Journal of Educational Research in Mathematics
2020, Special Issue, 229 ~ 248

Do Generic Proofs Improve Proof Comprehension?

Kristen Lew*†, Keith Weber**, Juan Pablo Mejía Ramos **

Professor, Texas State University, USA
**Professor, Rutgers University, USA*

ABSTRACT

Students often have difficulty understanding the proofs that they read. Some mathematics educators have suggested using generic proofs to improve students' proof comprehension. However, few empirical studies about students' perceptions of generic proofs or whether students understand generic proofs better than traditional proofs have been conducted. In this paper, we describe a qualitative interview study and a quantitative experimental study exploring these issues. Our first finding is that generic proofs are popular with students, who believe they have the potential to improve proof comprehension. Our second finding is that in a randomized controlled experiment, students did not learn more from reading a generic proof than a traditional proof. The significance of these findings is discussed at the end of the paper.

Key Words: Generic proof; Proof; Proof comprehension

INTRODUCTION

In advanced mathematics lectures, professors typically spend a substantial amount of time presenting proofs to their students (Weber, 2004). For instance, in her small-scale study, Mills (2011) found that in the advanced mathematics lectures that she observed, professors spent half of their lectures presenting proofs. A similar focus on proof appears in textbooks for advanced mathematics courses (e.g., Raman, 2004). One reason that proofs are presented to students is to enhance students understanding of how to prove certain types of statements (e.g., Weber, 2002) and of why certain statements are true (e.g., Hersh, 1993; Hanna, 2018).

Given the time and importance placed on proof in advanced mathematics courses, it is natural to wonder whether proofs achieve their goal of enhancing understanding. There is a growing body of evidence that suggests a negative answer to this question. Mathematicians have generally lamented that students frequently do not understand the proofs that they read. The mathematician Carl Cowen (1991) addressed the issue as follows:

> If you need evidence that we have a problem, let one of your B students … explain the statement and proof of a theorem from a section in the book that you have skipped. My students, at least, do not have the innate ability to read and understand what they have read. When I ask them to read a problem and explain it to me, the majority just recite the same words back again (p. 50).

Conradie and Frith (2000) expressed similar sentiments, noting that in their experience, students had difficulty when they were asked comprehension questions about proofs that they just read. Mathematics educators too have remarked that

† corresponding author
Email: kristen.lew@txstate.edu, keith.weber@gse.rutgers.edu, jpmejia@math.rutgers.edu

Please cite this article as: Lew, K., Weber, K., & Ramos, J. P. M. Do generic proofs improve proof comprehension? *Journal of Educational Research in Mathematics* (2020). https://doi.org/10.29275/jerm.2020.08.sp.1.229

students find the proofs that they read to be generally confusing or pointless (e.g., Harel, 1998; Porteous, 1986; Rowland, 2001). These observations tend to be anecdotal, but they are supported indirectly by some empirical studies. A number of studies have found that when undergraduates with experience in advanced mathematics courses are asked to determine whether a given argument constitutes a proof, they perform poorly (e.g., Alcock & Weber, 2005; Inglis & Alcock, 2012; Ko & Knuth, 2013; Selden & Selden, 2003; Weber, 2010; see also Knuth, 2002, for an illustration that secondary mathematics teachers with advanced mathematics training find this task difficult). Of course, understanding a proof and checking a proof for correctness are different activities (Mejía-Ramos & Inglis, 2009). Nonetheless, if students cannot distinguish a correct proof from a flawed argument, this suggests that they might not understand the proofs they read all that well. In our recent studies in lectures in advanced mathematics courses, we have found that students in advanced mathematics courses cannot identify the most important points that their professors are trying to convey when they present a proof (e.g., Krupnik, Fukawa-Connelly, & Weber, 2018; Lew, Fukawa-Connelly, Mejía-Ramos, & Weber, 2016), which again suggests that these proofs will not play the roles of explanation, illustrating methods, and systematization (c.f., deVilliers, 1990) that their professors were hoping to convey when presenting the proofs.

In this paper, we discuss a particular way that mathematicians and mathematics educators have proposed to enhance students' understanding of the proofs that they read. Namely, we explore the issue of whether students can learn more from reading a generic proof (Rowland, 2001) than a proof that was presented in a traditional manner. We will elaborate on what generic proofs are shortly, but the main idea is that a generic proof of a universally quantified statement (i.e., a "for all" statement) is an argument that shows the statement is true for a carefully chosen representative object, where the calculations and reasoning are presented in a transparent manner that allows the reader to see that the same types of calculations and reasoning would work for any object within the scope of the quantification. Many researchers have endorsed generic proofs as a means to enhance students' understanding (e.g., Leron & Zaslavsky, 2013; Mason & Pimm, 1984; Rowland, 2001; Weber, Housman, & Porter, 2008). However, despite the enthusiasm for generic proofs, we argue that there is little empirical evidence in the literature demonstrating that generic proofs actually improve students' comprehension. The goal of this paper is to present an exploratory study investigating two issues: Do undergraduate mathematics students believe proofs can engender more understanding than traditional proofs? To what extent do generic proofs actually improve undergraduate mathematics students' understanding of proofs, when compared to traditional proofs?

We organize the paper as follows: First, we present a literature review of alternative styles of proof presentation that mathematics educators have endorsed to improve students' proof comprehension. A theme from this literature is that although alternative styles of proof presentation are often lauded by mathematics educators and are sometimes popular with students, when their efficacy is empirically tested, researchers have found that they generally do not improve students' comprehension. Next, we present a theoretical model of what it means to understand a proof and how this understanding can be assessed (Mejía-Ramos et al., 2012). In the Interview Study section, we present an exploratory study in which 10 students were asked to read a generic proof and then discuss its strengths and weaknesses. In the Quantitative Comparative Study section, we present a controlled experiment with 106 university mathematics students in which half read a traditional proof of a theorem and the other half read an analogous generic proof of the same theorem, and both groups were given a post-test measuring the comprehension of the proofs that they read. Our main findings from the interview and quantitative comparative studies were that although generic proofs were popular with students, they did not improve students' comprehension of the proof. Finally, we discuss what can be concluded from these exploratory studies and suggest directions for future research.

LITERATURE REVIEW

As we noted in the introduction, mathematics educators have often complained that students do not understand the proofs that they encounter. This is a particularly pressing issue in advanced mathematics, where proofs play a prominent role in mathematics lectures (e.g., Mills, 2011) and textbooks (e.g., Raman, 2004). Indeed, as we have argued elsewhere, proofs are the dominant types of instructional explanations that students receive in their advanced mathematics courses (Lai, Weber, and Mejía-Ramos, 2012).

Many mathematicians and mathematics educators have claimed that students' lack of success comprehending the proofs that they read is due to the traditional manner in which proofs are presented. For instance, some scholars claim that the linear nature of traditional proofs prevents students from seeing the structure of the proof, or obscures the main ideas or methods of the proof, making the proof seem unmotivated or mysterious (e.g., Anderson, Boyle, & Yost, 1986; Davis & Hersh, 1981; Leron, 1983; Selden & Selden, 2003). Others claim the use of formal syntax and jargon in a proof can be intimidating to students and mathematicians alike (e.g., Davis & Hersh, 1981; Hersh, 1993; Thurston, 1994). Finally, many of the important insights of a proof may be implicit, and students may not have the background, orientation, or skill to unearth these implicit ideas (e.g., Gabel & Dreyfus, 2017; Konior, 1993; Weber, 2015; Weber & Alcock, 2005). To ameliorate these difficulties, some researchers have proposed alternative means for presenting proofs.

1. e-Proofs

e-Proofs are an electronic aid to proof comprehension developed by Lara Alcock. The idea of e-Proofs was to augment a traditional proof by focusing students' attention on key parts of a proof that were identified as important in the mathematics education research literature, such as how the assumptions and conclusions of the proof were related to the structure of the theorem statement. (This critical issue of the importance of the 'proof framework' was highlighted by Selden & Selden, 1995, 2003). To focus students' attention on important issues, the relevant parts of the proof were highlighted, the broader parts of the proof were grayed out, and a short audio file with an explanation of the key phenomenon in question was played to the reader. Students using the software could work at their own pace and replay the audio file as often as they wished (Alcock & Wilkinson, 2011).

