Running head: Self-Explanation Training Improves Proof Comprehension

Self-Explanation Training Improves Proof Comprehension

Mark Hodds, Lara Alcock and Matthew Inglis

Mathematics Education Centre, Loughborough University

Address for correspondence:

Mathematics Education Centre

Loughborough University

Loughborough

LE11 3TU

UK

M.Hodds-07@alumni.lboro.ac.uk

Abstract

In this paper we report three experiments demonstrating that a simple booklet containing *self-explanation training*, designed to focus students' attention on logical relationships within a mathematical proof, can significantly improve their proof comprehension. Experiment 1 demonstrates that students who receive the training generate higher quality explanations and perform better (effect size $d = 0.950$) on a comprehension test. Experiment 2 demonstrates that self-explanation training increases students' cognitive engagement and the frequency with which they move their attention around a proof. Experiment 3 demonstrates that a 15-minute in-lecture self-study intervention improves students' proof comprehension, and that the effect persists over time. Thus we argue that 'transition to proof' courses should incorporate self-explanation training.

Self-Explanation Training Improves Proof Comprehension

Introduction

Proof is central to mathematics, and consequently mathematics educators have devoted substantial research effort toward understanding how students engage with proof and proving (for a review see Reid & Knipping, 2011). Until recently, however, research has tended to focus on proof construction rather than the proof comprehension (Mejía-Ramos & Inglis, 2009). This is potentially problematic because, in traditional instruction at least, undergraduates are often introduced to new mathematical ideas and techniques through reading proofs in class and in textbooks (Rav, 1999; Selden & Selden, 2003; Weber, 2004).

It is thus of concern to see converging evidence that undergraduate students are not reliably able to judge whether a proof is valid or invalid (Alcock & Weber, 2005; Inglis & Alcock, 2012; Ko & Knuth, 2009, 2013; Selden & Selden, 2003; Weber, 2010). Although researchers have proposed pedagogic strategies to address this issue (e.g., Leron, 1983; Rowland, 2001a, b), there is scant research on the efficacy of these strategies; what evidence does exist indicates that such approaches may be less effective than hoped (Fuller, Mejía-Ramos, Weber, Samkoff, Rhoads, Doongaji, & Lew, 2011) or, in some cases, actually inferior to standard approaches (Roy, Alcock & Inglis, 2010). In this paper we adapt a pedagogical technique – *self-explanation training* – that has been shown to be effective in other contexts (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989), and demonstrate that it facilitates comprehension of mathematical proofs.

Theoretical Background

Research on proof comprehension

It has long been of concern that some students do not understand what it means for an argument to constitute a mathematical proof. Such students confuse the roles of evidence and proof, offering examples instead of deductive arguments (Chazan, 1993; Recio & Godino, 2001). They accept logically invalid deductions (Alcock & Weber, 2005; Ko & Knuth, 2009, 2013) and even whole arguments (Selden & Selden, 2003; Weber, 2010). They focus too much attention on surface features of an argument, over-valuing texts that contain large amounts of algebraic manipulation or that otherwise appear deductive (Harel & Sowder, 1998; Healy & Hoyles, 2000; Segal, 2000), and concentrating on algebraic manipulations at the expense of attending to logical claims captured elsewhere in the text (Inglis & Alcock, 2012; Selden & Selden, 2003). As a consequence, some researchers have argued that students need to develop more mature *proof schemes* (Harel & Sowder, 1997; 2007) and need to be exposed to teaching designed specifically to facilitate this learning (Stylianides & Stylianides, 2009).

More recently, however, attention has focused on whether these undesirable behaviors really indicate a fundamental misunderstanding of the relationship between evidence and proof, or whether they might be better understood as attempts to act appropriately that fail because the student lacks either the skill to construct a deductive argument or the will to thoroughly evaluate one (Inglis & Alcock, 2012). It certainly seems that undergraduate students are often ineffective readers of mathematical texts: they tend to focus on worked examples without reading potentially helpful explanatory material, they take insufficient care to ensure that they understand technical terms, and even those with appropriate mathematical backgrounds and good general reading skills do not read mathematics well enough to be able to apply new information in closely related problems (Lithner, 2003; Shepherd, 2005;

Shepherd, Selden & Selden, 2012). However, the fact that students often do not read effectively does not imply that they cannot do so, or that their standards of proof differ substantially from those of mathematicians. Healy and Hoyles (2000), for instance, found that high school students often selected empirical arguments as being most like their own, but that most were aware of the limitations of these arguments. Stylianides and Stylianides (2009) found similar awareness among pre-service teachers. Segal (2000) found that during their first undergraduate year, students curtailed their tendency to accept empirical arguments as proofs, and Weber (2010) found that mathematics majors judged an empirical argument to be both invalid and unconvincing.

Appropriate rejection of empirical arguments, of course, is not the same as appropriate evaluation of deductive arguments. Segal's (2000) participants, for instance, became more likely to accept 'proofs' that used general notation and appeared deductive; unfortunately, the same occurred for both a valid and an invalid proof with this appearance. Selden and Selden (2003), Inglis and Alcock (2012) and Weber (2009, 2010) all found that some undergraduate students performed poorly when evaluating even short purported proofs, for instance by failing to reject an argument that addressed only the converse of the target statement. It could be that the problem here is fundamental and depressing. Perhaps many students simply do not have the cognitive capacity to reason correctly about mathematical arguments. We do not believe that this is the case, and suggest that a detailed reading of this literature suggests an alternative possibility.

First, Selden and Selden (2003) suggested that students focused too much on surface features of purported proofs. Inglis and Alcock (2012) confirmed using eye tracking that undergraduates, compared with expert mathematicians, devoted more of their attention to algebraic parts of proofs and less to the surrounding text (in which logical claims are often made explicit). Second, Weber (2009) reported that students rarely spent more than two minutes evaluating an argument, and that some did not seem to consider it their responsibility to resolve any confusion that arose, or were willing to accept an argument that they did not find fully convincing or even in which they had found a logical flaw. Weber (2010) reported that his participants made errors because they did not check whether assumptions used in a proof were appropriate and because they focused on correctness of assertions but not on principles used to deduce new assertions. Inglis and Alcock (2012) confirmed (again using eye-tracking) that during proof comprehension attempts, undergraduates shifted their attention around less often than mathematicians; that mathematicians made significantly more moves from one line of a proof to another in a manner consistent with an attempt to understand the logical relationships between these lines.

It could therefore be that students fail to identify logical errors simply because they do not read proofs very thoroughly; that they *do* have the capacity to reason correctly about mathematical arguments, and that the failures at this level are not of understanding but of execution. In this case, as Inglis and Alcock (2012) suggested, considerable improvement might be possible with a relatively light-touch intervention that encourages students to conduct their reading more thoroughly. In order to evaluate such an intervention, and to compare the rationales behind different interventions, it is helpful to clarify what we expect a reader to achieve by way of proof comprehension; we turn to this next.

A theoretical model for testing proof comprehension

Mejía-Ramos, Fuller, Weber, Rhoads, and Samkoff (2012) recently proposed a theoretical model of proof comprehension. This model is built on the work of Conradie and Frith (2000) and Yang and Lin (2008), and is consistent with other comments by mathematicians on questions that one might set in order to establish whether a student has understood a proof (e.g., Cowen, 1991; Solow, 2005). It comprises seven dimensions, the first of which (*meaning of terms and statements*) assesses the basic knowledge required to correctly interpret technical terms and single statements. The next (*justification of claims*) assesses comprehension of logical relationships among individual lines in a proof. Specifically, if a proof is understood as a series of statements $A_1, A_2, \ldots A_n, B,$ where B is the to-be-proved theorem (Rav, 1999), then many of the steps $A_i \rightarrow A_{i+1}$ require a *warrant* (Toulmin, 1958): a justification that allows the reader to conclude that A_{i+1} follows logically from some subset of the preceding lines, together with agreed axioms, definitions or theorems. As Weber and Alcock (2005) pointed out (see also Solow, 2005, p.15), not all of these warrants will be explicitly articulated in the proof text; some will be left for the reader to infer, and this dimension tests a reader's ability to do so. The third dimension (*logical structure*) assesses a similar skill in a different way, asking about logical dependence relationships between specific lines. Four further dimensions – higher-level ideas, general method, application to examples, and identifying modular structure – assess understanding of holistic aspects of the proof's structure. All seven dimensions are listed in Table 1, together with samples of the question types that Mejía-Ramos et al. suggest are appropriate for testing each dimension.

This model constitutes an important methodological contribution to the field, because it permits researchers to assess the success of a reading attempt, and thus to rigorously evaluate interventions that attempt to improve students' proof comprehension. It also permits theoretical discussion of the characteristics of earlier interventions, a point we turn to next.

Table 1. The seven dimensions for assessing undergraduate proof comprehension suggested by Mejía-Ramos et al. (2012).

Dimension	Definition	Example question
Meaning of terms and statements	Understanding the meaning of symbols, terms and definitions.	What does the symbol ∃ mean?
Justification of claims	Understanding how new assertions in the proof follow from previous ones.	In the proof, which justification best explains why…?
Logical structure	Understanding the logical relationship between lines or components of a proof.	What is the logical relationship between lines (LX) and (LY)?
Higher-level ideas	Identifying a good summary of the overarching approach of the proof.	Which of the following is the best summary of…?
General method	Applying the methods within the proof to a different context.	Could the method of the proof applied in line X be used to prove…?
Application to examples	Using the ideas in the proof in terms of a specific example.	Using the logic of the proof, which best exemplifies why $x = 5$ is not a solution to $f(x) = 30$?
Identifying modular structure	Understanding the main components and modules within a proof and the logical relationship between them.	Which of the following explains why…was included in the proof?

Improving comprehension: changing the presentation

Among attempts to improve students' proof comprehension, one can distinguish two broad approaches: changing the presentation of the proof, and changing the way a student engages with it. These approaches share an assumption that students need additional or different explanations in order to fully understand a typical proof, but they differ in their assumptions about who should provide these explanations. When the presentation is changed, explanations are usually provided by the instructor, and in this section we review three theoretical suggestions that take this approach, together with empirical research investigating their efficacy.

