Undergraduates Proof Comprehension: A Comparative Study of Three Forms of Proof Presentation

Somali Roy, Lara Alcock, Matthew Inglis

Mathematics Education Centre,

Loughborough University,

UK

Abstract

In this study we investigated the effectiveness of the three different forms of proof presentation on mathematics undergraduates' proof comprehension: (i) written textbook proof, (ii) proof presented by a lecturer in a standard undergraduate lecture, (iii) a computer based e-Proof. A comprehension test was designed to examine different facets of proof comprehension as described by Yang and Lin (2008). The data suggested that the live lecture had a significantly greater impact on undergraduates' proof comprehension than the other forms of proof presentation and the e-Proof was least effective.

Introduction

The mathematics education literature suggested that students of all levels, school to university, have serious problems with proof. Especially at university level, proofs are common in mathematics courses, but students often face: difficulties in using and understanding mathematical language (Moore, 1994; Harel & Sowder, 1998); difficulties in using definitions to derive a proof (Moore, 1994); difficulties in unpacking the meaning of a mathematical statement or propositions

(J. Selden & Selden, 1995); difficulties in inferring warrants to check whether a line in a proof is valid or not (Weber & Alcock, 2005) difficulties in checking whether an argument is valid proof or not (Weber, 2009; A. Selden & Selden, 2003) and so on.

The two main setting in which students can comprehend proof are in mathematics classrooms or by independent study. When presenting proofs, lecturers typically provide verbal explanations which go beyond the written proof, both about the validity of each line and about the overall structure of the argument. But in real time, students might not be able to focus their attention as the lecturer intends, so they might not see the explicated links and structures. As a result, a large part of a student's task is to understand these proofs through independent study. However, there is often minimal instruction about how to do this effectively, and research indicates that students find independent proof comprehension a difficult task. Alcock (2009) attempted to address the problem of supporting students in proof comprehension by constructing e-Proofs. An e-Proof is a sequence of computer-based slides, each with accompanying audio commentary that can be replayed as many times as the student wishes. Each screen shows the whole proof, with large parts "greyed out" to focus attention on relevant sections. Arrows and boxes are superimposed to indicate the relationships between the lines, and the commentary explains the warrants that must be inferred and why we introduce objects or make certain inferences.

This study was designed to compare students' proof comprehension following three different presentations of a standard proof from real analysis. Those three different forms were:

- a written textbook proof (with no further explanation),
- a proof presented by a lecturer in a standard undergraduate lecture and
- a computer based e-Proof.

An unseen proof was selected for the study and a comprehension test was designed based upon it. The aim of the study was to investigate the effect of those three forms of proof presentation on undergraduates' proof comprehension.

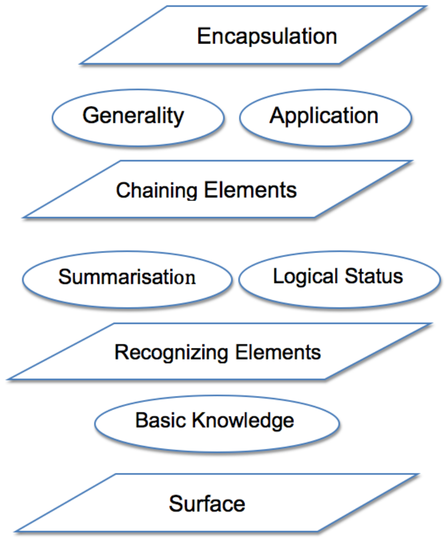

Figure 1. The hypothetical model with four levels and five facets of reading comprehension of proof, proposed by Yang and Lin (2008, p. 71)

Reading Proof Comprehension

There is limited research on students' reading comprehension of proof. But many mathematics educators have suggested various aspects that are necessary for good proof comprehension. Selden and Selden (1995) argued that it is important for students to comprehend the statement of the theorem before trying to comprehend a whole proof. Proofs are often known as "logical arguments", and this indicates that logic has a significant role in a proof. Weber and Alcock (2005) argued that, "if warrants were not considered when reading the proof, one's understanding of proof would be limited" (p. 38).