In developing e-Proofs, Alcock capitalized on the mathematics education literature to find aspects of proof comprehension that mathematics educators deemed important but students found problematic; she also designed her software to conform to the guidelines of developing multimedia educational resources in the research literature. e-Proofs could potentially confer a number of benefits, including drawing students' attention to the most important aspects of the proof, making explicit reasoning that otherwise might remain implicit, and allowing students to revisit instructional explanations that are often ephemeral and not in students' notes when they study lecture proofs at a later time (a phenomenon documented by Fukawa-Connelly et al., 2017). Both students and course instructors who used e-Proofs evaluated them very positively (as reported in Alcock et al., 2015, and Roy, Inglis, & Alcock, 2017).

In a controlled experiment, Roy, Inglis, and Alcock (2017) used Mejía-Ramos et al.'s (2012) model for assessing proof comprehension to compare students' comprehension of a proof, with an experimental group who read the proof as an e-Proof and a control group who read the proof in a textbook. The results were negative: There was no significant difference in student comprehension after initially reading the proof, and the experimental group forgot more of the proof than the control group as measured by a delayed post-test given two weeks later. Based on the results from a subsequent eye-tracking study, Roy, Inglis, and Alcock (2017) illustrated how the audio files contained in e-Proofs allowed students to passively hear about connections in the proofs that they read without actively making those connections themselves. As active learning contributes to greater comprehension and retention, Roy, Inglis, and Alcock conjectured this lack of active processing might contribute to the poor retention of the e-Proofs that their students read. Alcock et al. (2015) viewed these negative results as a "humbling reminder that good pedagogical intentions do not always translate into effective interventions" and "a salutary lesson on the limitations of our own understandings of the process of learning from mathematical text" (p. 744).

2. Structured proofs

Leron (1983) proposed a novel way to present proofs in terms of levels, where each level is an independent module of the proof. The highest level (Level 1) gives a summary of the main ideas of the proof, but does not provide detail on how these main ideas will be carried out. The next level (Level 2) provides a summary of how each of the main ideas will be implemented. Successively lower levels fill in the details of the implementation of higher levels of the proof. A notable additional feature in some structured proofs is a section called an "elevator," which is located between levels and provides a rationale for why the proof is proceeding the way that it is. Leron (1983) argued that the structured presentation allows students to see the big ideas of the proof that are often masked by the traditional linear presentation and the "elevator" between levels can motivate the steps that follow, which a reader might otherwise find unmotivated or mysterious.

Leron's structured proofs are frequently cited in the mathematics education literature as having the potential to improve students' comprehension of proofs (e.g., Alibert & Thomas, 1990; CadwalladerOlsker, 2011; Hersh, 1993; Mamona-Downs & Downs, 2002; Selden & Selden, 2003, 2008), with Mamona-Downs and Downs (2002) claiming that structured proofs are influencing the way that proof is taught. Melis (1994) asserted that "Uri Leron shows how proofs are better comprehensible by structuring them into different levels" (p. 2).

Fuller et al. (2014) evaluated the efficacy of structured proofs, first with a qualitative study exploring what students perceived to be their strengths and weaknesses and then with a quantitative controlled experiment in which an experimental group read a structured proof and a control group read an analogous traditional proof of the same theorems and were tested on their understanding of the proof using Mejía-Ramos et al.'s (2012) proof comprehension model. In the qualitative interviews, students appreciated the summaries, but were perturbed that the structured proof "jumped around" too much. The justifications at lower levels would reference ideas at higher levels but the higher-level summaries were far removed from their lower level details in the proof's layouts, making it difficult for students to coordinate between levels. In the quantitative study, the participants in the experimental group who read the structured proofs were better able to identify an accurate summary of the proof (which was made explicit in Level 1 of a structured proof) but generally performed worse on other comprehension questions from Mejía-Ramos et al.'s (2012) model for comprehension, such as justifying steps in the proof, transferring the ideas of the proof to a new context, and applying the general method of the proof to a specific example (although the differences between the two groups was often not statistically significant).

3. Generic proofs

Generic proofs are the target of investigation in this paper. A generic proof is an argument that shows why a general claim is true for a specific example; however the reasoning applied to that example can be applied to any other relevant example as well (Leron & Zaslavsky, 2013; Rowland, 2001). Consequently the reader can infer that the general claim will hold for all examples. Rowland (2001) argued that a generic proof should have the following elements:

- If the theorem being proven is of the form, "for every n, n has property P", the generic proof should begin with a particular n_0.
- The particular example, n_0, should be neither too trivial nor too complicated,
- Steps of reasoning are not rooted in the mathematical objects, n_0, themselves, but in the properties that are shared by all objects in the scope of the universal quantifier, and
- The reasoning should be constructive.

Yopp and Ely (2016) offered a useful clarification and addition about what a generic proof is. In checking whether a generic proof is, in fact, generalizable, the reader needs to ensure the (possibly implicit) warrants for the calculations in the proof do not rely on any properties that are not shared by all objects in the scope of the universal quantifier (for more discussion of implicit warrants, see Weber & Alcock, 2005).

As a common example of a generic proof that is cited in the literature, consider the claim that every perfect square has an odd number of factors (see, e.g., Leron & Zaslavsky, 2013, p. 24). In this generic proof of this "for all" statement we choose a particular example of a perfect square, 36, that has enough factors to not be trivial, but not so many factors to be overly complicated. The proof proceeds by constructively lining up the factors of 36 in pairs: 1 x 36, 2 x 18, 3 x 12, 4 x 9, and 6 x 6. The critical step is noting that all factors line up as pairs of different numbers, except the last, which is 6 x 6. The number 36 can be represented as a product of the form exactly because 36 is a perfect square; a perfect square, by definition, will have a factor whose divisor is itself. Thus the warrant to show that 36 will have one more than an even number of factors only involves the squareness of 36, satisfying Yopp and Ely's (2016) criterion for a valid generic proof. Each of the steps involve our example of 36, but are not rooted in it, and can apply to any other perfect square as well.

Mathematics educators commonly cite generic proofs as a way of presenting proofs that may improve proof comprehension (e.g., Leron & Zaslavsky, 2013; Mason & Pimm, 1984; Rowland, 2001; Selden & Selden, 2008; Weber, Housman, & Porter, 2008), with Leron and Zaslavsky (2013) arguing that "generic proofs can help understanding by enabling students to engage with the main ideas of the complete proof in an intuitive and familiar context, temporarily suspending the formidable issues of full generality, formalism and symbolism" (p. 27). However, empirical evidence supporting the position that generic proofs actually improve comprehension is sparse.

Rowland (2001) surveyed his own students when generic proofs were used in his own classrooms. Most of his students agreed that generic proofs improved comprehension and were not lacking in rigor. However, we note here that Alcock's e-Proofs were also popular with students, but did not actually improve comprehension. As Alcock et al. (2015) emphasized, and Nardi and Knuth (2017) concurred, we should avoid conflating popularity with efficacy. In a small-scale quantitative study, Malek and Movshovitz-Hadar (2011) compared the effects of using generic proofs[1] and traditional proofs in a linear algebra class with ten students. Malek and Movshovitz-Hadar concluded that under the right conditions, generic proofs led to greater comprehension than traditional proofs. These conditions included the students being unlikely to construct a proof of the theorem on their own, the proof involving non-routine techniques, the ideas of the proof could be transferred to prove another theorem, and the proof being short enough to be used in an interview setting. In providing a first study empirically assessing the efficacy of generic proofs for improving proof comprehension, Malek and Movshovitz-Hadar made a valuable contribution to the literature. However, in the cases in which generic proofs were shown to be effective, only three or four students actually read the generic proofs. Consequently, the generality of their results is limited.