Leron (1983) suggested that instead of a standard linear presentation – proceeding from hypothesis to conclusion in a unidirectional manner – one might present *structured proofs*. A structured proof is arranged in levels, with the main ideas and approach given at the top level, and subsequent levels giving details and justifications of each of the steps in the preceding levels. In terms of Mejía-Ramos et al.'s (2012) framework, a structured proof is designed to facilitate understanding of *higher-level ideas* and *identifying modular structure*, though it does so at the expense of separating some claims from their supporting data and warrants. These changes are reflected in empirical research on the efficacy of structured proofs. Fuller et al. (2011) found that, compared to reading a traditional proof, students who read structured proofs were more successful at summarizing the key ideas of the proof, but that they performed slightly, albeit not significantly, worse on other aspects of proof comprehension.

Rowland (2001) suggested (in the context of number theory) that instead of a fully general proof, one might present a *generic proof*. In such a proof, operations and arguments are applied, not to general algebraic notation (as in traditional proofs), but to a generic example that is used throughout. This generic example is chosen so that analogous operations and arguments would apply to every other member of the general class under consideration. In terms of Mejía-Ramos et al.'s framework, an entire generic proof may thus be understood as an *application to an example*. One might also argue that a generic proof facilitates understanding of logical relationships of all types as they pertain to that example, but does so at the cost that the reader is responsible for thinking about the invariance of those relationships in the general case. Rowland reported on the use of generic proofs in classroom settings, and his comments reflect these issues: he noted that some trainee teachers were able to see such proofs as generic, but some were not, and some were too ready to generalize from specific examples without attention to which relationships were, indeed, invariant.

Alcock and Wilkinson (2011) suggested that instead of a simple proof text, one might present an augmented text in the form of a computer-based *e-Proof*. An e-Proof consists of a series of computer slides showing the theorem and whole proof, together with annotations that vary from slide to slide (lines or algebraic expressions might, for example, be highlighted using boxes, linked with arrows, or grayed out). Each slide is accompanied by a replayable audio account explaining the relevant section, and an e-Proof is characterized by Alcock and Wilkinson as an attempt to capture the extra audio and visual explanations that a lecturer might give when presenting a proof. In terms of Mejía-Ramos et al.'s framework, the audio commentary might naturally include information relating to *meaning of terms and statements*, and both the audio and the annotations might draw attention to *justification of claims*, *logical structure*, *higher-level ideas* and *identifying modular structure*.

Roy et al. (2010) compared the understanding developed by students working with an e-Proof with that of students reading the un-augmented proof text for the same amount of time. They found that those who studied the e-Proof did not retain their knowledge as well as those who read the un-augmented version. Roy et al. suggested that this might be because e-Proofs reduced the need for the reader to think in depth about the proof by reducing the impetus to generate their own explanations. This suggestion is obviously related to our earlier point: perhaps students can provide their own explanations and would perform better if allowed and encouraged to do so.

As noted above, 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. This suggestion is consistent with findings from the general education literature on example-based learning: a meta-analysis by Wittwer and Renkl (2010) found that the benefits of adding additional instructional explanations is limited, and not necessarily more effective than supporting students to generate their own explanations. In the next section, we take up Roy et al.'s (2010) suggestion by considering approaches in which the extra or different explanations are generated not by an instructor but by the student.

Improving comprehension: changing engagement

The second broad approach to improving proof comprehension is to change a student's engagement with a proof. One way to do this is to provide *self-explanation training*, a strategy that has been shown to improve reading comprehension in other subject areas and in mathematics at lower levels. This is the approach taken in our experiments, and we briefly review the relevant literature here.

The term 'self-explanation' originated when Chi et al. (1989) asked undergraduate students to read chapters in a book on Newtonian mechanics. Students were encouraged to create their own explanations of the material, and were subsequently given 19 problems on related content. Chi et al. reported that the 'good' students (who had a mean success of 82% on the problems) produced more self-explanations – more interpretations of what had been read that involved information and relationships beyond those literally contained in the text – than did the 'poor' students (who had a mean success of 46% on the problems). The researchers raised the possibility that constructing self-explanations is one of the signatures of effective reading comprehension.

Later work investigated whether students can be taught to self-explain. Chi, de Leeuw, Chiu, and LaVancher (1994) found that eighth grade students who received self-explanation training and then read a text on the circulatory system learned significantly more than students who were simply asked to read the text twice. Similar results have since been found in other content domains such as history (Leinhardt, 1993), programming, and multimedia (Roy & Chi, 2005).

Studies also indicate that self-explanation training might be helpful in mathematics (e.g., Aleven & Koedinger, 2002; Durkin, 2011). In Rittle-Johnson's (2006) study, for instance, third- to fifth-grade students learned about solving unfamiliar types of problem concerning mathematical equivalence. Rittle-Johnson used a 2×2 design in which the students either invented their own methods or were instructed in standard methods, and in which they either did or did not receive self-explanation training. In both immediate and delayed post-tests, students in the self-explanation groups showed significantly higher procedural accuracy and significantly higher scores on transfer problems.

Wong, Lawson, and Keeves (2002) asked high-achieving 9th grade students to participate in sessions on circle theorems. Students from both a self-explanation and a control group had individual think-aloud training, and those from the self-explanation group additionally heard recorded examples of self-explanations. Each was then asked to apply their training to read a booklet on a particular geometry theorem and associated example; for those from the self-explanation group, the booklet contained additional self-explanation prompts. In a follow-up session, participants took a post-test that required application of the target theorem, and of other recently-reviewed circle theorems, to solve various problems. The results showed that the self-explanation group scored significantly higher on the post-test, with the effect being particularly pronounced for far-transfer items that involved using multiple theorems, constructing new figures and developing representations for word problems.

Of course, effective instruction might take many forms, and it is not obvious how self-explanation training should be linked with other forms of mathematical instruction. Matthews and Rittle-Johnson (2009), for instance, showed that giving children self-explanation prompts alongside conceptual instruction was no more effective than giving the conceptual instruction alone. Renkl (2002) argued that self-explanations are often faulty or otherwise inadequate, and developed an approach in which the learner does as much self-explanation as possible but also has access to additional instructional explanations. Nevertheless, empirical results on self-explanation effects are compelling. They indicate that self-explanation training can enhance students' comprehension of mathematical material and of texts at the undergraduate level more broadly.

Further, we believe there are reasons to anticipate that self-explanation training might be particularly effective in the domain of mathematical proof comprehension. These rest upon the particular features of mathematical proofs (of the kind encountered in undergraduate mathematics) as compared with other texts. Specifically, proof texts are extremely dense in deductive links (Mayans, 2004; Stewart & Tall, 1977): reading them requires one to reconstruct the author's thought processes (Houston, 2009; Solow, 2005), but this should be possible because each claim (if it does not, for example, simply define a new object) will follow deductively from theorem premises, previous claims and agreed true results. This is not the case in non-mathematical texts about other subjects, even in science. In the biology texts used in some self-explanation studies (e.g., Chi et al., 1994; McNamara, Kintsch, Songer & Kintsch, 1996), for instance, the material is primarily factual, and the order in which it is presented is not dictated by logical relationships between the claims that constitute the text. One can certainly construct self-explanations about such material: a person with appropriate knowledge might well be able to construct reasonable explanations for biological phenomena based on analogy or abductive reasoning. But one could not routinely *deduce* new valid statements from existing statements about such phenomena, or be certain purely by reference to logical reasoning that one's inferred warrants were appropriate and did lead to deductively valid conclusions. This means that proof texts might be particularly amenable to comprehension via self-explanation: a mathematically competent individual who looks for links within a proof will find unusually many links in an unusually simple form.

In view of these considerations, and our knowledge of the difficulties students have when reading mathematical proofs, we hypothesized that self-explanation training would be an effective method of improving undergraduate students' proof comprehension. Our primary goal in this paper is to test this suggestion experimentally. In order to frame our empirical work, the next section reviews the methods by which reading processes and outcomes can be empirically investigated.

Methodology

Measures

Learning outcome measures

Studies on self-explanation effects usually seek to demonstrate increases in learning due to self-explanation effects by reporting outcome measures. We take this approach, using proof comprehension tests constructed according to Mejía-Ramos et al.'s (2012) framework. In common with other self-explanation studies (e.g., Heijltjes, Van Gog & Paas, 2011; O'Reilly, Best, & McNamara, 2004; Ziegler & Stern, 2011), we administered these tests immediately after reading attempts and (in Experiment 3) in a delayed post-test.

Process measures: explanation quality

Some studies additionally seek to provide insights into the mechanisms behind self-explanation effects by reporting measures that capture some aspect of the reading process. Two such process measures are *explanation quality* (discussed in this section), and *reading behavior* as indexed by the loci of readers' attention (discussed in the next).

Some researchers interested in self-explanation effects, including Ainsworth and Burcham (2007), have studied explanation quality as a measure of what takes place during the reading process. Explanation quality is typically documented by recording think-aloud protocols as individual students read a text, and by coding their utterances. Ainsworth and Burcham used eight categories of utterances, which they also grouped into two supercategories: *explanations* and *non-explanations*. Seven[1] of their categories are listed in Table 2; because these were the categories used in Experiment 1, the definitions in the table are slightly adapted for the context of proof comprehension.

Explanation quality can then be related to comprehension outcomes. Ainsworth and Burcham showed that that students' comprehension of a biology text was related to the types of explanation they generated: students who produced more false explanations scored lower on the comprehension test, and students who produced more *positive monitoring* and *principle-based* statements performed better. In our context, and in relation to our suggestion above that one might expect proof texts to be particularly amenable to comprehension by self-explanation, we note that there is a relationship between this categorization of self-explanations and the earlier discussion about assessing proof comprehension. Specifically, a *principle-based* explanation might involve articulating the meanings of terms and statements

---

[1] Ainsworth and Burcham (2007) also included an eighth category, "elaborative explanations", which constituted self-explanations using elaborated metaphors or analogies. We found no such instances in our data. In Table 2 we also include a "no comment" category of non-explanation.

or inferring a warrant, a *goal-driven* explanation is likely to capture some aspect of the holistic structure of the proof, and *noticing coherence* might involve either of these.

Table 2. The categories of comments listed by Ainsworth and Burcham (p. 293; 2007), adapted slightly for the context of proof comprehension.

