Improving proof comprehension in undergraduate mathematics

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: [https://dspace.lboro.ac.uk/2134/14964](https://dspace.lboro.ac.uk/2134/14964)

Publisher: © Mark Hodds

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
Improving proof comprehension in undergraduate mathematics

by

Mark Hodds

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of
Doctor of Philosophy of Loughborough University

in the
Mathematics Education Centre
School of Science

June 20, 2014

© by Mark Hodds, 2013
Declaration of Authorship

I, Mark Hodds, declare that this thesis titled, “Improving proof comprehension in undergraduate mathematics” and the work presented in it are my own. I confirm that:

■ This work was done wholly while in candidature for a research degree at this University.

■ Neither the thesis nor the original work contained therein has been submitted to this or any other institution for a degree.

■ Where I have consulted the published work of others, this is always clearly attributed.

■ Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

■ I have acknowledged all main sources of help.

Signed: ______________________________

Date: ________________________________
“Proof is a tool in service of research, not a shackle on the mathematician’s imagination”

Ruben Hersh, 1997
When studying for a mathematics degree, it has been shown students have great difficulty working with proof (Moore, 1994). Yet, to date, there has been surprisingly little research into how we could improve the way students study mathematical proofs. Furthermore, there is relatively less research on students’ proof comprehension skills when compared with that of their proof construction skills (Mejía-Ramos and Inglis, 2009).

The aim of this thesis was therefore to build upon the existing proof comprehension literature to determine methods of improving undergraduate proof comprehension. Previously, text based manipulations (e.g. Alcock, 2009; Leron, 1983a; Rowland, 2002) have been tested as a way of improving proof comprehension but these have often not been as successful as we would have liked. However, an alternative method, called self-explanation training, has been shown to be successful at improving comprehension of texts in other fields (Ainsworth and Burcham, 2007; Chi et al., 1989; Rittle-Johnson, 2006; Wong et al., 2002).

This thesis reports three studies that investigate the effects of self-explanation training on proof comprehension. The first study confirmed the findings of previous self-explanation training research in other fields. Students in the study who received the self-explanation training showed a significantly greater understanding of the proof text compared to that of a control group.

Study 2 used eye-tracking analysis to show that self-explanation training actually changed the way students in the study read proofs; they concentrated harder on the proof (as measured by mean fixation durations), and made more between-line transitions.

The final study revealed that self-explanation training can be implemented into a genuine pedagogical setting with relative ease and also showed the positive effects on proof comprehension last for a longer term of three weeks.
From the findings of the research reported in this thesis it can be concluded that many students who participated in these studies appeared to have the knowledge required to understand proofs, it is perhaps they just needed some guidance on how to apply their knowledge. Self-explanation training appears to do this as it significantly improved proof comprehension in the short-term as well as offering longer-term benefits. More research will be needed to confirm these findings, given that the studies here involved participants from only one UK university on what would be considered as typical mathematics degree courses for the UK. However, these findings are promising and provide the foundation for improvements in undergraduate proof comprehension.
Acknowledgements

Whilst working towards completing this thesis, there are many people I have to thank and acknowledge for helping me - there is not enough space to thank you all! Firstly, I want to thank my supervisors, Dr Lara Alcock and Dr Matthew Inglis, for providing all of the support, advice and guidance I needed to produce this thesis. They were there for me whenever I needed anything and I thank them for all of the time, energy and effort they have put in to listening to my ideas, reading my work and pushing me to produce the best thesis I could.

Secondly, I want to thank everyone who has been a part of my experience of living in John Phillips Hall at Loughborough University. I often needed a way to let off steam and have some relaxation time and those were the people who helped me do that. I had a fantastic time on the hall committee whilst completing my thesis, both as media secretary and as the hall chairman, and all of the committees I have been on have been a pleasure to work with. Without JP, I may have been reduced to madness! I also want to thank my fellow subwardens and the warden, Dr Paul Leaney, in John Phillips Hall and Hazlerigg-Rutland Hall with whom I have shared the responsibility of looking after over 500 students on a weekly basis since 2012. We have shared some great times together and I have gained so much life experience and many practical abilities from taking up the role.

I of course want to thank my fellow PhD students in the Mathematics Education Centre for providing further support and an ear to bend when things needed to be discussed. We all have had to undergo the same process and to be able to share our issues, ideas and problems with students in the same situation is a blessing.

Several people in particular need to be acknowledged for always being there when I needed them. Katie Livermore, who has been there as a best friend ever since our East Norfolk Sixth Form College days, has provided opportunities to leave Loughborough on the odd weekend for a visit and was always available for a good chat. Also Tasha Kirkup, who has really helped me through a difficult time in 2013 and has become one of my closest friends, and Mike Hurst, with whom we have dominated university bowling! Jon Garrett has also been brilliant and spent the time going through my work, offering a fresh pair of eyes and a new perspective on the work.

I have left the most important people in my life until last and they are of course my family. Without them, I would be nothing and I wouldn’t be the person I am today.
They have always pushed me to be the best that I can be and believed in the paths I have chosen to take. Although I have enjoyed being in Loughborough and taking part in everything that is on offer there, home will always be a special place with truly special people that mean the world to me. Loughborough also doesn’t have a beach...
Contents

Declaration of Authorship i
Abstract iii
Acknowledgements v
List of Figures ix

1 Introduction 1
1.1 Improving proof comprehension 5
1.2 Theoretical and methodological framework 6
1.3 The significance of the research conducted 7
1.4 The structure of this thesis 8

2 Literature Review 9
2.1 What is proof? 10
2.2 Student beliefs about proof 14
2.3 Proof comprehension and proof construction 17
2.4 Logic, argumentation and proof 18
2.5 Assessing proof comprehension and the reading processes of students in proof comprehension tasks 26
  2.5.1 Frameworks for assessing undergraduate proof comprehension 27
  2.5.2 Eye-tracking and reading comprehension 30
2.6 Methods of improving proof comprehension 35
2.7 Summary 50

3 Methodology 53
3.1 The philosophy of science and experimental design 53
3.2 Quantitative methodologies 56
3.3 Comprehension test design 62
3.4 Eye-tracking methodologies 64
3.5 Ethics 68
3.6 Studies reported in this thesis 71

