"CAN A PROOF AND A COUNTEREXAMPLE COEXIST?" A STUDY OF STUDENTS' CONCEPTIONS ABOUT PROOF*

Andreas J. Stylianides — University of Cambridge, U.K.

Thabit Al-Murani — University of Oxford, U.K.

Despite the importance of proof and refutation in students' mathematical education, students' conceptions about the relationship between proof and refutation have not been the explicit focus of research thus far. In this article, we investigate whether high-attaining secondary students have the misconception that it is possible to have a poof and a counterexample for the same mathematical statement. The data consisted of 57 student surveys augmented by follow-up interviews with 28 students. While analysis of the survey data alone offered considerable evidence for the existence of the misconception among several students in our sample, subsequent analysis with the inclusion of the interview data showed no evidence of the misconception. Implications for methodology and research are discussed in light of these findings.

INTRODUCTION

Despite the fundamental role that proof and refutation play in mathematical inquiry (e.g., Lakatos, 1976) and the growing appreciation of the importance of these concepts in students' mathematical education (e.g., Lampert, 1992; Reid, 2002), students' conceptions about the relationship between proof and refutation have not been the explicit focus of research thus far. The lack of research that aimed to investigate specifically students' conceptions in this area creates a gap in the field's understanding of how students perceive the standards of evidence in mathematics. Yet, existing research literature on proof and refutation allows us to make a hypothesis about students' conceptions regarding the possible coexistence of a proof and a counterexample for the same statement.

Specifically, research studies identified two student conceptions whose combination gives rise to the hypothesis that some students believe that it is possible to have a proof and a counterexample for the statement. The first conception that some students have is that counterexamples do not really refute: students tend to treat valid counterexamples to general statements as exceptions that do not really affect the truth of the statements (Balacheff, 1988). The second conception that some students have is that proofs do not really prove: students have difficulties to understand that a valid proof confers universal truth of a general statement thus making further checks superfluous (Fischbein, 1982). However, we point out that the hypothesis that some

* The data reported in this article were collected and analyzed with support of an Economic Social and Research Council (ESRC) grant to the first author (RES-000-22-2536). The opinions expressed in the article are those of the authors and do not necessarily reflect the position of the ESRC.

students believe that a proof and a counterexample can coexist was derived by us considering findings from different studies that used different samples and methods and that were conducted in different cultural settings. So, the hypothesis is not attributed to any of those studies and should become the explicit focus of research.

In this article, we aim to contribute to this domain of research by reporting findings from an investigation of the possible existence of the aforementioned misconception among high-attaining secondary students. In this investigation, we used survey data from 57 students and follow-up interviews with 28 of them. With the interviews, we aimed to clarify some student responses to the survey and to test the tentative conclusions we had drawn from our analysis of the survey data.

BACKGROUND

The research was part of a *design experiment* (see, e.g., Schoenfeld, 2006) that was conducted in two Year 10 classes in a state school in England. The school had 165 Year 10 students (14 to 15 years old) who were set in seven classes according to their performance on a national assessment they took at the end of Year 9. A total of 61 students from the two highest attaining classes participated in the research.

Motivated in part by studies that showed that even high-attaining secondary students tend to have limited understanding of proof (Coe & Ruthven, 1994; Healy & Hoyles, 2000; Küchemann & Hoyles, 2001-03), the design experiment aimed to generate research knowledge about possible ways in which classroom instruction can help these students develop their understanding of proof. The design experiment involved development, implementation, and analysis of the effectiveness of a collection of lesson sequences that extended over one to three 45-minute periods. Each lesson sequence was intended to promote issues of proof in the context of mathematical topics and student learning goals that were consistent with the provisions of the English national curriculum, treating proof as a vehicle to mathematical sense making. As far as proof was concerned, the lesson sequences aimed to offer students opportunities to develop their understanding of the limitations of empirical arguments and of the importance of proof in mathematics, to construct proofs for true mathematical statements, and to formulate counterexamples for false mathematical statements. However, the issue of the possible coexistence of a proof and a counterexample for the same statement was not explicitly discussed in the classes.

The definition of proof that guided the work on proof within the two classes was an adapted version of the conceptualization of proof elaborated in Stylianides (2007, pp. 291-300). The following definition was used in the first lesson sequence in each of the two classes as part of students' introduction to the notion of proof.