Category	Type	Definition (in the context of proof comprehension)
Principle-Based	Explanation	Participant gave an explanation based upon definitions, theorems or ideas not explicitly written in the proof, e.g., "...this is because, by the definition of triadic, ..."
Goal-Driven	Explanation	Participant gave an explanation that related to the structure of the proof (how it is used in order to reach the goal of proving the theorem).
Noticing Coherence	Explanation	Participant gave an explanation that related to an idea used earlier in the proof, e.g., "...this is because in line 5 we introduced..."
No Comment	Non-explanation	Participant spoke no words for the line.
False Explanation	Non-Explanation	Participant gave an incorrect explanation.
Paraphrasing	Non-Explanation	Participant simply repeated the line or part of the line using similar words or the same words.
Positive Monitoring	Non-Explanation	Participant stated "I understand this", "OK, this makes sense" or similar.
Negative Monitoring	Non-Explanation	Participant stated "I don't understand this", "How is this true?" or similar.

Process measures: reading behavior

Another way to investigate the mechanisms behind self-explanation effects is to directly observe readers' loci of attention. Few studies have attempted this, although eye-movement methods have been used extensively to study numerous other phenomena including reading (e.g., Rayner, 1998) and logical reasoning (e.g., Ball, Lucas, Miles & Gale, 2003). Eye-movement studies rely on the so-called *eye-mind hypothesis* (Just & Carpenter, 1980), which states that gaze direction is closely related to attention location, especially during effortful tasks (e.g., Ball et al., 2003; Deubel & Schneider, 1996; Rayner, 2009). Measures used in such studies rely on the fact that eye movements during reading consist of *fixations* – stationary periods during which information is processed (typically lasting around 150ms to 500ms) – and *saccades* – rapid moves to new locations, during which no information can be processed (e.g., Matin, 1974). Records of eye movements consist of information on the duration and location of each fixation, and a variety of meaningful measures can be derived from these.

First, longer fixations are associated with more effortful cognitive processing (see e.g., Just & Carpenter, 1976; Poole & Ball, 2006; Rayner, 1977), so *mean fixation durations* provide a measure of an individual's cognitive engagement with a task: a higher mean fixation duration on a given location indicates a higher level of cognitive engagement with the information at that location.

Second, fixation locations allow us to track the way in which an individual moves their attention around during a task. It is less obvious how to use these data to differentiate reading behavior of relevance for proof comprehension, but Inglis and Alcock (2012) reported an approach in which they counted and compared the numbers of *between-line saccades* made by undergraduate students and by mathematicians. Specifically, they counted all saccades that started with a fixation on one line of a purported proof and finished with a fixation on another line (indicating that the reader has moved their attention from line to line, rather than simply reading along the current line). Using this approach, they reported that mathematicians made significantly more between-line saccades than did students, indicating that expert readers move their attention around more than novices during proof comprehension attempts.

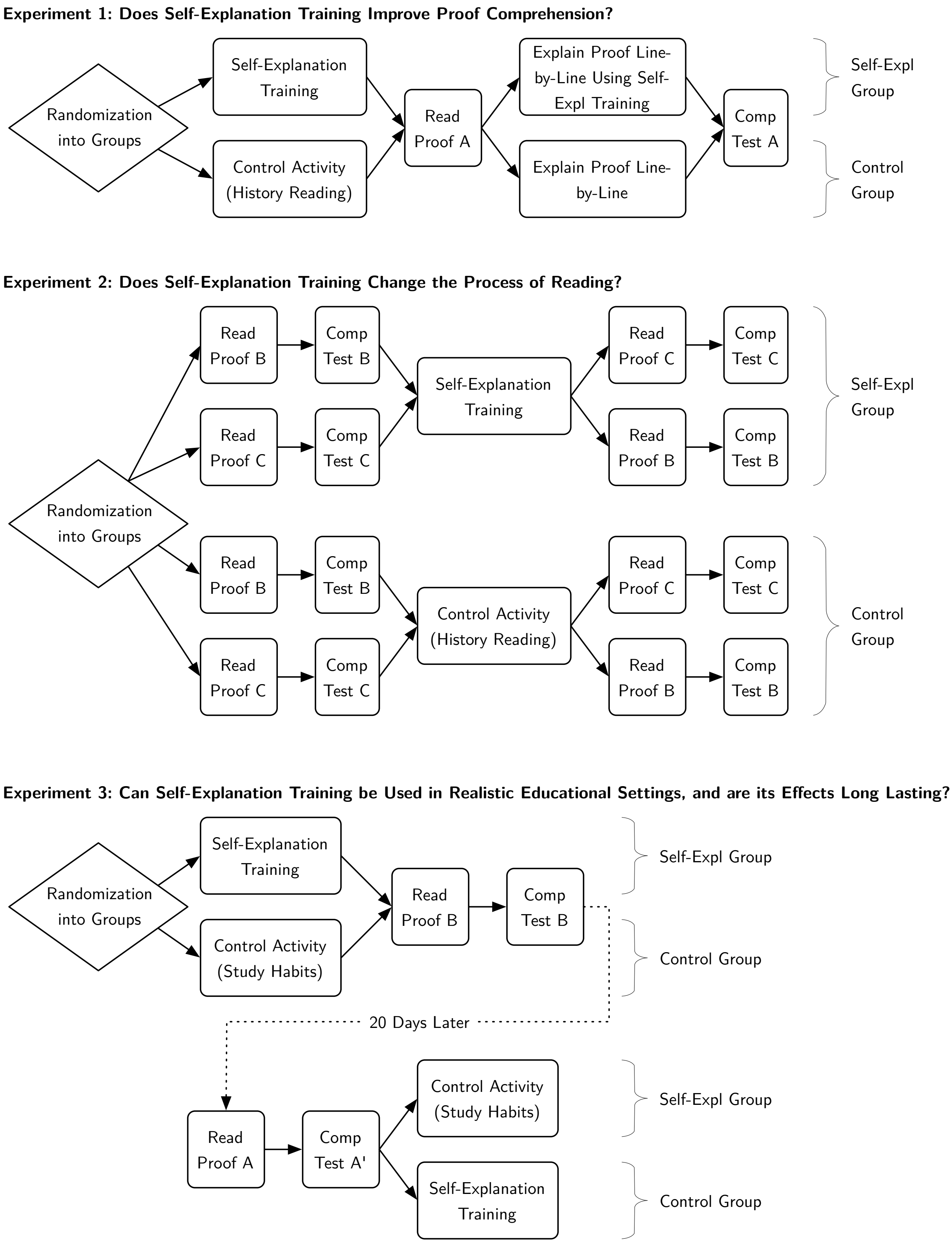
In any such work, it is important to note that eye-movement data are noisy, making it dangerous to interpret any single fixation as having a particular meaning (Inglis & Alcock, 2013) or otherwise to over-interpret a single sequence of eye movements. However, any eye-movement record contains data on hundreds of fixations (over 100 per minute in the data used in our studies), so these aggregate measures do provide a reliable way of comparing overall reading behavior both within and between participants. We used both types of eye-movement measure in Experiment 2.

Research Design

In the experiments reported in this paper, we asked three main questions. First, does studying self-explanation training improve students' comprehension outcomes, as measured by comprehension tests? Second, does studying self-explanation training change the *process* by which students read mathematical proofs? Third, can self-explanation training be used in genuine pedagogical settings and, if so, will it lead to long-lasting gains in comprehension skills? We now briefly summarize the designs of the three experiments reported in the remainder of this paper.

In Experiment 1, we used an experimental design to investigate the effects of self-explanation training on learning outcomes and on the nature of explanations produced by students. Specifically, we asked whether self-explanation training improves the quality of explanations students give and leads to improved proof comprehension. In Experiment 2, we used eye-tracking methods to investigate the effects of self-explanation training on reading processes. In Experiment 3, we investigated the effects of self-explanation training in a genuine pedagogical setting and over a longer time period. The designs of all three studies are summarized in Figure 1, and are described in detail in the following sections.

Figure 1. An illustration showing the designs of the three experiments reported in this paper.



Experiment 1

In Experiment 1 we investigated whether self-explanation training affects (a) the quality of explanations students give when reading a proof and (b) the level of resulting proof comprehension. To this end we conducted a true experiment by randomly allocating students into two conditions: those who received self-explanation training, and those who took part in a control activity.

Methods

Participants

Participants were 76 mathematics undergraduates at Loughborough University, who took part in exchange for an £8 (about $13) stipend. All participants were studying mathematics in 3-year single- or joint-honors degree programs, meaning that at least half (and often all) of their time was taken up with proof-based mathematics courses. In order to investigate whether self-explanation training differentially influenced students with different levels of experience, we sampled from all three years of the degree program: participants were 26 students in their first year of study, 26 students in their second year of study and 24 students in their third year of study. In order to determine whether there was a causal relationship between our intervention and any differences on our outcome measures, we randomly assigned each participant to either the control group (38 participants; 10, 15 and 13 from Years 1, 2 and 3 respectively) or the self-explanation group (38 participants; 16, 9 and 13 from Years 1, 2 and 3 respectively). Participants took part individually in a quiet room.

Materials

Self-explanation training.

The self-explanation training consisted of a series of computer slides adapted from earlier materials used by Bielaczyc, Pirolli, and Brown (1995) and Ainsworth and Burcham (2007). The slides explained the benefits of self-explanation training, and elucidated the key principles: identifying key ideas in each line of a proof, and explaining each line in terms of previous ideas presented in the proof or in terms of previous knowledge. The slides then demonstrated the self-explanation strategy via an example proof. Finally, a practice proof was displayed and participants were asked to generate self-explanations in response to it. No feedback was given. The training slides are provided in the Appendix. Participants in the self-explanation group took an average of 19.9 minutes (SD = 2.0 minutes) to work through the training materials. Those in the control group were slightly quicker to complete a control activity, described below (16.8 minutes, SD = 1.49 minutes).

Proof comprehension task.

For the comprehension task, we used Mejía-Ramos et al.'s (2012, p. 8) proof – referred to here as Proof A and reproduced in the Appendix – that there exist infinitely many triadic primes (participants were also provided with definitions of monadic and triadic). We were confident that all potential participants would possess the background knowledge required to successfully engage with this material because it assumes knowledge only of prime numbers, divisibility and routine algebraic manipulation.

To assess participants' success in comprehending this proof, we constructed a 14-item proof comprehension test using the questions suggested by Mejía-Ramos et al. (2012). In total, we chose two items for each of their seven dimensions of proof comprehension. These questions are provided in the Appendix. The order of questions was randomized for each participant.