4 Study 1: Using self-explanation training to improve proof comprehension 73
4.1 Methods 73
4.2 Results 87
4.3 Discussion 90
List of Figures

1.1 A diagram of the logical steps in a general, simple argument according to an adaptation of the Toulmin scheme by Krummheuer (1995) . . . . . . . 3
1.2 A simplified diagram of the logical steps in a proof using the adapted Toulminian scheme, according to Aberdein (2005) . . . . . . . . . . . . . 3

2.1 Toulmin’s scheme of argumentation . . . . . . . . . . . . . . . . . . . . . 20
2.2 Toulmin’s example of an argument . . . . . . . . . . . . . . . . . . . . . 20
2.3 Krummheuer (1995) adapted version of Toulmin’s scheme of argumentation 20
2.4 Example proof used by Aberdein (2005) to show that the Toulmin scheme is too simplistic (Toulmin, Richard and Allan, 1979, Fig. 7.4, p.89) . . . . 21
2.5 A model of mathematical argumentation according to Aberdein (2005), adapted and simplified for the purposes of this thesis . . . . . . . . . . . . . . 22
2.6 A visualisation of the model of Reading Comprehension of Geometry Proof (RCGP) suggested by Yang and Lin (2007) . . . . . . . . . . . . . . 28
2.7 Leron (1983b, p.175) comparison of linear proofs and structured proofs . . 37

4.1 Study 1: Diagram taken from Hodds, Alcock and Inglis (2014), Fig. 1, p.74 74
4.2 Study 1: The average score for participants in each group by year of study 88
4.3 Study 1: The median number of verbal “explanations” and “non-explanations” for each group . . . . . . . . . . . . . . . . . . . . . . . . 89

5.1 Study 2: Diagram taken from Hodds et al. (2014), Fig. 1, p.74 . . . . . . 96
5.2 Study 2: Areas of interest for Proof B from the eye-tracking program . . . 103
5.3 Study 2: Areas of interest for Proof C from the eye-tracking program . . . 104
5.4 Study 2: The mean total number of between-line transitions for the proof read second, split by condition. Error bars show ±1SE of the mean . . . . 107

6.1 Study 3: Diagram taken from Hodds et al. (2014), Fig. 1, p.75 . . . . . . 111
6.2 Study 3: Mean comprehension test scores by condition for post-test and delayed post-test . . . . . . . . . . . . . . . . . . . . . . . . 114
Dedicated to the Hodds and Middleton families
Chapter 1

Introduction

Students struggle with mathematical proofs. Not only do they find it difficult to construct proofs but it has also been shown they have difficulty in understanding them, often because they misunderstand a theorem or concept and then, consequently, misapply it (Weber, 2001). It is important for students to obtain the skills required to understand mathematical arguments since such arguments play a central role in explaining the mathematics that helps us to better understand the world we live in (Chazan, 1993). Furthermore, the skills required to understand mathematical proofs help students to “...think more clearly and effectively about mathematics” (Hanna, 2000, p.7). It is therefore important to conduct research in the field of mathematics education that aims to improve the way proofs are taught, constructed and understood.

The research conducted on proof activities in the field of mathematics education has tended to focus more on proof construction tasks rather than proof comprehension tasks (Mejía-Ramos and Inglis, 2009) and yet one might argue that in order for one to be able to construct mathematical proofs, one needs to understand existing proofs and how these proofs are constructed first. For example, one needs to know the different types of proof one could construct and also understand how, and when, these different types of proof should be used to prove a theorem. It is therefore somewhat surprising that there is comparatively less research into proof comprehension than proof construction. Furthermore, much of the published work offers theoretical analyses only, leading to a rather large gap in our empirical understanding of the subject.

It is possible that the majority of research carried out in the field of proof comprehension is not supported by empirical evidence because assessing the proof comprehension skills
of students effectively and accurately could be considered difficult. Some courses that involve the teaching and learning of proof do not directly ask comprehension questions and instead ask students to construct the proof of certain theorems (Weber, 2001). These proofs can of course be rote memorised, leading to little or no understanding. Recently however, several papers have suggested ways in which proof comprehension can be assessed or measured in a more valid fashion. Conradie and Frith (2000) first suggested a method of testing proof comprehension which was then adapted by Yang and Lin (2008) who suggested a model of reading comprehension of geometry proof (RCGP) for high school students. Later, Mejía-Ramos, Weber, Rhoads, Samkoff, Search and Fuller (2012) suggested a method of assessment of undergraduate proof comprehension, based upon the work of Yang and Lin (2008), using seven underlying factors, or “dimensions”. The authors claimed that by using the seven dimensions suggested, one could create questions that tested a student’s overall level of proof comprehension. These questions would find any areas of weakness a student had that would need improving upon to increase their overall understanding, since an overall understanding would require a combination of the seven different skills. If this claim were true, it would perhaps then enable tutors to focus on improving the different skills of individual students in order to increase their understanding.

One such skill that is required in order to understand proofs is the ability to understand the underlying logic of the proof. Indeed, it has been argued that students with poor logical ability appear to have a lack of understanding of proof (Selden and Selden, 1995). In order to investigate and discuss the process of understanding the underlying logic of a proof, it is useful to use the Toulmin (1958) model of argumentation (Aberdeen, 2005; 2006; 2009). In the simple scheme suggested by Krummheuer (1995), which is an adaptation of the scheme suggested by Toulmin, an argument begins with some data (D) and ends with a conclusion (C). Connecting the data with the conclusion are some warrants (W). These warrants are the logical connections that give justification for the conclusions being valid, given the data provided. For a general argument, a simplified diagram of the logical steps in the argument looks like Figure 1.1.

However, for a proof, there is usually more than one step. One cannot usually go from a theorem premise to a conclusion in the way described in Figure 1.1 (there are of course some exceptions). A proof will usually have several conclusions that explain each step on the path to the overall conclusion. The conclusion of one step is the data for the next step and so on. Figure 1.2 shows a diagram of this process that was first suggested by Aberdeen (2005).
Again, this is just a simplified version of the full model, which will be discussed and critiqued later in the literature review. However it is still clear from these simplified models that understanding the logical argumentation of a proof is a skill required of learners in order to fully understand a proof.

