Do Instructors Value Examples in Proof?

Jordan Kirby
Francis Marion University

During the transition from procedure-based mathematics courses to proof-based mathematics courses, many students struggle to understand the purpose of examples. A productive use of examples can help both the generation and understanding of proofs. Although recent advances in research have argued the importance of productive uses of examples, there is little research investigating the alignment of research with current practices of instructors or of how the instructors view research on example-use. Findings from this study indicate instructors are aligned with mathematics education research on differing levels of example-use. However, many instructors may be hesitant to see examples be used by their students in written work. Despite this hesitation, instructors still encourage students to use examples to aid their understanding of proof.

Keywords: Proof, Examples, Instructors

Introduction

Student difficulties in understanding and producing proofs are well documented in mathematics education research (e.g., Ellis et al., 2019; Harel & Sowder, 2007; Stylianides, 2007). One such difficulty with proving comes from the effective use of examples to aid proof production and understanding (e.g., Aricha-Metzer & Zaslavsky, 2019; Epp, 2003; Zaslavsky & Knuth, 2019). Although there is much research in understanding how and why students use examples when proving, little research exists studying instruction in proof classes based around examples (Zaslavsky & Knuth, 2019). Zaslavsky and Knuth (2019) noted, "Very little research has focused on the nature and design of instructional practices that facilitate the development of students' abilities to strategically think about and productively use examples as they engage in proving-related activities" (p. 243).

This paper begins the process of addressing Zaslavsky and Knuth's (2019) call for research by investigating the perceptions of students' use of examples held by introduction to proof (ITP) instructors. To develop effective instructional interventions, it is imperative to understand what practitioners currently do in their classrooms (Desimone & Garet, 2015). This study aims to lay groundwork necessary for future research to implement instructional interventions by ascertaining the current perceptions and reactions held by ITP instructors across the United States. To accomplish this goal, I seek to answer the following research question: **How do instructors of ITP classes perceive students' understanding and use of examples when proving?**

Theoretical Framework

I draw upon the work of Aricha-Metzer and Zaslavsky (2019) in distinguishing between productive and unproductive use of examples when proving. Aricha-Metzer and Zaslavsky defined productive examples when the example helped the prover make progress or gain insight into key aspects of the proof. An unproductive use of examples in proof was when the prover used an example but no progress towards developing or understanding a proof was made.

One of the aspects of proof in my work is how examples are used and potentially included in the formal written work provided by students in an ITP course. Balacheff (1987) described categories of example-use in proof; these categories include: naïve empiricism, crucial experiment, and generic example. Table 1 summarizes these categories.

Table 1 Balacheff (1987) example-use categories

Name of Category	Definition (From Balacheff 1987, p. 19-20)
Naïve Empiricism	Drawing from the observation of a small number of cases the certainty of the truth of an assertion
Crucial Experiment	A process of validation of an assertion in which the individual explicitly poses the problem of generalization and solves it by betting on the realization of a case that he recognizes as being as unspecific as possible
Generic Example	Explanation of reasons for the validity of an assertion by carrying out operations or transformations on an object present not for itself, but as a characteristic representative of a class of individuals

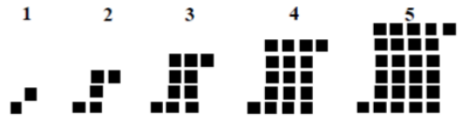
Aricha-Metzer and Zaslavsky (2019) explain that when examples are used akin to the generic example level described by Balacheff (1987), the examples used by students were typically classified as productive. In defining generic example, I build on the work provided by Rø and Arnesen (2020) citing Reid and Vallejo Vargas' (2018) two criteria of a generic example: evidence of awareness of generality and mathematical evidence of reasoning. In short, the argument must conclude with a general claim about the original problem to be proven as well as contain reasoning directly linked from the example to be classified as a generic example.

Methodology

Participants of this study include university professors across the Southeastern United States actively teaching their university's version of an ITP course. Eleven professors agreed to participate in the study with varying demographic information such as Carnegie classification of the university, type of Ph.D. received, tenure at the university, and self-described teaching style. One-hour semi-structured interviews were conducted with all participants online through Zoom. Participants were emailed two questions before the study shown in Figure 1.

Growing S Pattern Task

Question 1. Consider the pattern below. How many square tiles would there be in the eighth step of this pattern?



Question 2. Write an expression for the total number of square tiles in the figure at an arbitrary step (n) of the pattern.

Question 3. Prove that your expression is a valid representation of the number of tiles at step n of the pattern. You may use drawings, words, numbers, and/or symbols for your proof.