We conclude by arguing that there is reason to be skeptical about the efficacy of generic proofs to improve comprehension. First, as we documented, other alternative ways to present proofs did not demonstrate learning gains when they were tested in controlled experiments. Second, as Yopp and Ely's (2016) analysis shows, a deep understanding of a generic proof involves a careful analysis of the implicit warrants contained with a proof and research has shown that students often do not infer warrants when they are reading proofs (Alcock & Weber, 2005; Inglis & Alcock, 2012). Third, while examples certainly play a role in helping mathematicians understand proofs (see, for instance, Mejía-Ramos & Weber, 2014; Weber, 2008; and Weber & Mejía-Ramos, 2011), this does not imply that examples will be useful for students, who often attend to examples in less sophisticated ways than mathematicians do (e.g., Iannone et al., 2011).

Roy, Inglis, and Alcock (2017) and Nardi and Knuth (2017) argued that if mathematics educators are going to make instructional recommendations for how to improve students' comprehension of proofs, then as a community, we have an obligation to assess the efficacy of our recommendations. This paper represents an exploratory attempt to assess the efficacy of generic proofs.

THEORETICAL PERSPECTIVE

Based on interviews with mathematicians about how and why they read proofs (Weber & Mejía-Ramos, 2011) and why they present proofs to their students (Weber, 2012), Yang and Lin's (2008) proof comprehension model in secondary geometry, and the broader mathematics education research literature (e.g., deVilliers, 1990), Mejía-Ramos et al. (2012)

[1] Malek and Movshovitz-Hadar (2011) objected to the use of the term "generic proofs", preferring "transparent pseudo-proofs". We use generic proofs here for the sake of consistency with the other research literature cited in this article.

proposed a model for assessing a students' comprehension of a proof in advanced mathematics. This model included seven aspects of proof comprehension, and the corresponding types of questions one could use to assess them. Below, we discuss the four aspects of proof comprehension assessed in this study.

- Part of understanding a proof involves knowing how new steps in a proof are logical consequences from previous assertions. In some cases, the reasons for how a new step in a proof follows are not explicitly given, but need to be inferred by the reader (Weber & Alcock, 2005). A *justification question* asks a student to provide a mathematical reason why a new statement in a proof logically follows from previous statements.
- Part of understanding a proof is having a global grasp of the main ideas of the proof and how they fit together. A *summary question* presents the student with several possible summaries of the proof and asks the student to select the summary that best captures the main ideas of the proof.
- One of the main reasons that mathematicians read proofs (Mejía-Ramos & Weber, 2014; Rav, 1999; Weber & Mejía-Ramos, 2011) and present proofs to their students (Weber, 2012) is so students can use the methods in the proof to prove other statements. In a *transfer question*, students are asked to identify how the ideas used in the proof that they read could be used to prove another theorem.
- Understanding a proof that a "for all" claim works can involve relating the ideas of the proof to a specific object. In an *application to examples question*, students are asked how the general methods described in the proof can apply to a specific example.

INTERVIEW STUDY

For the first study, students read a generic proof and a traditional proof of the same theorem and provided their feedback on the format of the proof in individual interviews. Here we sought to replicate Rowland's (2001) claims that generic proofs were evaluated as popular and sufficiently rigorous with university mathematics students. Further, we sought to gain a better understanding of why university mathematics students thought generic proofs were valuable and how they might be limited.

1. The generic proof.

The generic proof that we used in this study is presented below:

Generic Proof

Theorem. Let S_n be the set of ordered finite sequences of positive integers that sum to n. (For example, $(2,1,2), (1,2,2)$, and (5) are distinct elements of S_5). Then the number of elements in S_n is 2^{n-1}. In other words, there are 2^{n-1} ways to express n as an ordered sum of positive integers.

Proof: (By induction).
Base case. If $n = 1$, the only element in S_1 is (1). Hence there are $1 = 2^0 = 2^{1-1}$ elements of S_1.
Inductive case. Assume that the theorem is true for $n = k$; that is, S_k has 2^{k-1} elements. We will illustrate how each element of S_k generates two new elements of S_{k+1} using the example $k = 3$.

For each element of S_3, we generate an element of S_4 either by increasing the last entry in the sequence by 1 or by appending a 1 at the end of the sequence.

To illustrate, $(2,1)$ is an element of S_3 and generates two elements of S_4. We add 1 to the last entry to get $(2,2)$ or we append a 1 at the end of the sequence to get $(2,1,1)$. Using this process,
(3) generates (4) and $(3,1)$
$(1,2)$ generates $(1,3)$ and $(1,2,1)$
$(2,1)$ generates $(2,2)$ and $(2,1,1)$
$(1,1,1)$ generates $(1,1,2)$ and $(1,1,1,1)$
Each generated element is in S_4.

Next, we show each element of S_4 was generated from one element of S_3. If the last entry of the sequence in S_4 is a 1, simply eliminate the 1 from the sequence. For instance, $(1,2,1)$ is in S_4 and was generated by $(1,2)$ in S_3.

Otherwise, if the last entry of the sequence in S_4 is greater than 1, decrease the last entry of the sequence by 1. For instance, $(2,2)$ is in S_4 and was generated by $(2,1)$ in S_3.

Since every element in S_3 generates two elements in S_4 and every element of S_4 is generated by exactly one element of S_3, there are twice as many elements in S_4 as there are in S_3.

The logic illustrated above can be applied to any S_k and S_{k+1}. Thus, S_{k+1} always has twice as many elements as S_k. By the inductive hypothesis, there are 2^{k-1} elements in S_k. Therefore, there are 2^k elements in S_{k+1}.

This particular theorem was chosen because it is accessible to mathematics undergraduate majors without having taken a course in number theory. We claim this proof meets the criteria listed by Rowland (2001), Malek and Movshovitz-Hadar (2011), and Yopp and Ely (2016) for a good generic proof. Following Rowland, we looked at a universal statement for all integers, we considered at $k = 3$, where the number of cases considered is not trivially small, but not so large as to be overwhelming, the steps would generalize to other cases in a straightforward manner, and the reasoning is constructive. Following Malek and Movshovitz-Hadar (2011), the proof employed a non-routine technique, the inductive doubling scheme can be used to prove other theorems, and the proof was short enough to be used in an interview setting. Following Yopp and Ely (2016), the warrants in the proof did not appeal to our choice of *k*.

2. Methods

1) Participants.

This study took place at a large state university in the United States. We solicited participation for the study by inviting fourth year mathematics majors and recent mathematics graduates who were enrolled in a masters program in mathematics education. (The university where this study occurred required all secondary mathematics teachers to complete the requirements for a standard mathematics degree). Ten students agreed to participate in this study and were paid for their participation. Five participants were male and five participants were female. We will use feminine pronouns to describe all ten participants.

2) Procedure.

Each participant met individually with a member of our research team for a video-recorded semi-structured interview. Participants were first given a brief description about what a generic proof was and how it should be read. These instructions are provided in the Appendix. Next, participants were handed the generic proof and asked to read the proof until they felt that they understood it. Participants were then asked to complete a six question open-ended proof comprehension test. (This phase of the interview was to build the comprehension tests that we would use in the quantitative study reported in the next section and will not be discussed further in this section). After reading the proof, participants were asked to report: (1) on a scale of 1 through 5, how well they felt they understood the proof, (2) on a scale of 1 through 5, how convincing they found the argument, (3) whether they thought the method of the argument could be generalized to any positive integer, and (4) what they thought about the format in which the proof was presented. Next, participants were presented with the traditional version of the proof below:

Given this traditional proof to compare, participants were asked the following questions:

Traditional Proof

Theorem. Let S_n be the set of ordered finite sequences of positive integers that sum to n. (For example, $(2,1,2), (1,2,2)$, and (5) are distinct elements of S_5). Then the number of elements in S_n is 2^{n-1}. In other words, there are 2^{n-1} ways to express n as an ordered sum of positive integers.

Proof: (By induction).
Base case. If $n = 1$, the only element in S_1 is (1). Hence there are $1 = 2^0 = 2^{1-1}$ elements of S_1.
Inductive case. Assume that the theorem is true for $n = k$; that is, S_k has 2^{k-1} elements. We show that each element of S_k generates two elements of S_{k+1}.

Let $(a_1, a_2, \ldots, a_m)$ be an element of S_k. By definition of S_k, we know that $a_1 + a_2 + \cdots + a_m = k$.