Items on the proof comprehension test were allocated one, two or three points according to their complexity, giving a total possible score of 28. Details of the scoring scheme are given in the Supplementary Materials (available online). Participants' responses were graded, using this scheme, by one of two mathematics postgraduate students, both of whom were naïve to the purpose of the experiment and blind to each participant's group assignment. Once one of the graders had scored a paper, they gave it to the other to moderate. No disagreements about the assigned scores were highlighted by the graders.

Procedure

Throughout the experiment, all materials other than the comprehension test were displayed on a computer screen and all audio responses were recorded for subsequent analysis. The experiment had three phases.

In the training phase, participants in the self-explanation group were given the self-explanation training to study at their own pace. Those in the control group were asked to study a passage on the history of the mathematics of right-angled triangles and to answer questions about it. This ensured that participants in both groups had approximately equal time on task.

In the reading phase, all participants were given Proof A to study silently and at their own pace. Once they had studied the proof to their satisfaction, they were given the proof again but with the first line highlighted. Control participants were asked to verbalize comments about the highlighted line that helped them understand the proof. Self-explanation participants were asked to do the same, using their training to guide them. Pressing the space bar moved the highlight to the next line of the proof.

In the test phase, all participants were asked to complete the proof comprehension task on paper, with the proof still available for viewing. When participants had completed all parts of the experiment, which took between 20 and 55 minutes, they were thanked and dismissed.

Results

In this section we report three main results. First, we justify the use of a single measure of proof comprehension by analyzing the dimensionality of our proof comprehension test. Second, we report the effects of self-explanation training on the nature of the participants' verbal explanations. Third, we report the effects of self-explanation training on proof comprehension.

The dimensionality of proof comprehension

Recall that Mejia-Ramos et al. (2012) identified seven dimensions of proof comprehension. If these dimensions were independent, it would be problematic to sum participants' scores on items from different dimensions, and therefore impossible to justify the use of an overall proof comprehension score for each participant. To address this issue we subjected participants' scores to a Principle Components Analysis (PCA, Lorenzo-Seva & Ferrando, 2006). Cattell's Scree Test indicated that a single factor should be extracted, suggesting that it is reasonable to treat proof comprehension as a one-dimensional construct. The split-half internal reliability of the test was .73, suggesting an acceptable level of internal reliability (e.g. Kline, 1999). This internal reliability figure suggests that around 45% of the variance in comprehension test scores involved knowledge/skills other than simple proof comprehension. This is in some sense unsurprising: clearly questions in any such test will implicitly draw to different extents upon a participants' background knowledge and skills.

Effect of Self-Explanation Training on the Nature of Explanations

To investigate the effect of self-explanation training on the nature of students' explanations, participants' verbal comments during the test phase were coded using the scheme given in Table 2. No comment, false explanation, paraphrasing, and positive and negative monitoring statements (those classified as non-explanations) have meanings directly analogous to those used in previous self-explanation studies. We illustrate principle-based, goal-driven and noticing coherence statements (classified as explanations) in the mathematical context of our experiment now (L1 indicates a comment made about line 1, etc.).

First, the following two comments were classified as *principle-based* because they invoke definitions that are not explicitly written in the proof:

> (L1) "…so therefore we conclude that the product of two monadic numbers is again monadic because it satisfies that definition."
>
> (L4) "I think the reason [it is] defined in this way is because it takes the form of a triadic number, um, $4k + 3$, as we see, but instead of using k, we've replaced it using the product of p_2, p_3, p_4 through to p_n then plus…3."

Second, the following comment was classified as *goal-driven* because it explains why setting up M as triadic helps reach the goal of proving the theorem:

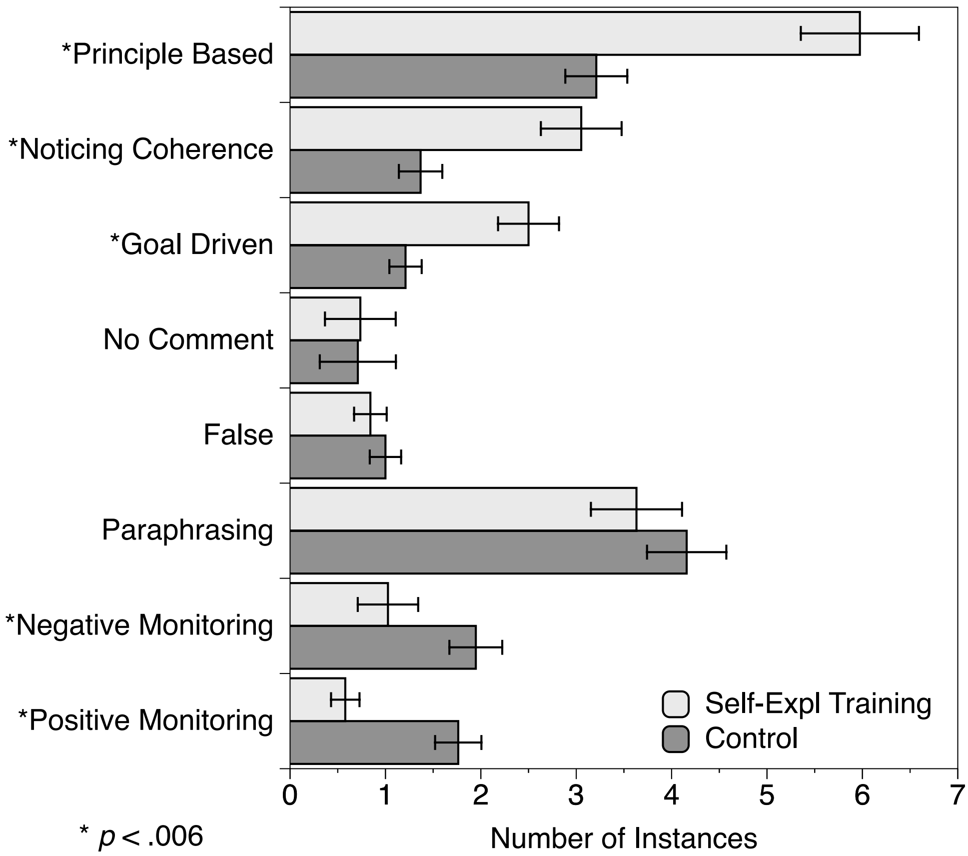
> (L8) "So this proves that M must be a monadic, which is a contradiction to the initial statement, explaining why we set up where M was triadic."

Finally, the following comment was classified as *noticing coherence* because it explains a result by relating it back to information from earlier lines:

> (L6) "Okay, so now we've just, well I've just, I think I just said no triadic primes can divide M. And so, from what we've shown from [line] four and [line] five, you can see that M can't be divided by any of the other triadic primes."

We compared the median number of each type of comment given by students in each group, shown in Figure 2, using a series of Bonferroni-corrected Mann Whitney U tests. We found that participants in the self-explanation group gave significantly more explanations; that is, they gave more comments categorized as principle-based, $U = 386$, $p < .001$, noticing coherence, $U = 399$, $p = .001$, or goal-driven, $U = 407$, $p = .001$. Indeed, they gave an average of 11.5 explanations of these types, whereas those in the control group gave an average of 5.8, $t(52.5) = 4.434$, $p < .001$, $d = 1.017$. In particular, participants in the self-explanation group gave an average of 6.0 principle-based explanations, whereas those in the control group gave an average of 3.2, indicating that those who received the training inferred approximately twice as many implicit warrants as those who did not.

Figure 2. The mean number of instances of each type of comment, separated by condition. Error bars show ±1SE of the mean.

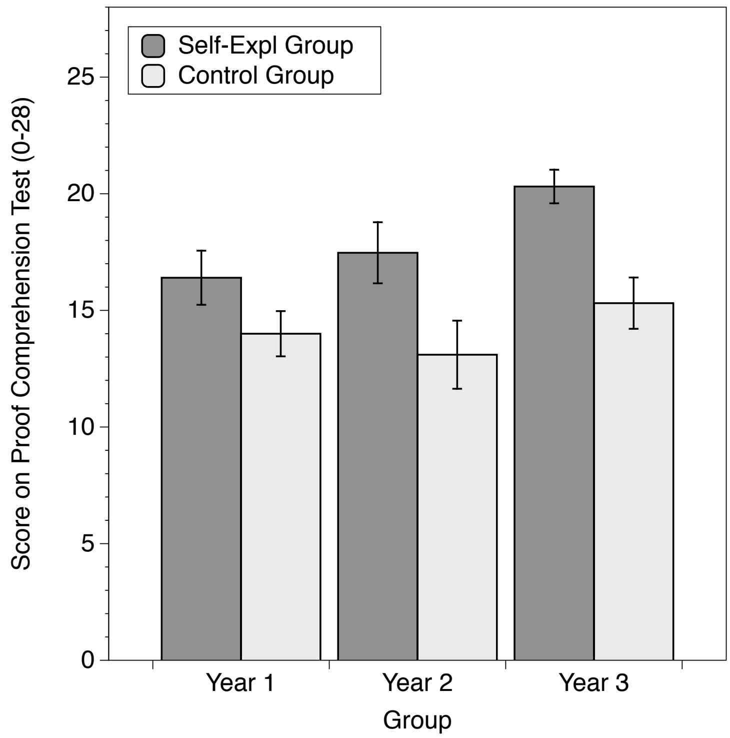


Participants in the self-explanation group also gave significantly fewer non-explanations, in particular fewer that were categorized as positive monitoring, $U = 400$, $p < .001$ and negative monitoring, $U = 440$, $p = .002$. The between-groups differences in the remaining categories (all non-explanations) did not reach significance. Overall, these results indicate that receiving self-explanation training increased both the number and proportion of high quality explanations given by participants during their proof comprehension attempts.

Effect of self-explanation training on proof comprehension outcomes

To investigate the effect of self-explanation training on proof comprehension, participants' proof comprehension scores were subjected to an ANCOVA with one between-subjects factor (condition: self-explanation training, control). Because the time participants spent studying the proof was correlated with their proof comprehension scores, $r = .439$, $p < .001$, we included proof study time as a covariate. This analysis revealed a significant effect of condition, $F(1,76) = 13.315$, $p < .001$, $\eta_p^2 = .154$. Those who received self-explanation training scored an average of 18.2 (SD = 4.2) on the comprehension test; those who did not scored an average of 14.2 (SD = 4.0). This difference corresponds to a very large effect size, $d = 0.950$. Because we included study time as a covariate, this difference cannot be accounted for by the groups spending different amounts of time studying the proof (a similar result was found when time was not included as a covariate, $t(74) = 4.142$, $p < .001$).