Warrants may be given explicitly in the proof text, defined to be an “explicit warrant” in this thesis. Alternatively, a warrant may be left out for the reader to infer for themselves. The warrant intended by the author remains within the logical reasoning of the proof but does not appear explicitly in the proof text and is thus defined as an “implicit warrant” in this thesis. Therefore if one can understand the explicit warrants and also infer the implicit warrants in a proof, it is likely one will be able to understand the proof and also make correct validity judgements about it. Using these ideas, Inglis and Alcock (2012) suggested that students need to undertake the following three-step process in order to validate and understand a proof. Firstly, the reader must determine when a warrant is required. Clearly, some lines within a proof do not require a warrant (for example,
setting a parameter to be a defined value) so it is important for a reader to distinguish between lines that require a warrant from those that do not. Secondly, the reader must infer the warrant required. The warrant inferred should be the warrant intended by the author of the proof but there is a possibility the reader may infer a different warrant. In order to understand the proof, one therefore needs to understand the warrant that was intended by the author. Finally, the reader must judge the validity of the warrant. For the purposes of this thesis, the proofs used are all considered to be valid. Thus the research reported investigates ways to improve students’ abilities to complete the first two steps.

Understanding warrants is clearly not the only aspect required for good proof comprehension but previous attempts at improving proof comprehension have often tried to make it easier for learners to understand warrants. These attempts have often focused on changing the way a proof is presented. For example, it has been suggested a proof could be split into “modules”, each with their own “sub-levels” explaining the purpose of including that particular module within the proof (Leron, 1983a). Other suggestions include generic proofs (Rowland, 2002), which explain the proof using a carefully chosen generic example, and electronic proofs with added voice and textual explanations known as e-Proofs (Alcock and Wilkinson, 2011; Roy, Alcock and Inglis, 2010). By changing the presentation of the text, the researchers hoped to make the proofs easier to understand and remove some of the difficulties students encounter when working with proofs. However, these ideas have often been somewhat unsuccessful.

An alternative method of improving proof comprehension would be to change the type of engagement a student has with a proof. This method considers what students are able to do when working with a proof and helps guide them to use the knowledge they have in a useful way. One simple method of changing the engagement would be to change the way a student reads a proof so that they can better understand the logical links and argumentation between lines and sections of the proof. Indeed, it has been shown that students do not read proofs in the same way that experienced mathematicians do (Inglis and Alcock, 2012), perhaps providing an explanation for some students’ lack of understanding. One suggestion from outside the field of proof that seeks to change students engagement with texts is to use a method called self-explanation training. Self-explanation training teaches a student to use their existing knowledge to link new ideas presented in a text. Previous studies conducted in other fields such as biology (Ainsworth and Burcham, 2007; Chi, de Leeuw, Chiu and LaVancher, 1994; McNamara, 2001), physics (Chi, Bassok, Lewis, Reimann and Glaser, 1989), history (Leinhardt, 1993) and mathematics (Wong, Lawson and Keeves, 2002) have all shown promising
results. Hence there was good reason to believe that self-explanation training could also improve the proof comprehension skills of undergraduate students, providing motivation to conduct the research reported.

1.1 Improving proof comprehension

The aim of the research conducted in this thesis is to investigate how we might improve proof comprehension at the undergraduate level and to provide some empirical evidence for, or indeed against, some of the theoretical arguments contained in the literature. There are three studies reported in this thesis, each using mainly quantitative methods. These three studies look at the claim that self-explanation training can improve undergraduate proof comprehension, with each study building upon the results of the previous one.

The first study focuses on the suggestion that changing students’ engagement with proofs by providing self-explanation training to them may be a useful method for improving proof comprehension. As stated previously, the majority of studies investigating the effects of self-explanation training on text comprehension in other fields have shown promising results. For example, students in physics and biology have been shown to improve their understanding of a text with prompts by the teachers to self-explain (Chi et al., 1989) but also without prompts (Chi et al., 1994). Furthermore, after receiving a self-explanation training booklet, students have been shown to produce higher quality verbal explanations when asked to discuss their understanding of a text (Ainsworth and Burcham, 2007). Wong et al. (2002) even showed that self-explanation training also works with texts on the topic of geometry, albeit with high-school students. Hence, conducting a study using self-explanation training seemed sensible as a way of investigating a potential method for improving proof comprehension. However, proof texts are very different to the texts provided in other fields. Proof texts can be considered special in terms of their structure and the logical understanding required to obtain comprehension. The special nature of proof texts will be discussed further later; however it is clear that one could not automatically assume that the same results reported in other fields would appear in a study with proof texts. Despite this, the results of the first study confirmed the previous findings in other fields and show that comprehension of a proof text is significantly improved through the use of self-explanation training.
Study 2 builds upon the work of Study 1 and looks at how self-explanation changes students’ engagement with proof texts. The study uses eye-tracking technology and eye-tracking methodologies to see how self-explanation training changes the way students read proof texts. By asking each student in the study to read two proof texts (one before being given self-explanation training or a control activity and one after), the analysis revealed significant differences between the groups and also changes in individual reading patterns, showing that self-explanation training actually changes how students read mathematical proofs.

Study 3 investigates the effects of self-explanation training in a genuine pedagogical setting, both in the short and longer term. The study took place over two sessions, roughly three weeks apart, with first-year students undertaking a calculus course. In session one, half of the students were given a self-explanation training booklet whilst the other half received a control activity on their time management skills. All students then received a proof followed by related comprehension questions. In session two, all students were given a different proof followed by another comprehension test specific to that particular proof. By comparing the comprehension scores of the two groups, the effects of self-explanation training in the longer term were determined. Once again, the results show self-explanation training significantly improves proof comprehension but importantly over the longer-term and in a genuine pedagogical setting.

1.2 Theoretical and methodological framework

In order to meet the goals and answer the questions set by this thesis, the research reported uses several theoretical suggestions and adopts a mainly quantitative methodology. The studies reported investigated a proposition suggested in the general reading literature regarding self-explanations, which have been tested in other fields: that self-explanation training can improve one’s understanding of a proof text. It is this proposition that is used to guide the research conducted for this thesis.

