

Creating a Need for Proof

Gabriel J. Stylianides
University of Oxford, UK

Andreas J. Stylianides
University of Cambridge, UK

A student's *justification scheme* (Harel & Sowder, 1998) denotes what counts as a proof for the student. Prior research showed that even advanced secondary students (e.g., Coe & Ruthven, 1994; Healy & Hoyles, 2000) and university students including prospective elementary teachers (e.g., Goetting, 1995; Harel & Sowder, 1998) have the *empirical justification scheme*. These students tend to consider that *empirical arguments* are proofs of mathematical generalizations, whereas empirical arguments are in fact invalid arguments that offer inconclusive evidence for the truth of a generalization by verifying its truth only for a proper subset of all possible cases. This misconception is a major stumbling block for students' learning about proof: unless students realize the limitations of empirical arguments, they cannot see an "intellectual need" (Harel, 1998) to learn about proof as understood by mathematicians.

In this chapter, we present an instructional sequence that our research showed can be used (after minor adaptation) to help prospective elementary teachers (Stylianides & Stylianides, 2009) and secondary school students (Stylianides, 2009) begin to overcome the misconception that empirical arguments are proofs. Thus the chapter can be of interest to both teacher educators and secondary school teachers. We originally developed the sequence in a four-year design experiment that we conducted in an undergraduate mathematics course for prospective elementary teachers, and it is in the context of this course that we will ground our discussion in this chapter.

Major Features of the Course

In this section we discuss four major features of the undergraduate course within which the instructional sequence was developed and implemented.

Feature 1: Proof as a vehicle to sense making in mathematics

Students' engagement with proof in the course was part of their engagement with the activity of *reasoning-and-proving* (Stylianides, 2008), a term that is used to describe a family of activities involved in the investigation of whether and why things make sense in mathematics. Such activities include generalizing mathematical relations and developing arguments for or against these generalizations (e.g., Mason, 1982). The course treated reasoning-and-proving as a process (strand) that permeated students' mathematical work.

Feature 2: Close connection between instructional sequences and learning trajectories

This feature relates to the notions of *hypothetical learning trajectories* (Simon, 1995) and *actual learning trajectories* (Leikin & Dinur, 2003). The first notion refers to the learning routes students

This chapter is adapted from G. J. Stylianides & A. J. Stylianides (2009), Facilitating the transition from empirical arguments to proof, *Journal for Research in Mathematics Education*, 40, 314–352.

were anticipated to follow as a result of the implementation of an instructional sequence, while the second refers to the learning routes students seemed to have actually followed during the implementation. Over the multiple research cycles of our design experiment, we aimed to achieve close matching between hypothetical and actual learning trajectories by deepening our understanding of students' learning routes and by refining accordingly the instructional sequences in order to better support achievement of the intended learning goals.

Feature 3: Cognitive conflict as a mechanism for supporting progression in students' learning

The main assumption underpinning the “cognitive conflict” approach to mathematics teaching is that, by creating situations that contradict some of students' current mathematical understandings (justification schemes, beliefs, etc.), students will recognize the importance (and intellectual necessity) of modifying these understandings in order to resolve the contradictions. A common challenge faced by this teaching approach, though, is that students often treat contradictions (presented, for instance, in the form of counterexamples) as exceptions and do not experience cognitive conflict (e.g., Zazkis & Chernoff, 2008). In our research, we identified two complementary conditions that increase the likelihood of a counterexample to create cognitive conflict for students in the domain of proof.

Condition 1: The counterexample is in accord with students' immediately accessible justification schemes.

The condition requires that alternative justification schemes that students can exhibit be arranged in a hierarchy according to their level of mathematical sophistication as in table 1.1. This arrangement indicates what may be the “immediately accessible” justification scheme for a student who holds a certain justification scheme at a given time.

The *naïve empirical* and *crucial experiment* justification schemes in table 1.1 relate respectively to the following forms of empirical arguments (Balacheff, 1988): (1) *naïve empirical arguments* are those in which the examined cases for validating a mathematical generalization are selected either without any particular reason or on the basis of practical convenience; (2) *crucial experiments* are those in which the examined cases are selected based on some kind of rationale (e.g., a strategy for discovering possible counterexamples). Empirical arguments of the second form are considered to be more advanced than those of the first, even though they are still invalid and do not meet the standard of proof. The third justification scheme in the hierarchy, the *non-empirical*, recognizes that empirical arguments (of any form) cannot be proofs.

