Article presented at the 11th Conference on Research in Undergraduate Mathematics Education,

San Diego, California, 2008:

ENHANCING UNDERGRADUATE STUDENTS' UNDERSTANDING OF PROOF

Gabriel J. Stylianides*

Andreas J. Stylianides**

* *University of Pittsburgh, USA*

E-mail: gstylian@pitt.edu

** *University of Oxford, UK*

E-mail: andreas.stylianides@education.ox.ac.uk

Author Note

The two authors had an equal contribution to the preparation of this article, which was supported by funds from the Spencer Foundation (Grant Numbers: 200700100 and 200800104). The opinions expressed in the article are those of the authors and do not necessarily reflect the position, policy, or endorsement of the Spencer Foundation.

Introduction

Although the concept of proof has long been viewed as the cornerstone of mathematical activity, students show limited understanding of it. For example, the *empirical justification scheme* (Harel & Sowder, 2007) is pervasive among school students including high-attaining secondary students (e.g., Healy & Hoyles, 2000) and university students including mathematics majors (e.g., Selden & Selden, 2003; Sowder & Harel, 2003). In other words, these students tend to formulate or accept empirical arguments as proofs of mathematical generalizations. By *empirical arguments* we mean arguments that provide inconclusive evidence for the truth of a generalization by verifying its truth in a proper subset of all the possible cases covered by the generalization, whereas by *proofs* we mean logical arguments that provide conclusive evidence for the truth of a generalization by treating appropriately all cases covered by the generalization.

To date, there has been limited research knowledge about how university instructors can help undergraduate students develop their knowledge about proof. In this article, we take a step toward addressing this need for research by focusing on the following question:

> What might be a promising *instructional sequence* (i.e., a series of tasks and associated instructor actions in implementing the tasks) that university instructors can implement to help undergraduate students *start to develop* an understanding of the limitations of empirical arguments and an appreciation of the importance of proofs as means for establishing the truth of mathematical generalizations?

Our work to address this research question was part of a four-year design experiment that aimed to develop, implement, and analyze the effectiveness of instructional sequences in an undergraduate mathematics course. This course was prerequisite for admission to the masters-level elementary teacher education program at the University of Pittsburgh and was attended primarily by juniors who pursued different majors. Also, because the course was the only required mathematics (content) course for the students who ultimately entered the program, it covered a range of topics/ideas in all major mathematical domains (algebra, geometry, etc.).

Theoretical Perspective

The theoretical perspective that underpinned the development of the focal instructional sequence and other instructional sequences in the course had four major features. The first feature concerned the place of proof within a broader scheme of mathematical work. Students' engagement with proof in the course was part of their engagement with the activity of *reasoning-and-proving* (Stylianides, 2008), a hyphenated term that encompasses a family of activities involved in the investigation of whether and why 'things work' in mathematics (e.g., making generalizations and developing arguments for these generalizations that may qualify as proofs). The course treated reasoning-and-proving as a vehicle to sense making and as a process (strand) that underpinned students' mathematical work on all topics and across mathematical domains. The development of proofs within this broader scheme of mathematical work served as a means for: (1) promoting mathematical understanding through *explanation* (e.g., Hanna & Jahnke, 1996); and (2) arriving at *conviction* (e.g., Bell, 1976) at both the individual and social levels.

The second feature of our theoretical perspective concerned the construct of *learning trajectories* (Simon, 1995). Clements and Sarama (2004, p. 83) argued that the power of this construct stems from the *inextricable relationship* between developmental progressions of learning (i.e., advancements in the level of understanding of a particular mathematical topic/idea) and instructional sequences designed to move students along these developmental progressions to achieve specific learning goals. Thus, our approach to developing instructional sequences in the course was based on the premise that learning trajectories cannot be studied in isolation from main components of instruction, notably, the tasks with which students interact and the teacher actions associated with the implementation of these tasks in the classroom.

The third feature of our theoretical perspective concerned the use of *cognitive conflict* (e.g., Zaslavsky, 2005) as a major mechanism to create and support an *intellectual need* (Harel,

1998) for developmental progressions in students' mathematical knowledge in general and knowledge about proof in particular. The main idea of the 'conflict teaching' approach in our course was to help students reflect on their current mathematical understandings, confront contradictions that arose in situations where some of these understandings no longer held, and recognize the importance of modifying these understandings to resolve the contradictions. In order to address the issue that students often possess contradictory understandings without however facing or acknowledging a cognitive conflict, we incorporated into the instructional sequences strategic selections of: (1) *pivotal examples* (Zazkis & Chernoff, in press), that is, examples (including counterexamples) that our teaching experience showed to be successful in creating a dissonance in students' incorrect or incomplete understandings of a particular topic or idea; and (2) *conceptual awareness pillars*, a theoretical construct that we introduce to describe instructional activities that aimed to direct students' 'attention' (cf. Mason, 1998) to their understandings of a particular topic or idea in order to increase, for example, the possibility that students would experience a cognitive conflict when encountering examples intended to create a dissonance in their existing understandings (thereby making these examples *pivotal* for them).

