This is a pre-publication version of:

Can a proof and a counterexample coexist?

Students’ conceptions about the relationship between proof and refutation

Andreas J. Stylianides\textsuperscript{a} and Thabit Al-Murani\textsuperscript{b}

\textsuperscript{a}Faculty of Education, University of Cambridge, Cambridge, UK; \textsuperscript{b}Department of Education, University of Oxford, Oxford, UK

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. Nevertheless, the combined consideration of findings from different studies suggests that some students believe it is possible to have a proof and a counterexample for the same assertion. In this article, we investigated the possible existence of this misconception among high-attaining secondary students. We used survey data from 57 students and follow-up interview data with 28 of them. Our analysis of the survey data alone offered considerable evidence for the existence of the misconception among several students. Yet, our analysis of data from the interviews, which aimed to test the tentative conclusions we had drawn from the survey data analysis, showed no evidence of the misconception. Implications for methodology and research are discussed in the light of these findings.

**Keywords:** conceptions; counterexample; proof; surveys; interviews; validity

*Corresponding author. Email: as899@cam.ac.uk*
Introduction

Proof and refutation play a fundamental role in mathematical inquiry. The process of validating assertions (and mathematical knowledge more broadly) often follows a ‘zig-zag’ path between attempts to generate proofs for the truth of the assertions and the discovery of counterexamples that refute the assertions and necessitate their refinement before they can be subjected to new proving attempts (e.g., Lakatos 1976; Schoenfeld 1983). A fundamental idea that underpins this validation process is that it is not possible to have a proof and a counterexample for the same assertion: the discovery and acceptance of a counterexample to an assertion makes absurd any further attempts to prove its truth because the assertion cannot be both true and false.¹

Despite the growing appreciation of the importance of proof and refutation in students’ mathematical education (e.g., Carpenter, Franke, and Levi 2003; Lampert 1992; Reid 2002; Zack 1997), students’ conceptions about the relationship between proof and refutation have not been the explicit focus of research thus far. The lack of research in this area has created a gap in the field’s understanding of how students perceive the standards of evidence in mathematics. Yet, prior mathematics education literature allows us to make a hypothesis regarding students’ conceptions about the relationship between proof and refutation.

Specifically, research studies identified two student conceptions whose combination gives rise to the following hypothesis: some students believe it is possible to have a proof and a counterexample for the same assertion. The first conception that some students have is that counterexamples do not really refute: students tend to treat counterexamples to general assertions as exceptions that do not really affect the truth of the assertions (Balacheff 1988). The second conception that some students have is that proofs do not really prove: students have difficulty understanding that a proof confers the universal truth of a general assertion, thus making further
checks superfluous (Fischbein 1982). However, the aforementioned hypothesis was derived based on findings from different studies, which used different samples and methods and which 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 contribute to this domain of research by reporting findings from an investigation of whether (and, if so, to what extent) high-attaining secondary students have the misconception² that it is possible to have a proof and a counterexample for the same assertion. In this investigation, we used survey data from 57 students and follow-up interview data 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 we report herein was part of a school-based design experiment in two high-attaining Year 10 (14–15 year olds) classes in England – for discussion of design experiment methodology, see, e.g., Schoenfeld (2006). This was a follow-up study to a university-based design experiment that was conducted by Gabriel Stylianides and the first author over a four-year period in an undergraduate mathematics course in the United States (see, e.g., Stylianides and Stylianides 2009). Even though the two design experiments were conducted in different settings, they had similar goals: both of them aimed to generate theoretical and practical knowledge about ways in which instruction can help students to develop their understanding of proof and related concepts (refutation, generalisation, etc.). The school-based design experiment used and adapted theoretical and practical tools that had been developed in the university-based design experiment.

The focal (school-based) design experiment was motivated in part by the findings of the Longitudinal Proof Project in England (Küchemann and Hoyles 2001–03). This large-scale
study surveyed annually, from the end of Year 8 to the end of Year 10, 1512 high-attaining students from randomly selected schools within nine geographically diverse regions in England, and showed that there were only modest improvements in students’ understanding of proof and related concepts (including refutation) over time. The findings of this study were consistent with those of earlier studies involving high-attaining secondary students in England (e.g., Coe and Ruthven 1994; Healy and Hoyles 2000), thereby painting a bleak picture of these students’ understanding of proof and related concepts. These research findings raise a concern about high-attaining secondary students’ potential to develop competency in these concepts, and may suggest an even more pessimistic prospect of less advanced or younger students’ potential to develop such competency.