Take, for example, a student who holds the naïve empirical justification scheme. Condition 1 says that a counterexample that would challenge not only this justification scheme but also the more advanced justification scheme of crucial experiment would not have the best potential to create a cognitive conflict for the student. Such a counterexample would ambitiously expect the student to make a “conceptual leap” from the naïve empirical justification scheme to the non-empirical justification scheme, skipping the intermediate crucial experiment justification scheme.

Table 1.1
A hierarchy of justification schemes by increasing level of mathematical sophistication

	Justification scheme		
	<i>Naïve empirical</i>	<i>Crucial experiment</i>	<i>Non-empirical</i>
Description	Empirical arguments of the form of naïve empiricism can be proofs.	Empirical arguments of the form of crucial experiment can be proofs.	Empirical arguments of any form cannot be proofs.

Condition 2: Students become more aware of their understandings about proof that the counterexample aims to challenge.

The underlying assumption in this condition is that the more aware students are of their existing understandings of a particular topic (in this case, proof), the more likely they will be to experience a cognitive conflict when they encounter a counterexample that contradicts these understandings. To help increase our students' awareness, we used certain instructional activities, which we call *conceptual awareness pillars* (or simply *pillars*), to direct students' attention to the understandings we wanted them to attend to or reflect on at any given point. We used pillars both prior to and following each counterexample that aimed to create cognitive conflict for students. The pillars that came prior to the implementation of a counterexample aimed to prepare students for the conflict, whereas the pillars that followed aimed to focus students' attention on the issues raised by the counterexample.

Feature 4: Means for supporting the resolution of cognitive conflicts

The instructor had a critical role in supporting the students to resolve cognitive conflicts and develop new understandings that better approximated conventional mathematical knowledge. Toward this end, the instructor used scaffolding strategies (e.g., he asked probing questions) and implemented conceptual awareness pillars. Less frequently the instructor offered the students direct access to conventional knowledge that the students saw the intellectual need for but were unable to develop on their own (due to time constraints, conceptual barriers, etc.).

The Instructional Sequence and Its Implementation

The instructional sequence was developed iteratively in the design experiment. Herein we focus on its final iteration as implemented in a section of a mathematics course for prospective elementary teachers that was attended by eighteen students and taught by the first author. The implementation lasted less than three hours (spread over three class sessions) and aimed to support student learning progression (by means of two cognitive conflicts) along the trajectory of justification schemes in table 1.1: from the naïve empirical, to the crucial experiment, to the non-empirical justification scheme.

The sequence comprised three different tasks, as presented in figure 1.1. The figure also outlines major instructor actions during the implementation of the sequence, including the position in the sequence of four pillars that helped fulfill condition 2 for creating cognitive conflicts for students, as described earlier.

Table 1.2 supplements the information in the figure by summarizing major features of the tasks and corresponding elements of the classroom community's learning trajectory, including information about how the counterexamples in two of the tasks fulfilled condition 1 for creating cognitive conflicts for students. Our analysis showed close matching between hypothetical and actual learning trajectories and so the table does not specify one or the other.

Next we discuss the implementation of the instructional sequence in sufficient detail to support its potential use in other classrooms at the teacher education or secondary school levels. While some adaptation of the sequence will likely be necessary in order to accommodate the circumstances of individual classrooms, we caution against major modifications that would fail to preserve essential features of the intervention. Indeed, our research suggests that major modifications of this kind can compromise the intended learning outcomes (for elaboration, see Stylianides & Stylianides, 2009, pp. 331–333).