The fourth feature of our theoretical perspective concerned two primary means for supporting the resolution of cognitive conflicts experienced by students and the development of shared knowledge in the class that better approximated conventional mathematical knowledge: (1) social interactions among the class members (cf. Simon & Blume, 1996), and (2) an active role of the instructor who was viewed as the representative of the mathematical community in the classroom (cf. Yackel & Cobb, 1996). The instructor had the responsibility to not only scaffold students' work (by asking probing questions, etc.) and help them become (more) aware of their current understandings (cf. conceptual awareness pillars), but also to offer students access to conventional knowledge for which they saw the intellectual need but were unable to develop on their own (due to conceptual barriers, etc.). The latter aspect of the instructor's responsibility was consistent with Hanna and Jahnke's (1996) remark in the domain of proof at the school level that "[a] passive role for the teacher … means that students are denied access to available methods of proving" and that "[i]t would seem unrealistic to expect students to rediscover sophisticated mathematical methods or even the accepted modes of argumentation" (p. 887).

Method

Design Experiment Methodology

To develop the focal and other instructional sequences in the course we followed design experiment methodology (see, e.g., Schoenfeld, 2005), which included the following:

(1) *A design feature*: the design of instructional sequences (using theory and research as described earlier) intended to support learning trajectories toward the development of specific aspects of knowledge about mathematics in general and proof in particular.

(2) *An empirical feature*: the study of the implementation of the instructional sequences through collection and analysis of data on both students' *actual learning trajectories* (Leikin & Dinur, 2003), that is, the learning routes that students seemed to follow as a result of the implementation of the instructional sequences, and the ways in which the instructional sequences seemed to support these trajectories.

(3) *An iterative feature*: research cycles of design, implementation, analysis, and refinement, so as to better understand students' learning trajectories toward the development of specific aspects of knowledge and to adapt the instructional sequences to these trajectories thereby achieving closer matching between actual and hypothetical learning trajectories, whereby *hypothetical learning trajectories* (Simon, 1995) we refer to the learning routes that students were anticipated to follow as a result of the implementation of the instructional sequences.

In regard to the third feature, we conducted five research cycles over the last four years that included eight enactments of the course in two U.S. institutions with at least one of us as instructor. Our 'critical reflection' on the implementation of the instructional sequences (cf. Zaslavsky, 2005), especially in relation to the actual and hypothetical learning trajectories, fed into cycles of analysis and refinement of the instructional sequences. Due to space constraints, we only present the final version of the focal instructional sequence as implemented in cycle #5 of our design experiment. The presentation is intended to exemplify the theoretical perspective that underpinned the design of the sequence and it suggests the promise of the sequence in supporting the intended learning goals in the domain of proof.

Research cycle #5 was conducted in two sections of the course at the University of Pittsburgh. The two sections were taught by the first author and were attended by a total of 39 students. The data for the article relate to the implementation of the focal instructional sequence in one of the two sections of the course that was attended by 18 students. The data include: video records of relevant classroom episodes, fieldnotes that focused on students' small group work, and students' written responses to questions that the instructor posed at several instances during the implementation of the instructional sequence to elicit their thinking and provoke reflection.

The Focal Instructional Sequence

Figure 1 summarizes the *interconnection* (cf. feature #2 of our theoretical perspective) between major elements of the focal instructional sequence (tasks and conceptual awareness pillars [CAPs]) and corresponding elements of the classroom community's hypothetical learning trajectory (developmental progressions) that the instructional sequence was intended to support. As will be shown in the next section, the implementation of the instructional sequence in research cycle #5 achieved close matching between the community's hypothetical and actual learning trajectories. We clarify that our focus in this article is on the learning trajectory of *the class as a community of learners* rather than on learning trajectories of individual students.