> An argument that counts as proof [in our class] should satisfy the following criteria:
>
> 1. It can be used to convince not only myself or a friend but also a sceptic. It should not require someone to make a leap of faith (e.g., "This is how it is" or "You need to believe me that this [pattern] will go on forever.")

2. It should help someone understand why a statement is true (e.g., why a pattern works the way it does).
3. It should use ideas that our class knows already or is able to understand (e.g., equations, pictures, diagrams).
4. It should contain no errors (e.g., in calculations).
5. It should be clearly presented.

The definition was discussed and referred to by both classes several times during the course of the design experiment, and it can be considered to reflect the classes' "idealized" shared understanding of the criteria for an argument to qualify as a proof.

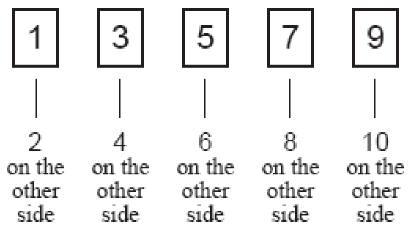
METHOD

Data Sources

The data for the article are derived from: (1) 57 student responses to a survey that we administered to the two classes at the end of the third lesson sequence of the design experiment (some students were absent the day we administered the survey), and (2) follow-up interviews with 28 students. The students completed the survey part way through the design experiment, after they had been given learning opportunities to develop understanding of different issues related to proof as described previously.

Five cards have the odd numbers 1, 3, 5, 7 and 9 printed on one side, and the even numbers 2, 4, 6, 8 and 10 printed on the other side.

The cards are dropped on the floor and spread out.



Amina, Ben, Carol and Davor are discussing whether this statement is true:

When two of the visible numbers are even, the five visible numbers add up to 27.

Ben's answer

I tried all odd numbers first and got 25:

1 + 3 + 5 + 7 + 9 = 25.

If I change one odd number to an even number, the total will be 1 bigger.
So if I have two even numbers, the total will be 2 bigger.
So the total will be 27.

So Ben says it's true

Carol's answer

I wrote down these numbers:

1, 2, 3, 4, 9.

Two of the visible numbers are even but the total is 19. So you do not always get 27.

So Carol says it's not true

Figure 1: A mathematical problem and two sample solutions to the problem

Survey

The survey presented the students with a true statement contextualized in a mathematical problem, four sample solutions to the problem, and some open-ended and multiple-choice questions (figures 1 and 2); in this article we focus on students' evaluations of only two solutions (Ben's and Carol's).

Open-ended questions:

1. Whose answer is closest to what you would do? Explain your answer.
2. Whose answer would get the **highest** mark from your teacher? Explain your answer.
3. Whose answer would get the **lowest** mark from your teacher? Explain your answer.

Multiple-choice questions:

For each of the following, circle whether you agree, don't know, or disagree.

The statement is:

When two of the visible numbers are even, the five visible numbers add up to 27.

Ben's answer ...	agree	don't know	disagree
shows you that the statement is **always true**	1	2	3
only shows you that the statement is true for **some** examples	1	2	3
shows you **why** the statement is true	1	2	3

Carol's answer ...	agree	don't know	disagree
shows you that the statement is **not true**	1	2	3
shows you **why** the statement is not true	1	2	3

Figure 2: Open-ended and multiple-choice questions.

The survey derived from one used in the Longitudinal Proof Project (Küchemann & Hoyles, 2001-03; Technical Report for the Year 8 Survey, pp. 93-94). We added the third open-ended question and the probes inviting students to explain their answers. We hoped these additions would increase the survey's potential to reveal student thinking about the possible coexistence of a proof and a counterexample. While this issue does not seem to have been one that Küchemann and Hoyles aimed to explore (ibid, pp. 6-7), we thought the survey offered an excellent opportunity to do this: some students might not notice the (subtle) mistake in Carol's solution and consider it a valid counterexample to the statement, while at the same time recognize the value of Ben's deductive argument and consider it a proof for the statement.

Interviews

We interviewed 28 students based on their responses to the multiple-choice and open-ended questions in the survey. Most interview sessions began with us asking the students to review their responses to the survey and then to explain which survey question they found the hardest. This general interview probe was followed by specific probes for the students to elaborate on particular responses in their scripts.

Procedure and Analysis

Patterns in students' responses were identified and used to formulate hypotheses about their conceptions. Interview data were then used to test/refine the hypotheses.