Each of the studies reported contain a form of test that is administered to the students. Their responses to the questions in these tests are scored and then subjected to statistical analyses. These analyses then provide empirical evidence for, or indeed against, the theory that is being tested. In Study 1, there is also an added qualitative element whereby the phrases spoken by the students in the study are classified into seven different groups and analysed. However, these seven groups are also subjected to statistical
analyses to determine which groups of phrases are used by the students in both of the experimental conditions.

1.3 The significance of the research conducted

The studies described above provide a detailed investigation of one approach to answering the question “How can we improve the understanding of mathematical proof at the undergraduate level?” By conducting research that aims to answer this question, the investigations reported in this thesis are significant for the following reasons.

Firstly, as mentioned previously, there is a lack of research conducted in proof comprehension in comparison to proof construction (Mejía-Ramos and Inglis, 2009). Therefore, the research conducted for this thesis will add to the growing body of research on undergraduate proof comprehension. Indeed, this thesis reports original research that tests a method specifically designed to improve proof comprehension: self-explanation training. Self-explanation training has been shown to be successful in other fields but has never previously been used to train students to improve their proof comprehension. Therefore the research reported takes a previously successful pedagogical method, which requires little or no change in the way students are taught mathematical proofs, and tests it in a new field for the first time.

The research undertaken also aims to create links with other fields outside of mathematics education. By focusing on improving proof comprehension through consideration of different aspects of proof, such as the proof text itself and the way in which proofs are read, the results of these studies can be used by researchers in other fields to investigate ways of improving their students’ comprehension of texts in their subjects. For example, computer science uses a lot of logic-based texts, similar to proofs, and modern foreign languages may contain unfamiliar symbols in a new language, similar to the way mathematics uses Greek letters and symbols. Although it is true to say that proof texts can be considered special in terms of their structure, content, and layout, the basic reading processes required to begin to understand them are also used to understand texts in other subjects. Hence, ideas that work for proof texts could work for general texts and vice-versa. The research reported could therefore have implications not only for mathematical proof, but for mathematics and other fields more generally.
1.4 The structure of this thesis

The thesis begins with a detailed literature review that can be broadly split into three main categories. The first category discusses the philosophical and cognitive issues faced when studying proofs. There is a review of the beliefs of philosophers, educators and students on what a proof actually is and an in-depth discussion on the knowledge and skills required to successfully teach and understand proofs. The second category discusses the methods for measuring students’ proof comprehension, which is crucial for assessing students’ proof comprehension skills and analysing the results of the studies reported in this thesis and beyond. The final category discusses the theoretical and empirical studies that have already investigated methods of improving undergraduate proof comprehension.

After the conclusion of the literature review there is a discussion of the general methodology used in all of the studies reported. The methodology chapter considers the use of quantitative and qualitative methods to conduct research into improving proof comprehension, the ethics behind conducting the studies reported and other general methodological issues, such as different study designs available. There is also a discussion of the specific methodology relating to the use of eye-trackers and eye-tracking methodology, which features heavily in Study 2.

Following the methodology, the three studies are reported. Each study has three sections: methods, where the methods specific to each study are addressed, results, where the statistical results are reported, and discussion, providing some initial offerings on the meaning of the results for undergraduate proof comprehension.

The thesis concludes by providing a general summary of the research conducted with some of the main limitations of the studies. Some specific directions for possible future study are suggested before a final discussion of the implications of the studies reported for proof comprehension and mathematics education more generally.
Chapter 2

Literature Review

The literature review in this thesis is split into several sections discussing different points that come under three main areas of discussion.

The first area explored is the philosophical question of what a proof actually is and what may be required in order to understand a proof. There are many different ideas and suggestions provided by the literature on what a mathematical proof is and these are addressed. Furthermore, philosophers and mathematicians have differing views on what constitutes “formal” and “informal” proofs; a distinction that is made clear for the reader in order to avoid any confusion later on in the thesis. Also discussed within this first area are some of the skills required, and knowledge needed, to understand proofs. This begins with a comprehensive review of the role logic and argumentation plays within mathematical proofs, focusing mainly on the work of Toulmin (1958). Toulmin suggested a model of argumentation that described the processes undertaken when one is trying to convince another in an argument that something is true. The model has been cited extensively throughout the mathematics education literature and has even been adapted to describe the logical steps used in a mathematical argument by Aberdein (2005).

The second area explored is methods of measuring and investigating undergraduate proof comprehension. Clearly, if one could find an accurate model that describes undergraduate proof comprehension then educators could focus on improving individual students’ areas of weakness, as shown up by the model. Therefore some previous suggestions on how to measure, assess, and attempts to model proof comprehension at various levels are discussed. Within this second area there is also a discussion of the use of eye-tracking methodologies in proof-related research, since the use of an eye-tracker features heavily
in one of the studies conducted for this thesis. Although eye-tracking methodologies do not explicitly help to show whether a student understands a proof, using an eye-tracker does provide insight into what a student is focusing their attention on and, thus, what they are devoting their cognitive effort towards during a reading attempt. It is therefore important to review the previous uses of eye-tracking methodologies in proof-related tasks.

The final area of this literature review considers previous suggestions on how to improve proof comprehension. These previous attempts can be split into two main categories. The first category looks at the different ways one could present a proof text. It has been suggested that, for example, a proof text could be split into sections, or “modules”, each providing extra detail for the reader to help understanding (Leron, 1983a; 1983b). Other suggestions discussed include using generic proofs (Rowland, 2002), which use examples to aid understanding, and e-proofs (Alcock and Wilkinson, 2011), which are electronic proofs that have added verbal and written explanations. The second category looks at how we could change students’ engagement with proofs, focusing on the use of self-explanation training. As the name suggests, self-explanation training requires the reader to explain to themselves any new concepts or logical links within a text using their own prior knowledge. Previous studies in other fields that have investigated the use of self-explanation training in order to better understand a text have been promising, indicating that the same could be true if self-explanation training was provided to students studying proofs.