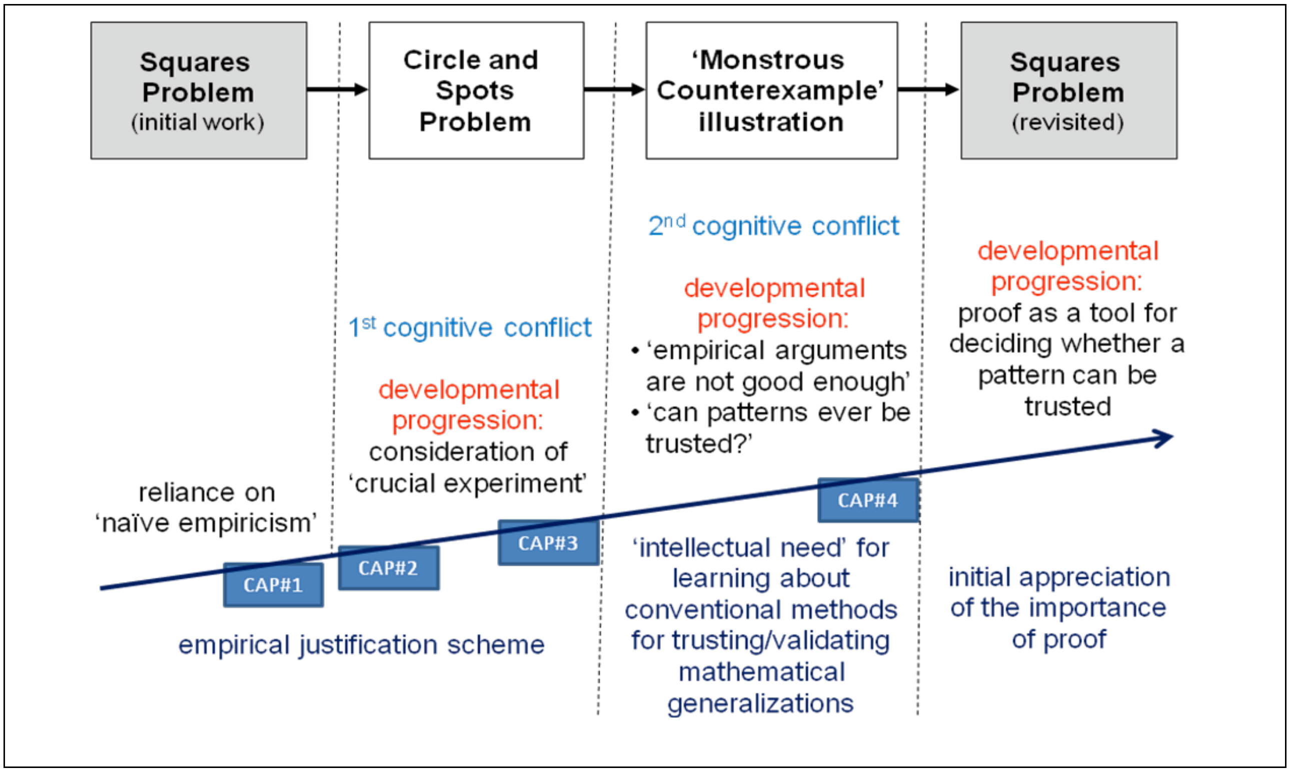
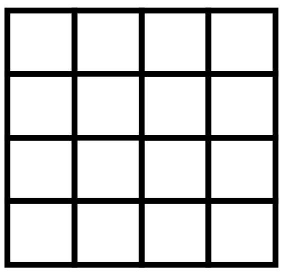


Figure 1. Major elements of the focal instructional sequence and corresponding elements of the classroom community's hypothetical learning trajectory.

Next we describe our rationale for the design of the instructional sequence (including explanation of important changes we made in earlier enactments of the sequence), with particular attention to our considerations related to selecting and sequencing the tasks in the sequence.

There were three main reasons for which we selected the Squares Problem (figure 2) to be the first problem in the instructional sequence. First, it had the potential to support meaningful engagement with the concept of proof in the broader context of *reasoning-and-proving* (cf. feature #1 of our theoretical perspective). Second, it gave rise to an interesting numerical pattern whose proof was not straightforward, thus having the potential to provoke students' *empirical justification schemes* (Harel & Sowder, 2007) when they were asked to say what would happen in a large case that was practically difficult to check empirically (see part #3). Students' justifications were likely to take the form of *naïve empiricism* (Balacheff, 1988), whereby 'naïve empiricism' we refer to the special kind of empirical argument that validates a generalization (e.g., a pattern) based on the confirming evidence offered by few cases (usually the selected cases are the first few terms of the corresponding sequence). Third, the development of a proof for the pattern that would *convince* and *explain* (cf. feature #1 of our theoretical perspective) was within students' conceptual reach after some scaffolding from the instructor.



1. Find the number of all different squares in the 4-by-4 square on the left.
2. What if this was a 5-by-5 square?
3. What if this was a 60-by-60 square? How would you work to find how many different squares there would be? How would you make sure that you found them all?

Figure 2. The Squares Problem.