Figure 3. The mean score on the proof comprehension test, separated by condition and year group. Error bars show ±1SE of the mean.



We also explored whether the effect of self-explanation training differed by year of study. Mean scores for each year group in the two conditions are shown in Figure 3. A 3 (year) × 2 (condition) ANCOVA (again with time as the covariate) revealed, unsurprisingly, a main effect of year, $F(2, 69) = 3.456$, $p = .037$, with those in Year 3 ($M = 17.8$, $SD = 4.2$)

outperforming those in Years 2 (M = 15.8, SD = 5.2) and 1 (M = 14.9, SD = 3.9). Importantly, however, there was no significant year-by-condition interaction, $p > .2$, indicating that the self-explanation training did not appear to substantially benefit one year group more than any other.

Summary and discussion

Experiment 1 investigated the suggestion by Inglis and Alcock (2012) that self-explanation training might lead to an increase in proof comprehension. We found that, compared with those in the control group, participants who had received the self-explanation training generated higher quality explanations and performed on average performed almost one standard deviation higher on the comprehension test. Our analysis ruled out the possibility that this finding was due to between-group differences in the length of time spent studying the proof, suggesting that it was due to the higher-quality reading by those in the self-explanation group. Because we randomly allocated participants to groups, our study was a true experiment, so these results demonstrate a causal relationship between receiving self-explanation training and better subsequent proof comprehension.

Experiment 2

Experiment 1 focused on the influence of self-explanation training on the nature of students' explanations, and on their consequent learning outcomes. In Experiment 2 we directly investigated the impact of self-explanation training on actual reading processes when participants were not required to verbalize their explanations. To this end we recorded students' eye-movements as they read proofs before and after receiving self-explanation training. As discussed in the Methodology section, eye tracking allows the nature and location of attention to be studied while dispensing with the requirement that participants articulate their self-explanations out loud (cf. Inglis & Alcock, 2012). This approach therefore put participants in conditions more similar to those they would encounter when independently studying a proof for the first time.

We addressed three main questions. First (as in Experiment 1), does self-explanation training improve students' proof comprehension? Second, does self-explanation training change the level of cognitive engagement with mathematical proofs, as measured by mean fixation durations? Third, does self-explanation training increase the extent to which students attend to the logical relationships in proofs, as measured by the number of between-line eye movements?

Because there are large individual differences in eye-movement measures (Rayner, 1998), we opted to increase statistical power by adopting a within-subjects approach. Specifically, we investigated whether self-explanation training changed individual participants' behaviors by comparing their eye movements before and after they had received either self-explanation training or a control activity. If self-explanation training is an effective method, we would expect students who had received it to engage more deeply with the text of mathematical proofs and to attend more to logical relationships. In other words, we would expect students who receive self-explanation training to subsequently show longer mean fixation durations and make more between-line saccades. A summary of the experimental design is shown in Figure 1.

Methods

Participants

Participants were 32 undergraduate students studying for mathematics degrees at Loughborough University, who took part in exchange for an £8 (about $13) stipend (none of these participants had taken part in Experiment 1). The participants were tested on an individual basis in a research lab and were randomly assigned to one of four experimental groups (eight participants per group). Because of recording problems in four cases, data from 28 participants were included in the analysis.

Materials

The self-explanation slides and control-group passage on the history of right-angled triangles were the same as those used in Experiment 1. Two proofs were used: Proof B and Proof C (given in the Appendix). For each proof, we constructed a 10-item multiple-choice proof comprehension test, again based on the Mejía-Ramos et al. (2012) framework (the test is provided in the Supplementary Materials). The order of questions was randomized for each participant, and each item was allocated 1 point, giving a maximum possible score of 10.

Procedure

Experiment 2 had three phases. In Phase 1, each participant read either Proof B or Proof C on a screen and answered the corresponding comprehension test on paper (with the proof still visible). In Phase 2, each participant read either the self-explanation materials or the passage on the history of right-angled triangles. In Phase 3, each participant read and answered questions on the proof they had not seen in Phase 1.

Participants' eye movements were recorded using a Tobii T120 eye tracker set to sample at 60Hz. This is a remote eye tracker that consists of two hidden binocular infrared cameras underneath a 17-inch TFT monitor. Stimuli are displayed on a screen that participants view (without head restriction) from approximately 60 cm away. For each participant, prior to the start of the study the eye tracker was calibrated with a 9-point display. This setup is typical for eye-movement studies (e.g., Inglis & Alcock, 2012).

Participants completed the study at their own pace; as before, they moved through the self-explanation training or history-of-triangles passage by clicking a mouse button to proceed to the next slide. The experimenter sat in the room; participants were told that he was there only to rectify any technical issues and to provide materials. Once participants had completed all parts of the experiment, which took between 20 minutes and 48 minutes, they were thanked and dismissed.

Results

We report the results of the experiment in three sections. First we consider the effect of self-explanation training on proof comprehension scores. Second, we investigate whether self-explanation training affected students' cognitive engagement. Finally, we look at the effect of self-explanation training on students' attention movements. In each case we analyzed the dependent measure from participants' reading of their second proof using an ANCOVA with two between-subject factors – group (self-explanation training, control activity) and proof order (Proof B read second, Proof C read second) – and one covariate – the dependent measure from participants' reading of the first proof. This structure allowed us to control for individual differences in eye-movement behavior while maximizing statistical power (Van Breukelen, 2006).

The Effect of Self-Explanation Training on Proof Comprehension

There was no significant difference between the mean times the two groups spent reading either of the proofs, *p*s < .1. To investigate the change self-explanation training produced in proof comprehension, we subjected the proof comprehension scores from the second reading attempt to an ANCOVA with two between-subjects factors (condition: self-explanation, control; proof read second: Proof B, Proof C), and one covariate (proof comprehension scores from the first reading attempt). The results showed a main effect of condition, $F(1,27) = 8.850$, $p = .006$, $\eta_p^2=0.247$, but no significant effect of proof order and no significant condition-by-proof-order interaction, both *F*s < 1. Those in the self-explanation condition achieved a mean comprehension score of 7.56 ($SD = 1.9$), compared to 5.56 (*SD* =

1.5) for those in the control condition. These results therefore replicate the finding from Experiment 1 that self-explanation training improves proof comprehension performance.

The Effect of Self-Explanation Training on Cognitive Effort

To investigate the effects of self-explanation training on cognitive effort, we used mean fixation durations as an index (longer fixations are associated with more effortful cognitive processing; see e.g., Duchowski, 2007; Inglis & Alcock, 2012; Just & Carpenter, 1976; Poole & Ball, 2006; Rayner, 1977). Mean fixation durations for the second reading attempt were subjected to an ANCOVA with two between-subjects factors (condition: self-explanation training, control; proof read second: Proof B, Proof C). Mean fixation durations for the proofs read first were included as a covariate because we would expect large individual differences in this measure. In other words, we would expect that a student's effort while reading the second proof would be strongly related to their effort while reading the first; and we are interested in the change in effort. The analysis revealed a significant main effect of condition, $F(1,23) = 14.234$, $p = .001$, $\eta_p^2 = .382$. Those who received self-explanation training had average mean fixation durations of 301ms (SD = 33.5) while reading their second proof; those who did not had average mean fixation durations of 276ms (SD = 30.0). There was no significant main effect of proof order and no significant condition-by-proof-order interaction (*p*s > .3).

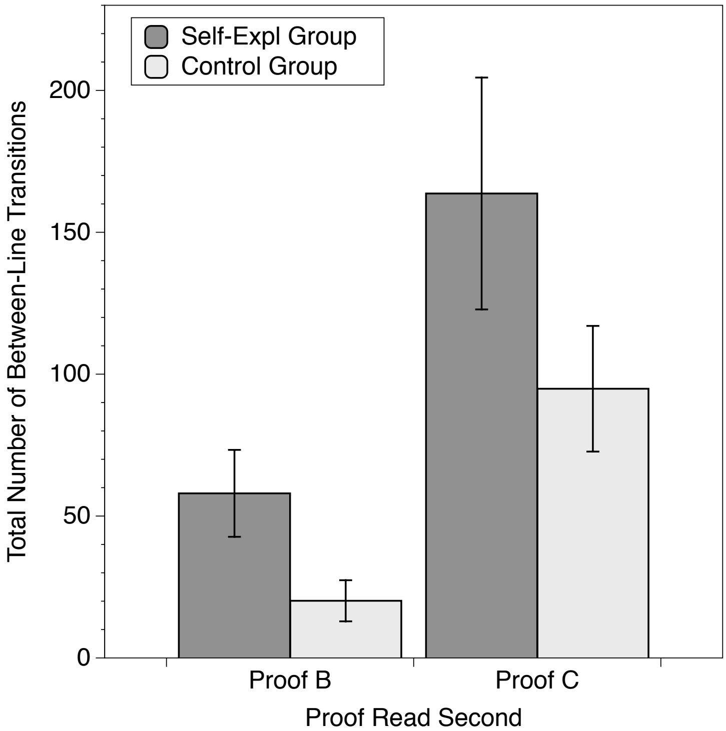
Because the only significant effect was that of condition, we can conclude that the self-explanation training caused changes in reading behavior (it was not the case, for instance, that it interacted in a more complex way with the text, causing greater changes for one proof than for the other). These findings are consistent with our prediction that self-explanation training would lead to deeper engagement with mathematical proofs.

Effect of Self-Explanation Training on Between-Line Transitions

To investigate the effects of self-explanation training on attention to logical relationships within a proof, we studied the way in which participants moved their attention around during their proof comprehension attempts. Specifically, we counted the number of *between-line saccades* in their eye-movement records; that is, the number of times they moved their attention from one line of the proof to another. An ANCOVA was conducted on the numbers of between-line saccades during the second proof reading attempt, with two between-subjects factors (condition: self-explanation training, control; proof read second: Proof B, Proof C), and two covariates (number of between-line saccades made during the first proof reading attempt, and the overall duration of the second proof reading attempt). The overall duration of the reading attempt was included as a covariate because we would expect a longer reading time to correspond to more between-line transitions, and we were interested in changes in transition behavior over and above this.