Yet, classroom-based studies with primary students (e.g., Ball and Bass 2003; Carpenter et al. 2003; Lampert 1992; Maher and Martino 1996; Reid 2002; Stylianides 2007; Zack 1997) showed that, in supportive classroom environments, even young children can engage successfully with proof, refutation, generalisation, and other related concepts. The findings of these studies, together with those of psychological studies on the cognitive development of students’ ability for deductive reasoning and proof (see Stylianides and Stylianides [2008] for a review of this psychological literature), suggested that the bleak picture of high-attaining secondary students’ understanding of proof and related concepts ought to be attributed less to innate cognitive/developmental constraints and more to the likelihood that prevalent classroom instruction offered these students inadequate opportunities to develop competency in these concepts. This raised a need for research studies that would guide the design of effective classroom instruction to enhance students’ understanding of these concepts.
The focal design experiment took a step towards addressing this need for research by focusing on the teaching and learning of high-attaining Year 10 students in two classes. Specifically, the design experiment involved the development, implementation, and analysis of the effectiveness of a collection of lesson sequences (one to three 45-minute periods each) to provide the students with opportunities to enhance their mathematical knowledge in general, and knowledge about proof and related concepts in particular. The intention was to integrate the learning of these concepts with the learning of other mathematical topics in the English national curriculum. Some learning goals promoted by the lesson sequences were: to help students understand the limitations of empirical arguments and the importance of proof in mathematics (see Stylianides [2009] for discussion of a lesson sequence that focused on this goal); to construct proofs about true mathematical assertions; and to generate counterexamples to refute false mathematical assertions. However, the issue of whether it is possible to have a proof and a counterexample for the same assertion was not a focus of any of the lesson sequences.

The lesson sequences were planned, in the most part, by the first author, on the basis of prior discussions with the two class teachers about the mathematical topics they wanted to cover in their classes, the students’ anticipated proficiency level with respect to those topics, curricular and institutional constraints, etc. The draft lesson plans and the rationale for their development were then discussed with the two teachers, and occasionally with the second author. These discussions led to modification and adaptation of the lesson plans according to the particular needs of each class prior to their implementation by the teachers. The planning of the lesson sequences was underpinned by an adapted version of the theoretical framework that Gabriel Stylianides and the first author had developed in the university-based design experiment mentioned previously (see Stylianides and Stylianides 2009, 316–324). Due to space limitations,
and given that this article is not concerned with the ‘interventionist’ aspect of the school-based design experiment, we do not describe here the theoretical framework that underpinned the planning of the lesson sequences.

The definition of proof that guided the work in the two Year 10 classes consisted of five criteria, as described in the following quotation from a PowerPoint slide used during the first lesson sequence:

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.

This definition derived from a conceptualisation of proof elaborated in Stylianides (2007, 291–300). Criteria 1 and 2 in the definition are particularly important. They reflect, respectively, the following two related functions that the development of proofs served in the mathematical work of the Year 10 classes: (1) as a means for justifying that an assertion was true, thereby promoting conviction in its truth (e.g., Mason 1982) at both the individual and social levels; and (2) as a means for explaining why an assertion was true, thereby promoting understanding of the underlying mathematical topics (e.g., Hanna and Jahnke 1996). The focus on these two
functions of proof – justification and explanation – aimed to support students’ engagement with mathematics as a sense-making activity. According to Harel and Sowder (2007): “[m]athematics as sense-making means that one should not only ascertain oneself [individual level] that the particular topic/procedure makes sense, but also that one should be able to convince others [social level] through explanation and justification of her or his conclusions” (pp. 808–809). 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’ ‘idealised’ shared understanding of the criteria for an argument to qualify as a proof.