When implementing the Squares Problem in the early research cycles of our design experiment, we were encouraging the students to develop an argument that would guarantee and explain the validity of the identified pattern in the 60th case, but the students were resistant to abandon their empirical arguments (see Simon and Blume [1996] for a description of a similar situation with prospective teachers). Their resistance was presumably due to that they did not see an *intellectual need* (Harel, 1998) for doing so. In later research cycles, we decided to go along with students' propensity to operate empirically and allow them to conclude (tentatively) their work on the Squares Problem without a proof for the pattern. Yet, we helped students develop awareness of the naïve empirical basis on which they trusted the pattern (see *conceptual awareness pillar* [*CAP*] #1 in figure 1) in order to make it more likely that they would experience a cognitive conflict in the next activity in the instructional sequence.

Place different numbers of spots around a circle and join each pair of spots by straight lines. Explore a possible relation between the number of spots and the greatest number of non-overlapping regions into which the circle can be divided by this means.

When there are 15 spots around the circle, is there an easy way to tell for sure what is the greatest number of non-overlapping regions into which the circle can be divided?

Figure 3. The Circles and Spots Problem (adapted from Mason et al. [1982]).

The Circle and Spots Problem (figure 3) was intended to challenge students' reliance on naïve empirical arguments by: (1) engaging them in the identification of a numerical pattern that they were expected to trust and apply in a large case that would be practically difficult for them to check empirically (like in part #3 of the Squares Problem), and (2) creating a *cognitive conflict* for them (cf. feature #3 of our theoretical perspective) with the discovery of a counterexample to the pattern. The two conceptual awareness pillars (CAPs #1 and #2) that preceded students' engagement with the Circles and Spots Problem were expected to support students' awareness of a cognitive conflict, thereby making the counterexample *pivotal* for students (Zazkis & Chernoff, in press) and creating an intellectual need in them for learning about conventional (secure) methods for trusting/validating mathematical generalizations. However, our experience in implementing the Circle and Spots Problem showed that several students resolved the conflict by adopting a more sophisticated form of empirical argument than naïve empiricism, namely, *crucial experiment* (Balacheff, 1988). In crucial experiment, a generalization is validated based on the confirming evidence offered by few cases that are selected based on some kind of rationale. This new conception of what it means to trust/validate a generalization was formed (in part) because the counterexample in the Circle and Spots Problem corresponded to the sixth term of the relevant number sequence, and thus a more careful and thorough examination of cases could reveal the exception and prevent trusting the pattern for use in larger cases. In order to challenge and support a new developmental progression in students' conception of what it means to trust/validate a mathematical generalization, we introduced a new *cognitive conflict* in students' activity with the help of the 'Monstrous Counterexample' illustration (figure 4).[1]

> Consider the following statement:
>
> The expression $1+1141n^2$ (where n is a natural number)
> *never* gives a square number.
>
> People used computers to check this expression and found out that it does ***not*** give a square number for any natural number from 1 to 30,693,385,322,765,657,197,397,207.
>
> BUT
>
> It *gives* a square number for the next natural number!!!

Figure 4. The 'Monstrous Counterexample' illustration (adapted from Davis [1981]).

The Monstrous Counterexample illustration was selected because it presented a pattern that held for a huge number of cases (of the order of septillions) but ultimately failed, thus challenging the legitimacy of using any kind of empirical argument to trust/validate a mathematical generalization. In order to make this counterexample *pivotal* for students, we inserted into the instructional sequence two additional conceptual awareness pillars (see CAPs #3 and #4 in figure 1), which aimed to direct students' attention to their current conceptions of what it means to trust/validate a mathematical generalization and to the implications of the Monstrous Counterexample illustration for the veracity of these conceptions.

Our expectation was that the Monstrous Counterexample illustration would help students realize the limitations of empirical arguments and would create an intellectual need in them for learning about conventional (secure) methods for trusting/validating mathematical generalizations (in this case, patterns). A byproduct of students' engagement with the illustration could be that some of them would start to develop distrust in any pattern and raise doubts about the possibility of finding a method for validating mathematical generalizations. Yet, such

---

[1] The name 'Monstrous Counterexample' was not mentioned in the class. We use this name in the article for purposes of easy reference.

concerns could be seen as another indication of their intellectual need to access conventional mathematical knowledge. The instructor as the representative of the mathematical community in the classroom (cf. feature #4 of our theoretical perspective) would help students resolve the new conflict by explaining to them the power of proof as a tool for deciding whether a pattern can be trusted in mathematics and helping them produce a proof for the pattern they identified earlier in the Squares Problem. The development of this proof would give students an image of a proof and would support them to begin to appreciate the importance of proof in mathematics.