We generate two elements of S_{k+1} as follows. The first one is generated by increasing the last term of the sequence by 1 to yield $(a_1, a_2, \ldots, a_m + 1)$. This is an element of S_{k+1} since

$$a_1 + a_2 + \cdots + (a_m + 1) = (a_1 + a_2 + \cdots + a_m) + 1 = k + 1.$$

The other element is generated by appending a 1 at the end of the sequence to yield $(a_1, a_2, \ldots, a_m, 1)$. This is also an element of S_{k+1}.

Next, we show that each element of S_{k+1} was generated from one element of S_k. Let $(b_1, b_2, \ldots, b_m)$ be an element of S_{k+1}.

If $b_m = 1$, simply eliminate b_m from the sequence. In this case, $(b_1, b_2, \ldots, b_m)$ was generated by $(b_1, b_2, \ldots, b_{m-1}) \in S_k$.

Otherwise, if $b_m > 1$, decrease the last entry of the sequence by 1. In this case, $(b_1, b_2, \ldots, b_m)$ was generated by $(b_1, b_2, \ldots, b_m - 1) \in S_k$ [note that $b_m - 1$ is a positive integer because $b_m > 1$].

Since every element of S_{k+1} is generated by exactly one element of S_k, this ensures that the process above yields S_{k+1} and does not double count.

Because every element of S_k generates two elements in S_{k+1}, S_{k+1} has twice as many elements as S_k. By the inductive hypothesis, there are 2^{k-1} elements in S_k. Thus, there are 2^k elements in S_{k+1}.

- Here is a more traditional version of the same theorem. If you were in a class, would you prefer the traditional version or the proof you just read?
- Was there anything better about this new version of the proof that might have helped you understand the proof better?
- Was there anything about this new version of the proof that might have made it more difficult to understand?

Interviews ranged from 25 to 60 minutes.

3) Analysis.

Interviews were transcribed and were coded using an open coding scheme (Strauss & Corbin, 1990). In the first pass through the data, memos were made to highlight each instance in which a participant commented on a perceived attribute or deficiency of the generic proof or the format of generic proofs overall. This initial analysis yielded six preliminary codes: three codes describing positive attributes and three codes describing criticisms of generic proofs.

Students lauded generic proofs for reducing the complexity of the argument of the proof, eliminating technical jargon from the proof, and enhancing student understanding of the argument of the proofs. Students were critical of whether generic proofs were true proofs, the rigor of the arguments made in generic proofs, and the generality of the arguments made in the generic proof. Each of these categories are described in more detail in the following subsection.

Following the identification of these six categories, we returned to the data to verify that each appropriate instance of a participant's evaluation of the generic proof was coded and that instance was correctly categorized: checking that (1) the students' quotes exemplified the categorization and (2) the quotes were not better described by another category not present in our codes. We note that our analysis allowed for multiple codes to be attributed to a single instance of a participant discussing the generic proofs. For instance, when G1 was asked what they thought about the generic proof format, she said

> It helps to show specific cases as well and […] I do like this because it gives you a better idea and it doesn't just go straight into the theoretical and just the variable notation. […] I don't think you lose anything by adding on – 'here's another example' or here's another example of exactly why the proof works numerically.

This quote highlights G1's praise that the generic proof helps the reader to see the argument using *specific cases* to *giving a better idea* than if it had simply given the theoretical, or abstract, argument. This suggests that G1's believes that the generic proof format reduces the complexity of the argument of the proof. Meanwhile, G1 also highlighted the avoidance of *the variable notation* achieved by the generic proof format and praised the idea of using examples to show *exactly why the proof works numerically*. Thus, this instance of G1 discussing the format of the generic proof is also coded as eliminating jargon.

3. Results

When asked directly which proof format (generic or traditional) they preferred, five students indicated that they preferred the generic proof, two reported they lacked a strong preference, and three students preferred the traditional proof. The five students who preferred the generic proof suggested they did so because the generic proof was easier to understand and did not believe the generic format detracted from the proof. The two students who lacked a strong preference suggested that the generic proof format was easy to follow, but they also expressed concerns about the lack of rigor. The three students who preferred the traditional proof expressed concerns about the generalizability of the generic proof and found the generic format confusing. For instance, G10 found the generic proof too wordy, "I think sometimes when you are reading a proof, the more words gets to be a little bit more confusing so when it's clear and concise like the traditional one I felt it was a lot easier to understand." Meanwhile, despite G10's preference for the traditional proof over the generic proof, G10 also saw the potential of generic proofs suggesting that the example would be helpful to more novice students' understanding, explaining "if it was for someone who wasn't a math student I think the example was very helpful to give an example of exactly what they mean by adding a 1."

Like participant G10, most of the students (seven of the ten) expressed both positive and negative comments about the generic proof they read. *Table 1* summarizes participants' comments about the generic proofs, showing the students' comments were generally positive but some of the students were also critical of the generic proof format. The student feedback suggests that students value the reduced complexity and elimination of notation and jargon, as well as note how generic proofs can aid student comprehension. On the other hand, we also found

Table 1. Participants' comments on generic proofs

	Number of Participants	Specific Comments	Number of Participants
Students reporting positive comments on generic proofs	9	Reduce complexity	4
		Eliminate notation and jargon	7
		Potential to improve student understanding	8
Students who were critical of generic proofs	6	Generic proofs are not true proofs	1
		Lack rigor	2
		Lack generality	5

that students had some criticisms about the generic proof suggesting they were not true proofs, or that the proof's argument lacked rigor or generality.

1) Positive comments about generic proof format.

Nine of the ten participants expressed some positive comment about the generic proof. Three participants exclusively offered positive comments and indicated that there were no negative attributes related to generic proofs.

(1) Reducing complexity.

Four participants suggested that generic proofs reduced the complexity of the argument of the proof. When comparing the generic and traditional versions of the proof, G5 said:

> This one the traditional proof, looks scary and confusing. But this one you can actually […] see what they're doing. Rather than, I mean it is the same thing but it's easier to see what they're doing with this one rather than the traditional proof.

In comparing the two versions of the proof, G5 described the traditional version of the proof as scary and confusing. Moreover, the phrase "this one you can actually […] see what they're doing" suggests G5 may have reservations about her ability to understand the complexity of the argument presented in the traditional proof. However, the explicit computations in the generic proof allow G5 to make sense of the argument.

After similarly indicating a preference for the generic proof, G6 explained the generic argument helped them understand the proof better:

> Because it was more concrete and they were showing one particular case and since they explicitly said that you're adding 1 and appending 1, I was able to more quickly draw a conclusion since you're doing two possible things to each element.

In this quote G6 is attributing her better understanding of the proof to the transparency of the generic example. Seeing the generation of the sets explicitly reduced the complexity of the argument for G6, allowing them to more quickly understand the proof.

(2) Eliminating notation and jargon

Seven participants commented that generic proofs eliminated notation and jargon, emphasizing that they liked the implementation of numbers in the generic example. When comparing the two proofs, G8 quickly indicated that they preferred the generic proof because "all the variables like a_1 and a_2 it gets hard for me to keep in my head". Similarly while explaining why she believed the generic proof format helped their understanding, G2 explained:

> Seeing the example of it makes a lot more sense than just having dummy variables $b_1, b_2, \cdots, b_m$. And like seeing the example here now S_4 and this is S_3 and see how it changes, as opposed to this is just what you have to do.

Here, G2 explains that the generic example is easier to understand than general notation and subscripts. Moreover, she emphasized that the example helped them to see how it changes, whereas the traditional proof and notation *highlights what to do*.

In addition to participants like G8 and G2 praising the lack of notation, three participants also reported the notation and jargon used in the traditional format to be intimidating and confusing. For instance, G5 said:

> This [traditional proof] looks like the kinds of proofs that we had to write up that I'd always mess up with the variables or subscripts or whatever we were dealing with. I'd always lose some number or, just kind of get lost with all the different variables that we had to keep track of… This one looks scary and confusing. But this [generic proof] you can actually like, you can see what they're doing.

This excerpt suggests that G5 has experience with and, in turn, may be intimidated by proofs that require multiple variable representations or variables with subscripts. Such experiences and intimidation are not surprising, but highlight one reason why a student may prefer a generic proof over a traditional proof of the same claim.

(3) Improving student understanding.