2.1 What is proof?

Since this thesis investigates improving undergraduate proof comprehension, this literature review begins with a discussion on what various groups, such as educators, philosophers, researchers and students believe constitutes a proof. As yet, there is no agreed definition of what a proof actually is (Selden and Selden, 2008) and different people are persuaded by different types of “proof” (Harel and Sowder, 2007). This section considers how proof has been defined more broadly in the literature and makes clear how a proof is defined for the purposes of this research.

According to Hersh (1997), “…proof is math and math is proof” (p.59) and therefore without proof, there would be no such thing as “mathematics”. But what exactly is a
mathematical proof? The literature provides many answers to this question. For example, Rav (2007) suggested that “... a mathematical proof is perhaps just a sequence of logical steps, following some (implicit) rules, to be judged on the legitimacy of these steps” (p.293). Pelc (2009) stated that “by proofs we ... mean the arguments used in mathematical practice in order to justify the correctness of theorems” (p.85) whilst Selden and Selden (2003) simply stated that proofs are “... arguments that prove theorems” (p.4). Combining these suggestions, it appears that a mathematical proof is a mathematical argument that is used to convince others that a theorem is true. Indeed, “a proof confirms truth for a mathematician the way experiment or observation does for the natural scientist” (Griffiths, 2000, p.2). Azzouni (2004) adds to this idea by suggesting that “... proofs really are devices that mathematicians use to convince one another...” (p.84). However, proofs are not just used to convince. They are used for inquiry, pedagogy, and knowledge transfer (Aberdein, 2013). Therefore the reality of what constitutes a proof is perhaps more complex and should involve the consideration of other different aspects.

One such aspect that has been linked to what constitutes a proof is the social aspect. Indeed, according to Heinze (2010), “social processes particularly play an important role in the acceptance of ... proofs” (p.101). Furthermore, Manin (1977) stated “a proof becomes a proof after the social act of accepting it as a proof” (p.48). Hence a proof is accepted by the mathematical community as a proof once there is social agreement that it is indeed a proof; one needs to convince others that one’s proof is acceptable. The social aspect seems to play an important role in determining what is, and what is not, considered as a mathematical proof.

Indeed, convincing oneself and others has led proof to also be considered as a discursive practice (Sfard, 2000). When one tries to create a proof, one needs to convince oneself that a proof is indeed a proof. When one tries to convince others that a proof is indeed a proof, there is also a discourse, distinct to the discourse with oneself. Similarly, Mason, Burton and Stacey (1982) suggested that to create a proof one needs to convince oneself first, then convince a friend and then convince an enemy. However, in order to convince the three characters suggested by Mason et al., one must also consider a person’s individual proof schemes.

What persuades a person, or a community of people, that a proof shows a claim to be true determines their individual proof scheme (Harel and Sowder, 2007). According to Harel and Sowder, there are three types of proof scheme. The first is external
conviction, where a person or community relies on authority (e.g. a teacher or a book),
the appearance of the proof, or on symbolic manipulations in order to believe a proof
proves a claim. The second proof scheme is the empirical proof scheme. Those with
this proof scheme rely on evidence from examples or internal perceptions to believe a
proof proves a claim. The final proof scheme is the deductive proof scheme. With this
proof scheme, a person or community believes a proof proves a claim based on logical
inferences and accepted principles (axioms). Harel and Sowder suggested that the de-
ductive proof scheme is the ideal scheme for students to have, since it is in line with
the proof scheme of mathematicians. They also suggested that students’ proof schemes
can be changed and refined over time. Although this may be true, the consideration of
individual proof schemes may perhaps be more of an issue for validation tasks rather
than comprehension tasks since it is what convinces you that will determine whether
you think a proof is valid or not. In this thesis, all the proofs are considered to be valid
because the focus is on comprehension.

Another aspect to consider is the nature of the proof; whether the proof is “formal”
or “informal”. Philosophers and mathematicians have differing views on which proofs
come under each heading. Philosophers suggest a formal proof has a logical structure
and each line logically follows from the next (Panza, 2003), often requiring the full use of
available axioms. On the other hand, an informal proof, according to philosophers, may
not have an explicit linear logical structure. For example, the proof may even include
(or be made entirely from) diagrams. Philosophers also suggest that an informal proof is
a “real (mathematical) proof” that students produce themselves and/or convinces them
(Panza, 2003) and so textbook proofs may be considered as informal by some philoso-
phers. Conversely, some mathematicians and mathematical experts would argue that a
textbook proof is actually formal (Aberdein, 2009) as it provides all of the information
a student needs to obtain a basic understanding of the proof, even though some of the
logical links may not be explicitly written in the text. Since the proofs used in the
studies reported in this thesis are mainly taken from lecture notes and textbooks, this
distinction is important as readers from different fields will view the proofs described in
different ways.

There is some agreement that textbook proofs are useful for learning and teaching at
the higher levels of mathematics. Mathematics educators have shown that providing
students with a proof which gives them sufficient understanding is different to providing
a rigorous proof showing an argument is true (Alibert and Thomas, 1991) so the latter is
perhaps not suitable for aiding students in their initial understanding of proof (Alibert
and Thomas, 1991; Balacheff, 1988). Furthermore, it has been suggested that in the real
world a formal proof can never be complete and calculations will invariably be omitted (Davis, 1986) whereas informal proofs are often convincing and can lead to new discoveries (Tymoczko, 1986). A mathematician, who recall may describe a textbook proof as formal, may argue that calculations are omitted to prevent a proof from running on for pages and pages. Furthermore, the reader should reasonably be expected to determine some calculations and steps in the proof for themselves. Textbook proofs offer the reader an opportunity to do such steps and calculations and are therefore considered useful for the teaching and learning of proof at the higher level.

Moreover, traditional mathematicians “...demand that proof should not only be logical, but that there should be some over-riding principle that explains why the proof works” (Tall, 1991, p.16) which textbook proofs can offer through an explanation that enlightens the reader on why the proof proves the theorem. Hanna (1991) states: “Formalism should...be seen as...an important tool for clarification, validation and understanding. When a need for justification is felt, ...met with an appropriate degree of rigour, learning will be greatly enhanced” (p.61) and since there is some agreement by both philosophers and mathematicians that textbook proofs offer this, using textbook proofs for the research conducted in this thesis seems appropriate.