The Implementation of the Focal Instructional Sequence in Research Cycle #5

Squares Problem (Initial Work)

In order to solve the first two parts of the Squares Problem, most students counted, one-by-one, the number of squares of different sizes in the 4-by-4 and 5-by-5 squares and noticed that the answers were the sums of the first four and five square numbers, respectively. Based on this observation, many students identified the pattern that the total number of squares in any square, say of size n, was given by the sum of the first n square numbers. They trusted the pattern and were sure that the answer to part #3 of the problem would be the sum of the first 60 square numbers, without feeling the need to explain whether or why the pattern would hold in this case.

Before the end of the class session, the instructor asked all students to explain in writing *whether* and *why* they could be sure that the answer to part #3 would be the sum of the first 60 square numbers (cf. CAP #1 in figure 1). Analysis of the 18 students' responses to the prompt (see Table 1 for examples of responses) showed that 17 of them trusted the pattern and expressed their conviction that the sum of the first 60 square numbers would give the correct answer to part #3 of the problem (code 'T' in Table 1). The remaining student did not give a response that would clearly indicate her view on the issue (code 'U'). Out of the 17 students who trusted the pattern, two offered an incomplete or superficial explanation for why the pattern would give the correct answer (code 'T&IE'). Thus it is possible that these two students thought they had a proof. The other 15 students showed an *empirical justification scheme* (code 'T&EJS'). Specifically, three of them suggested checking also the pattern in one or more strategically selected cases (*crucial experiment*; code 'T&EJS$_{CE}$') and the remaining 13 students trusted the pattern based on the few cases already checked in the class (*naïve empiricism*; code 'T&EJS$_{NE}$').

Table 1.
Students' Responses to the Prompt: "Can we be sure that this expression ($1^2+2^2+3^2+\ldots+59^2+60^2$) will give us the right answer for the 60-by-60 square? Why?"

Code	Student	Response
T&IE	Michelle	Yes, I think it would work because it worked for the 4x4 and 5x5, and nothing changes about the square except for the number of smaller squares in it, and the equation accommodates this by increasing the numbers accordingly.
T&EJS$_{CE}$	Laura	(1) It is a valid expression because the pattern was verified for the 4x4 and 5x5 squares. As long as all we do is add 1 unit of squares then we should be able to use the expression. (2) I'd probably choose a large square like 25x25 to verify that the expression remains true.
T&EJS$_{NE}$	Victor	Yes, because it has worked for several previous problems. I would expect for it to also work in a 60x60 square problem.
T&EJS$_{NE}$	Aleara	Yes, it is sure. For each size square, it goes by the same pattern. The last square is always the size of the square. For this, a 60x60 square, the last one would be 60^2. Then just add all of the squares together.
U	Maria	You could be sure that it would be the correct answer by continuing to actually show the work using a chart or mapping it out, or dividing the square root 60^2.

Note. The student names in the table and elsewhere in the article are pseudonyms.

Circle and Spots Problem

The next class session started with the instructor summarizing what happened in the previous session, commenting also on the students' responses to the prompt about whether and why the sum of the first 60 square numbers would give the correct answer for the number of different squares in the 60-by-60 square (cf. CAP #2 in figure 1):

Stylianides: I received different responses but a considerable number of you said something along the following lines: "Yes, we can be sure that this expression will give the right answer for the 60-by-60 square, because we found a pattern by checking smaller squares." Again, this is not what everybody said, but it captures well what many of you said. We will come back to this issue and to the Squares Problem later, but first I'd like us to work on a different problem.

He then presented the Circle and Spots Problem and asked the students to work in their groups to answer the question in italics. Almost all students began exploring simpler cases, trying to find a pattern that they could use to answer the question about 15 spots. Many of them noticed that the maximum number of non-overlapping regions doubled with each additional spot:

Laura: As you add a dot, you take the number of sections and multiply it by 2. So, for example, 2 dots: 2 sections, 3 dots: 4 sections, 4 dots: 8 sections, 5 dots: 16 sections.

Like it happened in the Squares Problem, many students trusted the pattern based on naïve empiricism and were sure that the maximum number of non-overlapping regions with 15 spots would be 16384, which they found by using the 'times 2' method.

However, some other students checked what happens with 6 spots and discovered a counterexample to the pattern: the maximum number of non-overlapping regions was 31, not 32 as predicted by the pattern. This counterexample helped the class realize that there was no easy way to tell what would be the maximum number of non-overlapping regions into which the circle could be divided with 15 spots. The instructor then asked the students to think in their small groups the following question: *What does this problem teach us?* (cf. CAP # 3 in figure 1)

Below is the response of Sherrill's small group to the question as presented in the class:

Sherrill: Um, patterns aren't always consistent. Cause you saw that the 'times 2' thing [pattern] stopped once you went past 5. And you kind of always assume that patterns are going to continue.