Question 1. Given the number of vertices in a polygon, find a method to calculate the number of diagonals of the polygon. Explain why your answer is correct.

Figure 1. *Growing S pattern task and polygon diagonal task*

Participants were not expected to solve these problems before the meeting. The problems were chosen to be easy enough to solve for most faculty members during the interview so time could be spent discussing student use of examples rather than proving the claim. Four participants mentioned they would use or have previously used one or both of the questions shown to them in the interview. Ten of the eleven participants expressed belief these two questions were appropriate for early semester in an ITP course.

During the interview, participants were shown three student work samples for each of the two questions provided. These student work samples were chosen to exemplify the three levels of Balacheff's (1987) framework on example-use. Figure 2 details one of these student work samples: Carla at the crucial experiment level solving the growing S pattern task.

Carla

Question 2. Write an expression for the total number of square tiles in the figure at an arbitrary step (n) of the pattern.

$$[(n+1)*(n-1)]+2$$

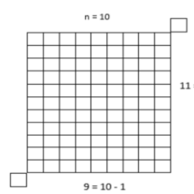
Question 3. Prove that your expression is a valid representation of the number of tiles at step n of the pattern. You may use drawings, words, numbers, and/or symbols for your proof.

n = 1

[(1+1)*(1-1)] +2 = 2

I tried the smallest case and a larger case. Since this formula holds for both, it makes sense for all cases.

n = 10



11 * 9 = 99 tiles

99 + 2 = 101 for the total tiles

Figure 2. *Carla solving the growing S pattern task at the crucial experiment level.*

During the interviews, the order of the two questions was randomized as was the order the student work samples were shown. Student work samples were always kept clustered so a participant would either receive the three growing S pattern task solutions first or the three polygon diagonal task solutions first. When presented with student work, participants were asked to mention anything they noticed about the student work sample. Participants were informed this may include if they believe the student is correct, if they liked what the student did, if they had comments on any of the student work, if they had seen something similar to this solution in their classroom, or any questions they would have as instructors if a student in their class presented them with a solution like this as a homework assignment.

After responding to all six student solutions, participants were asked to group the student solutions into as many or as few categories as they deemed appropriate and name each category. Finally, participants were asked to rank the six student responses from 1 to 6 with a response of 1 signaling the best attempt at a proof and a response of a 6 signaling the most work needed to be done for a prof. After finalizing their rankings, participants were asked to provide instructional feedback to advance the students forward whom they ranked best and worst. This manuscript will focus on the ranking task and transcript from across the responses to student solutions.

Analysis

Video data from Zoom interviews with faculty were transcribed by talk turn. I define a talk turn as all utterances spoken by one participant until interrupted by the interviewer. This data was then open coded for common remarks about students' example use among the faculty participants. In order to be considered as a potential code, the code theme had to be included at least twice within the same participants' transcript and across three different participants. For instance, Dr. Amanda's transcript showed a potential code for *mathematical induction*. This was coded when Dr. Amanda remarked discussing Carla's work, "Okay, so here's some induction ideas. You've got the base case. And a small case for the largest." Similarly, Dr. Amanda remarked on Nancy's work (naïve empiricism, growing S pattern task), "[Nancy] tried a smallest case. And the larger case, is suggesting that there's maybe some more idea there of like, maybe I can interim check intermediate cases, which sort of feels like you're building to induction." Dr. Amanda continued to note similar remarks about two other participants' work. This was sufficient within Dr. Amanda's transcript to be a potential code. If this code showed up in another transcript from another participant, I included the code as a theme. For the ranking task, screen captures were taken and compiled to look for common rankings. The rankings were compiled into a table with the feedback given for each participant coded as described earlier.

Results

I will answer my research question in two parts. First, I will list how participants ranked the student responses. Then, I will discuss one of the common themes that emerged from transcribing the data.

Rankings

Participants were asked to rank the six student work samples from a 1, signaling the best attempt at a proof, to a 6, signaling the worst attempt at a proof. The six student work samples were split between the two questions listed in Figure 1 with one student work sample at each of the levels described by Balacheff (1987). Table 2 shows the rankings of each student argument across the 11 faculty participants.