Eight participants noted how generic proofs could improve student understanding. Seven of these participants mentioned that the generic example helped them to generally understand the proof. For instance, G9 claimed the generic proof was "probably easier to understand than any other proofs I've read." G1 argued that the generic example is useful in that "it [gave her] a general idea of what [she] should be working towards". This observation suggests that G1 sees not only the utility and the ease of seeing an example rather than a formal proof, but also that G1 sees the generic aspects of the example as intended.

In particularly, some of the participants mentioned that generic proofs could be used as an aid to understanding a traditional proof of the same theorem. For instance, G4 noted:

> If I was first learning it, I would want my professor to go over this [the generic proof] and then go to something like this [the traditional proof], more rigorous."

Similarly G2 said, "I understood this [traditional proof] more because I read this one [generic proof]." In these quotes, we see the participants' beliefs that the generic proof may be a useful precursor to a traditional proof. Some participants also noted that generic proofs may be particularly helpful for novice students, as seen in G10's comment above. Similarly, G1 believed the generic proof "would be easier to teach to people who aren't very, very adept at math" suggesting that the generic proof format could be especially valuable to students who were first becoming acquainted with proof.

2) Reservations about the generic proof

Six participants were also critical of the generic proof format in some way. However, in general, these participants noted the possible learning gains that can be achieved from reading generic proofs, despite their criticisms.

(1) Not a true proof.

When first presented the generic proof, G7 questioned, "This is a real proof?" Later, G7 continued in her disbelief adding, "I think that it's a good first step, but I don't think it's a real proof." Here, G7 did remark on the potential utility of a generic example noting that "one example is a good way to try to figure out how to write the proof from there", but reiterated that she did not believe the generic proof was a *real proof.* (Of course, it should be noted that some proponents of generic proofs, such as Malek and Movshovitz-Hadar, 2011, and Leron and Zaslavsky, 2013, agree with the students and have explicitly emphasized that generic proofs are not genuine proofs in important respects as well).

(2) Lacking rigor.

Two participants questioned the rigor of generic proofs. For instance, G4 expressed concern that a mathematician would not be convinced by the generic proof and later added that "this definitely helps but it's just not a rigorous way to prove it". Similarly, G3 found the generic proof to be "just a little bit unjustified". As such, we see both of these students doubting the rigor of a generic proof. Later in their interview, G3 expanded on this concern:

> I think being able to abstract things is also a pretty powerful skill. So if you saw all of your proofs where they did it like this, I think you probably would have less experience abstractly proving something. Because you'd have less examples to work from, so that could be a potential downside.

This quote shows that G3's dismissal of the generic proof based on the lack of rigor is a thoughtful one. Recognizing the importance of rigor and abstraction, G3 is concerned with the potential repercussions of working with generic proofs, if one is still expected to construct traditional proofs.

(3) Lacking generalizability

Five participants questioned whether the generic proofs were sufficiently general. Despite preferring the generic proof format, when first presented with the generic proof G9 said "it's only proven from S_3 to S_4. I guess I'm not hugely convinced […] that it would go the same pattern like from to S_5 to S_6." Participant G5 similarly expressed doubt that the generic argument would apply to other natural numbers, "Because examples aren't infinite and just because it worked for the one example I showed doesn't show it works for everything." In these quotes, we see students – even G5 and G9 who both indicated a preference for generic proofs and found the format illustrative – have some doubts about the generalizability of the generic example employed in the proof and question whether the argument will transfer to larger natural numbers. These students failed to see that the reasoning in the generic proof applied to not only 3, but also all natural numbers. We do note that while each student was provided with written instruction on generic proofs, it is possible that the participants did not view the examples as generic, but as specific examples.

In a similar vein, participant G7 expressed concern that the argument lacked the universal quality of proofs, "it's supposed to be, it doesn't matter what number it is, it should be for every single possible number and this only uses one example." Here, we see G7 doubting the generalizability of the proof, but also noting that the use of an arbitrary natural number k would avoid this question of generalizability. G7 further noted that using an example could be a productive avenue to "figure out how to write the proof" but, once again, pointed out that a generic proof is not a real proof.

4. Summary

In summary, our results are somewhat consistent with those from Rowland's (2001) survey, but not entirely so. Like the students in Rowland's survey, our participants collectively viewed the generic proofs were collectively more positively than negatively, and nearly all participants believed generic proofs such as this one could be a useful way to improve comprehension. Unlike the student's in Rowland's survey, there was some skepticism about the generality and rigor of the generic proofs and some students preferred the traditional format. There are many possible reasons for these differences, including Rowland's students having more extensive experience with the generic proof format. Of course, any differences that we observed could have been due entirely to the small sample size in our study, a point we start to address in the next study. Nonetheless, a main theme from both our interviews and Rowland's (2001) survey is that university mathematics students believe generic proofs have potential to improve student understanding.

QUANTITATIVE COMPARATIVE STUDY

In the discussion of the qualitative study, we highlighted that students provided positive feedback after reading generic proofs based on the reduction of complexity and the lack of notation. However, these data are limited for two reasons. First, the study included only ten participants from the same program in the same university. This small sample of students reduces the generalizability of our study. Second, as Nardi and Knuth (2017) warned and as Roy, Alcock, and Inglis (2017) demonstrated, students having positive opinions with regard to alternative proof format does not imply that the alternative format leads to improved comprehension.

As such, the goal of this larger quantitative study was to replicate the trends observed in the qualitative study and to seek evidence that reading generic proofs may indeed aid student comprehension. Specifically, using Mejía-Ramos et al.'s (2012) theoretical model for proof comprehension, we investigated the extent to which reading generic proofs may impact students' abilities to: see how a proof relates to specific examples, transfer the ideas of the proof to another theorem, summarize the proof, and see how particular statements are justified within the proof.

Before proceeding, we offer an important caveat framing our study. When authors such as Rowland (2001) and Leron and Zaslavsky (2013) claim that generic proofs can improve proof comprehension, they are not offering a particular pedagogical suggestion, but rather are providing a panoply of pedagogical suggestions. There are many different ways that generic proofs can be introduced to students and the decision of how generic proofs are introduced can very likely influence how effective generic proofs are at improving student comprehension. Consequently, even if students do not learn much from the generic proof in this study, it is quite possible that learning gains could have been realized if the generic proofs were introduced in a more effective manner, or students had more experience with the format, or a different generic proof was used, amongst other factors.

In this study, we chose to do a straightforward comparison in which some students read a generic proof, others read a traditional proof, and all were given a proof comprehension test; our analysis focused on whether there was a difference between the two groups on their performance on the test. We believe this is consistent with how some proponents of generic proofs envision them being used to improve proof comprehension. For instance, Malik and Movshovitz-Hadar (2011) used a similar design in their small-scale study illustrating the efficacy of generic proofs and Rowland (2011) suggested that a teacher might generic proofs can be used in lieu of traditional proofs.

1. Methods

1) Participants

The participants in this quantitative internet study were recruited from mathematics majors from top universities in the United States and Canada. Students received the recruitment email explaining the purpose of the experiment via the secretaries of their institution's mathematics department. Third and fourth year mathematics majors and minors[2] were invited to visit the experimental website to participate. The analysis reported here is based on the participation of 106 students.

2) Procedure.

Each student was randomly assigned to one of two groups: 54 participants were placed in the generic group and 52 in the traditional group. Participants reported their program (math major, math minor, or other) as well as their year of study (1st year undergraduate, 2nd year undergraduate, 3rd year undergraduate, 4th year undergraduate, postgraduate, or other). Next, each student was presented with instructions for the study. Participants in the generic group also received brief instruction on generic proofs, which is included in the Appendix.

Each student was presented with a single proof; the generic or traditional proof depending on their group assignment. After reading the proof, participants used a five-point Likert scale to report how well they understood the proof and to what extent they were persuaded the claim is true given the information presented in the proof. Students were also asked whether they found the result applied generally to any natural number k and whether they believed the proof was valid. Next, participants completed the six comprehension assessment questions in a randomized order. Finally, students were asked to report on a five-point Likert scale whether they liked the format in which the proof was presented and were given space to give any additional comments.