With regard to students' conceptions about the coexistence of a proof and a counterexample, our analysis of the survey data focused on those scripts that contained evidence to suggest the potential existence of the misconception. Specifically, we focused on the scripts that contained evidence of one or more of the following "inconsistencies": (1) the student found a mistake in Carol's solution and said that she would get the lowest mark from the teacher but agreed with the sentence that Carol's solution showed that the statement was not true; (2) the student said that the highest mark from the teacher would go to both Ben's and Carol's solutions; and (3) the student agreed both with the sentence that Ben's solution showed that the statement was always true and with the sentence that Carol's solution showed that the statement was not true.

We coded the type of evidence that was present in the scripts into two categories – strong or weak – depending on the degree of confidence that it gave us as researchers for the existence of the misconception. Specifically, we considered that strong evidence was offered by those scripts that had either "agree" or "don't know" in the first multiple-choice questions for both Ben's and Carol's solutions, and that included no relevant disconfirming evidence in the open-ended questions. The scripts that we considered offered weak evidence for the existence of the misconception had again either "agree" or "don't know" in the first multiple-choice questions for both Ben's and Carol's solutions, but included some relevant disconfirming evidence in the open-ended questions (e.g., they offered evidence that the student was aware that Carol's solution had a mistake in it). For each of the strong or weak evidence categories we used the interview data to examine the extent to which there was, overall, evidence to suggest that the students actually had the misconception. Also, we used the interview data to seek possible explanations (from the students' point of view) for the "inconsistencies" that we identified in their scripts.

RESULTS

General Findings

Our analysis of the survey scripts showed that 16 out of the 28 students interviewed exhibited some evidence to suggest the existence of the misconception that a proof and a counterexample can coexist. Of these, ten scripts showed strong evidence and six showed weak evidence for the misconception. Our subsequent analysis of the interview data revealed that the students in each group (i.e., strong or weak evidence group) tended to offer similar justifications for their choices.

Regarding the strong evidence group, our interview data suggested that the inconsistencies in students' responses derived from them considering Ben's and Carol's solutions in isolation from one another when they were completing the survey. While discussing their responses with the interviewers, however, all the students in this group became aware of the potential inconsistency between their evaluations of Ben's and Carol's solutions, presumably because the interviewers' questions directed (implicitly or explicitly) students' attention to the relationship between their evaluations. Yet, the manner in which the students became aware of this inconsistency and how the awareness played out in the interviews varied.

On the one hand, some students realized the mistake in Carol's solution without any prompting from the interviewers and immediately dismissed her solution. As a result of this dismissal, there was no opportunity for the interviewers to explore further whether these students would experience any sense of conflict that a proof and a counterexample can coexist. On the other hand, some other students needed explicit prompting from the interviewers to reflect on whether or how their evaluations of Ben's and Carol's solutions fitted together before they appreciated the potential inconsistency between these evaluations. Believing that Carol had found a genuine counterexample, these students attempted to resolve the emerging conflict by assuming there was a flaw in Ben's argument, which however they were unable to identify. The interviewers then helped these students see the mistake in Carol's solution and realize it was not a genuine counterexample. As a result of this realization, the students subsequently rejected Carol's solution, but this rejection was not always accompanied with endorsement of Ben's solution as a proof.

Regarding the weak evidence group, our interview data suggested that the students in this group seemed to be aware of the following 'inconsistency' we identified in their scripts: the students pointed out the mistake in Carol's solution in their response to the open-ended questions, but in the first multiple-choice question for Carol's solution they agreed that the solution showed the statement was not true.

During the interviews, the students argued, with different degrees of clarity, that, in spite of the mistake in Carol's solution, her reasoning should be valued because her logic was correct and she had disproved a statement, albeit a different one from that in the problem. Consequently, none of these students changed their minds about their evaluations of Carol's solution during the interview. The issue of the misconception was not pursued further by the interviewers, as the students were already aware that Carol's argument was not a counterexample to the particular statement.

To sum up, there is no evidence from our interviews to suggest that any of the 16 students we originally identified as potentially having the misconception actually had it. Furthermore, the interview data showed that any potential conclusions that could be drawn from the survey data alone would be insecure, as students appeared to have good reasons for ‘inconsistencies’ we identified in their scripts. For this reason we do not report findings with students we did not interview.