It can therefore be seen that what the individual reader of a proof finds convinces them and whether they prefer a formal or informal proof style determines how they interpret a proof. To ensure there is no confusion, the proofs used in this thesis are all the general types of proofs encountered by general undergraduate students every day within the course of their work. These proofs are traditional proofs that come from textbooks or lecture notes and cover a variety of proof types that undergraduate students are familiar with. However, it can clearly be seen that determining exactly what a makes a mathematical proof is not straightforward. There is, as yet, no precise definition of proof and this may cause confusion for students. As Selden and Selden (2008) state: “the [fact the] concept of proof in mathematics is a difficult one for students is not surprising given the various everyday uses of the word proof” (p.2). Therefore, the next section of this literature review discusses students’ beliefs about proofs and some of the difficulties they appear to encounter when working with mathematical proofs.
2.2 Student beliefs about proof

The previous section mainly focused on the debate among mathematicians and philosophers on what they believe constitutes a mathematical proof. It is nevertheless important to also consider what students believe a proof is. Being aware of what students believe about proof could help direct research into improving proof comprehension and improve the support offered by educators. In this section, several issues are addressed relating to students’ beliefs about proof, showing that some of their misconceptions have led to research on improving proof comprehension to focus on highlighting, or even trying to remove, these misconceptions. Firstly, it has been shown that some students use empiricism when working with proof (Balacheff, 1988; Weber, 2001) which could be considered by some as far from ideal. Secondly, it has been shown that some students use diagrams when working with proof or are convinced by proofs that are made entirely from diagrams (Almeida, 2000; Chazan, 1993). Given the discussion in the previous section on formal and informal proofs, this again could be considered as an imperfect scenario. Finally, one also needs to be aware that individual students have their own conceptions about proof (Almeida, 2000), with some being unsure about what a proof is or does (Healy and Hoyles, 2000). These issues will be discussed in turn, showing there is conflicting evidence for each issue within the literature. By considering these issues described, it will be argued that perhaps one should concentrate on positive aspects, such as what students do understand about proof, and embrace and enhance these aspects to improve their proof comprehension, rather than the traditional approach of focusing on the misconceptions.

It has been suggested that students first encounter proof, and indeed proving, in Euclidean geometry at the high school level (Stylianides, 2007) and Bell (1976) stated that a proof for a high school student, according to teachers, “...is what brings him conviction” (p.24). It has been shown that one way a student obtains conviction is through testing a theorem holds using examples; a form of (naive) empiricism. Empiricism is defined as “…asserting the truth of a result after verifying several cases” (Balacheff, 1988, p.218). Chazan (1993) illustrated this with an example. The theorem given in Chazan’s paper is that two triangles created when a median is drawn in any possible triangle have an equal area. Students who are convinced by empiricism will measure the areas of just some triangles to establish the theorem is true for all cases. Chazan stated that some students think writing a proof based only on measuring certain examples can constitute a proof that provides certain conclusions about infinite sets. These students see evidence as proof and therefore do not know the difference between evidence and
There is conflicting evidence in the literature on whether students do use empiricism when working with proof. It had previously been shown that many students at all levels gain conviction by verifying a general theorem holds in one or several instances which provides for them enough evidence to constitute a proof (Weber, 2001). Hoyles and Küchemann (2002) also found high school students have three proof strategies based on their beliefs: empirical strategy, focused empirical strategy and deductive proof strategy. It was argued that the deductive proof strategy was the ideal strategy but they found students’ progress from having an empirical strategy towards having a deductive strategy was only slight. More recently however, Weber (2010) found some slight contradictions to the previous findings. He conducted a study with 28 mathematics majors who had recently completed a transition-to-proof course in the US. The students were presented with 10 arguments which varied along several dimensions such as the type of argument, the format of the argument and the mathematical content. Three judgements were to be made on each argument by the students. Firstly, the students had to rate on a five point scale the extent to which they understood the argument. Secondly, the students had to rate on a five point scale how convinced they were by the argument. Finally, the students had to decide whether the argument constituted a rigorous proof. The results showed that students did not find empirical arguments convincing. This result is slightly in contrast the findings of Segal (2000) who found university students in her study saw empirical arguments as personally convincing but not publically valid. Instead, Weber (2010) reported that students found valid deductive arguments convincing and attended to the argument’s mathematical content in forming their evaluations. Although the students in Weber’s study were indeed older than those reported in some of the previous studies, perhaps this shows that students can learn how to behave in ways that are considered “mathematically appropriate” by mathematicians. Indeed, it has been shown some students do appear to have a good idea of what a proof is when they begin their degrees and have some ability to correctly evaluate their proficiency on proof tasks (Iannone and Inglis, 2010).

Another issue often found when looking at students work with proofs is the use of diagrams. Mathematicians would generally agree that a diagram in itself does not constitute a mathematical proof. However, Weber (2010) also reported that the majority of students in his study believed diagrams can constitute a mathematical proof. In contrast, it had previously been shown that students believe a proof is valid only if it follows a traditional format (Harel and Sowder, 1998). Despite this, many studies have shown that the use of diagrams can enhance students’ understanding of proofs and students...
use diagrams themselves to assist their understanding (e.g. Barker-Plummer and Bailin, 1997; Gibson, 1998; Hanna, 1991; Jamnik, 2001).

Of course, not all students find the use of diagrams in proofs useful as each student will have their own individual preferences. Indeed, a study conducted by Healy and Hoyles (2000) showed that individual students hold simultaneously two different conceptions about proof. These were conceptions about which proofs the students themselves believed would give them the best mark, and conceptions about what proof techniques they would adopt for themselves. Almeida (2000) was able to categorise students in his study into four groups based on their beliefs. The first group of students accepted the need for formal proof but they struggled with it in their own proof practices. The second group of students again accepted the need for formal proof but used informal proof practices, such as diagrams, until they were adept at formal proof practices. The third group of students accepted intuitive arguments as proof and view formal proof as something required to pass examinations. The final group of students accepted the need for formal proof but became disengaged when they struggled due to rote learning techniques. Both studies show individual students have individual preferences and ideas when working with proofs, perhaps suggesting that some form of tailored learning is a potential way of improving individual students’ proof comprehension abilities.