Table 2. Participant rankings of student work samples from 1 to 6

	1st	2nd	3rd	4th	5th	6th
Gina's argument (generic example)	10	1				
Jacob's argument (generic example)	1	9	1			
Carla's argument (crucial experiment)		1	9	1		
Eric's argument (crucial experiment)					8	3
Aaron's argument (naïve empiricism)			1	5		5
Nancy's argument (naïve empiricism)				5	3	3

Table 2 shows how frequently each student was ranked in positions 1 through 6 in the 11 interviews. This table gives evidence to the participant's inclination to rank students' arguments in alignment with the levels of example-use described by Balacheff (1987). Ten of the 11 participants ranked Gina's argument as the best attempt at a proof. Many participants expressed although they similarly liked Jacob's argument, the design of the growing S pattern task led to a seemingly more appealing argument leading to Gina's ranking above Jacob's. Although Eric (crucial experiment – polygon diagonal task) was positioned to exhibit a more robust use of examples than students at the naïve empiricism level, some of the wording used by Eric caused concern for many participants. Eric included statements such as, "This led me to a pattern" and "If my pattern holds," without every explaining what his pattern was. Most participants expressed concern with these phrases when ranking Eric and decided to rank him in the bottom two for lack of clarity.

Theme of Acceptance of Example-Use

A common theme that emerged from the interviews was the theme of acceptance or rejection of example-use. Of the 11 interviews conducted, 5 participants expressed concern with seeing examples of any form included in the written work shown by students. The other six participants alternatively expressed interest and praise for the use of examples to aid proof production and understanding.

The theme of acceptance or rejection was common amongst all 11 participants. A participant was classified as either accepting or rejecting examples entirely based on unprovoked remarks of the first phase of the interview responding to student work samples. All participants expressed their preference of acceptance or rejection of examples at least twice within their interview. The acceptance category included comments promoting student reasoning and praise of the use of examples to aid the understanding or production of the proof. The rejection category instead expressed concern in every instance for the example included in the written work and felt the proof was more lacking with the inclusion of the example. There were no differences in demographic information between the acceptance and rejection groups.

One instance of an acceptance category comes from Dr. Tucker. When responding to Jacob (generic example – polygon diagonal task), Dr. Tucker remarked, "So, I do like the, I guess he kind of start out with, you know, thinking of a fixed number. Just to kind of get an idea of what to do with it." Later, Dr. Tucker similarly commented on both Gina (generic example – growing S pattern task) and Carla's (crucial experiment – growing S pattern task) work praise for the pictures drawn and an interest in having more students use this type of reasoning. These comments were sufficient to include Dr. Tucker in the acceptance of example-use category. Participants in the acceptance category did not accept all forms of examples as sufficient for proof. Five of the six participants in this category expressed concern for both Nancy (naïve empiricism – growing S pattern task) and Aaron's (naïve empiricism – polygon diagonal task) use of examples in their proof. Dr. Tucker commented on both Eric (crucial experiment – polygon diagonal task) and Aaron's work,

> I guess generally, a pitfall I see with students is trying to just do a couple of examples. And then that kind of just, I know this doesn't say prove (referring to the wording of the question given), but kind of thinking that counts as a proof. And yeah, it feels like they just kind of need to bring more justification of why would this work in general.

Although Dr. Tucker and others in the acceptance of example-use category appreciated the work shown by students when proving, this did not mean they accepted incomplete proofs. Participants in the acceptance of example-use category expressed appreciation for the productive uses of examples as described by Aricha-Metzer and Zaslavsky (2019) and concern for the unproductive use of examples.

Participants in the rejection category responded to student work regardless of Balacheff's (1987) associated level with explicit concern for the written inclusion of examples in the product produced. These comments were made for both productive uses of examples as well as unproductive uses of examples. For instance, Dr. Hubert remarked about Gina's work, "I mean, it's reasoning by example. Which is always rough. I would say the n minus one, n plus one part could have helped me earlier on." Dr. Hubert here refers to how Gina starts with a drawing of a specific case. From this case, Gina gathers information about the nature of the problem and correctly answers the problem producing a generic formula to solve the growing S pattern task along with a short algebraic proof. Dr. Hubert expressed concern that Gina left the reasoning with her example in the work instead of removing this first.

Later in the interview, Dr. Hubert again remarked about Gina's work, "So I'm struggling right now with if you can do this (referring to Gina's formal answer at the bottom of the page), how does it not stick? I mean, I ought to be able to just state that part earlier right?" Dr. Hubert was struggling in the interview to understand why if Gina knew the complete answer she could not remove the discussion about how she arrived at the answer and instead state her proof more formally. This idea of formality was raised by Dr. Sarah about Jacob's work with examples commenting, "The fact that Jacob does still have, like the thinking about his example. To me, I would not want to see evidence of it in the proof." The rejection of examples category was largely defined by the difference in how formality was viewed as a necessary condition in an ITP course.