**Method**

**Data sources**

The focal design experiment was conducted in an English state school which had 165 Year 10 students, set in seven classes according to their performance in a national assessment they took at the end of Year 9. All 61 students from the two highest attaining Year 10 classes participated in the research. The data for the article are derived from: 57 student responses to a short survey that we administered to the two Year 10 classes at the end of the third lesson sequence of the design experiment (some students were absent the day we administered the survey); and 29 follow-up interviews that we conducted with 28 students (one student was interviewed twice). The students completed the survey part-way through the school year, after they had been given opportunities to develop understanding of different issues about proof and refutation (but not explicitly of their relationship), as described previously. The interviews were conducted over two days towards the end of the school year.
Survey: description and analysis

The survey, which the students completed individually, presented a mathematical problem and four sample solutions to the problem (Figure 1a), which were followed by some open-ended and multiple-choice questions (Figure 1b). The survey was an extended version of a section from a longer survey developed by the Longitudinal Proof Project (Küchemann and Hoyles 2001–03) and administered to Year 8 students in England (see Technical Report for the Year 8 Survey, 93–94). In particular, we added the third open-ended question about which sample solution the students thought would get the lowest mark from their teacher, and also the probes for students to explain their answers to all three open-ended questions. We speculated that these additions would increase the survey’s potential to reveal student thinking, and could facilitate triangulation of inferences from other data sources.

[Figures 1a and 1b]

The statement in the mathematical problem presented in Figure 1a is true. The variation in the sample solutions, and the students’ evaluations of these solutions (as reflected in their responses to the questions in Figure 1b), allowed us to explore a wide range of issues. Specifically, we explored students’ understanding of the limitations of different kinds of empirical arguments (see the solutions of Amina and Davor), their appreciation of a deductive argument that could potentially be considered a proof (see Ben’s solution), their ability to identify a deviation from the permissible conditions of a problem in the development of an argument (see Carol’s solution), their understanding of different linguistic forms used in the arguments (compare for example “I tried” in Amina’s solution and “I thought” in Davor’s solution), etc.

In this article, we focus on students’ evaluations of Ben and Carol’s solutions as a means for investigating students’ conceptions about whether it is possible to have a proof and a
counterexample for the same assertion. Although students’ conceptions about this issue do not seem to have been a focus for the developers of the longer survey (see the Longitudinal Proof Project Technical Report for the Year 8 Survey; Küchemann and Hoyles 2001–03, 6–7), the survey offered a good opportunity for investigation of these conceptions. It was anticipated that some students would not notice the (subtle) mistake in Carol’s solution, thereby considering it to be a counterexample to the statement, but at the same time they would also recognise the mathematical value of Ben’s deductive argument, thereby considering it to be a proof for the statement. These contradictory conclusions that students could draw from evaluating the two sample solutions had the potential to raise the issue of whether it is possible to have a proof and a counterexample for the same assertion.

The process of reducing the data from the completed surveys into a form appropriate for analysis involved two distinct stages, due to the difference in question types. While the multiple-choice section of the survey allowed the students’ answers to be pre-coded, the coding for the open-ended questions was developed after consideration of the students’ answers. The latter involved randomly selecting half of the surveys and dividing these equally between the two authors, who independently constructed their own coding frames. The separate frames were later consolidated and refined into a single coding frame, whose validity was checked by using it to code the other half of the surveys.

Our analysis concentrated on those surveys that contained evidence to suggest the potential existence of the misconception that it is possible to have a proof and counterexample for the same assertion. In particular, we focused on the surveys that contained evidence of one or both of the following ‘inconsistencies’ (from a mathematical standpoint): (1) the student who completed the survey had either “agree” or “don’t know” for the sentence about whether Ben’s
solution showed that the statement was always true and the sentence about whether Carol’s solution showed that the statement was not true (see the first multiple-choice questions for Ben and Carol’s solutions); and (2) the student said that the highest mark from the teacher would go to the solutions of both Ben and Carol (see the second open-ended question).

We coded the type of evidence present in the surveys into two categories – strong or weak – depending on the degree of confidence it gave to us as researchers for the existence of the misconception. Specifically, we considered that strong evidence was offered by those surveys that contained inconsistency 1 and included no relevant disconfirming evidence in the open-ended questions (e.g., the surveys also contained inconsistency 2, or contained no evidence to contradict the presence of this inconsistency). The surveys that we considered as offering weak evidence of the existence of the misconception again contained inconsistency 1, but additionally included some relevant disconfirming evidence in the open-ended questions (notably, the surveys contained evidence that the students were aware that Carol’s solution had a mistake in it).