A final issue found when investigating students’ beliefs about proof is that they are often confused about the nature and purpose of proof. Healy and Hoyles (2000) found that 50% of high school students surveyed in their study believed the purpose of proof was to provide truth, with 35% believing proofs are provided for explanation. Over a quarter of the students (28%) in the study had no idea what a proof was for. Indeed, it has been shown in several other studies that this is the case and students are often confused about the nature and uses of proofs (Fischbein and Kedem, 1982; Martin and Harel, 1989). Furthermore, it has been suggested that some undergraduate students do not appear to have an accurate conception of what constitutes a mathematical proof (Weber, 2001). Clearly if students are unsure what constitutes a proof they will have difficulties with proof, including focusing too much on algebraic manipulations, termed in the literature as the “surface features” of a proof (Selden and Selden, 2003), execution of nonsensical operations (Harel, 1998) and systematic misapplication of theorems and concepts due to a lack of understanding (Weber, 2001).

It appears therefore that students, like mathematicians, philosophers, and researchers, have many different beliefs about what a proof is. However, students’ perceptions of
proof have been described as inadequate (Selden and Selden, 1995) and clearly these perceptions do not help their understanding, possibly adding to the difficulties with proof they encounter. The combination of such beliefs, the issues described above, and there being no formal definition of what a proof is makes the understanding of proofs much harder because there is perhaps no clear direction for students to focus their learning. Furthermore, it has been claimed students believe mathematics is learned best when rote memorised (Schoenfeld, 1989) and they may therefore reduce the effort they put in to understanding a proof. Indeed, a recent study by Weber and Mejía-Ramos (2014) showed that when reading a good proof, students believe they are not expected to provide their own justifications even though they were shown to also believe that in order to understand a proof requires one to justify each step of the proof. Although “how to” textbooks on proofs have previously appeared (e.g. Cupillari, 1989; Franklin and Daoud, 1988; Solow, 1990) and emphasis has been put on reforming proof instruction (National Council of Teachers of Mathematics, 2000), students continually have difficulties when working with proofs. Research into improving proof comprehension has tended to focus on highlighting students’ misconceptions, or trying to change what they do when working with proof through changes to the presentation of the proof text, designed to make comprehension easier. However, as stated previously, these changes have mainly not had the results we would have hoped for, suggesting one should perhaps instead embrace what students do know about proof but also provide clearer direction. It is these issues that provided motivation for conducting the research reported which explores new avenues of investigation to improve students’ understanding of proofs, focusing on improving proof comprehension.

2.3 Proof comprehension and proof construction

The research undertaken in this thesis concentrates on improving proof comprehension in undergraduate mathematics rather than proof construction. There are two motivations for concentrating on improving proof comprehension. Firstly, within the mathematics education literature, little research has been conducted on proof comprehension, with the majority focusing on the problems faced by students in proof construction tasks (e.g. Mejía-Ramos and Inglis, 2009). Secondly, in order to construct a proof, one needs to have a general understanding of mathematical proofs. Hence a key step in constructing new proofs is to understand the proofs that already exist and apply the methods and logic used to one’s own constructions. Therefore it is important to distinguish between these two major types of proof activity in the mathematics education literature in order
to fully appreciate the research already undertaken. This section therefore briefly confirms what is meant by proof comprehension and proof construction and the differences between the two.

It is generally understood that proof construction tasks simply relate to the creation of an argument that attempts to prove a given theorem. In contrast, it is generally understood that proof comprehension relates to understanding a proof which has already been constructed. Proof comprehension can involve the overall understanding of a proof, the logic behind the proof, why the author of the proof constructed it in that particular way or deciding whether a proof is valid or not. Although proof comprehension and proof construction are clearly two distinct tasks undertaken by students, it has been argued that the research showing the difficulties students face in proof construction tasks are similar to the difficulties faced by students in proof comprehension tasks (Weber, 2010). Therefore, the following sections of this literature review will draw from both proof comprehension and proof construction activities.

The validation of proofs, described as part of the wider proof comprehension process above, could also be considered as a third, separate activity. However, Selden and Selden (1995) state that validation is the “...process an individual carries out to determine whether a proof is correct and actually proves the particular theorem it claims to prove” (p.127). Hence, validation tasks require understanding of the proof first in order to make correct judgements about the validity of the proof in question. Research into proof validation tasks can therefore be useful when considering proof comprehension issues, and will be used alongside the literature from proof construction and proof comprehension research to discuss some of the issues reported later in this literature review.

### 2.4 Logic, argumentation and proof

Aside from the issue of students not having an accurate concept about what a proof actually is, it has been shown that some students lack the knowledge and skills to obtain good understanding of proofs. Alternatively, they may have the knowledge and skills required, but be unable to apply that knowledge and those skills. This section will discuss the evidence for these claims by describing one of the factors that can effect one’s understanding of proofs: logical reasoning and argumentation. Proofs are mathematical arguments so being able to understand the logic that underpins the proof is crucial for
good understanding. Therefore, the literature on logical reasoning and argumentation, based mainly on the works of Toulmin (1958) and Aberdein (2005), are discussed.

Logical reasoning plays a big part in mathematics in general, but especially in proofs. Indeed, it has been suggested one of the reasons why students have difficulty with proof is because they lack logical maturity and the skills needed for deductive reasoning (Selden and Selden, 1995). Without logic, theorems could not be proved as the arguments used in the proof would not have structure or meaning. Theoretical models of argumentation have been suggested in the literature and adapted for use in mathematics. What follows is a discussion on the development of these models; how they have been used to investigate and discuss the process of understanding the underlying logic of a proof, and how using these models can show where in a proof students need to focus their attention in order to obtain a good understanding.