Sherrill's comment is a verbalization of the *cognitive conflict* that she and other students in the class experienced: there was a contradiction between students' earlier experiences with patterns that never failed and their current experience in the Circle and Spots Problem with an inconsistent pattern. The instructor elicited the thinking of other students in the class:

Beth: [I]f we just were to stop at 5 and assume that the pattern was right, you wouldn't really know it was wrong when it gets to, um, 15. So I guess the right thing to do is to go all the way through.

Stylianides: Michelle?

Michelle: Um, pretty much just that the problem teaches us that if we come up with a pattern, we need to test it more extensively.

Stylianides: So then, how many should we check? [to the class:] She [Michelle] said we should check more examples.

Joan: See, I think if I was a student, I would probably have stopped at 5, because that would have taken… 5 times 3 is 15. So I can just take the pattern for the first 5 and apply it; just doing it 3 more times – 3 times 4. So then, is the Squares Problem… Does that [the pattern that the class found in the Squares Problem and was still written on the chalkboard] apply through 60?

Stylianides: You see, so that's a very good [point]… so this is where we are going.

Laura: So when can you make a conjecture about something? Is it, do I test 50% of the possible, you know, solutions and then, you know, say "based on 50% of actual proof, I can now make this conjecture"? Is there a rule for that?

These remarks suggest that, like Sherrill, other students started to reconsider their previous approaches to trusting patterns and become aware of the limitations of these approaches. Specifically, Joan wondered whether the pattern that the class found in the Squares Problem would indeed apply through the 60-by-60 square, expressing her concern that perhaps that pattern would also fail. The students' remarks suggest also that they started to look for better (more secure) approaches to deciding when to trust a pattern. For example, Beth suggested checking the pattern all the way to the focal case (case 15 in the Circle and Spots Problem), whereas Michelle was vaguer in her suggestion, saying that a pattern should be tested 'more extensively.'

Laura's comment suggested an 'intellectual need' on her part to learn about the conventional approach to the issue of trusting patterns. Joan's comment indicated that her faith in the class's answer to the Squares Problem was shaken. So the instructor had, at this moment, an opportunity to give a closure to the work of the class on Circle and Spots Problem, skip the

'Monstrous Counterexample' illustration, and explain to the students that mathematicians would deal with the issue of trusting patterns by trying to develop a *proof*.

Yet, although Laura and Joan appeared to be ready to 'hear' about the role of proof in trusting patterns, many other students did not reach that stage yet. Also, experience from previous research cycles suggested that, if the instructor introduced the class to the notion of proof at this point, this would create a discontinuity in students' developmental progression, primarily because the issue of *crucial experiment* as a means for establishing the truth of a pattern had not come up yet in the whole group discussion. Therefore, the instructor decided to proceed with the next part of the instructional sequence as planned originally, offering the students further opportunity to reflect on what it meant to trust a pattern in mathematics. Specifically, he asked the students to reflect on the following fictitious pupil statement about the Circle and Spots Problem (CAP #3 continued): "This problem teaches us that checking 5 cases is not enough to trust a pattern in a problem. Next time I work with a pattern problem, I'll check 20 cases to be sure."

After some work in small groups, the class convened for a whole group discussion.

Stylianides: Okay, so let's see, what are your reactions to this student's thinking?

Monica: Um, I just said that instead of trying, like, 20 cases, like right in a row… if there was a certain number you were trying to figure out; like, if in the Squares Problem we were trying to figure out 60, you could try a number lower than 60… um, because we said that as you went up in the Squares Problem… [inaudible] Like not try all numbers. Try a number below 60 and a number above 60… [inaudible]

Monica's proposed selection of cases had some obvious weaknesses. For example, why would one check a more complicated case than the focal case and not just check the focal case itself? But still, her comment raised a reasonable concern about the limitations of *naïve empiricism* (checking 20 cases in a row) and reflected her attempt to propose an alternative way of validating the emerging pattern. The way of validation that she proposed involved strategic selection of cases, one above and one below the focal case, thus meeting the definition of *crucial experiment*.

Given the importance of Monica's comment in relation to the progression of the class along the learning trajectory summarized in figure 1 and the instructor's goal to help the students reflect on the limitations of empirical arguments (including crucial experiments), the instructor explored whether other students had similar thoughts with Monica.

Joan: My concern is that, let's say you do the first 20 in a row and then 21 is the place where it changes – like for us, we did 5 and then it changes at 6. And if you choose 20 random numbers and you don't see that there's a pattern, what do you do? Because there's no pattern to figure out the problem; and if you're taking a test and there's a time constraint, and you don't have a calculator, how is it possible to check it when you still have 20 other problems to do besides this one?