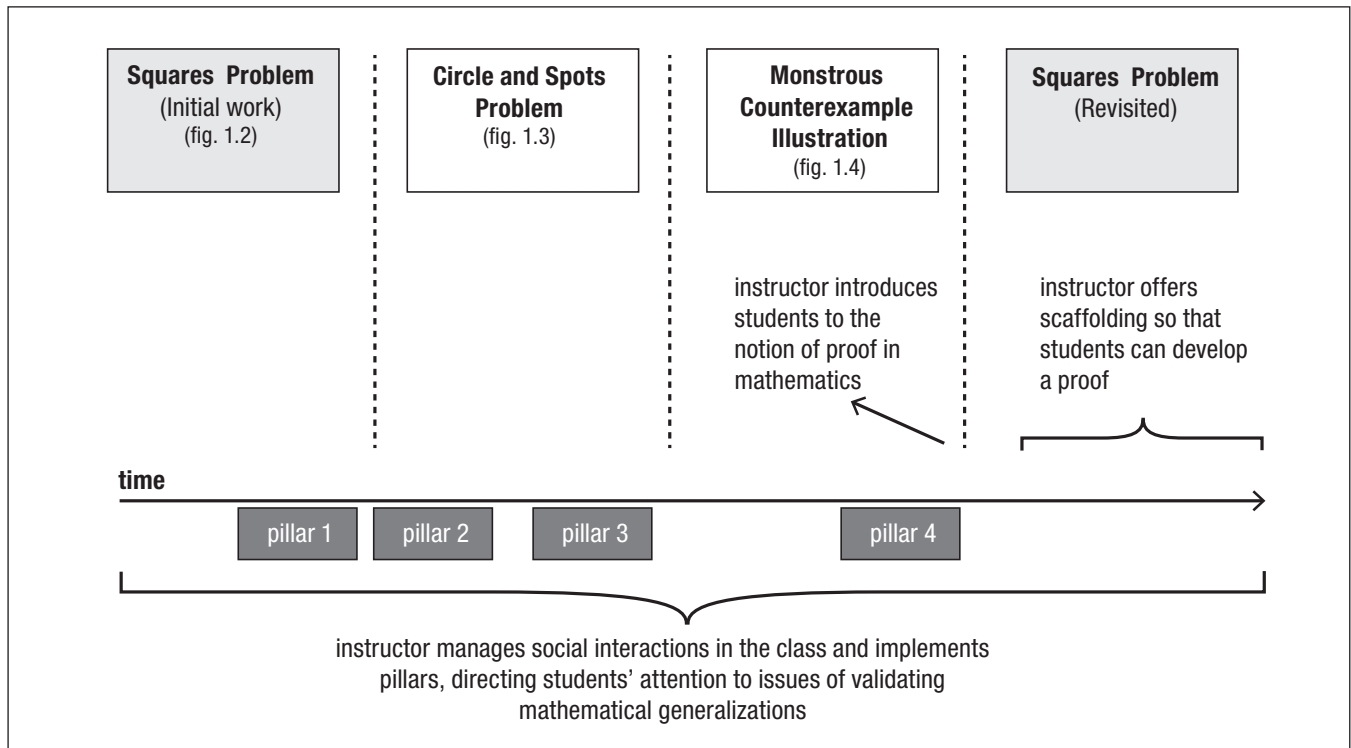


Fig 1.1. Outline of the tasks that comprised the instructional sequence and corresponding planned instructor actions over the time of their implementation

Table 1.2

Major features of the tasks that comprised the instructional sequence and corresponding major elements of the classroom community's learning trajectory

Task	Task features	Elements of the learning trajectory
Squares Problem (Initial work)	Gives rise to a numerical pattern (true): $\alpha_n = \sum_{k=1}^n k^2$	Students explore particular cases and identify a pattern.
	Asks for the value of the corresponding number sequence for a large case (for $n = 60$) that is practically difficult to check	Students validate the pattern on the basis of empirical arguments in the form of naïve empiricism.
Circle and Spots Problem	Includes five cases (the first five) that give rise to a numerical pattern ($\alpha_n = 2^{n-1}$) that does not apply for later cases	Students explore particular cases and identify a pattern (false).
	Asks for the value of the corresponding number sequence for a large case ($n = 15$) that is practically difficult to check	Students validate the pattern on the basis of empirical arguments in the form of naïve empiricism.

Table 1.2—*Continued*

Task	Task features	Elements of the learning trajectory
Circle and Spots Problem	Includes a counterexample that aims to create cognitive conflict for students by challenging the naïve empirical justification scheme: <ul style="list-style-type: none"> The counterexample in the pattern (for $n = 6$) falls right outside the domain of cases students with the naïve empirical justification scheme would likely check (the first four or five cases in the pattern). 	Students experience cognitive conflict 1. Corresponding progression in students' learning: <ul style="list-style-type: none"> Students consider more advanced empirical arguments in the form of crucial experiment.
Monstrous Counterexample Illustration	Includes a counterexample that aims to create cognitive conflict for students by challenging the crucial experiment justification scheme: <ul style="list-style-type: none"> The counterexample falls outside the domain of cases that students with the crucial experiment justification scheme would check. 	Students experience cognitive conflict 2. Corresponding progressions in students' learning: <ul style="list-style-type: none"> Students recognize empirical arguments of any form as insecure methods for validation. Students see an intellectual need to learn about secure methods for validation. Students are introduced (by the instructor) to the notion of proof in mathematics.
Squares Problem (Revisited)	Offers the opportunity for the development of a proof for the numerical pattern	Students develop (with the help of the instructor) a proof: <ul style="list-style-type: none"> Students see what a proof can look like. Students see that it is possible to validate (securely) a mathematical generalization that involves a large number of cases.