The analysis revealed a significant effect of condition, $F(1,22) = 10.394$, $p = .004$, $\eta_p^2 = 0.321$, showing that self-explanation training lead students to make more between-line transitions. It also revealed a significant effect of proof order, $F(1,22) = 8.449$, $p = .008$, $\eta_p^2 = 0.277$, indicating, perhaps unsurprisingly, that the proof itself influenced reading behavior (there were significantly more between-line saccades for Proof C than for Proof B). There was, however, no significant interaction between condition and proof order, $p = .742$, indicating that self-explanation training increased the number of between-line transitions for both proofs. Figure 4 shows the mean numbers of between-line transitions made by students for their second proof, split by condition.

Figure 4: The mean total number of between-line transitions for second proof read, split by condition. Error bars show ±1SE of the mean.



Because the total reading time for the second proof was included as a covariate, we can rule out the possibility that self-explanation training merely increases the time a student spends reading a proof (and therefore the number of between-line saccades). Instead, these findings are consistent with our prediction that self-explanation training would encourage students to search for the logical connections while reading proofs.

Experiment 3

Experiments 1 and 2 showed that self-explanation training has positive effects on both reading processes and learning outcomes, but they did so under lab conditions, and only for proofs read immediately after self-explanation training. In this final experiment, we investigated whether our short self-explanation training materials could improve proof comprehension in a genuine pedagogical setting. Because there is evidence from other contexts that self-explanation training effects can persist over time (e.g. O'Reilly, Best, &

McNamara, 2004; Rittle-Johnson, 2004), we also asked whether our training materials had lasting effects on students' proof comprehension skills.

Method

Participants

Participants were first-year undergraduate mathematics students in a calculus course at Loughborough University (none of these participants had taken part in Experiment 1 or Experiment 2). In England, students begin to specialize at the age of 16, studying only three or four subjects between 16 and 18 (for some, two of these subjects will be mathematics and further mathematics). Such students study differential and integral calculus, roughly equivalent to the material in typical Calculus I and Calculus II courses in a US college/university. The students who took part in Experiment 3 were all studying mathematics as either a single-honors or joint-honors program at university, meaning that at least half (and perhaps all) of their time was taken up with mathematics courses. The Loughborough calculus course covers advanced techniques in differentiation and integration, together with associated formal definitions and an introduction to limits. Thus, students who took part in this experiment would be roughly equivalent in their mathematical experience to US sophomore mathematics majors.

Participants took part during two normal scheduled lectures, 20 days apart. A total of 139 students took part in the first session and 122 in the other; we analyzed data only from those 107 who were present in both. Students were offered the opportunity of opting out of the study, but none decided to withdraw.

Materials

Self-explanation and control materials

The self-explanation training provided was the same in structure as that used in Experiments 1 and 2. However, the materials in this experiment were provided in a paper booklet rather than on a screen, and the example proof and the practice proof were changed in order that the self-explanation group did not spend extra lecture time studying calculus proofs. The replacement example proof and practice proof are provided in the Supplementary Materials.

The control materials were also provided in a paper booklet, which this time consisted of information on time management (for ethical reasons we wanted to ensure that both groups would spend lecture time studying material relevant to study skills for their course). The booklet asked students to provide written answers to questions on their current time management (e.g., How long do you spend working on tutorial sheets each week? How many hours of lectures do you attend each week?), to read information on how to improve their time management, and to provide written answers to final questions on how they thought they could apply the information provided.

Proof comprehension tasks

Experiment 3 used Proof B and the associated multiple-choice questions from Experiment 2, and Proof A from Experiment 1 along with a newly-constructed 10-item multiple-choice proof comprehension test, given in the Supplementary Materials. Again, the order of the comprehension test questions was randomized for each participant, and each item was allocated one point to give a maximum total score of 10 for each proof.

Procedure

The experiment took place during two scheduled lectures in the first semester of the academic year. Participants were split into a control group (54 participants) and a self-explanation group (53 participants) based on their randomly-assigned student identification numbers (even identification numbers in the control group and odd identification numbers in the self-explanation group).

In the first lecture, those in the control group studied the time management booklet while those in the self-explanation group studied the self-explanation training booklet. As a post-test, both groups then read and answered questions on Proof B. Participants worked individually in silence, and had 15 minutes for each task.

In the second lecture (twenty days later), as a delayed post-test, both groups read and answered questions on Proof A. Then participants who had previously received the self-explanation training booklet studied the time management booklet, and vice versa. Again participants worked individually in silence, and had 15 minutes for each task. The design of the experiment is summarized in Figure 1.

After the testing and data analysis was complete, students were able to access all materials, including answers to the multiple-choice questions, via the calculus course page in the University's virtual learning environment. The course page also provided links to further information on time management, and to a feedback document explaining the results of the experiment and offering advice on the implications of these results for undergraduate mathematics students.

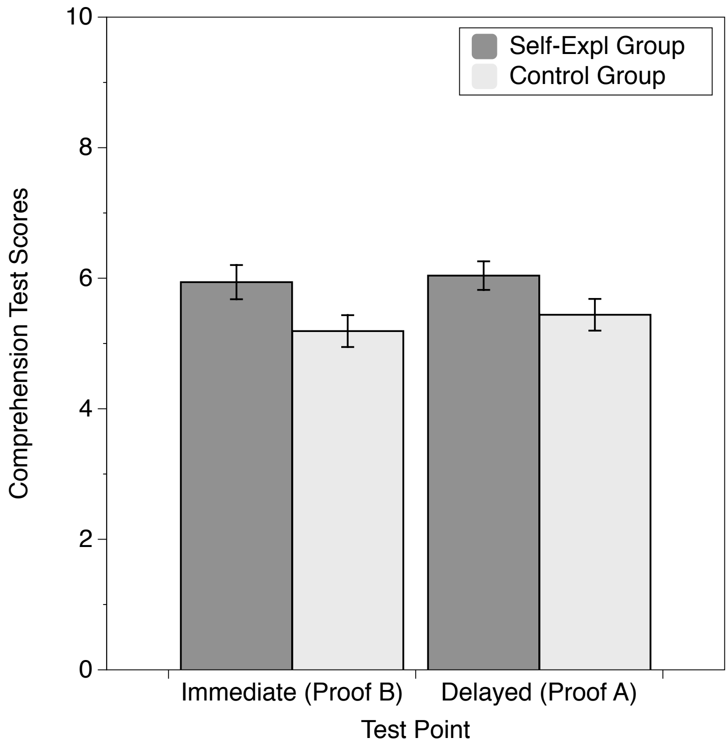
Results

As in Experiment 1, we investigated the dimensionality of the comprehension tests used in Experiment 3 (this analysis was omitted in Experiment 2 because the sample size was insufficient for a PCA to be conducted). PCAs indicated that, in both cases, a single factor should be extracted, suggesting again that it is reasonable to treat proof comprehension as a one-dimensional construct. Next we calculated the split-half reliability coefficients for each test. At .637 and .694 respectively, they were slightly lower than the coefficient found in Experiment 1, presumably because of the reduced test length (necessary due to the practical constraints of the classroom context).

The mean scores of the two groups on the proof comprehension post-test and delayed post-test are shown in Figure 4. These data were subjected to an ANOVA with one within-subjects factor (time: post-test, delayed post-test) and one between-subjects factor (condition: self-explanation, control). This analysis showed a main effect of condition, $F(1,105) = 6.024$, $p = .016$, $\eta_p^2 = 0.054$, but no significant effect of time and no interaction between condition and time, $F < 1$ in both cases. The differences corresponded to effect sizes of $d = 0.410$ at post-test and $d = 0.350$ at delayed post-test.

These results show that self-explanation training in a typical pedagogical environment improves proof comprehension significantly in the short term, and has lasting effects.

Figure 4: The mean comprehension test scores at post-test and delayed post-test, split by condition.



Discussion

Pedagogical Implications

The collective results of our three experiments show that self-explanation training improves the quality of students' reading of mathematical proofs, that it leads to increased proof comprehension, and that this effect persists over time. These results have an obvious and straightforward pedagogical implication: self-explanation training should be incorporated in the teaching and learning of mathematical proof at the undergraduate level. We also believe that both mathematics educators and instructors should be encouraged by these findings, for at least two reasons.

First, mathematics educators may note that our results are consistent with the suggestion that students fail to identify logical errors in proofs simply because they do not read very thoroughly; that students *do* have the capacity to reason correctly about mathematical arguments, and that the failures at this level are not of understanding but of execution. One could interpret this negatively (students are lazy and do not read carefully) but we believe that this would be a mistake, and that our results call for optimism. Although students apparently do not read as well as they might, they have capacity for considerable improvement without the need for extensive instruction directed at changing their beliefs about proof. Perhaps all that many students need is encouragement to believe that they can take responsibility for checking deductions and for remedying their own confusion (cf. Weber, 2009).

Second, instructors may note that self-explanation training is generic: it directs the student to explain ideas within a proof in terms of previous ideas from the theorem/proof and in terms of their own knowledge; it does not specify what form these explanations should take, and it does not include any prompts specific to a particular proof. It is also short, taking only 15-20 minutes of individual study. This contrasts favorably with pedagogical approaches in which self-explanations are prompted in the moment by an instructor, and means that its implementation can be very resource-light. Furthermore, unlike approaches that involve changing the presentation of many individual proofs (Alcock & Wilkinson, 2011; Leron, 1983; Rowland, 2001), it requires minimal instructor time, because it focuses on changing not the presentation but the engagement, by training students to interact effectively with standard instructional materials. This is not to say that proof presentation is unimportant: there are various factors that should arguably be considered in any written instructional presentation (Alcock & Inglis, 2010; Lai, Weber & Mejía-Ramos, 2012; Weber, 2012). But it does mean that there might be no need to invest large amounts of time and effort in attempting to perfect instructional resources; it might instead be more effective to teach students to function better as independent learners.

Our results have particular potential relevance for modern educational environments, in which both flexible online learning (e.g., Allen & Seaman, 2011) and peer instruction (e.g., Crouch & Mazur, 2001; Slavich & Zimbardo, 2012) are increasingly common. Under such models of education the learner has substantial responsibility for digesting written information, either alone or in collaboration with other students. There is reason to be concerned about the extent to which mathematics can be taught effectively when there is minimal student-instructor interaction – drop-out rates in online mathematics courses are higher than in comparable face-to-face mathematics courses and higher than in online courses in other disciplines (Mensch, 2010; Smith & Ferguson, 2005; Xu & Jaggers, 2011). But the pace of change is fast and the direction is not likely to alter, so it will be appropriate for all instructors and institutions to consider how best to prepare students so that they become effective independent learners. In mathematics, self-explanation training might well be a valuable part of this preparation. Future research could productively consider whether self-explanation training materials could be effective in distance learning contexts, and if so how they could best be integrated into online courses.