Each of the comprehension questions appeared on a new screen and participants were asked not to move back in their browser to review the proof or change their answers for previous questions. Participants were informed that if they did go back to a previous question, their responses would not be included in the analysis.

3) Comprehension assessment items

The generic proof and traditional proof were the same proofs presented in the qualitative study reported above. Regardless of their assignment to the generic or traditional proof, each participant completed the same six proof comprehension assessment questions in a randomized order. These six questions, developed based on Mejia-Ramos et al.'s (2012) model, assess students' abilities to apply the proof to examples, transfer the ideas of the proof, summarize the proof, and identify the justification for specific statements within the proof. The six items were piloted in the qualitative study and the multiple-choice responses, when offered, were based on the responses provided by the participants of the qualitative study.

4) The use of an internet study

This study was conducted online in order to maximize our sample size. The validity and reliability of this type of study have been extensively discussed in the research methods literature (e.g. Gosling et al., 2004; Reips, 2000). To ensure validity, we took multiple safeguards when conducting this study: 1) Each student reported whether they were seriously participating in the study, 2) the instrument recorded participant IP addresses, and 3) the instrument recorded the order in which pages of the

[2] In the United States and Canada, mathematics majors are undergraduates who are studying for a university degree in mathematics. Mathematics minors complete a subset of the courses that mathematics majors complete, but are typically required to take some proof-oriented courses.

study were viewed. Before analyzing the data, we first discarded data when there was evidence (repeated IP addresses) of a student participating in the experiment multiple times (24 instances). We then removed any participant that revisited pages while completing the study (7 instances), did not complete the survey (24 instances), or were not seriously participating (1 instance). This follows the methodology of Inglis and Mejia-Ramos (2009) to deal with the common threats to validity for this type of study. The cleaned data left 54 participants in the generic group and 52 participants in the traditional group. Data was subsequently analyzed for aggregate performance on the comprehension assessment questions for each group.

2. Results

1) Evaluation of the proofs.

When analyzing the students' reports on their beliefs of generality and validity, we found some noteworthy results. Both the traditional group and the generic group had the same views about how general their arguments were, with 17% of the traditional group and 18% of the generic group questioning whether the proofs were sufficiently general. However, the generic group (39%) was significantly more likely than the traditional group (17%) to challenge the validity of the proof (Fisher exact, p=.018). These results are summarized in *Table 2*.

Table 2. Participants' responses to the proof evaluation questions

Student Evaluation of the Proofs	Generic group (N=54)	Traditional group (N=52)
How well do you feel you understand this proof? (Scale 0-4)	3.481	3.462
Now say to what extent you are persuaded that the claim is true, given the information, and only the information that is contained in the proof. (Scale 0-4)	2.667	3.077
Do you think this proof applies to any? (Yes/No)	82% / 18%	83% / 17%
Do you think this proof is valid? (Yes/No)	61% / 39%	83% / 17%

2) Participant opinions on generic proof format

The students who read the generic proofs were asked whether they liked, disliked, or were neutral to the format. The results here are consistent with the results of our qualitative interview study presented earlier in the paper. Of the 54 participants in the generic group, 38 claimed to like the format, 12 disliked the format, 3 were neutral, and 1 did not respond. Comments varied in length and content, but some conveyed a particular enthusiasm for generic proofs: "Genius!" Of the 38 who were favorable towards the generic proofs, 20 left comments. Eight of those participants commented on why they liked the format reporting on the reduced complexity and the use of examples. For instance, one participant left the comment that the "example based proof makes the argument more transparent". Similarly, another participant noted that "The plethora of examples used to illustrate specific cases of the proof helped in understanding what the proof was trying to say."

Of the participants who indicated they had a favorable view of the generic proof, 12 participants also left comments suggesting they had reservations about the validity of a generic proof or were left wanting a more rigorous proof, as we illustrate below:

> Although I really like the idea of illustrating the formal proof with specific examples, it is no substitute for a formal proof by induction. Examples are excellent tools and should be used when writing/reading proofs, but specific examples do not prove a theorem. A formal inductive proof is needed.

This suggests that some participants may appreciate the value of generic proofs in respect to comprehension, yet still value deductive proofs for other purposes, such as validity and generality. (Again, this viewpoint is consistent with some mathematics education scholars who endorse generic proofs).

Of the 12 participants who responded unfavorably to the generic format, ten of these participants left comments suggesting that they were not convinced by the proof or found the generic example to be unnecessary. For instance, one participant noted "The format of this proof seems unconvincing, it seems like there is other possible elements we can get from

S_{k+1} from an element of S_k." This participant is questioning the generality of the argument, doubting that the generic example is representative of any arbitrary natural number. Meanwhile, another participant left the following comment, "I ended up thinking of each '3' as a 'k' anyway, so it took a little extra effort." This comment indicates that the participant in question took the generic example of S_3 and mentally translated this example to an arbitrary k. In such cases, reading the generic proof can be seen as adding to the cognitive demand of the reader.

3) Comprehension assessment results.

Participants' performance on the comprehension assessment questions is presented in *Table 3*. As *Table 3* documents, overall participants in the Generic Group did not do better on the post-test than participants in the Traditional Group. In fact, they did slightly worse, although this effect was not statistically reliable. Following Fuller et at. (2014), we split the test items into two groups: those directly pertaining the feature of the proof highlighted by the alternative proof format (in this case the examples questions) and those not directly pertaining to that format. We note two trends in the data. First, the Generic Group performed better on the two example questions. Second, the Generic Group performed worse on three of the four questions that did not pertain to examples. That the Generic Group performed worse on the transfer question is the opposite effect of what Malek and Movshovitz-Hadar (2011) found in their study.

Table 3. Participants' performance on the comprehension test.

Question	Generic Group (N=54)	Traditional Group (N=52)
Example 1	91%	83%
Example 2	94%	85%
Transfer	30%	42%
Summary	65%	62%
Justification 1	37%	63%
Justification 2	74%	87%
Test score	65%	70%

In *Figure 1*, we perform the analysis that highlights this difference. We categorize the assessment items by example items and non-example items. In this analysis, the Generic Group performed statistically reliably worse than the Traditional Group on non-example item with $t(104) = 2.33$ and $p < .05$ ($p = .0216$). The data also suggests that the Generic Group outperformed the Traditional Group on example items, although not statistically reliably so. These findings suggest that reading a generic proof may be helpful in learning how to apply ideas to specific examples, but harmful in considering more abstract ideas about the proof.

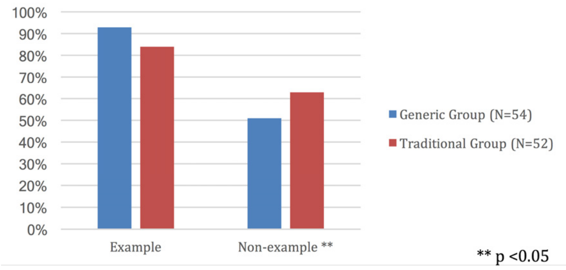

Figure 1. Participant performance on example and non-example items

3. Summary

The results about students' perceptions of generic proofs are again consistent with Rowland (2001) as well as our qualitative study; the illustrate that students generally have a positive opinion of generic proofs, although even some students who view these proofs favorably value some features of a deductive proof that generic proof lacks. However, we failed to find evidence that students understood the generic proof better than the traditional proof.

DISCUSSION

We begin by summarizing the two main results from this paper. First, consistent with Rowland (2001), we have found that most students believe generic proofs can be valuable for improving their understanding of the proofs that they read. This result was found both in our qualitative interviews and in a larger scale study in which 54 students gave their opinion on a generic proof that they read. Second, in our quantitative comparison study, we found no evidence that students actually learned more from reading the generic proof than an analogous proof presented in a traditional format.

To reiterate an important caveat we provided in Section 5, our results do not imply that the use of generic proofs cannot improve proof comprehension. It is certainly possible that with different types of pedagogy, generic proofs can improve proof comprehension. What our study does highlight, however, is that mathematics educators' enthusiasm for generic proofs far outpaces empirical evidence that generic proofs can improve proof comprehension. We call for more empirical research demonstrating that generic proofs actually deliver on the promise that their proponents cite.