With this comment Joan raised two important and interrelated points about the identification and validation of patterns by checking particular cases. First, her experience in solving the Circle and Spots Problem helped her realize that, by checking a certain number of cases in a row, one can identify a pattern but, at the same time, one cannot be sure that the pattern will not fail in a case that one did not check. Joan's second point ("And if you choose 20 random numbers and you don't see that there's a pattern, what do you do?") was an implicit criticism of the idea (expressed in some form earlier by Monica) that one can validate a pattern by checking non-consecutive cases (as is sometimes the case in crucial experiment), because in this situation one may not even be able to notice the existence of a pattern.[2] Joan also raised some practical issues that reflected her intellectual need to learn an appropriate/efficient way to validate patterns; the only way she could think of was to check the pattern all the way to the focal case (like Beth suggested earlier). Joan's comment shows a development in her thinking on the issue of trusting patterns: earlier in the episode she said that, if she were a student, she would trust the pattern after checking the fifth case and then she would apply it to other cases.

[2] Joan seemed to use 'random' as an antonym of 'consecutive.'

Testing all the cases through the focal case was clearly not an ideal method for validating a pattern. But, still, the instructor did not consider that this was the right moment for him to explain to the students how mathematicians would address the limitations of this method. He knew that for many students an empirical approach was still a viable method for validating a pattern. Thus, he proceeded with the next activity in the instructional sequence.

'Monstrous Counterexample' Illustration

The contradiction between students' earlier practice to trust empirical arguments and the evident inadequacy of such arguments to establish certainty in the truth of a pattern, as shown by the Monstrous Counterexample illustration, provoked a new *cognitive conflict* among the students, which was expressed primarily in the form of utterances of surprise and amazement.

The instructor invited the students to use the new information offered by the illustration to reflect on their earlier discussions about what they learnt from the Circle and Spots Problem (cf. CAP #4 in figure 1). The discussion in the class that followed showed that more students started to become aware of the limitations of empirical arguments, including crucial experiment:

Joan: You can never check enough cases.

Also, some students started to wonder whether one could ever trust patterns in mathematics, and whether and how they could check the pattern they found in the Squares Problem:

Victor: I guess you can never really be sure of anything in math. […]
Stylianides: […] Is there any way that we can be sure? Lindsey? Is there any way that we can be sure that this is correct [pointing to the expression for the 60-by-60 square]?
Lindsey: [inaudible] So, I'm not really sure.
Stylianides: About what?
Lindsey: Um, I just, you know, checking it would be the right thing to do.
Stylianides: For this one?
Lindsey: For the 60-by-60 [square]. Yeah, checking and making sure that…

The only way that Lindsey and the other students in the class could think of to check their answer to part #3 of the Squares Problem was to count one-by-one all the different squares in the

60-by-60 square. The instructor considered that this was the right moment for him to talk to the students about the role of proof in mathematics, primarily because the classroom community seemed to have reached a conceptual barrier the students could not overcome by themselves:

Stylianides: If you find a way to explain your pattern, to *prove* your pattern, to see where this pattern comes from, then you trust it. [pause]
Does this make sense? [students nod in agreement] So then in order to… our alternative here, instead of counting the 60-by-60 square and seeing whether this expression is correct, we need to figure out a way to *explain* it. Where does this expression come from? Okay? And this is not a trivial question. […]

With his last comment ("And this [explaining/proving the pattern] is not a trivial question") the instructor laid the ground for the scaffolding that he would provide to help the students develop a proof for the pattern in the Squares Problem.

Squares Problem (Revisited)

With scaffolding from the instructor, the students were able to develop a proof for the pattern they identified originally and, thus, become certain that the pattern would indeed give the correct answer to part 3 of the Squares Problem. Due to space constraints, we will not describe how the proof was developed in the class. Below is an outline of the proof:

(1) In a 60-by-60 square we have squares of sizes k-by-k, where $1 \leq k \leq 60$.
(2) To find the total number of different squares in a 60-by-60 square we need to add the numbers of squares of all different sizes.
(3) The formula $(60-k+1)^2$ gives the number of squares of each size. (The class established why the formula would work for all possible values of k.)
(4) We add up together all the numbers given by the formula for each value of k to get the total number of different squares in a 60-by-60 square. This gives: $1^2+2^2+3^2+\ldots+59^2+60^2$.

It was important for the instructor to scaffold students' work to produce a proof for the pattern for two main reasons: (1) to address the misconception developed by some students that one could never be sure about the truth of a mathematical generalization that involved an infinite (or even a large) number of cases, and (2) to provide students with an image of a proof so that they would start to appreciate the importance of proof in establishing the truth of generalizations.