Yang and Lin (2008) proposed a hypothetical model of reading comprehension of geometry proof in which they identified four levels and five facets of proof comprehension (see Fig 1). The four levels of proof comprehension are: surface level, recognizing elements, recognizing chaining and encapsulation and five facets are basic knowledge, logical status, summary, generality and application (see Table 1).

In this study our interest was to investigate how far undergraduates comprehend an unseen mathematical proof. Though the notion of understanding could be different for different people: as Hersh (1997) stated, "we recognize understanding, though we can't say precisely what it is" (p.

Table 1: Facets of proof comprehension

Facet	Descriptions
Basic Knowledge	Recognizing the meaning of a symbols or notations and recognizing/ explaining the meaning of a property used in a proof.
Logical status	Identify the logical order of statements and recognizing which properties are applied in a proof.
Summarization	Identifying critical procedures, premises or conclusions and identifying critical ideas in a proof.
Generality	Judging the correctness and identifying what is validated by a proof.
Application	Knowing how to apply the premises in other situations.

163). Conradie and Frith (2000) argued that even if students successfully reproduce a proof in exam it dose not necessarily reflect understanding, because often poor students try to memorize the entire proof word by word, with possibly no understanding at all. To remove this problem, they proposed a "comprehension test" where students are presented a proof and then required to answer questions on specific features of the proof. These authors argued that this method could provide the opportunity to deeply explore the nature of the students understanding.

The Comparative Study

We designed a comparative study to examine undergraduates' proof comprehension by presenting three different forms of proof-presentation to them. To achieve the goal a comprehension test was designed to reflect different facets of students proof comprehension addressed by Yang and Lin (2008).

Participants

Participants in the main study were undergraduate students of Loughborough University. All were enrolled on a module that covered material in real analysis and were in either the second or third year of a mathematics degree (either single or joint honours). Participants were informed that the comprehension test was only for research purposes and would not affect their final module marks.

Material and Methodology

The proof of Cauchy's Generalised Mean Value Theorem (GMVT) was selected, because when the study took place participants had not yet seen the theorem. The next step was to use a method that can reflect undergraduates' proof comprehension on an unseen proof. Conradie and Frith (2000) argued that a comprehension test is an alternative way of testing student's understanding of a theory. Therefore, the comprehension test was employed as a methodological tool for this study.

Comprehension Test Design. A comprehension test was designed on Cauchy's GMVT to examine students' proof comprehension abilities The test (see below) initially contained nine questions. To answer each question correctly would require different facets or levels of proof comprehension. For example, to answer Question 8 students would need basic knowledge of mathematical manipulation and for Question 4 they need to know how to differentiate a function. Questions 2 and 5 reflected students' understanding of mathematical logic. On the other hand, understanding the critical ideas that are used in the proof, are essential for answering Questions 1 and 3 (facet of summarization). Question 6 and Question 7 were reflected to the facet of generality and the facet of application respectively.

Comprehension Test Questions.

1. (a) How does line (1)[1] contradict the theorem premise?

   (b) Why do we need the conclusion from the contradiction?

2. Which of the following properties are used in the proof?

   If functions p and q are continuous on $[a,b]$ and differentiable on (a,b) then

   (a) $p+q$ is continuous on $[a,b]$ and differentiable on (a,b).

   (b) pq is continuous on $[a,b]$ and differentiable on (a,b).

   (c) mp is continuous on $[a,b]$ and differentiable on (a,b), where m is a constant.

   (d) p/q is continuous on $[a,b]$ and differentiable on (a,b).

[1] Here line (1) indicates the first line of the short version of Cauchy's GMVT that students received at the beginning of the comprehension test.