Like Alcock et al.'s (2015) work with e-proofs, our study offers another example in which we could demonstrate that students found an alternative proof presentation to be popular, but we failed to find evidence that this improved their learning. We agree with Nardi and Knuth (2017) that the popularity of a proof is not a good measure of its efficacy in improving proof comprehension. One possible account is that students prefer working in representation systems and engaging in types of reasoning in which they have familiarity. Generic proofs are therefore preferable to traditional proofs to novice students, as the former are based on the types of concrete computations that align with their previous experience. In particular, the justification for claims within a proof is often given in the form of a computation that is easy for the student to follow. Traditional proofs are written abstractly. This abstract representation and reasoning has the virtue that it requires the reader to focus on the generality of the claim being made. In other words, we suggest it might be possible that students find generic proofs preferable to traditional proofs exactly because it allows the students to avoid the difficult issue of generalization. Of course, proponents of generic proofs can counter that students should be reading the generic proofs with an eye toward generalization; in other words, students should treat the example used in the proof generically. We know mathematicians do this, but we also know that some students engage with examples in less sophisticated ways than we would like (e.g., Iannone et al., 2010). The preceding account is admittedly very speculative, but aligns with Roy, Inglis, and Alcock's (2017) finding that alternative proof formats may be popular with students because it allows the students to avoid the difficult cognitive processing that is needed to make sense of the proof.

Our final comments concern the broader research on proof presentation. Like most mathematics educators who investigate proof, we believe that students' failure to understand the proofs that they read is a significant problem. However, we are skeptical of whether presenting proofs in a different manner will be productive for improving their comprehension of the proofs that they read. There have been a number of promising alternative ways to present proofs to students, such as e-proofs (Alcock & Wilkerson, 2011), structured proofs (Leron, 1983), and generic proofs (Rowland, 2001). However, when the efficacy of these alternative proof presentations has been tested empirically in studies with moderate or large sample sizes, the researchers were not able to document learning gains (see Roy, Inglis, & Alcock, 2017; Fuller et al., 2014; and the studies in this paper respectively). Hodds, Alcock, and Inglis (2014) summarized the situation as follows:

> All three of these approaches involve instructor provision of different or extra explanations: A structured proof involves restructuring the proof text, a generic proof involves changing its content, and an e-Proof involves augmenting the proof with annotations and commentary. Changing the presentation in such ways requires substantial instructor effort, and the underwhelming empirical results suggest that this may not be effort well spent (p. 67).

What Hodds, Alcock, and Inglis (2014) proposed as an alternative was helping students learn to read proofs more effectively. We believe that they are correct. Further, in contrast to presenting proofs in different formats, Hodds, Alcock, and Inglis (2014) have shown that their "self-explanation training" actually led to large and sustainable learning gains for students using Mejía-Ramos et al.'s (2012) measure for proof comprehension. Given the lack of success to date in demonstrating the efficacy in alternative modes of proof presentation (and the costs of presenting proofs in this way) and the demonstrable success that Hodds, Alcock, and Inglis have obtained, we think their approach is the more promising one to improving students' comprehension of proofs.

References

Alcock, L., Hodds, M., Roy, S., & Inglis, M. (2015). Investigating and improving undergraduate proof comprehension. *Notices of the American Mathematical Society, 62*(7), 742-752.

Alcock, L. & Weber, K. (2005). Proof validation in real analysis: Inferring and evaluating warrants. *Journal of Mathematical Behavior, 24*(2), 125-134.

Alcock, L. & Wilkinson, H. (2011). e-Proofs: Design of a resource to support proof comprehension in mathematics. *Educational Designer,* 1(4). Retrieved from: http://www.educationaldesigner.org/ed/volume1/issue4/article14/

Alibert, D. & Thomas, M. (1991). Research on mathematical proof. In D. Tall (Ed.), *Advanced Mathematical Thinking* (pp. 215-230). Kluwer: The Netherlands.

Anderson, J., Boyle, C., & Yost, G. (1986). The geometry tutor. *Journal of Mathematical Behavior, 5*, 5-20.

CadwalladerOlsker, T. (2011). What do we mean by mathematical proof. *Journal of Humanistic Mathematics, 1*(1), 33-60.

Conradie J., & Frith, J. (2000). Comprehension tests in Mathematics. *Educational Studies in Mathematics, 42*(3), 225-235.

Cowen, C. C. (1991). Teaching and testing mathematics reading. *The American Mathematical Monthly, 98*(1), 50-53.

Davis, P. J. and Hersh, R. (1981*). The mathematical experience*. New York: Viking Penguin Inc.

de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras, 24*, 17-24.

Fukawa-Connelly, T., Weber, K., & Mejía-Ramos, J. P. (2017). Informal content and student note-taking in advanced mathematics classes. *Journal for Research in Mathematics Education, 48*(5), 567-579.

Fuller, E., Weber, K., Mejia-Ramos, J. P., Rhoads, K., & Samkoff, A. (2014). Comprehending structured proofs. *International Journal of Studies in Mathematics Education, 7*(1). Downloadable from: https://seer.pgsskroton.com/index.php/jieem/article/view/84. Last downloaded March 2, 2020.

Gabel, M., & Dreyfus, T. (2017). Affecting the flow of a proof by creating presence—a case study in Number Theory. *Educational studies in mathematics, 96*(2), 187-205.

Gosling, S.D., Vazire, S., Srivastava, S., & John, O.P. (2004). Should we trust web-based studies? A comparative analysis of six studies about internet questionnaires. *American Psychologist, 59*, 93-104.

Hanna, G. (1990). Some pedagogical aspects of proof. Interchange, 21(1), 6-13.

Hanna, G. (2018). Reflections on proof as explanation. In A. Stylianides & G. Harel (Eds). *Advances in mathematics education research on proof and proving* (pp. 3-18). Springer, Cham.

Harel, G. (1998). Two dual assertions: The first on learning and the second on teaching (or vice-versa). *American Mathematical Monthly, 105*, 497-507.

Hersh, R. (1993). Proving is convincing and explaining. *Educational Studies in Mathematics, 24*(4), 389-399.

Hodds, M., Alcock, L., & Inglis, M. (2014). Self-explanation training improves proof comprehension. *Journal for Research in Mathematics Education, 45*, 62-101.

Iannone, P., Inglis, M., Mejía-Ramos, J. P., Simpson, A., & Weber, K. (2011). Does generating examples aid proof production?. *Educational Studies in Mathematics, 77*(1), 1-14.

Inglis, M. & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education, 43*(4), 358-391.

Inglis, M. & Mejia-Ramos, J.P. (2009). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction, 27*, 25-50.

Kline, M. (1973). *Why Johnny can't add: The failure of New Mathematics*. New York, NY: St. Martin's Press.

Ko, Y. & Knuth, E. (2012). Validating proofs and counterexamples across domains: Practices of importance for mathematics majors. *Journal of Mathematical Behavior, 32,* 20-35.

Konior, J. (1993). Research into the construction of mathematical texts. *Educational Studies in Mathematics, 24*(3), 251-256.

Krupnik, V., Fukawa-Connelly, T., & Weber, K. (2018). Students' epistemological frames and their interpretation of lectures in advanced mathematics. *The Journal of Mathematical Behavior, 49*, 174-183.

Knuth, E. J. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for research in mathematics education, 33*, 379-405.

Lai, Y., Weber, K., & Mejía-Ramos, J. P. (2012). Mathematicians' perspectives on features of a good pedagogical proof. *Cognition and Instruction, 30*(2), 146-169.

Leron, U. (1983). Structuring mathematical proofs. *American Mathematical Monthly, 90*(3),174-184.

Leron, U., & Zaslavsky, O. (2013). Generic proving: Reflections on scope and method. *For the Learning of Mathematics, 33*(3), 24-30.

Lew, K., Fukawa-Connelly, T. P., Mejia-Ramos, J. P., & Weber, K. (2016). Lectures in advanced mathematics: Why students might not understand what the mathematics professor is trying to convey. *Journal for Research in Mathematics Education, 47*(2), 162-198.

Malek, A. & Movshovitz-Hadar, N. (2011). The effect of using Transparent Pseudo-Proofs in linear algebra. *Research in Mathematics Education, 13*, 33-58.