Illustrative Case 1: The Case of Emily

The first case is of a student we call Emily, whose responses to the survey showed strong evidence of the misconception. Emily’s script had “agree” in the first multiple-choice question for Carol’s solution and “don’t know” in the corresponding question for Ben’s solution. Furthermore, in response to the second open-ended question, Emily wrote that both Ben and Carol would receive the highest marks from the teacher and justified her thinking as follows:

> Ben: It [Ben’s solution] is carefully thought out and written down in an understandable and clear manner.
>
> Carol: She has shown when it [the statement] is not true.

During the interview Emily explained her thinking about Carol’s solution as follows:

> The question was saying [that] when two of them [the visible numbers on the cards] were even that the answer is always 27, but she proved that it’s not, so she answered the question that was being asked.

In regard to Ben’s solution, Emily said:

> It [Ben’s answer] was very, like, well set out and easy to understand and I think that was how I would have done it cause the other answers are like gabbling on a bit and they don’t really explain why it’s [the statement is] true or false.

She explained further that her “don’t know” response in the first multiple-choice question for Ben’s solution was because Ben “didn’t show that it’s always true, he only showed it for some numbers.” When asked whether she thought Ben had a proof, Emily said that Ben “needed to maybe expand it [his solution] a bit more to convince people that it was true” and noted that Ben could come up with a proof if he worked a bit harder on his solution.

After summarizing what Emily said about the two arguments, the interviewer asked Emily how her two evaluations fitted together. Realizing the inconsistency between the evaluations, Emily laughed and said: “they don’t [fit together] because Carol’s proved that it’s wrong and so it’s impossible to prove that it’s true… cause it’s not true!” Asked what she thought was going on with the two arguments, Emily asserted:

> They [Ben and Carol] have both tried different ways and got different answers, so if they kept working at it, if Ben kept working on his [solution], he would eventually figure out that it’s not true.

The interviewer then helped Emily to see the mistake in Carol’s solution. Once Emily realized the mistake, she exclaimed: “Oh, so she [Carol] could be wrong… so hers is wrong then.” On reviewing her original responses to the multiple-choice questions for Carol’s solution, Emily decided to change her response to the first question from “agree” to “disagree,” because, as she said, Carol “hasn’t followed the instruction.”

Emily concluded that Ben's solution "might be true" but she decided not to change her responses to the multiple-choice questions for his solution.

Illustrative Case 2: The Case of Evans

The second case is of a student we call Evans, whose responses to the survey showed weak evidence of the misconception. Evans' script had "agree" in the first multiple-choice questions for both Ben's and Carol's solutions, an indication of the existence of the misconception. Furthermore, Evans' responses to the first two open-ended questions showed particular appreciation of Ben's solution: he wrote that Ben's solution would be close to what he would do and that the solution would get the highest mark from his teacher "[b]ecause [it] shows working and offers convincing proof." Yet Evans' response to the third open-ended question offered disconfirming evidence of the existence of the misconception as it indicated that he was aware of the mistake in Carol's solution and said that Carol's answer would get the lowest mark from the teacher. In a series of two interviews, we tried to understand the reasoning for the apparent contradiction in Evans' evaluation of Carol's solution.

Evans was aware that Carol's solution had a mistake in it, but on the basis that she applied a correct mathematical method and that this application warranted recognition, he consciously agreed that she had shown the statement was not true.

> Well what she [Carol] has done is like impossible because 1 and 2 can't be seen at the same time, so then I would have disagreed because that can't be true. But seeing as though she has shown that she's thought it through and like, with her own reasoning she's come to an answer, then I would have put she technically has [shown the statement is not true] but she's got it wrong. […] Carol tried to prove the statement wrong, so one counterexample was enough. She had the logic right but she didn't succeed to come up with a correct counterexample.

This interview excerpt shows that Evans evaluated Carol's solution from her own point of view and that he understood the fundamental idea that a single counterexample suffices to refute a general statement. Evans considered that Carol's solution embodied understanding of the latter idea, even though the counterexample she offered did not satisfy, as he observed, the problem's conditions.

When pressed by the interviewer to explain his thinking further, Evans described the different evaluation standards that he perceived existed in exams and in class work:

> In an exam you don't get marks for the proof, do you? You get marks for showing your working and actually getting the answer in the end. But it [Carol's solution] does show the proof and everything. I don't know, it depends on what sort of question it is… if it's like what we're doing proof and stuff [referring to the proof work in class] then that [Carol's solution] would probably get the highest mark if that was what it was marked on… but in the exam it would be marked differently because it's not about how you are thinking, it's about getting the answer and getting the working and everything right.