By reflecting on the coded data from a number of different perspectives, we generated some alternative hypotheses to the hypothesis that students responded in the ways they did because they had the misconception. We also considered the following hypotheses that might explain (independently or in combination) the inconsistencies we identified in students’ surveys: the students (1) were confused by the complex linguistic structure involved in disagreeing with the sentence that “Carol’s answer shows you that the statement is not true” (see first multiple-choice question for Carol’s solution); (2) had not understood Ben’s argument; (3) were confused about the number of counterexamples needed to disprove a statement; and (4) answered the first multiple-choice question for Carol’s solution from Carol’s perspective (Carol believed she
generated a counterexample), rather than from their own. This process amounted to a content analysis of the surveys.

*Interviews: procedure and analysis*

The follow-up interviews we conducted aimed to clarify some student responses to the survey in relation to the whole range of issues we intended to examine, as we explained previously. The possible existence of the misconception that it is possible to have a proof and counterexample for the same assertion was one of these issues. Our analysis of the survey data showed that 26 students could have the misconception: ten students belonged to the strong evidence group and 16 belonged to the weak evidence group. Of these 26 students, we selected 16 to interview: we selected all ten students from the strong evidence group (due to the increased chance of them having the misconception) and six students from the weak evidence group (due to the special opportunity they offered us for further examination of the alternative hypotheses we described earlier). In what follows, we explain the procedure and method of analysis we followed for our interviews with these 16 students.

In order to test and refine our hypotheses and cast light on the possible mechanisms through which these students rationalised their survey responses, we collected interview data that would complement the survey data (Denzin 1988; Robson 1993). Using the content analysis as a basis, we constructed a preliminary interview protocol containing questions of a type that we hoped would elicit student responses to inform our hypotheses. This resulted in us preparing two to four questions (with relevant probes) for each student we wanted to interview. In the preliminary interview protocol, students were initially asked to review their completed surveys. In an attempt to make them check thoroughly both the survey questions and their responses to them, without however drawing their attention to the particular items we were interested in, we
informed them that our first question would be, “Which question did you find the hardest, and why?” On the occasions when this question failed to elicit responses relevant to the issue(s) we had identified as being of interest, we asked the students to elaborate on specific survey questions and their corresponding responses. The particular interview questions, their sequencing during the interviews, and the subsequent probes, were dependent on the specific pattern of responses in the students’ surveys. For example, we would ask students who had chosen “don’t know” for Carol’s solution to explain what they thought about it, and how this led to their particular response to the first multiple-choice question. Our probing would then aim to examine whether these students thought that Carol produced a counterexample to the statement.

The preliminary interview protocol was piloted with four students, in order to ascertain whether the interview data would be suitable to test out hypotheses, and to obtain a sense for the approximate length of each interview and the amount of data it would generate. Following the pilot, we decided that in subsequent interviews we would leave out the first question, because the interviews were taking too long and were generating some data that were unrelated to our research objectives. Our goal was for each interview to take approximately 20 minutes.

As can been seen from the previous examples, the questions we asked during the interviews were open and both specific and non-specific in nature. These questions, together with the prior content analysis, meant our interviews were ‘semi-structured’ and ‘focused’ (Merton and Kendall 1946). While the pre-prepared questions that each interviewee was asked had the potential to expose common trends, individual understanding was probed further through the evolving exchanges, with us trying to ask indirect questions, as these are generally considered to elicit relatively reliable responses from interviewees (Tuckman 1972).
For example, after reviewing their responses to the survey at the beginning of the interview, some students became aware of the inconsistencies in their responses to the survey items concerning Ben and Carol’s solutions. In such cases, the interviewer encouraged the students to elaborate on the nature of these inconsistencies, their mathematical implications, and why the students thought they had initially responded to the survey items in the way that they did. In other cases, where the students did not attend to the inconsistencies themselves, we first asked them to explain their responses to the survey items concerning Ben and Carol’s solutions, and then we probed them to consider how these responses fitted together. In doing this, we wanted to investigate whether the students’ previous lack of perception of the inconsistencies was due to them considering their responses to Ben and Carol’s solutions in isolation from one another, or whether it was because they had the misconception. As such, our probing was not intended to teach or instruct the students, but rather to allow them to expound on their responses and to help us test our hypotheses.