3. Write a short paragraph to explain why we set

$$h(x) = f(x) - \frac{f(b)-f(a)}{g(b)-g(a)}g(x).$$

4. With h defined as in the proof, what is $h'(x)$?
5. Where does the proof show that h satisfies the conditions for Rolle's Theorem?
6. (a) Imagine that the premises of Cauchy's GMVT read:

   Suppose that f and g are continuous on $[a,b]$ and differentiable on (a,b) and that $\forall x \in (a,b), f'(x) \neq 0$.

   What would the conclusion say?

   (b) If you were proving the new version (i.e. (a)) directly, how would you define $h(x)$?
7. Find an interval where $f(x) = 1+x^2$ and $g(x) = x^2 - 1$ satisfy the premises of Cauchy's Generalised MVT.
8. Show that
$$f(a) - \frac{f(b)-f(a)}{g(b)-g(a)}g(a) = \frac{f(a)g(b)-g(a)f(b)}{g(b)-g(a)}$$
9. Write down in your own words a short summary of the proof.

Procedure of The Main Study

The main study took place during the sixth week of the module. Participants were divided randomly into three groups, called the reading group, the lecture group and the e-Proof group. The participants in the reading group received the written textbook proof of Cauchy's GMVT, the e-Proof group sat in a computer lab to read the e-Proof of Cauchy's GMVT and lecture group attained the live lecture on the theorem, which was given by a lecturer from the school of Mathematics of Loughborough University. Both the reading group and the e-proof group had 15 minutes to read the proof. After 15 minutes the reading group were asked to return the written textbook proof and for e-proof group their computers were logged off automatically. The lecture group had 21 minutes to listen to the lecturer, as the lecturer overran by six minutes. In the next stage, each of the participants received the comprehension test question paper which contained nine questions. In addition, they received a short version of Cauchy's GMVT and an additional information sheet which contained definitions of continuity and differentiability and statements of a few relevant theorems, for example Rolle's theorem (see appendix). The additional information was provided to the participants, so that they could integrate the information while comprehending the proof, because the focus of the study was to explore to what extent students comprehend an unfamiliar proof, but not to investigate their existing subject knowledge.

To find out the longer term effects of different representations on students' proof comprehension, a delayed post test was conducted after two weeks, during a lecture, where all participants received the same question paper along with the short-version of the Cauchy's GMVT and the additional information sheet (as they received in post-test). This time they had only 20 minutes to answer the questions. There were 80 participants (reading group= 28, lecture group=31, e-proof group=21) who took part in both the post-test and the delayed post-test.

All the answer sheets for the post-test and the delayed post-test were checked and marked (except question 9 — "Write down in your own words a short summary of the proof", which is not discussed here), yielding a score out of 18.

Results

Participants' scores in both tests were subjected to an analysis of variance (ANOVA) with one within-participant factor, time (post-test and delayed post-test), and one between-participant factor (group) which had three levels (reading group, lecture group and e-proof group). Overall, in the post-test and the delayed post-test the lecture group performed significantly better than the reading group and the e-proof group. The mean average (across both tests) for the lecture group was 12.6, whereas the reading group and the e-Proof group scored 10 and 9.8 respectively. The mean scores in the post-test and the delayed post-test for each group are shown in the Table 2. Comparing the decrement in overall score in both tests, it was found that for the e-Proof group the decrement was 16% in delayed post-test, whereas the decrement was 4% for the lecture group and 7.5% for the reading group decrement (see Figure 2).

The analysis revealed that the main time effect was highly significant, $F(1,77) = 33.124, p <$

Table 2: Mean scores of each group in both tests (a score out of 18)

Group	Post-test	Delayed Post-test	Total
Reading group	10.7	9.4	10
Lecture group	13	12.2	12.6
e-Proof group	11.2	8.4	9.8

0.001 with mean total score of 11.7 for the post-test compared to 10.2 for the delayed post-test. The interaction effect for time $\times$ group was also significant, $F(2,77) = 33.124, p < 0.001$, indicating differential drop offs between the post-test and the delayed-post test.