The Squares Problem (Initial Work)

After individual reading of the Squares problem (see fig. 1.2) and brief whole-group discussion of the meaning of key terms, the students worked on the problem, first individually and then in groups of three or four. In the whole-group discussion that followed, Monica was the first to present the work of her group on part 1 of the problem. (All student names are pseudonyms.)

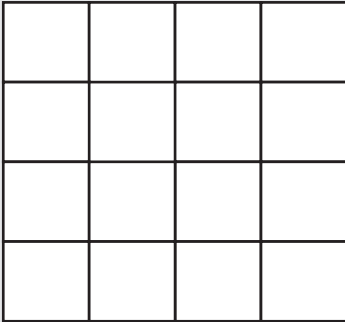
	<ol style="list-style-type: none"> Find the number of all different squares. What if this was a 5-by-5 square? What if this was a 60-by-60 square? How would you work to find how many different squares there would be? How would you make sure that you found them all?
---	--

Fig. 1.2. The Squares problem (adapted from Zack, 1997)

Monica: Well, there are only 4 different sizes [of squares], so we just counted how many there were of each size and got 30.

Monica also reported the number of squares of each size as follows: sixteen 1-by-1 squares, nine 2-by-2 squares, four 3-by-3 squares, and one 4-by-4 square. All groups agreed with these numbers, and so the discussion continued to how the numbers were derived. Kira showed first a nonsystematic way of counting the different squares of certain sizes, but then Sharon offered a systematic way. Sharon started with the first square of a given size in the top-left corner of the “big” square (in this case the 4-by-4 square) and moved across that “row” to count all the squares of that size; she did the same for each subsequent row and added the numbers. The class concluded that a systematic way helped ensure that no squares were omitted or counted multiple times.

Then the discussion turned to part 2 of the problem about the number of different squares in a 5-by-5 square. In this discussion, students began to describe the emerging pattern:

Helen: There’s a pattern, so you don’t have to count anymore. There will be 55 squares: 5^2 1-by-1 squares, 4^2 2-by-2 squares, 3^2 3-by-3 squares, 2^2 4-by-4 squares, and the large square.

Stylianides: Are you sure? Why? Did you count them?

Helen: No, we didn’t count them. We just went from the 4-by-4 case to the 5-by-5 case.

This was the first occurrence of a naïve empirical approach: Helen’s group concluded that the total number of squares in the 5-by-5 square would equal the sum of the first 5 square numbers because the students trusted the pattern that emerged from their consideration of the 4-by-4 square. A bit later the instructor asked Helen whether she could tell the number of different squares in tiled squares of different sizes. Helen responded as follows:

Helen: Yeah, but I wouldn’t know a formula [for the sum of the square numbers]. I’d have to just write out all of those terms [the square numbers in the sum].

Helen’s comment suggested again her conviction in the pattern. A similar conviction was evident in the remarks of other students such as Lisa about their solutions to part 3 of the problem:

Lisa: For the 60-by-60 square, you can follow the pattern and say that there will be $1^2 + 2^2 + 3^2 + \dots + 59^2 + 60^2$ different squares.

Before the session ended, the instructor asked all students to respond individually and in writing to the following prompt, which was pillar 1 in the sequence:

Stylianides: Can we be sure that this expression ($1^2 + 2^2 + 3^2 + \dots + 59^2 + 60^2$) will give us the right answer for the 60-by-60 square? Why?

Our analysis of the eighteen students’ written responses to pillar 1 indicated that seventeen of them trusted the pattern and expressed conviction that the sum of the first 60 square numbers would give the correct answer to part 3 (the remaining student’s response was unclear). The student responses in table 1.3 illustrate this finding. Of these seventeen students, two offered an incomplete explanation for why the pattern would give the correct answer (Natalie and Michelle). The other fifteen students showed an empirical justification scheme: three suggested checking also the pattern in one or more strategically selected cases (crucial experiment justification scheme; Laura) and twelve trusted the pattern based on the two cases they checked in parts 1 and 2 of the problem (naïve empirical justification scheme; Victor and Aleara). Thus, the classroom community’s dominant justification scheme at this point was naïve empiricism.