All interviews were audio-taped. Each audio file was listened to in its entirety several times, and summaries of the individual interviews were written. The summaries consisted of the focused questions that each student was asked, any additional questions asked and relevant student responses, and a description of the episodes relating to the specific hypotheses. These data were tabulated for comparison and used to inform our interpretation of students’ conceptions. We looked for patterns and also for variations. Following this, we constructed a composite table, incorporating the data from the surveys and those elicited from the interviews, in order to construct a synthetic account of students’ conceptions.
Results

**General findings**

As we noted previously, our survey analysis showed that the surveys of 16 of the 28 students we interviewed contained evidence to suggest the existence of the misconception that it is possible to have a proof and a counterexample for the same assertion. Of these, ten surveys showed strong and six weak evidence of the misconception. Our subsequent analysis of the interview data revealed that students tended to rationalise their survey responses in a manner that was similar among students within the same group (strong evidence group or weak evidence group). Next, we describe the patterns of reasoning we identified within each group.

**Strong evidence group**

Our interview data suggested that the inconsistencies in this group’s responses to the survey originated in them considering their evaluations of Ben and Carol’s solutions in isolation from one another when they were completing the survey. While discussing their responses with the interviewers, all the students in this group became aware of the potential inconsistency between their evaluations of Ben and Carol’s solutions. Yet, the manner in which the students became aware of this inconsistency, and how the awareness played out in the interviews, varied.

Some students realised the mistake in Carol’s solution without any prompting from the interviewers, and they immediately dismissed her solution on the grounds that it was flawed. As a result of the dismissal of the purported counterexample, there was no opportunity for the interviewers to explore further whether these students would experience any sense of conflict in having a proof and a counterexample for the same assertion.

Other students needed explicit prompting from the interviewer to reflect on how their evaluations of Ben and Carol’s solutions fitted together before they appreciated the inconsistency
between these evaluations. Believing that Carol found a genuine counterexample, these students attempted to resolve the emerging conflict in having a proof and counterexample for the same assertion by assuming there was a flaw in Ben’s argument, which, however, they were unable to identify. The interviewers then probed these students to see the mistake in Carol’s solution, in order to investigate students’ subsequent reaction to the situation. Once the students realised Carol did not have a genuine counterexample, they rejected Carol’s solution. Nevertheless, this rejection was not always accompanied by endorsement of Ben’s solution as a proof.

**Weak evidence group**

As we noted earlier, the students in the weak evidence group responded to the multiple-choice questions in a way that suggested they believed Ben had produced a proof and Carol a counterexample. However, in addition to containing this inconsistency, the surveys included some relevant disconfirming evidence: students’ responses to the open-ended questions showed they were aware that Carol’s solution had a mistake in it. Our interview data cast light on the ways in which the students rationalised their evaluation of Carol’s solution, thereby resolving the issue of the inconsistency we identified in their surveys, and suggesting that the students did not have the misconception. Specifically, the students argued with different degrees of clarity during the interviews that, in spite of her mistake, Carol’s solution demonstrated understanding of refutation by counterexample, and so they thought the solution deserved to be acknowledged as a sample of good thinking. These students did not really think that Carol’s solution refuted the particular statement in the problem, and so the issue of whether it would be possible to have a proof and a counterexample for the same statement was non-applicable.
Summary of general findings

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 conclusions based on the survey data alone would be problematic, as students appeared to have good reasons for the inconsistencies we identified in their surveys. Accordingly, we do not report findings with students we did not interview. Next we report two illustrative cases in order to exemplify the patterns of reasoning we identified within each group. The student names are pseudonyms.

Two illustrative cases

Strong evidence group: The case of Emily

Emily’s responses to the survey showed strong evidence of the misconception. Emily’s survey 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, justifying 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.

In the interview with Emily we tried to understand the reasoning for the apparent contradiction in her evaluation of the two solutions.