The main effect of group was also highly significant, $F(2,77) = 7.482, p = 0.001$. In addition, a post hoc test was carried out to further explore the difference between the groups. This test confirmed that there was a significant difference between the reading group and the lecture group $(p = 0.004)$ and also between the lecture and the e-Proof group $(p = 0.006)$.

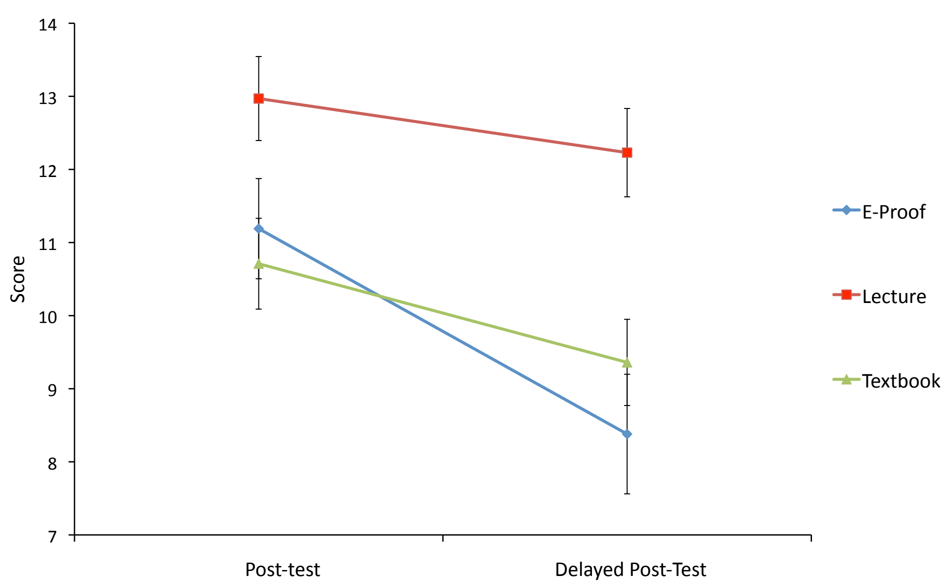

Figure 2. Groups' performances in post-test and delayed post-test

Moreover, 2×2 ANOVA was conducted to look at the interaction between the reading group and the e-Proof group (see the bottom section of Fig 2) and the analysis showed that the time × group interaction was significant, ($F(1,47) = 28.213, p < 0.001$). This result indicates that the e-Proof group had significantly worse retention than the reading group in the delayed post-test.

Discussion

Comparing lecture with other forms of proof-presentation

On analysing data it became clear that the lecture group performed significantly better than other two groups, but no significant difference was found between the reading group and the e-Proof group. Participants of the lecture group had 21 minutes to listen to the lecture, whereas the other two groups had only 15 minutes reading time. The extra six minutes time could be a reason behind the better performances of the lecture group. There could have been several other reasons, for example, students were very familiar with the live lecture situation whereas during lecture-time, sitting in a lab-class and learning through an e-Proof was a different experience for the students. In the lecture, the explanations of the proof that students received were from an authority and that factor could played a role engaging students to focus on the lecture. Conversely, in the both situations of e-Proof and textbook proof reading, proof comprehension was a self-learning process and that process reasonably depends on students' motivations and their engagement with the task. However these are only conjectures and to support those views further study would be needed.