Mamona-Downs, J., & Downs, M. (2002). Advanced mathematical thinking with a special reference to mathematical structure. In L. English (Ed.), *Handbook of international research in mathematics education* (pp. 165-196). Lawrence Erlbaum Associates: Mahwah, NJ.

Mason, J., & Pimm, D. (1984). Generic examples: Seeing the general in the particular. *Educational Studies in Mathematics, 15*(3), 277-289.

Mejía-Ramos, J. P., & Inglis, M. (2009). Argumentative and proving activities in mathematics education research. In F.-L. Lin, F.-J. Hsieh, G. Hanna, & M. de Villiers (Eds.), *Proceedings of the ICMI Study 19 conference: Proof and Proving in Mathematics Education* (Vol. 2, pp. 88-93). Taipei, Taiwan.

Mejía-Ramos, J.P., Fuller E. ,Weber, K., Rhoads, K., & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics, 79*, 3-18.

Mejia-Ramos, J. P., & Weber, K. (2014). Why and how mathematicians read proofs: Further evidence from a survey study. *Educational Studies in Mathematics, 85*(2), 161-173.

Melis, E. (1994). How mathematicians prove theorems. In *Proceedings of the 16th Annual Cognitive Science Conference*. Atlanta, GA.

Mills, M. (2011). Mathematicians' pedagogical thoughts and practices in proof presentation. In S. Brown, S. Larsen, K. Marrongelle, & M. Oehrtman (Eds.), *Proceedings of the 14th Annual Conference on Research in Undergraduate Mathematics Education, 2*, 283-297. Portland, OR.

Nardi, E., & Knuth, E. (2017). Changing classroom culture, curricula, and instruction for proof and proving: how amenable to scaling up, practicable for curricular integration, and capable of producing long-lasting effects are current interventions?. *Educational Studies in Mathematics, 96*(2), 267-274.

Porteous, K. (1986). Children's appreciation of the significance of proof. *Proceedings of the Tenth International Conference of the Psychology of Mathematics Education* (pp. 392-397). London, England.

Rav, Y. (2007). A critique of a formalist-mechanist version of the justification of arguments in mathematicians' proof practices. *Philosophia Mathematica, 15*(3), 291-320.

Reips, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (pp. 89–117). San Diego: Academic Press.

Rowland, T. (2001). Generic proofs in number theory. In S. Campbell and R. Zazkis (Eds.), *Learning and teaching number theory: Research in cognition and instruction.* (pp. 157- 184). Westport, CT: Ablex Publishing.

Roy, S., Inglis, M., & Alcock, L. (2017). Multimedia resources designed to support learning from written proofs: An eye-movement study. *Educational Studies in Mathematics, 96*(2), 249-266.

Selden A. & Selden J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for research in mathematics education, 34*(1) 4-36.

Selden, J., & Selden, A. (1995). Unpacking the logic of mathematical statements. *Educational Studies in Mathematics, 29*(2), 123-151.

Selden, A. & Selden, J. (2008). Overcoming students' difficulties with learning to understand and construct proofs. In M. P. Carlson and C. Rasmussen (Eds.) *Making the Connection: Research and Teaching in Undergraduate Mathematics Education*. (pp. 95-110). MAA Notes: Washington, D.C.

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. London: SAGE.

Thurston, W.P. (1994). On proof and progress in mathematics, Bulletin of the American Mathematical Society, 30, 161-177.

Weber, K. (2004). Traditional instruction in advanced mathematics courses: A case study of one professor's lectures and proofs in an introductory real analysis course. *Journal of Mathematical Behavior, 23*(2), 115-133.

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education*, 431-459.

Weber, K. (2010). Mathematics majors' perceptions of conviction, validity, and proof. *Mathematical Thinking and Learning, 12,* 306-336.

Weber, K. (2012). Mathematicians' perspectives on their pedagogical practice with respect to proof. *International Journal of Mathematical Education in Science and Technology, 43*(4), 463-482.

Weber, K. (2015). Effective proof reading strategies for comprehending mathematical proofs. *International Journal of Research in Undergraduate Mathematics Education, 1*(3), 289-314.

Weber, K., & Alcock, L. (2005). Using warranted implications to understand and validate proofs. *For the Learning of Mathematics, 25*(1), 34-51.

Weber, K., & Mejia-Ramos, J. P. (2011). Why and how mathematicians read proofs: An exploratory study. *Educational Studies in Mathematics, 76(*3), 329-344.

Weber, K., Porter, M., & Housman, D. (2008). Worked examples and concept example usage in understanding mathematical concepts and proofs. In M. Carlson & C. Rasmussen (Eds). *Making the connection: Research and teaching in undergraduate mathematics education*, 245-252. Mathematical Association of America: Washington, DC.

Yang, K.-L., & Lin, F.-L. (2008). A model of reading comprehension of geometry proof. *Educational Studies in Mathematics, 67*, 59–76.

Yopp, D. A., & Ely, R. (2016). When does an argument use a generic example?. *Educational Studies in Mathematics, 91*(1), 37-53.

Appendix: Instructions on generic proof and questions (final version from quantitative study)

Instructions on generic proof

In college math classes, theorems are traditionally stated and then proven in general, abstract terms. Some mathematicians have suggested another way of presenting proofs for theorems — by illustrating the proof with one or more specific examples. For instance, to justify a fact about the first n odd natural numbers, one might illustrate how the proof works with the first 3 odd natural numbers. Though the reasoning might only be shown for a certain set of examples, it would work in a similar way for any set of examples. One goal of this study is to see how students read proofs presented in this manner.

Questions

Summary

Which of the options below best expresses the main ideas used in the proof?

A. **Each element of S_k generates two elements of S_{k+1}, because we can add 1 to the last entry in the sequence or append a 1 at the end of the sequence. This generates all elements of S_{k+1}.**

B. *Base case.* If $n = 1$, the only element in S_1 is (1). Hence there are $1 = 2^0 = 2^{1-1}$ elements of S_1. *Inductive case.* Suppose S_k has 2^{k-1} elements. Then S_{k+1} has 2^k elements.

C. S_{k+1} is obtained from sequences of S_k either by removing the last entry in the sequence or by subtracting a 1 from the last entry. Hence, every element of S_{k+1} generates two elements of S_k.

D. I don't know.

Justification 1

How do we know that the method used in this proof did not produce the same element of S_{k+1} twice?

A. Each element of S_k produces exactly two elements of S_{k+1}.

B. **We can tell which unique element of S_k produced a particular element of S_{k+1}.**

C. Adding 1 to the last entry of a sequence cannot yield the same result as appending a 1 at the end of the sequence.

D. I don't know.

Justification 2

If (a, b, c, d) is an element of S_k, how do we know that $(a, b, c, d, 1)$ is an element of S_{k+1}?

Example 1
$(2, 3, 1)$ is an element of S_6. Using the ideas of the proof, what two elements of S_7 can be formed from $(2, 3, 1)$?

Example 2
$(3, 1, 2)$ is an element of S_6. Using the ideas of the proof, what element of S_5 was used to form $(3, 1, 2)$?

Transfer
Consider the following theorem: There are 2^n subsets of $\{1, 2, \ldots, n\}$. Two students came up with the following approaches to prove this theorem. Which of their approaches is most consistent with the ideas of the proof?

A. **Take a subset $\{b_1, b_2, \ldots, b_m\}$ of $\{1, 2, \ldots, k\}$. This forms two subsets of $\{1, 2, \ldots, k+1\}$ as follows. Take the original subset $\{b_1, b_2, \ldots, b_m\}$ or include $k+1$ in this subset to yield $\{b_1, b_2, \ldots, b_m, k+1\}$.**

B. Take a subset $\{b_1, b_2, \ldots, b_m\}$ of $\{1, 2, \ldots, k\}$. This forms two subsets of $\{1, 2, \ldots, k+1\}$ as follows. Add 1 to the last element of $\{b_1, b_2, \ldots, b_m\}$ to obtain $\{b_1, b_2, \ldots, b_m + 1\}$ or append a 1 at the end of the subset to obtain $\{b_1, b_2, \ldots, b_m, 1\}$.

C. I don't know.