Overall Commentary

Given the close interdependence among the various instructional sequences that we implemented in the course, it is hard to specify the level of the contribution of individual instructional sequences to the overall development of students' knowledge about proof. Yet, we believe that the instructional sequence we examined in this article had an important contribution to this development. The classroom interactions we discussed earlier show that the instructional sequence challenged students' misconceptions about the potential of empirical arguments to establish the truth of mathematical generalizations by helping many students start to realize the inadequacies of different kinds of empirical arguments. This realization set strong foundations for students' appreciation of the importance of proof in mathematics. Our belief in the important contribution of the focal instructional sequence to students' learning of proof is suggested also by the students' self-reflections in the context of a question in the last homework assignment of the course. This question asked the students to identify three activities in the course[3] that they felt contributed the most to their learning and write a paragraph for each activity explaining what exactly they found useful about it. Twenty-one out of the 39 students who attended the course during research cycle #5 included in their list at least one of the activities that belonged to the instructional sequence we examined in this article. Several students referred in their reflections to the element of surprise provoked in them by the inconsistent patterns in the Circle and Spots Problem and the Monstrous Counterexample illustration.

Conclusion

Despite the mounting body of research that documents the strong reliance of students on empirical arguments, this article shows that, with theoretically informed and carefully developed

[3] The term 'activity' in the question of the homework assignment was left open to include any kind of activity done in the course. The students could choose from a list of more than 40 activities done in the course.

instructional sequences (within the design-experiment research paradigm), university instructors can help undergraduate students begin to realize the limitations of such arguments and appreciate the importance of proof. Our four year experience from implementing earlier versions of the instructional sequence in mathematics courses for preservice elementary teachers at two different U.S. institutions suggests that the instructional sequence we presented in this article is robust under application to different groups of preservice elementary teachers, presumably because these groups face similar difficulties with proof. Given that these difficulties are persistent among undergraduate students more broadly, we conjecture that the instructional sequence will be robust also under application to other undergraduate student populations.

This article presented an example of how theory and research can support the teaching of hard-to-learn topics at the undergraduate level by conceptualizing and studying the interconnection between student learning trajectories and instructional sequences that aim to facilitate students' developmental progressions along those trajectories (Clements & Sarama, 2004). Teaching practices at the undergraduate level that aim to embody understanding of the interconnection between particular student learning trajectories and related instructional sequences have good potential to improve over time and achieve their intended goals.

References

Balacheff, N. (1988). Aspects of proof in pupils' practice of school mathematics. In D. Pimm (Ed.), *Mathematics, teachers and children* (pp. 216-235). London: Hodder & Stoughton.

Bell, A. W. (1976). A study of pupil's proof-explanations in mathematical situations. *Educational Studies in Mathematics, 7*, 23-40.

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning, 6*(2), 81-89.

Davis, P. J. (1981). Are there coincidences in mathematics? *American Mathematical Monthly, 88*, 311-320.

Hanna, G., & Jahnke, H. N. (1996). Proof and proving. In A. J. Bishop, K. Clements, C. Keitel, J.

Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 877-908). Dordrecht, Netherlands: Kluwer Academic Publishers.

Harel, G. (1998). Two dual assertions: The first on learning and the second on teaching (or vice versa). *American Mathematical Monthly, 105*, 497-507.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 805-842). Greenwich, CT: Information Age Publishing.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education, 31*, 396-428.

Leikin, R., & Dinur, S. (2003). Patterns of flexibility: Teachers' behavior in mathematical discussion. *Proceedings of the 3rd Conference of the European Society for Research in Mathematics Education.* Bellania, Italy.

Mason, J. (1998). Enabling teachers to be real teachers: Necessary levels of awareness and structure of attention. *Journal of Mathematics Teacher Education, 1,* 243-267.

Mason, J., Burton, L., & Stacey, K. (1982). *Thinking mathematically.* London: Addison-Wesley.

Schoenfeld, A. H. (2005). Design experiments. In P. B. Elmore, G. Camilli, & J. Green (Eds.), *Complementary Methods for Research in Education* (pp. 175-187). Washington, DC: American Educational Research Association.

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

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*, 114-145.

Simon, M. A., & Blume, G. W. (1996). Justification in the mathematics classroom: A study of prospective elementary teachers. *Journal of Mathematical Behavior, 15*, 3–31.

Sowder, L., & Harel, G. (2003). Case studies of mathematics majors' proof understanding, production, and appreciation. *Canadian Journal of Science, Mathematics and Technology Education, 3,* 251-267.

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

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*, 458–477.

Zaslavsky, O. (2005). Seizing the opportunity to create uncertainty in learning mathematics. *Educational Studies in Mathematics, 60*, 297-321.

Zazkis, R., & Chernoff, E. J. (in press). What makes a counterexample exemplary? *Educational Studies in Mathematics.*