Comparing e-Proof and textbook proof

Before the experiment, it was expected that the e-Proof group would perform better than the reading group. The expectation was logical, because in the e-Proof the more comprehensive explanations were given, in contrast the reading group received a written textbook proof without any additional explanation. (McNamara, 2001) shows that one of the key factors that influences text comprehension (if we consider a written proof as a text) is the structure of the text. A high-coherence text has fewer conceptual gaps in the text, therefore it requires fewer inferences to fill the gaps. A low-coherence text has bigger gaps in it, so requires more inferences. Clearly, the written textbook proof (that has no explanation) was a comparatively low-coherence text, and the e-Proof was an example of high-coherence text. This suggest that participants of the reading group would need to generate more inferences to comprehend the proof than the e-Proof group. Research (Chi, Leeuw, Chiu, & LaVancher, 1994; Ainsworth & Burcham, 2007) has shown that students who spontaneously self-explain when they study learn more than those who do not. Furthermore, self-explanations are usually more effective than explanations provided by others. One of features of the e-Proof is "explanations provided by others". This feature perhaps did not encourage participants inferring in proof comprehension and that could be reflected in the post test result (with 16% dropped in scores). Conversely, comprehending a textbook proof demands more self-explanation and perhaps the self-explanation factor had a long term effect on their proof comprehension. Further work is required to explore this hypothesis.

Conclusion

In this study, we compared undergradutes' comprehension of an unseen proof in several comditions: (i) reading the textbook proof, (ii) live lecture, (iii) computer based e-Proof. Our data suggested that live lecture was most effective, whereas e-proof was least effective with the largest drop off in the delayed post-test.

However, there are limitations of this study. When participants of the reading group and the e-Proof group read the proof, we were unaware of what kinds of strategies they used for reading and comprehending proof. We assumed that perhaps the reading group used self-explanation strategies, but there could have several other possible strategies that students employed. Therefore we suggest that a further study would be needed where students would be interviewed closely to explore their strategies of reading and comprehending proof.

References

Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction*, *17*(3), 286-303.

Alcock, L. (2009). e-Proofs: Student Experience of Online Resources to Aid Understanding of Mathematical Proofs.

Chi, M. T. H., Leeuw, N. D., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439-477.

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

Harel, G., & Sowder, L. (1998). Students'proof schemes: Results from exploratory studies. *Research in collegiate mathematics education*, *7*, 234-283.

Hersh, R. (1997). Prove–Once More and Again. *Philosophia Mathematica*, *5*(2), 153.

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*(1), 51-62.

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, *27*(3), 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.

Skemp, R. (1976). Instrumental understanding and relational understanding. *Mathematics Teaching*, *77*, 20–26.

Weber, K. (2009). Mathematics Majors' Evaluation of Mathematical Arguments and Their Conceptions of Proof.

Weber, K., & Alcock, L. (2005). Using warranted implications to understand and validate proofs. *For the Learning of Mathematics*, *25*(1), 34-38.

Yang, K. L., & Lin, F. L. (2008). A model of reading comprehension of geometry proof. *Educational Studies in Mathematics*, *67*(1), 59-76.

Appendix A

Cauchy's Generalised Mean Value Theorem: Textbook Version

Suppose that f and g are continuous on $[a,b]$ and differentiable on (a,b).

Suppose also that $\forall x \in (a,b)$, $g'(x) \neq 0$.

Then $\exists c \in (a,b)$ such that $\frac{f'(c)}{g'(c)} = \frac{f(b)-f(a)}{g(b)-g(a)}$.

Proof

Note that if $g(a) = g(b)$ then by Rolle's Theorem $\exists c \in (a,b)$ s.t. $g'(c) = 0$.

This contradicts the theorem premise so $g(a) \neq g(b)$.

Define

$$h(x) = f(x) - \frac{f(b)-f(a)}{g(b)-g(a)} g(x)$$

h is continuous on $[a,b]$ and differentiable on (a,b) (sum and constant multiple rules).

Also

$$h(a) = f(a) - \frac{f(b)-f(a)}{g(b)-g(a)} g(a) = \frac{f(a)g(b)-g(a)f(b)}{g(b)-g(a)},$$

and

$$h(b) = f(b) - \frac{f(b)-f(a)}{g(b)-g(a)} g(b) = \frac{f(a)g(b)-g(a)f(b)}{g(b)-g(a)},$$

so $h(a) = h(b)$.