Table 1.3

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

Student	Response
Natalie	Yes, you can be sure this method will give the correct answer. Within each square there are squares of various sizes and to find their areas, you must multiply each side (i.e., 5×5 , 49×49).
Michelle	Yes, I think it would work, because it worked for the 4×4 and 5×5 , 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.
Laura	(1) It is a valid expression because the pattern was verified for the 4×4 and 5×5 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 25×25 to verify that the expression remains true.
Victor	Yes, because it has worked for several previous problems. I would expect for it to also work in a 60×60 square problem.
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 60×60 square, the last one would be 60^2 . Then just add all of the squares together.

The Circle and Spots Problem

The next session started with pillar 2, whereby the instructor reported to the class the dominant validation method (naïve empiricism) based on students' responses to pillar 1. This helped increase students' awareness of their use of a method that would soon be challenged.

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.

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 nonoverlapping 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 nonoverlapping regions into which the circle can be divided?

Fig 1.3. The Circles and Spots problem (adapted from Mason, 1982)

The instructor then presented the Circle and Spots problem (see fig. 1.3), clarified some terms (e.g., "spots around a circle," which meant points on the circumference of a circle), and asked the students to work in their groups to answer the question in italics. Many of them noticed that the maximum number of nonoverlapping regions doubled with each additional spot. This is how Laura presented the "times 2 pattern" in the whole-group discussion that followed:

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.

Stylianides: Okay, let me write what you just said on the board. [. . .] [He creates the following table with the numbers that Laura provided.]

Number of spots	Maximum number of regions
1	1
2	2 $\times 2$
3	4 $\times 2$
4	8 $\times 2$
5	16 $\times 2$

Stylianides: [to Laura] And what did you say you observed there?

Laura: That every time you would add another spot, it increases the number of regions. You multiply the one before it by 2.

Stylianides: So, from here to here, times 2. [. . .] [Adds the “ $\times 2$ curved arrows” to the table.] [. . .] So then, what would be your answer to the question?

[. . .]

Laura: I would just keep multiplying it down.

Similar to Laura and her group, many other students in the class validated the “times 2 pattern” based on naïve empiricism and were sure that they could apply the pattern to find the maximum number of nonoverlapping regions for $n = 15$, where n represents the number of spots on the circumference of a circle. Yet, some other students in the class checked also what happens for $n = 6$ and realized that the maximum number of nonoverlapping regions they could find for this case was 31, not 32, as predicted by the pattern. This is indeed a counterexample to the pattern. For information, we note that the correct formula for the maximum number of nonoverlapping regions is $1 + {}_nC_2 + {}_nC_4$. (A proof of this result can be found in Hart, 2007.) Yet, the question for the students in the Circle and Spots problem was purposefully phrased so that discovering the aforementioned formula or proving it was not required for a response to the question. The inability of the class to generate 32 nonoverlapping regions for $n = 6$ created doubt among students about whether the “times 2 pattern” was correct, and this doubt entitled them to reject application of this pattern as “an easy way” to “tell for sure” what would be the answer for $n = 15$.

The counterexample helped the class realize that there did not seem to be “an easy way” to “tell for sure” what would be the maximum number of nonoverlapping regions for $n = 15$. The instructor then asked the students to consider in their small groups the following question, which was the first part of pillar 3 and aimed to direct students’ attention to the implications of the Circle and Spots problem for their learning:

What does this problem teach us?

Sherrill presented the response of her small group to the question as follows:

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 the other members of her group experienced: there was a contradiction between their earlier experiences with patterns that never failed and their current experience with a "failing pattern." Comments similar to Sherrill's followed, indicating that other students in the class also experienced cognitive conflict.

In preparation for the second part of pillar 3, the instructor asked the students to say what the problem has taught them specifically about validating mathematical generalizations.

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 [examples] 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 [she refers to the pattern that the class found in the Squares problem and that was still written on the chalkboard] apply through 60?

Stylianides: You see, so that's a very good [point]. . . .