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.
Emily said the following with regard to Ben’s solution:

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 summarising what Emily said about the two arguments, i.e., that with extra work Ben could prove that the statement was always true, and that Carol had already shown that the statement was not true, the interviewer asked Emily how these two evaluations fitted together. Realising the inconsistency between the two 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!” When asked what she thought was happening 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 above excerpt shows that Emily resolved the emerging conflict by assuming that Ben’s argument had a flaw in it.

The interviewer then probed Emily to see the mistake in Carol’s solution. Once Emily realised that it was not possible for Carol’s choice of numbers to appear on the cards, 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, like she’s picked some random numbers and she should have picked the right cards.” Emily concluded that Ben’s solution “might be true”, but she decided not to change her responses to any of the multiple-choice questions for his solution.

Weak evidence group: The case of Evan

Evan’s responses to the survey showed weak evidence of the misconception. Evan’s survey had “agree” in the first multiple-choice question for both Ben and Carol’s solutions, an indication of the existence of the misconception. Furthermore, Evan’s responses to the first two open-ended questions showed particular appreciation of Ben’s solution: he wrote that Ben’s solution was 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, Evan’s 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. Specifically, Evan wrote that Carol’s answer would get the lowest mark from the teacher “[b]ecause her example (1, 2, 3, 4, 9) cannot be shown by the cards.” In a series of two interviews, we tried to understand the reasoning for the apparent contradiction in Evan’s evaluation of Carol’s solution, as reflected in the combined consideration of his responses to the third open-ended question and the first multiple-choice question for Carol’s solution. The second interview was based on our review and analysis of the first interview, and aimed to clarify further issues related to Evan’s thinking.
Evan 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 deliberately 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 suggests that Evan evaluated Carol’s solution from her point of view (see in particular his phrase “with her own reasoning she’s come to answer…”). The excerpt further suggests that Evan understood a fundamental idea related to refutation of a false statement by means of a counterexample, namely, that a single counterexample suffices to refute a general statement. Evan considered that Carol’s solution embodied understanding of this 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, Evan, seemingly aware of the apparent contradiction in rewarding an incorrect answer, described the different evaluation standards he perceived existed between exams and their class work on proof:

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 explicitly raise the issue of whether it would be possible to have a proof and a counterexample of the same statement, for Evan was clearly aware that Carol’s argument was not a counterexample to the particular statement in the problem.

**Discussion**

Although our analysis of the survey data alone offered considerable evidence (weak and strong) for the existence of the student misconception that it is possible to have a proof and a counterexample for the same assertion, our analysis of data from the follow-up interviews cast considerable doubt on this evidence. 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: stricter definitions could have reduced the number of student surveys that suggested the existence of the misconception. Although adopting stricter definitions could have lessened the gap between the tentative conclusions we had drawn from the survey data alone and the conclusions we drew after the inclusion of the interview data, the gap would not have been eliminated. The existence of the gap reinforces and exemplifies the point that student responses to surveys may, by themselves, offer a rather limited insight into their conceptions, and that follow-up interviews with selected students are important for the construction of a more trustworthy picture of these conceptions (see, e.g., Reading and Reid 2007).

The latter statement is more than a reiteration of the well known methodological principle that triangulation of multiple data sources allows for a more nuanced examination of research questions than when using a single data source (e.g., Marshall and Rossman 1999). The statement is also a cautionary remark that conclusions about students’ conceptions 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 it may not be possible to illuminate satisfactorily students’ conceptions about multifaceted mathematical ideas based only on 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, such as the methods that were used to validate the survey, the kind of questions included in the survey (e.g., open-ended versus multiple-choice), the conditions under which the survey was administered, and the coding scheme used to analyse the survey data.4

In addition to the opportunity that they afforded us to test the tentative conclusions we had drawn from our analysis of the survey data, the interview data gave us a wealth of other useful information about students’ conceptions. For example, our interview with Evan revealed an interesting layer of thinking about students’ criteria for evaluating the sample solutions in the survey. Specifically, this interview showed that, in their evaluations of sample solutions, some students placed more weight on the underlying reasoning of a solution than on whether there was a mistake in it. Noteworthy is Evan’s comment that his evaluation criteria would have been different had he responded to the same problem in an exam setting. According to him, exams tend to place more emphasis on the correctness of an answer than on the reasoning that led to it.