Hence, by Rolle's Theorem, $\exists c \in (a,b)$ s.t. $h'(c) = 0$.

But

$$h'(c) = 0 \Rightarrow f'(c) - \frac{f(b)-f(a)}{g(b)-g(a)} g'(c) = 0 \Rightarrow \frac{f'(c)}{g'(c)} = \frac{f(b)-f(a)}{g(b)-g(a)},$$

So

$$\exists c \in (a,b) \quad \text{s.t.} \quad \frac{f'(c)}{g'(c)} = \frac{f(b)-f(a)}{g(b)-g(a)}.$$

Appendix B

Cauchy's Generalised Mean Value Theorem: Short Version

Suppose that f and g are continuous on $[a,b]$ and differentiable on (a,b).

Suppose also that $\forall x \in (a,b)$, $g'(x) \neq 0$.

Then $\exists c \in (a,b)$ such that $\frac{f'(c)}{g'(c)} = \frac{f(b)-f(a)}{g(b)-g(a)}$.

Proof

1) If $g(a) = g(b)$ then $\exists c \in (a,b)$ s.t. $g'(c) = 0$.

2) This contradicts the theorem premise so $g(a) \neq g(b)$.

3) Define

$$h(x) = f(x) - \frac{f(b)-f(a)}{g(b)-g(a)} g(x)$$

4) Then h is continuous on $[a,b]$ and differentiable on (a,b)

5) Also $h(a) = h(b)$.

6) Hence $\exists c \in (a,b)$ s.t. $h'(c) = 0$.

7) But

$$h'(c) = 0 \Rightarrow \frac{f'(c)}{g'(c)} = \frac{f(b)-f(a)}{g(b)-g(a)}$$

8) So

$$\exists c \in (a,b) \quad \text{s.t.} \quad \frac{f'(c)}{g'(c)} = \frac{f(b)-f(a)}{g(b)-g(a)}$$

Appendix C

Recent definitions and theorems

Definition: $f : A \to \mathbb{R}$ is continuous at $a \in A$ if and only if $\forall \varepsilon > 0\ \exists \delta > 0$ s.t. $x \in A$ and $|x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon$.

Definition: $f : A \to \mathbb{R}$ is differentiable at $a \in A$ if and only if $\lim_{x \to a} \frac{f(x)-f(a)}{x-a}$ exists.

Rolle's Theorem: Suppose that f is continuous on $[a,b]$ and differentiable on (a,b) and that $f(a) = f(b)$. Then $\exists c \in (a,b)$ such that $f'(c) = 0$.

Mean Value Theorem Suppose that f is continuous on $[a,b]$ and differentiable on (a,b). Then $\exists c \in (a,b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$.

Theorem (algebra for continuous functions) Suppose that $f, g : A \to \mathbb{R}$ are both continuous at $a \in A$ and that $c \in \mathbb{R}$. Then:

1. $cf : A \to \mathbb{R}$ is continuous at a.
2. $f + g : A \to \mathbb{R}$ is continuous at a.
3. $fg : A \to \mathbb{R}$ is continuous at a.
4. $f/g : A \to \mathbb{R}$ is continuous at a, provided $g(a) \neq 0$.

Theorem (algebra for derivatives) Suppose that $f, g : (a,b) \to \mathbb{R}$ are both differentiable at $x \in (a,b)$ and that $c \in \mathbb{R}$. Then:

1. $cf : (a,b) \to \mathbb{R}$ is differentiable at x.
2. $f + g : (a,b) \to \mathbb{R}$ is differentiable at x.
3. $fg : (a,b) \to \mathbb{R}$ is differentiable at x.
4. $f/g : (a,b) \to \mathbb{R}$ is differentiable at x, provided $g'(x) \neq 0$.