learning*, 12(4), 306–336.
- Winsløw, C., & Grønbæk, N. (2014). Klein's double discontinuity revisited. *Recherches En Didactique Des Mathématiques*, 34(1), 59–86.
- Zazkis, R., & Herbst, P. (Eds.). (2017). *Scripting approaches in mathematics education: Mathematical dialogues in research and practice*. Springer.
- Zazkis, R., Sinclair, N., & Liljedahl, P. (2013). *Lesson play in mathematics education: A tool for research and professional development*. Springer.