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

These remarks suggest that several students started to move beyond the naïve empirical justification scheme. For example, Beth pointed out the limitations of checking cases for $n \leq 5$, while Joan wondered whether the pattern that the class found in the Squares problem would indeed apply through the 60-by-60 square and expressed her concern that the pattern might fail. Also, the students started to look for more secure methods for validating patterns. For example, Beth suggested checking the pattern all the way to the case of interest ($n = 15$), an idea that enters the sphere of the non-empirical justification scheme. Michelle's suggestion was more vague: She said that a pattern should be tested "more extensively" before it can be accepted. Laura's pressing questions can be taken to suggest her intellectual need to learn not only about more secure methods for validating patterns but also about the conventional method. Laura may have been using the term "conjecture" to mean "theorem," but, essentially, she was asking about whether there is an established practice in validating generalizations.

Although Laura and perhaps other students in the class (including Beth and Joan) appeared to be ready to learn about the role of proof in validating mathematical generalizations, other students had not reached that point yet. We had anticipated this, and so the instructor proceeded with the plan and offered students another opportunity to think about validation methods. This was done in the context of the second part of pillar 3, which asked the students to discuss in their small groups the following (fictitious) student statement about the Circle and Spots problem:

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 a brief work in small groups, the class convened for a whole-group discussion. Monica said:

Monica: [I]nstead 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 limitations. For example, why would one check a more complicated case than the case of interest? However, she did raise a viable criticism of naïve empiricism—namely, checking many consecutive cases—and proposed an alternative method for validating a generalization that involved strategic selection of cases. In other words, Monica's comment reflected a progression toward the crucial experiment justification scheme.

Given the importance of Monica's comment, the instructor invited comments from other students. Joan said:

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 changed 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?

Joan raised three important points. Her first point was based on her realization that a pattern could fail in a case that one had not checked; in the Circles and Spots problem her group checked all cases up to $n = 5$, and the pattern failed for $n = 6$. Her second point was essentially a criticism of an idea that Monica expressed earlier: when one checks nonconsecutive cases, one may not be able to notice the existence of a pattern. Joan's third point raised some practical problems that derive from validating patterns by checking a large number of cases. Overall, Joan's comment can be viewed as a criticism of any kind of empirical validation method (i.e., as a step toward the non-empirical justification scheme) and as an indication of her intellectual need to learn about a better validation method. Interestingly, Joan had shown earlier a naïve empirical justification scheme: she said that if she were a student working on the Circle and Spots problem, she would trust the pattern after checking the first 5 cases and would then apply the pattern to other cases.

Although Joan and other students in the class seemed to have progressed at this point to the non-empirical justification scheme, there were others (e.g., Monica) who still held the crucial experiment justification scheme. The next task was intended to engineer another cognitive conflict that would allow more students to progress toward the non-empirical scheme.

The Monstrous Counterexample Illustration

The instructor presented a PowerPoint slide with the illustration in fig 1.4, without however using the name "Monstrous Counterexample Illustration" with the students. The presentation was done step by step to allow for clarification of terms and time for the students to check the statement for different values. When the instructor presented the last part of the illustration, many students uttered comments of surprise and amazement. Once the class calmed down, the instructor introduced pillar 4:

What does this illustration teach us? How does the illustration inform our previous discussions about the [fictitious] student statement in the Circle and Spots problem?

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!

Fig. 1.4. The Monstrous Counterexample Illustration (adapted from Davis, 1981)

The whole-class discussion that followed the small-group work on these questions showed that more students started to progress toward the non-empirical justification scheme:

Joan: [It teaches you that] you can never check enough cases.

[. . .]

Stylianides: Does everybody agree with what Joan said—that you can never check enough [cases]? Sherrill, do you agree?

Sherrill: Now I do.

[. . .]

Laura: [C]ouldn't you say that because the numbers . . . because you can't count to infinity, that there really is, like, you can never be totally sure?

[. . .]

Stylianides: [Refers to the whole class] So, what would be now your reactions to this question? [Shows the PowerPoint slide that included pillar 1] Can we be sure that we found a [the correct] pattern [for the Squares problem] by checking smaller squares? Victor?

Victor: I guess you can never really be sure of anything in math. I mean, after seeing that last problem [the Monstrous Counterexample Illustration] I would say “no”—I mean, I still think that's right [refers to the sum of the first 60 square numbers].

The students in the excerpt above seemed to have recognized empirical arguments as offering an insecure validation method, thereby progressing to the non-empirical justification scheme. At the same time, however, the students were in an intellectually challenging position, for they lacked knowledge of a method that would address the limitations of empirical arguments. Before helping the students address this issue, the instructor checked again whether they had any ideas about how to validate generalizations that involve a large or an infinite number of cases:

Stylianides: Is there any way that we can be sure that this is correct? [Referring to the pattern about the sum of the first 60 square numbers in part 3 of the Squares problem]

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.