The interviewer did not raise explicitly the issue of the possible coexistence of a proof and a counterexample, as Evans was clearly aware that Carol's argument was not a valid counterexample to the particular statement in the problem.

DISCUSSION

Although our analysis of the survey data alone offered considerable evidence (both weak and strong) for the existence of the misconception that a proof and a counterexample can coexist, our subsequent analysis with the inclusion of the interview data showed no evidence of the misconception. The size of the mismatch between the findings of the two analyses might have been influenced by what we considered as evidence for the possible existence of the misconception in our analysis of the survey data. Nevertheless, the existence of the mismatch reinforces and exemplifies the point that student responses to surveys may, by themselves, offer a rather limited insight into students' conceptions and that follow-up interviews with selected students are important for the construction of a more trustworthy picture of students' conceptions.

The latter statement is more than a reiteration of the well known methodological principle that triangulation of multiple data sources allows the examination of research questions in more nuanced ways than when using a single data source. The statement is also a cautionary remark that conclusions about students' conceptions that are based only on analysis of students' responses to surveys may be seriously misleading. This should not be taken as a criticism of the use of surveys in examining educational issues in general, but rather as a concern that the complexity that surrounds the particular issue of students' conceptions about multifaceted mathematical ideas may not be possible to be illuminated satisfactorily on the basis only of survey data. Of course this is not a black and white situation. The extent to which survey data alone can help illuminate complex issues depends on several factors: the methods that were used to validate a survey, the kinds of questions included in the survey, the conditions under which the survey was administered, the coding scheme used to analyse the survey data, etc.

In spite of the limitations in the conclusions that could be drawn based on the survey data alone, the survey offered a meaningful context in which we discussed during our interviews with students their ideas about the possible coexistence of a proof and a counterexample. This discussion was done with reference to Ben's deductive argument, which could be considered a proof, and Carol's purported counterexample. Carol's argument worked particularly well for the purposes of our research, as the subtle mistake in it passed unnoticed by several students, thereby helping us meet the challenge of presenting the students with a believable "counterexample" to a true statement. Ben's argument did not work as well as Carol's argument: students like Emily recognised the value of Ben's argument, but they did not accept it as a proof, primarily because they thought it needed "unpacking." The fact that some students did not consider that Ben's argument qualified as a proof gave them an "easy" way to resolve the problematic situation regarding the possible coexistence of a proof and a counterexample: these students suspected a mistake in Ben's argument and thus felt less hesitant to endorse Emily's counterexample. Given that the statement in the problem was true, it would not be difficult to strengthen Ben's argument in the survey so that more students would accept it as a proof; this modification in the survey would increase its potential to elicit students' conceptions about the possible coexistence of a proof and a counterexample.

Future research on students' conceptions in this area can use this modified version of the survey. Also, it would be useful if future research used an additional problem that asked students to evaluate a valid counterexample and a believable "proof" for a *false* statement. This would complement our examination in this study, thus contributing to the development of a more comprehensive approach to eliciting students' conceptions about the possibility of having a counterexample and a proof for the same statement. The fact that our research did not reveal this misconception does not mean that there are no students who have it; less advanced students, younger students, or students with fewer experiences with proof are more likely to have the misconception than the students who participated in our research.

REFERENCES

Balacheff, N. (1988). A study of students' proving processes at the junior high school level. In I. Wirszup & R. Streit (Eds.), *Proceedings of the Second UCSMP International Conference on Mathematics Education* (pp. 284-297). Reston, VA: National Council of Teachers of Mathematics.

Coe, R., & Ruthven, K. (1994). Proof practices and constructs of advanced mathematics students. *British Educational Research Journal, 20*, 41–53.

Fischbein, E. (1982). Intuition and proof. *For the Learning of Mathematics, 3*(2), 9-18, 24.

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

Küchemann, D., & Hoyles, C. (2001-03). *Longitudinal Proof Project* (Technical Reports for Year 8, 9, and 10 surveys). London: Institute of Education.

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery*. Cambridge University Press.

Lampert, M. (1992). Practice and problems in teaching authentic mathematics. In F. K. Oser, A. Dick, & J. Patry (Eds.), *Effective and responsible teaching: The new synthesis* (pp. 295-314). San Francisco, CA: Jossey-Bass Publishers.

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

Schoenfeld, A. H. (2006). Design experiments. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 193-205). Washington, DC: American Educational Research Association.

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*, 289-321.