INCONSISTENCIES IN STUDENTS' UNDERSTANDING OF PROOF AND REFUTATION OF MATHEMATICAL STATEMENTS

Orly Buchbinder*, Orit Zaslavsky**

*University of Maryland; **New York University
Technion – Israel Institute of Technology

The study reported herein is part of a larger study[1] that examined high-school students' understanding of the roles of examples in proving. Data is based on a series of students' interactions with specially designed mathematical tasks that elicit their thinking. The findings provide a complex account of students' conceptions and reveal inconsistences in their understanding. In particular, all students in our study exhibited indicators of understanding that for a universal statement to be true it has to hold for all cases. At the same time, some of these students remained convinced that a statement can be 'proven' through examination of several confirming examples.

BACKGROUND

NCTM (2000) Standards and Common Core State Standards for Mathematics (CCSSI, 2010) state that reasoning and proving are an integral part of school mathematics. In order for students to engage in proving they need to develop an understanding of the status of empirical evidence in proving and refuting mathematical statements (e.g., Harel and Sowder, 2007). However, studies consistently show that students at all grades and levels tend to rely on examples that satisfy a given statement as sufficient evidence for proving it. This phenomenon is referred to as *empirical proof scheme* (Harel and Sowder, 2007), *naïve empiricism* (Balacheff, 1988), or *example-based proof* (Healy and Hoyles, 2000). In addition, students often hold incorrect views with respect to counterexamples: they reject them or treat them as exceptions (Balacheff, 1988). These studies suggest that developing an understanding of the status of examples in proving (and refuting) is a non-trivial process.

Although, the Standards define instructional goals and outcomes, they do not specify the methods for achieving them (CCSSI, 2010). Stylianides and Stylianides (2009) maintain that there has not been enough research into the ways of supporting students in developing coherent understanding of the role of empirical evidence in proving. Thus, students are often left to develop this understanding on their own, with insufficient direct instructional support.

The goals of our study were to explore high-school students' understanding of the status of empirical evidence in proving and refuting mathematical statements, along with ways in which this understanding can be diagnosed and enhanced. To address these goals we developed a framework (Buchbinder and Zaslavsky, 2009) that captures our conceptualization of *what it means to understand the status of examples in*

[1] The study was supported by Israeli Ministry of Education.

determining the validity of mathematical statements. The framework provided a basis for constructing special tasks that elicit students' conceptions and for analyzing students' conceptions.

What does it mean to understand the status of examples in proving?

Every mathematical statement can be characterized by a domain *D* of mathematical objects to which it refers (e.g., *'all integers ending with 7'*) and a proposition that describes a certain property *P* (e.g., *'multiple of 7'*). A universal statement that is based on domain *D* and property *P* states that every *x* in *D* has the property *P* (e.g., the false universal statement: '*every integer ending with* 7 *is a multiple of* 7'). An existential statement that is based on domain *D* and property *P* states that there exists *x* in *D* that has the property *P* (e.g., the true universal statement: '*there exists an integer ending with* 7 that *is a multiple of* 7'). With respect to a given domain *D* and a property *P*, four types of mathematical objects can be defined, based on whether or not an object *x* belongs to the domain *D* or not, and whether it satisfies the given property *P* or not: 1. An object that belongs to *D* and has the property *P* (e.g. $x=77$). This is a *confirming* example, for both universal and existential statements; 2. An object that belongs to *D* and does not have the property *P* (e.g. $x=17$). This is *counterexample* or a *contradicting* example for the universal statement and a *non-confirming* example for the existential statement; 3. An object that does not belong to *D*, and has the property *P* (e.g. $x=70$); 4. An object that does not belong domain *D* and does not have the property *P* (e.g. $x=71$). Objects of types 3 and 4 are *irrelevant* to both kinds of statements (universal and existential). We separate them as they may be interpreted differently in terms of their logical status. Our framework describes the logical status of each type of example with respect to the two types of statements (Buchbinder and Zaslavsky, 2009). Thus, one confirming example is insufficient for proving a universal statement, but is sufficient for proving an existential statement. One counterexample is sufficient for refuting a (false) universal statement, but a non-confirming example is insufficient for refuting an existential statement. Irrelevant examples have no logical status in the sense that they do not support any proof or refutation of a statement.