The only way that Lindsey and the other students in the class could think of in order to check their answer to part 3 of the Squares problem was to count all the different squares one by one. However, what if one wanted to know whether a pattern would work for an infinite number of cases? Michelle raised this issue in the discussion that followed:

Michelle: Um, when there are an infinite number of cases, we can't be sure that the pattern is a pattern.

[. . .]

Helen: I don't think you can say that it's a pattern, because in the other example [she refers to the Circle and Spots problem] we thought we did [i.e., found a pattern], but we didn't. So you can't say for certain, but we are pretty sure.

The preceding discussions verified that the students had reached a conceptual barrier. This was the right moment for the instructor to offer to the students access to conventional mathematical knowledge about proof for which the students saw the intellectual need but were unable to develop on their own.

Stylianides: If you find a way to explain your pattern, to *prove* your pattern, to see where this pattern comes from, then you can trust it. [Pause] Does this make sense? [Students nod in agreement] So then in order to . . . our alternative here, instead of counting [all the different squares in] the 60-by-60 square and seeing whether this expression [sum of the first 60 square numbers] 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. [. . .]

This short talk served as a preliminary introduction of students to the notion of proof as a secure method for validating mathematical generalizations. Also, it was highlighting the explanatory function of proof (e.g., Hanna & Jahnke, 1996), which would be emphasized further later on in the course. With his last comment ("And this [explaining/proving the pattern] is not a trivial question"), the instructor laid the groundwork for the scaffolding he would offer the students to prove the pattern in the Squares problem.

To conclude, the Monstrous Counterexample Illustration created a cognitive conflict for students, as it challenged their faith in crucial experiment, which was evidenced in their earlier work on the Circles and Spots problem. As we had anticipated, the cognitive conflict helped more students to progress to the non-empirical justification scheme, and it created an intellectual need in them to learn about secure methods for validation.

The Squares Problem (Revisited)

The next step was for the class to revisit the Squares problem and, with scaffolding from the instructor, to develop a proof for the pattern they identified earlier. Here is an outline of the proof that the class developed:

1. In a 60-by-60 square, we have squares of sizes k -by- k , where $1 \leq k \leq 60$ ($k \in \mathbb{N}$).
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 (see what follows for elaboration on this step).
4. We add the numbers given by the formula for every value of k to get the total number of different squares in a 60-by-60 square. This gives us the sum of the first 60 square numbers: $1^2 + 2^2 + 3^2 + \dots + 59^2 + 60^2$.

We include fig 1.5 as a way to explain how the class established the third step in the proof. The class examined how many different k -by- k squares were in the “top row” of the 60 -by- 60 square (i.e., $60 - k + 1$) and then multiplied that number by the total number of such rows in the 60 -by- 60 square (i.e., $60 - k + 1$). To determine, for example, the number of different k -by- k squares in the top row, the class imagined the first k -by- k square in the top row and then calculated how many available places there were to move this k -by- k square to the right (since there were $60 - k$ available places, this meant that there were another $60 - k$ possible k -by- k squares in the top row in addition to the original square).

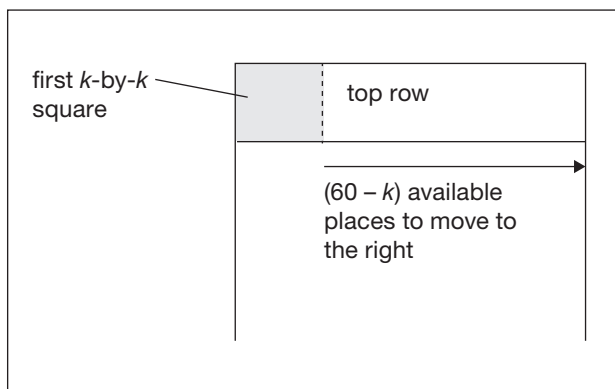


Fig. 1.5. Visual illustration of a step in the proof developed by the class for the Squares problem

The development of the proof served two main purposes in relation to addressing students' intellectual need to learn about secure validation methods: (a) it offered students an image of a proof, and (b) it helped them see that it is possible to validate a generalization that involves a large number of cases.