Discussion and Conclusion

This research looks to make progress in answering the call to action from Zaslavsky and Knuth (2019) to include more instructional tools about examples in proving. I set out to answer my research question: How do instructors of ITP classes perceive students' understanding and use of examples? Answering this question will assist future researchers in developing instructional interventions for practitioners of ITP courses by considering the current state of the field. I answered this research question in two ways. I first answered my research question through the participants' rankings of the provided student work based on the completeness of a proof. Second, I showed a theme for common responses to student work involving example-use through the acceptance or rejection of examples in proving.

Through the ranking task, participants are largely aligned with the levels of example-use described by Balacheff (1987). Although there was a discrepancy with Eric, most of the participants still ranked the student work largely by type of example used by the student. This finding can help future instructional interventions find a starting point for helping instructors of ITP courses aid students in using examples more productively as instructors seem to be implicitly aware of the categories of example-use without any background information given. Participants of the study were asked in a short debrief after the interview if they were familiar with any research on generic examples or Balacheff's framework. None of the participants were familiar with any of this research.

Through the coding, a potential theme of acceptance or rejection of example-use came to light. This theme should be considered when designing instructional interventions in the future for example-use in proof. There were no differences in any demographic information collected between the acceptance of example-use group and the rejection of example-use group. For professional development to be most effective, practitioners need to be met at their current level (Desimone & Garet, 2015). This study gives some evidence that for some practitioners, work may need to be done by mathematics educators on convincing their future professional development groups of the effectiveness and usefulness of generic examples. There is a potential conflict with the message sent to students if practitioners value examples to aid a students' understanding while in class but then do not want to see examples in the written work. Finally, the issue of formality of written work in an ITP course was repeatedly seen. Participants in the acceptance of example-use group did not bring up issues with formality of the presentation of a proof. Participants in the rejection of example-use group frequently brought up issues of how including work in a proof detracted from the argument.

I encourage future researchers to investigate this disconnect of example-use and acceptance in other areas of mathematics, other types of proof, and in more detail. The comments taken in this study came unprovoked as the interview protocol did not include questions about acceptance or rejection of example-use. Directed questioning towards types of examples and what a proof should look like would benefit future work in developing instructional interventions for ITP classes. Mathematics education researchers and practitioners should consider the idea of formality of proof and determine how to best approach future research in the field of example-use in proof.

References

Aricha-Metzer, I., & Zaslavsky, O. (2019). The nature of students' productive and non-productive example-use for proving. *The Journal of Mathematical Behavior, 53*, 304–322. https://doi.org/10.1016/j.jmathb.2017.09.002

Balacheff, N. (1987). Processus de preuve et situations de validation. *Educational Studies in Mathematics, 18*(2), 147–176. https://www.jstor.org/stable/3482413

Boyle, J. D., Bleiler, S. K., Yee, S. P., & Ko, Y.-Y. (2015). Transforming perceptions of proof: A four-part instructional sequence. *Mathematics Teacher Educator, 4*(1), 32–70. https://doi.org/10.5951/mathteaceduc.4.1.0032

Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society, and Education, 7*(3), 252-263

Ellis, A. B., Ozgur, Z., Vinsonhaler, R., Dogan, M. F., Carolan, T., Lockwood, E., Lynch, A., Sabouri, P., Knuth, E., & Zaslavsky, O. (2019). Student thinking with examples: The criteria-affordances-purposes-strategies framework. *Journal of Mathematical Behavior, 53*(January 2017), 263–283. https://doi.org/10.1016/j.jmathb.2017.06.003

Epp, S. S. (2003). The role of logic in teaching proof. *The American Mathematical Monthly, 110*(10), 886–899. https://doi.org/10.1080/00029890.2003.11920029

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 805–842). Information Age Publishing.

Reid, D., & Vallejo Vargas, E. (2018). When is a generic argument a proof? In A. J. Stylianides, & G. Harel (Eds.), *Advances in Mathematics Education Research on Proof and Proving* (pp. 239-251). Springer International Publishing. https://doi.org/10.10007/978-3-319-70996-3_17

Rø, K., & Arnesen, K. K. (2020). The opaque nature of generic examples: The structure of student teachers' arguments in multiplicative reasoning. *Journal of Mathematical Behavior, 58*(2020), 100755. https://doi.org/10.1016/j.jmathb.2019.100755

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*(3), 289–321. https://doi.org/10.2307/30034869

Zaslavsky, O., & Knuth, E. (2019). The complex interplay between examples and proving: Where are we and where should we head? *Journal of Mathematical Behavior, 53*(October 2018), 242–244. https://doi.org/10.1016/j.jmathb.2018.10.001