In the spirit of Borgen and Manu (2002) we conceptualize 'understanding' of the roles of examples in determining the validity of mathematical statements in operational terms as becoming fluent with types of inferences that can and cannot be drawn based on the four types of examples with respect to two types of statements. In this paper we focus on students' understanding of the status of confirming and contradicting examples in proving or refuting of *universal* statements. According to the conceptual framework such understanding entails: (1) recognizing the type of the statement (universal); (2) realizing that in order for it to be true the proposition has to hold for all the elements in the domain; (3) realizing that confirming examples are insufficient for proving; and (4) understanding that a single counterexample is sufficient for refuting a false universal statement.

THE STUDY

Instruments

Based on the conceptual framework presented above, we constructed a collection of 6 types of tasks that aim at revealing and enhancing students' understanding of the roles of examples in proving. Each type of task addressed various aspects of the framework, and the collection as a whole covered all aspects of the framework[2] (Buchbinder & Zaslavsky, in press).

The tasks drew on topics from the regular 9th and 10th grade mathematics curriculum in Israel. While we wanted to ensure that students have the relevant content knowledge to cope with the tasks, we tried to confront them with statements that were unfamiliar to them, and which had a potential to evoke uncertainty regarding their truth-value. Uncertainty is widely recognized as a powerful trigger for creating situations that promote students' intellectual need for proof (e.g., Zaslavsky, 2005). The process of resolving the uncertainty can both reveal and enhance students’ understanding. One type of task, inspired by Healy and Hoyles (2000) and by Zaslavsky and Ron (1998), which we term *"Who is right?"*, creates uncertainty by confronting students with a false universal statement followed by arguments of five hypothetical students stating their opinion on its truth-value. Student A uses multiple confirming examples to “prove” the statement; Student B refutes the statement with a single counterexample; Student C maintains that multiple counterexamples are needed; Student D maintains that the statement is false but does not accept counterexamples as sufficient, and requires a general argument; Student E maintains that since both confirming and contradicting examples exist, the truth value of the statement cannot be determined.

> Five students worked independently on determining whether the following statement is true or false: *For every natural number n,* n^2+n+17 *is a prime.*
> For each of the arguments raised by the students below, decide whether it is correct or not, and justify your decision.
>
> > *Tali:*
> > I checked the value of the expression for 10 different natural numbers (odd, even, prime) and in all cases the result was a prime. For example:
> > For $n=2$, I got *23*, which is a prime. For $n=3$, I got *29*, which is a prime.
> > For $n=11$, I got *149*, which is a prime. Thus the statement is **true**.
>
> > *Yael:*
> > I tried $n=16$ and got: $16^2+16+17=289$. *289* is not a prime since $17\cdot 17=289$.
> > Thus, the statement is **false**.

Figure 1: Two parts of the algebraic version of the task ‘*Who is right?*’

[2] For discussion of types of tasks and the underlying design principles, see Buchbinder & Zaslavsky (2012).

For each argument, participants were asked to determine whether it is correct or not and to justify their decision. Figure 1 shows 2 parts (Students A & B) of the algebraic version of the task.

Data Collection

Two parallel versions of the tasks (algebraic and geometric) were implemented with six pairs of top-level 10^{th} grade students from two distinct schools in the northern area of Israel. The group included 7 girls and 5 boys who volunteered for the study. Each pair of students participated in a series of six, one hour long, task-based interviews. Across all task types, each pair responded to 11 tasks involving universal statements. During the sessions, students coped with the different tasks with minimal intervention from the interviewer. There were no time constrains, so students could discuss the task with each other as much as they needed. Data collection included video recordings of the interviews, students' written work and researcher field notes.

Data Analysis

The data were analyzed using qualitative research methodology. Students' written work and utterances consistent with the framework were coded as 'indicators of understanding' (IOU). E.g., expressions stating that confirming examples are insufficient for proving. Students' responses inconsistent with the framework were coded as 'non-normative responses' (NNR). E.g., explicit acceptance of an example-based 'proof' as valid. Note that only *explicit* indicators of understanding (or mis-understanding) were coded.

Each task was chosen as a unit of analysis, even though multiple IOUs and NNRs could occur in it. Also, since students worked on the tasks in pairs, and it was not possible to distinguish between individual contributions, both types of indicators (IOU and NNR) were assigned to pairs, not to individuals.

FINDINGS

The findings provide a complex account of students' understanding. All students exhibited IOUs in each one of the aspects outlined by the framework. Note that each pair received 11 tasks involving universal statements, thus, there were 66 possibilities to exhibit IOUs, NNRs, or both.