Implications for Research

In the Theoretical Background section, we reviewed research showing that students sometimes appear not to understand what constitutes mathematical proof and that their reading is often ineffective. We argued, however, that a careful reading of a number of studies indicates that students may fail to appropriately evaluate deductive arguments not because they lack the cognitive capacity to reason correctly, but because they do not read proofs very thoroughly. We suggested that if this were the case then considerable improvement might be possible via self-explanation training. We reviewed evidence on the efficacy of this approach in higher education and in school-level mathematics, and argued that self-explanation training is theoretically well aligned with the particular requirements of proof comprehension because the density of deductive links within a proof text mean that explanations are there to be inferred.

We set out to test this theoretical account experimentally, and our results indicate that this argument was reasonable. We saw measurable changes in reading behavior: students exposed to self-explanation training were found (in Experiment 2) to invest more cognitive effort in their reading, and to read more in the manner of mathematicians (cf. Inglis & Alcock, 2012) in the sense that they moved their attention around more during comprehension attempts. They were found (in Experiment 1) to generate higher-quality explanations: compared with those in the control condition, students exposed to self-explanation training generated approximately twice as many explanations that involved inferring implicit warrants or otherwise commenting upon the logical relationships among lines in a proof. And these behavioral changes led to desirable learning outcomes: self-explanation training led to better proof comprehension.

Further, the high effect sizes we saw for differences in comprehension scores are consistent with our suggestion that proof texts might be particularly amenable to self-explanation. This suggests a first follow-up question for future theoretically-driven investigation: Is self-explanation indeed more effective for texts that have a higher density of deductive links? Answering this would present a methodological challenge, since it is far from obvious how one might construct comparable comprehension tests for different types of material. Asking for direct recall, for instance, would not function in the same way for a proof as it might for a primarily factual text, because recall of three facts in a random order might constitute evidence of relevant knowledge, but recall of three lines of a proof in a random order would not constitute evidence of proof comprehension: a proof functions as a unified logical entity in the sense that its components get an important part of their meaning from their relationships with each other, so order should not be violated (cf. Mayans, 2004). One way around this might be to use within-subjects designs such as that used here in Experiment 2, so that the same person's scores could be compared for different tests on different texts. In conjunction with such a study, it might also be interesting to consider different prompt types in addition to self-explanation training. For instance, we might ask whether students read differently when they are told that they will be tested for recall as opposed to comprehension.

Taking the research in a different direction, one might focus less on the text and more on the student, asking whether self-explanation training has different effects for different student groups. Our investigations made only a nod in this direction, by including year group as a factor in Experiment 1. We did not find a difference in efficacy for students of different year groups, but it remains the case that students with different initial capabilities on some other measure might benefit to a greater or lesser degree from self-explanation training – we would expect to find 'good' and 'poor' students in each year group, for instance. Indeed, this makes for an interesting research question, because it is not obvious who would benefit more. Perhaps 'good' students are already good readers in the sense that they already employ self-explanation skills, so that self-explanation training would not generate much change in their behavior; in this case, we might expect that 'poor' students would see greater gains. On the other hand, perhaps 'good' students have good background knowledge and logical reasoning skills, but do not employ these so well as they might, so that self-explanation training would teach them to put their knowledge to more effective use; in this case, we might expect that 'good' students would see greater gains. Understanding such effects, in either case, would be useful both theoretically and for designers of undergraduate mathematics programs.

Finally, a broader question is that of the effects of self-explanation training on actual study strategies. Although we have shown that the training improves comprehension, it is unclear whether the training led the students in Experiment 3 to change their study strategies between the two tests. We do not know whether self-explanation effects are limited to individual proofs, or whether they might extend to students' day-to-day work with longer and more varied types of mathematical material such as textbooks and lecture notes.

We do, however, have reason to believe that mathematics educators need not be too angst-ridden about students' failures to engage effectively with proofs. Our results support the view that these failures are due not to some inherent intellectual incapacity. They indicate that undergraduate students do have at least some of the skills and understanding they need in order to read proofs effectively, and that a light-touch intervention can lead to better mobilization of these skills and thus to considerably better proof comprehension.

References

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

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

Alcock, L., & Wilkinson, N. (2011). e-Proofs: Design of a resource to support proof comprehension in mathematics. *Educational Designer*, *4*. Retrieved Mar 21 2011 from http://www.educationaldesigner.org/ed/volume1/issue4/article14/index.htm.

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective meta-cognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26,* 147-179.

Allen, I.E., & Seaman, J. (2011). *Going the distance: Online education in the United States, 2011*. Newburyport, MA: The Sloan Consortium.

Ball, L. J., Lucas, E. J., Miles, J. N. V., & Gale, A. G. (2003). Inspection times and the selection task: What do eye-movements reveal about relevance effects? *Quarterly Journal of Experimental Psychology*, *56A*, 1053-1077.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*, 221-252.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, *24*, 359–387.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

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

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

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

Crouch, C.H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69, 970-977.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36*, 1827-1837.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*, Springer, London.

Durkin, K. (2011). The self-explanation effect when learning mathematics: A meta-analysis. Presented at the *Society for Research on Educational Effectiveness 2011,* Available online [accessed 6th August 2013]: http://eric.ed.gov/?id=ED518041.

Fuller, E., Mejía-Ramos, J. P., Weber, K., Samkoff, A., Rhoads, K., Doongaji, D., & Lew, K. (2011). Comprehending Leron's structured proofs. In S. Brown, S. Larsen, K. Marrongelle, and M. Oehrtman (Eds.), *Proceedings of the 14th Conference on Research in Undergraduate Mathematics Education* (Vol. 1, pp. 84-102). Portland, Oregon.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A.H. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Research in Collegiate Mathematics III*, pp. 234–282. Providence, RI : American Mathematical Society.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.), *Second handbook of research on mathematics* *teaching and learning*, Vol. 2, pp. 805–842. Greenwich, CT: Information Age Publishing.

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

Heijltjes, A., Van Gog, T., & Paas, F. (2011). Effects of instructions on economics students' critical thinking. In *Proceedings of the Fourteenth Biennial Conference of the European Association for Research on Learning and Instruction*. Retrieved Feb 19, 2013 from http://www.earli2011.org/media/Documents_EARLI2011/BookofAbstractsandSummaries.pdf .

Houston, K. (2009). *How to think like a mathematician*. Cambridge: Cambridge University Press.

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education, 43*, 358-390.

Inglis, M., & Alcock, L. (2013). Skimming: A Response to Weber & Mejía-Ramos. *Journal for Research in Mathematics Education*, *44*, 471-474.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes, *Cognitive Psychology, 8,* 441-480.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixation to comprehension. *Psychological Review, 87*, 329-354.

Kline, P. (1999). *The handbook of psychological testing* (2nd ed). London: Routledge.

Ko, Y.-Y., & Knuth, E. (2009). Undergraduate mathematics majors' writing performance producing proofs and counter-examples about continuous functions. *Journal of Mathematical Behavior, 28,* 68-77.

Ko, Y.-Y., & Knuth, E. (2013). Validating proofs and counterexamples across content domains: Practices of importance for mathematics majors. *Journal of Mathematical Behavior*, *32*, 20-35.

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

Leinhardt, G. (1993). Weaving instructional explanations in history. *British Journal of Educational Psychology, 63,* 46-74.

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

Lithner, J. (2003). Students' mathematical reasoning in university textbook exercises. *Educational Studies in Mathematics*, *52*, 29-55.

Lorenzo-Seva, U., & Ferrando, P.J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, Instruments and Computers*, *38*, 88-91.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81*, 899–917.

Matthews, P., & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of Experimental Child Psychology, 104,* 1-21.

Mayans, R. (2004). The future of mathematical text: A proposal for a new internet hypertext for mathematics. *Journal of Digital Information*, *5*(1). Retrieved Feb 15 2013 from journals.tdl.org/jodi/index.php/jodi/article/viewArticle/128/126.

Mensch, S. (2010). Issues in offering numeric based courses in an online environment. *Journal of Instructional Pedagogies, 3*, Retrieved December 5, 2012: http://www.aabri.com/manuscripts/09405.pdf

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1-43.

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.

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

O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-explanation reading training: Effects for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), Proceedings of the 26th Annual Meeting of the Cognitive Science Society (pp. 1053–1058). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. In C. Ghaoui (Ed.), *Encyclopedia of Human-Computer Interactions* (pp. 211-219). Hershey, PA: IGI Publishing.

Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica, 7,* 5-41.

Rayner, K. (1977) Visual attention in reading: Eye movements reflect cognitive processes, *Memory & Cognition, 5,* 443-448.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62,* 1457-1506.

Recio, A., & Godino, J. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics*, *48*, 83-99.

Reid, D. A., & Knipping, C. (2011). *Proof in mathematics education: Research, learning and teaching*. Rotterdam, The Netherlands: Sense Publishers.

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction, 12,* 529-556.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77,* 1-15.

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

Rowland, T. (2001b). Generic proofs: Setting a good example. *Mathematics Teaching, 177*, 40-43.

Roy, M., & Chi, M. T. H. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning,* (pp. 271–286). Cambridge, England: Cambridge University Press.

Roy, S., Alcock, L., & Inglis, M. (2010). Undergraduates' proof comprehension: A comparative study of three forms of proof presentation. In *Proceedings of the 13th Conference on Research in Undergraduate Mathematics Education*, Raleigh, NC, February, 2010.

Segal, J. (2000). Learning about mathematical proof: conviction and validity. *Journal of Mathematical Behavior*, *18*, 191–210.

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,* 4–36.

Shepherd, M. D. (2005). Encouraging students to read mathematics. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *15*, 124-144.

Shepherd, M. D., Selden, A., and Selden, J. (2012). University students' reading of their first-year mathematics textbooks. *Mathematical Thinking and Learning*, *14*, 226-256.