In spite of the aforementioned limitations in the conclusions that could be drawn based on the survey data alone, the sample solutions in the survey offered a meaningful context in which to discuss, during the interviews, students’ ideas about whether it would be possible to have a proof and a counterexample for the same assertion. This discussion was conducted with reference to Ben’s deductive argument, which could potentially be considered a proof, and Carol’s purported counterexample. Carol’s argument worked particularly well for the purposes of our research, for 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 assertion. Ben’s argument did not work as well as Carol’s argument: some students, like Emily, recognised the value of Ben’s argument, but did not accept it as a proof for various reasons (e.g., because they thought the argument needed ‘unpacking’, or because the argument did not feature some of the traits like symbolic notation that students at this stage typically perceive as ‘mathematical’). 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 possibility of having a proof and a counterexample for the same assertion: they suspected a mistake in Ben’s argument, and thus felt less hesitant about endorsing Carol’s counterexample. Given that the assertion 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 possibility of having a proof and a counterexample for the same assertion.

Future research on students’ conceptions about the relationship between proof and refutation could incorporate this modification into a revised version of the survey. Also, future research could expand the survey to include an item that would ask students to evaluate a valid counterexample and a believable ‘proof’ for a false assertion. Such an expanded survey, together with follow-up interviews to clarify students’ responses to it, would constitute a more comprehensive approach to eliciting students’ conceptions about the possibility of having a counterexample and a proof for the same assertion. Although our research did not identify this misconception, we cannot conclude that it is absent in the student population. Indeed, some students might show evidence of the misconception if it were investigated in a different way. It is also possible that less advanced students, younger students, or students with fewer experiences
with proof may have this misconception, and to a greater extent than the students who participated in our research.

Acknowledgements

The data reported in this article were collected and analysed with the support of a grant from the Economic Social and Research Council (ESRC) 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, policies, or endorsement of the ESRC. Part of an earlier version of the article was presented at, and will be published in the proceedings of, the Sixth Congress of the European Society for Research in Mathematics Education (CERME 6), in Lyon, France, in January 2009.

Notes

1. The use of the term ‘proof’ is restricted in this article to valid arguments that establish the truth of mathematical assertions.

2. Our use of the term ‘mis-conception’ should not be interpreted as lack of appreciation of students’ views about mathematics. Its use is intended to designate views that deviate from conventional views widely accepted nowadays in the field of mathematics.

3. Toward the end of the school year, the two teachers took more responsibility for planning the lesson sequences and the first author assumed more of a supportive role. This pattern of collaboration continued during the following school year (the two teachers taught the same groups of students in Year 11).

4. Although the survey we used was based in part on a section of a longer survey developed by the Longitudinal Proof Project (Küchemann and Hoyles 2001–03), the investigation of students’ conceptions about whether it is possible to have a proof and a counterexample
for the same assertion was not an apparent intention of the developers of the longer
survey. The methods we used for collecting and analysing the survey data were our own.

References

Balacheff, N. 1988. A study of students’ proving processes at the junior high school level. In
Proceedings of the Second UCSMP International Conference on Mathematics Education,
Mathematics.

companion to Principles and Standards for School Mathematics, ed. J. Kilpatrick, W.G.


Harel, G., and L. Sowder. 2007. Toward comprehensive perspectives on the learning and
teaching of proof. In Second handbook of research on mathematics teaching and learning,


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.

**Amina’s answer**
I tried this example:
1, 4, 6, 7, 9.

I then tried two more examples. Each had two even numbers and the total came to 27 each time. I could try other examples with two even numbers, they would come to 27 as well.

So Amina says it’s true

**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

**Davor’s answer**
I thought of these as the visible numbers:
1, 3, 6, 8, 9.

Two of them are even and when I add all the numbers I get 27.

So Davor says it’s true

Figure 1a. A mathematical problem and four sample solutions to the problem (as appeared in Küchemann and Hoyles 2001–03, Technical Report for the Year 8 Survey, 93).
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.

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

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

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

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

Figure 1b. Open-ended and multiple-choice questions (extended and slightly modified version of questions as appeared in Küchemann and Hoyles 2001–03, Technical Report for the Year 8 Survey, 93–94).