With respect to *confirming* examples, we recorded 16 IOUs (Table 1). This relatively low rate (only 24%) can be related to the fact that only *explicit* indicators of understanding were recorded. As shown in Table 1, all pairs provided at least one explicit IOU that confirming examples are insufficient for proving. At the same time, *all* pairs also exhibited at least one NNR, such as justifying a statement by checking several confirming examples, or accepting such justifications, made by others, as valid. Overall, the same number of IOUs and NNRs was documented for understanding the status of confirming examples, with only two pairs exhibiting more IOUs than NNRs.

	Understanding the status of Confirming examples in proving		Understanding the status of Counterexamples in refuting	
Student pairs	No' of IOU	No' of NNR	No' of IOU	No' of NNR
Neta and Ronit	2	3	8	3
Nurit and Limor	6	4	18	1
Omer and Yaron	1	1	11	0
Tami and Natalie	2	4	6	3
Keren and Ben	2	3	11	2
Paz and Ronen	3	1	8	0
Total	16	16	62	9

Table 1: Distribution of indicators of understanding (IOU) and non-normative responses (NNR) with respect to the status of examples and counterexamples in proving and refuting universal statements.

All students provided multiple evidence of understanding of the role of *counterexamples*. Overall, 62 such IOUs were documented. In other words, in 94% of tasks involving false universal statements, students provided explicit indicators of understanding that a single counterexample refutes a universal statement. The 9 cases of NNRs reflect the instances in which students required multiple counterexamples for refuting a false universal statement.

We illustrate our findings through the case of one pair of students' encounters with the parts of the task illustrated in Figure 1.

The case of Neta and Ronit

Neta and Ronit started by checking some small values of *n*, which appeared to confirm the statement. Then they turned to examine the hypothetical students' arguments:

Ronit: Is Tali's response correct? Yes. Why? ….According to her results…

Neta: [While writing] In addition to Tali, we tried several numbers and every time the result was a prime. Thus, Tali is right.

Ronit: Wait! Look at the response of Yael. [Reads it aloud]. *289* is not a prime…

Neta: She is right, what can I tell you…

…

Ronit: So, first of all, Tali is right. It [the statement] is true but not for all natural numbers. Because here, Yael proved that if we take *n=16*…..It's not that the statement is false…. It's like…. this statement is false. It's not for every natural n. So here, Yael is right and Tali not. Because she [Tali] didn't check all natural numbers. Perhaps some of them do not [satisfy the statement].

Neta: The statement is false.

Ronit: So, Tali says that the statement is true, because she tried different numbers and the resulting numbers are primes. She is right, like, in her way, but she is not right in that…. the statement is false.

Neta: So, both Tali and Yael are right.

Ronit: Yael is right. It is not "for every natural *n*".

Neta: Yes. [While writing] Yael is right because she found a proof that not every natural number that we substitute for *n* gives us a prime number.

Neta and Ronit did not change their written justification for Tali's utterance. They moved on with the task but later returned to Tal's response. It seems that they realized that their acceptance of both Tali's and Yael's arguments constitutes a contradiction. Following is their attempt to resolve the conflict:

Ronit: OK. Now we have to go back to Tali. [Reads Tali's response aloud]. She is right!

Neta: Definitely. She is right. We can tell that Tali is right since we do not know what happened earlier.

Interviewer: What do you mean?

Neta: We have met her [Tali] earlier. And she is right. For example, we meet Tali on Sunday, and she proves to us that the statement is true. She gives us examples, gives us the whole investigation that she made, and she shows us that she got it right. We read her report, and we see that she is right. The next day, we meet someone else - Yael, and she shows us that the statement is false. So the first girl was right, but the second girl is also right. Afterwards.

Ronit: We can say that it [the statement] is false based on what Yael did. It is false because we saw what Yael did and we found out that not for every natural number that we substitute for *n*, the result will be a prime.

Interviewer: Do I understand correctly, that if you would not have met Yael, you would say that Tali's response is correct?

Ronit: Exactly.

Neta: Yes.

DISCUSSION

Applying our framework to analyse Ronit and Neta's case we can see that they correctly identified the statement as universal and explained that it has to hold for all natural numbers. They accepted Yale's counterexample as refutation and used it to justify why the statement is false. At the same time, Neta and Ronit referred to Tali's example-based argument as valid, even after direct prompting. Their line of reasoning can be described as "the statement is true, unless shown otherwise". Outside mathematics it is common to regard repeating evidence as true unless contradicting evidence is presented; which, in turn, does not necessarily overthrow previous results. It is possible that Neta and Ront's reliance on confirming examples for justifying universal statements stems from such 'every-day logic'. This is consistent with Leron and Hazan (2009) who maintain that in case of conflict between mathematical reasoning and every-day logic, students often resolve the conflict in favour of the latter.