Conclusion

For completeness and as guidance to other instructors who will implement the sequence with their students, we describe briefly what happened after the teacher education class proved the Squares problem. The instructor engaged the students in an explicit discussion about the criteria that an argument needs to satisfy in order to qualify as a proof in their class. These criteria are discussed in Stylianides and Stylianides (2009, pp. 242–243). A similar list of criteria, adapted for use with secondary school students, is offered in the appendix in Stylianides (chapter 6 in this volume). The criteria guided the subsequent work of both the teacher education and secondary school classes in the area of proof and offered a backdrop against which the students themselves could judge whether different arguments met the standard of proof, without having to rely on the authority of the teacher. In both classes the criteria were applied in the context of a wide variety of problems, which offered students opportunities to refine the criteria and more fully understand them. In the teacher education class, the instructor also engaged his students in analysis and discussion of episodes from elementary classrooms (such as the episodes reported in Zack [1997] and in Stylianides [this volume]) so that the prospective teachers would get images of how the notion of proof could play out with elementary students.

As we already mentioned, a slightly modified version of the sequence was used successfully with secondary school students in England (Stylianides, 2009). Also, the sequence was trialed successfully with pre- and in-service secondary mathematics teachers and will be part of a teacher education curriculum on reasoning-and-proving currently under development (Smith, Boyle, Arbaugh, Steele, & Stylianides, 2014). The emerging potential of the sequence to be used in different classroom contexts can be attributed (at least in part) to (a) the robust research finding that the empirical justification scheme is pervasive among students of all levels and (b) the fact

that the design of the sequence was not contingent upon special characteristics of a particular student population.

The work we reported in this chapter can also be viewed as an example of how research can support the teaching of hard-to-learn topics by conceptualizing and studying the interconnection between student learning trajectories and instructional sequences that aim to facilitate students' progression along those trajectories. Mathematics educators can view this process in two ways: (a) as a way to improve their own teaching practice; and (b) as an approach to research that can contribute to theory development and curriculum design, thereby having broader implications for the improvement of the practices of others.

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, United Kingdom: Hodder & Stoughton.
- Coe, R., & Ruthven, K. (1994). Proof practices and constructs of advanced mathematics students. *British Educational Research Journal*, 20, 41–53.
- Davis, P. J. (1981). Are there coincidences in mathematics? *American Mathematical Monthly*, 88, 311–320.
- Goetting, M. (1995). *The college students' understanding of mathematical proof* (Unpublished doctoral dissertation). University of Maryland, College Park.
- 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, The Netherlands: Kluwer.
- 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. (1998). Students' proof schemes: Results from exploratory studies. In A. H. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Research in collegiate mathematics education III* (pp. 234–283). Providence, RI: American Mathematical Society.
- Hart, S. (2007). A regional problem. *Symmetry Plus*, 32, 8.
- 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*. Bellaria, Italy. Retrieved July 22, 2007, from http://www.dm.unipi.it/~didattica/CERME3/proceedings/Groups/TG11/TG11_Leikin_cerme3.pdf
- Mason, J. (with Burton, L., & Stacey, K.). (1982). *Thinking mathematically*. London, United Kingdom: Addison-Wesley.
- Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, 26, 114–145.
- Smith, M. S., Boyle, J., Arbaugh, F., Steele, M., & Stylianides, G. J. (2014). Cases as a vehicle for developing knowledge needed for teaching. In Y. Li, E. A. Silver, & S. Li (Eds.), *Transforming mathematics instruction: Multiple approaches and practices*, pp. 311–333. Basel, Switzerland: Springer.
- Stylianides, A. J. (2009). Breaking the equation “empirical argument = proof.” *Mathematics Teaching*, 213, 9–14.
- Stylianides, A. J., & Stylianides, G. J. (2009). Proof constructions and evaluations. *Educational Studies in Mathematics*, 72(2), 237–253.
- Stylianides, G. J. (2008). An analytic framework of reasoning-and-proving. *For the Learning of Mathematics*, 28(1), 9–16.
- Stylianides, G. J., & Stylianides, A. J. (2009). Facilitating the transition from empirical arguments to proof. *Journal for Research in Mathematics Education*, 40, 314–352.
- Zack, V. (1997). “You have to prove us wrong”: Proof at the elementary school level. In E. Pehkonen (Ed.), *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 291–298). Lahti, Finland: University of Helsinki.
- Zazkis, R., & Chernoff, E. J. (2008). What makes a counterexample exemplary? *Educational Studies in Mathematics*, 68, 195–208.