Slavich, G.M., & Zimbardo, P.G. (2012). Transformational teaching: Theoretical underpinnings, basic principles, and core methods. *Educational Psychology Review*, *24*, 569-608.

Smith, G. G., & Ferguson, D. (2005). Student attrition in mathematics e-learning. *Australasian Journal of Educational Technology, 21*, 323-334.

Solow, D. (2005). *How to read and do proofs (4th Edition)*. Hoboken, NJ: John Wiley & Sons, Inc.

Stewart, I., & Tall, D. (1977). *The foundations of mathematics*. Oxford: Oxford University Press.

Stylianides, G. J., & Stylianides, A. J. (2009). Facilitating the transition from empirical arguments to proof. *Journal for Research in Mathematics Education*, 314-352.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16,* 209-220.

Toulmin, S. (1958). *The uses of argument*. Cambridge: CUP.

Van Breukelen, G. J. P. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology, 59,* 920-925.

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,* 115–133.

Weber, K. (2009). Mathematics majors' evaluation of mathematical arguments and their conception of proof. In *Proceedings of the Twelfth sigmaa on rume Conference on Research in Undergraduate Mathematics Education*. Retrieved Feb 19, 2013 from http://mathed.asu.edu/crume2009/Weber1_LONG.pdf

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 practices with respect to proof. *International Journal of Mathematics Education in Science and Technology, 43,* 463-482.

Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review, 22,* 393-409.

Wong, R. M. F., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction 12,* 233-262.

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

Xu, D., & Jaggars, S. S. (2011). The effectiveness of distance education across Virginia's community colleges: Evidence from introductory college-level math and English courses. *Educational Evaluation and Policy Analysis, 33*, 360-377.

Ziegler, E., & Stern, E. (2011). The impact of self-explaining and example generation on learning similar concepts in mathematics. In *Proceedings of the Fourteenth Biennial Conference of the European Association for Research on Learning and Instruction*. Retrieved Feb 19, 2013 from http://www.earli2011.org/media/Documents_EARLI2011/BookofAbstractsandSummaries.pdf.

Appendices

Self-explanation training slides

Slide 1

Self-Explanation

- The “self-explanation” strategy has been found to enhance problem solving and comprehension in learners.
- To improve your understanding of a proof, there are a series of techniques you should apply after reading each line:
    - Try to identify and elaborate the main ideas of the proof.
    - Attempt to explain each line presented to you in terms of previous ideas, by saying aloud any self-explanations you make. These may be ideas from previous theorems/proofs or ideas from your own knowledge of the topic area.
    - You should raise any questions that may arise when presented with new information that may contradict your current understanding.

Slide 2

- Before proceeding to the next line of the proof you should ask yourself the following:
    - Do I understand the ideas used in that line?
    - Do I understand why that idea has been used?
    - How does this idea link to other ideas in the proof/other theorems/prior knowledge that I may have?
    - Does the self-explanation I have generated help to answer the questions that I am asking?
- On the following slide, you will find an example of possible self-explanations generated by students when trying to understand a proof presented to them.

- Please read the example carefully in order to help you understand how to use this strategy in your own learning.

Slide 3

An Example

Theorem: 3.14 is a rational number

Proof: To check that $x = 3.14$ is a rational number, it is enough to find two integers p and q with

$q \neq 0$ such that $x = {}^{p}/_{q}$.

Choose $p = 314$ and $q = 100$.

Both 314 and 100 are integers and $100 \neq 0$.

Thus, by definition, 3.14 is a rational number. □

- After reading this proof, one student made the following self-explanations:
  - To prove something is rational, we need to use the definition of rational numbers, which is used in the proof.
  - $q \neq 0$ because otherwise x would not exist.
  - $p = 314$ and $q = 100$ because ${}^{314}/_{100} = 3.14$.
  - Since p and q satisfy the definition of an integer, ${}^{314}/_{100} = 3.14$ satisfies the definition of a rational number.
  - 3.14 therefore must be a rational number.

Slide 4

- You must be aware that the self-explanation strategy is not the same as monitoring or paraphrasing. These two methods will not help your learning to the same extent as self-explanation.

Paraphrasing

"Both p and q have to be positive or negative whole numbers"

- There is no self-explanation in this statement. No additional information is added or linked. The student merely uses different words to describe the word "integer". You should avoid using such paraphrasing during your own text comprehension. Paraphrasing will not help your understanding of the text as much as self-explanation will.

Slide 5

Monitoring

"OK, I understand the proof sets $p = 314$ and $q = 100$."

- This statement simply shows the student's thought process. It is not the same as self-explanation where the student relates the sentence to additional information in the text or prior knowledge. Please concentrate on self-explanation rather than monitoring.
- A possible self-explanation of the same sentence would be:
  "OK, $p = 314$ and $q = 100$ because these are integers and $^{p}/_{q} = x = 3.14$. The proof could also have used $p = 628$ and $q = 200$."
- In this example the student identifies and elaborates the main ideas in the text. They use information that has already been presented to them to help with their understanding of how the proof is logically connected.
- This is the approach you should take after reading every line of a proof in order to improve your understanding of the material.

Slide 6

Practice

- Please now read this short proof and self-explain each line using the training you have been given.

Theorem: $(0,\infty)$ is not bounded.

Proof: Assume that the theorem is false and that $(0,\infty)$ is bounded.

Therefore, by assumption, there exists a constant $C > 0$ such that $(0,\infty) \subset [-C,C]$.

Note, $C + 1 > 0$ thus $C + 1 \in [-C,C]$.

This contradicts the assumption that $(0,\infty) \subset [-C,C]$.

Thus $(0,\infty)$ is not bounded.

Proofs

Proof A

Theorem: There are infinitely many triadic primes.

Proof: Consider a product of two monadic numbers: $(4j + 1)\ (4k + 1)$

$$= 4j \cdot 4k + 4j + 4k + 1$$

$$= 4(4jk + j + k) + 1$$

which is again monadic.

Similarly, the product of any number of monadic numbers is monadic.

Now, assume the theorem is false, so there are only finitely many triadic primes, say $p_1, p_2, \ldots, p_n$.

Let $M = 4p_2 \ldots p_n + 3$, where $p_1 = 3$.

$p_2, p_3, \ldots, p_n$ do not divide M as they leave a remainder of 3, and 3 does not divide M as it does not divide $4p_2, \ldots, p_n$.

We conclude that no triadic prime divides M.

Also, 2 does not divide M since M is odd.

Thus all of M's prime factors are monadic, hence M itself must be monadic.

But M is clearly triadic, a contradiction. □

Proof B

Theorem: If $n \in \mathbb{Z}$ and $n > 0$, then n is even if and only if $3n^2 + 8$ is even.

Proof: Let $n \in \mathbb{Z}$ and $n > 0$.

By definition, if n is even then $\exists k \in \mathbb{Z}$ such that $n = 2k$.

Then, $3n^2 + 8 = 3(2k)^2 + 8 = 12k^2 + 8 = 2(6k^2 + 4)$.

Therefore, $3n^2 + 8$ is even.

Now, assume n is odd and we will show that $3n^2 + 8$ is odd.

By definition, if n is odd $\exists j \in \mathbb{Z}$ such that $n = 2j + 1$.

Then, $3n^2 + 8 = 3(2j + 1)^2 + 8 = 3(4j^2 + 4a + 1) + 8 = 2(6j^2 + 6j + 5) + 1$.

Therefore, n is even. □

Proof C

Theorem: If p is prime and $n \in \mathbb{Z}$ and p is a divisor of $(4n^2 + 1)$, then $p \equiv 1(\text{mod } 4)$.

Proof: Clearly, p cannot be 2, so we need only show that $p \not\equiv 3(\text{mod } 4)$.

Suppose $p = 4k + 3$ for some $k \in \mathbb{Z}$.

Let $y = 2n$.

Then, by Fermat's Little Theorem, $y^{p\text{-}1} \equiv 1(\text{mod } p)$.

But $y^2 + 1 \equiv 0(\text{mod } p)$.

So, $y^{p\text{-}1} \equiv y^{4k+2} \equiv (y^2)^{2k+1} \equiv (\text{-}1)(\text{mod } p)$.

But this cannot be the case.

Therefore, $p \equiv 1(\text{mod } 4)$. □

Comprehension Questions for Proof A used in Experiment 1

1. Using the method of the proof you have been working with, what would be an appropriate value for *M* if you were writing proof for the theorem that there are infinitely many primes of the form $6k + 5$?
2. Could $M = 87$, where *M* is defined as in this proof, if there were only 2 triadic primes? If yes, state the values of these 2 triadic primes. If no, explain why.
3. In line 3, what is the purpose of assuming that the theorem is false and that there are only finitely many triadic primes?
4. Why does the proof include the sub-proof that the product of monadic numbers is monadic?
5. Which claim(s) in the proof logically depend on line 2 of the proof?

6. What does it mean for a number to be triadic?

7. What does it mean for a number to be a prime?

8. In line 5, why does the fact that 3 does not divide $4p_2, \ldots, p_n$ imply that 3 does not divide *M*?

9. Is the product of two triadic numbers triadic? Why, therefore, would this prevent the methods used in the proof you have been working with from being used to prove there are infinitely many monadic primes?

10. If 3, 7, 11 and 19 were the only triadic primes, what would the value of *M* be?

11. Which of the following summaries best capture the ideas of the proof? (Please circle the letter of your choice):

    a. The proof assumes there are infinitely many triadic primes and uses them to construct a triadic number *M* that has only monadic prime factors, which would imply *M* is also monadic. *M* cannot be monadic as *M* is triadic.

    b. The proof lets $M = 4p_2, \ldots, p_n + 3$, where p_i are prime numbers and p_i does not equal 3. Thus, 2 does not divide *M* because *M* is odd. Further, p_i does not divide *M* because it leaves a remainder of 3.

    c. The proof introduces monadic primes to be used later on in the proof. It lets $M = 4p_2, \ldots, p_n + 3$ and shows 2 does not divide *M*, since 2 is even and *M* is odd. However, this would not itself create an infinite triadic prime so the proof uses monadic primes to create an infinite triadic prime.

12. Summarize in your own words how the proof arrives at the conclusion that *M* itself must be monadic.

13. Do lines 3 & 7, which establish that *M* is not divisible by a triadic prime, depend on the statements made in lines 1 & 2, which establish that the product of monadic primes is monadic? Explain your answer.

14. What type of proof is this?