Our findings outline a complex picture of students' understanding of the roles of examples in determining the validity of mathematical statements. Specifically, we identified two types of inconsistencies. The first type of inconsistency is manifested as discrepancies between students' responses to different tasks. In particular, with respect to the status of *confirming* examples in proving, the students, as a group, exhibited the same number of non-normative responses as the number of indicators of understanding (Table 1). This means that while on some tasks the students stated explicitly that confirming examples are insufficient for proving, on other occasions (or even on the same task) they used confirming examples to justify that a certain universal statement is true.

Some students justified the use of confirming examples by maintaining that they have been chosen in a specific way - systematically or by random. Balacheff (1988) terms this type of reasoning - *crucial example*. Neta and Ronit justified their reliance on confirming examples by referring to the timing of occurrence of a counterexample. Though their reasoning was unique for our group of students, we hypothesise that it can occur with other students outside our group. Thus, our findings concur with the literature on students' difficulties to accept the limitation of empirical evidence as means for proving (Harel & Sowder, 2007, Healy and Hoyles, 2000).

Contrary to the literature on counterexamples (Balacheff, 1988, Zaslavsly and Ron, 1998) the students in our study exhibited strong understanding of the status of *counterexamples,* accepting them as refutations. The data in Table 1 and Neta and Ronit excerpts from Neta and Ronit's discussion illustrate this finding.

The second type of inconsistency in students' understanding of the roles of examples in determining the validity of universal statements is their apparent lack of connection between the roles of examples and counterexamples in this process. From a logical point of view, to understand that in order for a universal statement to be true it must hold for all elements in the statement’s domain and that a single counterexample is sufficient for refuting a false statement, implies that confirming examples are insufficient for proving and that a general justification is needed (Harel and Sowder, 2007, Stylianides and Stylianides, 2009). Our findings, and specifically the case of Neta and Ronit, suggest that students held two conceptions that logically are contradicting.

Implications for education

Supporting the development of students' understanding of proving, is a non-trivial task for mathematics educators. One approach to that involves designing instructional tasks that highlight limitations of empirical evidence by emphasizing the role of counterexamples (Buchbinder and Zaslavsky, 2012, Stylianides and Stylianides, 2009). The type of task *Who is right?* proved successful in evoking uncertainty, and in promoting students' awareness of their own conceptions. In most cases this led to enhanced understanding of the roles of examples in proving. However, as our data show, some students did not resolve the uncertainty in mathematically correct way. More research is needed to determine the types of tasks and instructional scaffolding needed to promote students' understanding of the roles of examples in proving.

References

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

Buchbinder, O. & Zaslavsky, O. (2009). A framework for understanding the status of examples in establishing the validity of mathematical statements. In Tzekaki, M., Kaldrimidou, M. & Sakonidis, C. (Eds.). *Proc. 22nd Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 2, pp. 225-232). Thessaloniki, Greece.

Buchbinder, O., & Zaslavsky, O. (in press). A Holistic Approach for Designing Tasks that Capture and Enhance Mathematical Understanding of a Particular Topic: The Case of the Interplay between Examples and Proof. To appear in *Proc. of the Int. Commission on Mathematical Instruction (ICMI), Study 22: Task Design in Mathematics Education.* Oxford, UK.

Borgen, K. L. & Manu, S. S. (2002). What do students really understand? *Journal of Mathematical Behavior, 21*, 151-165.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning.* (pp. 805-842). NCTM. Reston, Va.: Information Age Pub Inc.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education, 31(4),* 396-428.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

Leron, U., & Hazzan, O. (2009). Intuitive vs. analytical thinking: four perspectives. *Educational Studies in Mathematics, 71*, 263–278.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards: Mathematics.* Washington DC: Council of Chief State School Officers.

Stylianides, A. J., & Stylianides, G. J. (2009). Facilitating the Transition from Empirical Arguments to Proof. *Journal for Research in Mathematics Education, 40(3)*, 314-352.

Zaslavsky, O. (2005). Seizing the Opportunity to Create Uncertainty in Learning Mathematics. *Educational Studies in Mathematics, 60*, 297-321.

Zaslavsky, O., & Ron, G. (1998). Student's Understanding of the Role of Counter-Examples. In A. Oliver & K. Newstead (Eds.), *Proc. 22nd Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 4, pp. 225-232). Stellenbosch, South Africa: PME.