The model of argumentation cited most in the literature was constructed by Toulmin (1958). The model suggests that to argue a conclusion (C) from some data (D), on which the argument is based, requires a connection between the two called a warrant (W). Toulmin describes warrants as “...hypothetical statements, which can act as bridges, and authorise the sort of step to which our particular argument commits us” (p.98) and within proofs these may or may not be explicitly written within the proof text itself (Weber and Alcock, 2005). The warrant may have further supporting evidence called the backing (B) and the qualifier (Q) expresses the degree of confidence we have about the conclusion. The rebuttal (R) states any conditions for which the conclusion does not hold. An example given by Toulmin to explain this scheme is an argument that shows “Harry”, who was born in Bermuda, is a British subject. The data provided is Harry was born in Bermuda. The warrant provided is a man who is born in Bermuda will generally be a British subject. This warrant is given a backing, which is that statutes and legal provisions provide that a man born in Bermuda will generally be a British subject. Toulmin uses “presumably” as his qualifier before providing some rebuttals, such as both of Harry’s parents were aliens or he has become a naturalised American (p.97). Figure 2.1 shows a diagram of the advanced scheme suggested by Toulmin with Figure 2.2 showing how the advanced scheme is used to argue that “Harry is a British subject”.

In mathematics, a simplified version of the Toulmin model has been used to analyse mathematical arguments in the classroom (Krummheuer, 1995). The simple scheme
Figure 2.1: Toulmin’s scheme of argumentation

D → Q → C
W
B

Figure 2.2: Toulmin’s example of an argument

D → Q → C
W

Figure 2.3: Krummheuer (1995) adapted version of Toulmin’s scheme of argumentation

only requires the data, the warrant, and the conclusion, as shown in Figure 2.3.

Previous research has suggested that the full scheme is superior for categorising mathematical argumentation, despite previous studies (e.g. Alcock and Weber, 2005; Pedemonte, 2005; Yackel, 2001) using only the basic scheme (Inglis, Mejía-Ramos and Simpson, 2007). Inglis et al. (2007) argued that the full scheme is superior since it allows
analysis of the full range of mathematical argumentation, from informal reasoning to formal proof.

The Toulmin model of argumentation could be seen as a model of the processes one undertakes to convince another in an argument. Indeed, Aberdein (2005) investigated how closely Toulmin’s model of argumentation models the processes in mathematical proof and stated that there have been critics of the scheme of argumentation as a way of modelling mathematical proofs and the processes involved. The critics suggest that “...the layout...require[es] so much abstraction as to permit incompatible reconstructions” (Aberdein, 2005, p.287) and is “…ambiguous, making different, incompatible reconstructions possible” (Aberdein, 2005, p.289). In other words, the Toulminian scheme is perhaps too simple to allow a full overview of the processes involved in understanding the logic of a mathematical argument. Thus, according to Aberdein, one cannot fully appreciate the logical reasoning within a proof using the Toulmin model. Aberdein illustrates this claim with an example proof taken from Toulmin’s book itself (shown in Figure 2.4), arguing that the proof chosen by Toulmin is too simplistic to show the model can be used to reconstruct mathematical proofs. The proof only has one step containing one warrant. More complex proofs will have more steps, more warrants and a more complex structure, thus suggesting the Toulmin model is insufficient when trying to model the logical argument of a more complex proof.

![Figure 2.4: Example proof used by Aberdein (2005) to show that the Toulmin scheme is too simplistic (Toulmin, Richard and Allan, 1979, Fig. 7.4, p.89)](image)

Aberdein therefore explained that with proofs that require lots of logical steps to get from the original data to the overall conclusion, the Toulmin model would need to be adjusted.
He suggested that each step should have its own separate argument and all of these arguments put together form the mathematical proof. Each step contains an individual conclusion which becomes the data of the next step. Within each of the steps there may also be a particular warrant, qualifier, backing and rebuttal which help explain how the author of the proof goes from the data to the conclusion of that particular step. Clearly, not all steps within a proof will have a particular warrant. Hence determining when a warrant is required is a key skill that one must possess when understanding a proof, as will be discussed shortly. By using this method of modelling mathematical arguments however, one can obtain a good representation of the proof and how the argument fits together. Figure 2.5 below shows a diagram of the model suggested by Aberdein (2005).

Figure 2.5: A model of mathematical argumentation according to Aberdein (2005), adapted and simplified for the purposes of this thesis.

Toulmin’s model, in its various forms, has been used extensively throughout the mathematics education literature. For example, the model has been used to describe mathematical practices of instructors and students alike in a differential equations classroom (Stephan and Rasmussen, 2002), how gestures in a mathematics classroom can help with argumentation (Rasmussen, Stephan and Allen, 2004), and how middle school students produce different arguments to justify their own individual solutions to a single geometric problem (Forman, Larreamendy-Joerns, Stein and Brown, 1998). The difference between the present research and previous uses of the Toulmin model is that the research reported in this thesis uses the model to break down the components of a proof in order to discuss the issues for proof comprehension.

The Toulmin scheme is useful for explaining ideas regarding proof comprehension and the proof comprehension process but it is not a perfect model of proof comprehension. For example, it does not consider that some lines within a proof may rely on warrants and ideas from more than one line. Furthermore, proofs do not always proceed in a linear fashion; some ideas set up in the first couple of lines may be used later on in
the proof while others may not. The skills required of students when understanding proofs are perhaps more complex than the Toulmin model suggests and one needs to consider each line in much greater detail in order to improve understanding. Therefore, the Toulmin model should not be considered as a perfect model of proof comprehension but should perhaps instead be considered as a useful tool for mathematics educators in explaining some of the ideas surrounding the proof comprehension process.

The model has also been used to show that students’ difficulties with the warrants in mathematical proofs is a potential symptom of the fact that they lack the logical maturity necessary to understand proofs. Indeed, students may not even consider the warrants at all. For example, it has been shown that if students reject a proof as invalid because, according to them, an appropriate definition is not used, they may automatically dismiss the rest of the proof (Weber and Alcock, 2005). It has also been shown some students may spend too much time focusing on the “surface features”, such as the algebra and the mathematical symbols of the proof (Selden and Selden, 2003). Hence, when trying to work with proofs, warrants may not be considered by the students and they may therefore fail to gain conviction and understanding from the proof.

From the adapted Toulminian model of Aberdein (2005), one can see that in order to understand how the author of the proof proceeded from each individual step to the next, one needs to understand the warrants for each step (when applicable). The creator of the proof needs to be able to convince others that the conclusion drawn from some data is true by using these warrants. However, before discussing the importance of warrants in mathematical proofs in more detail, it should be noted that Toulmin’s use of the word “warrant” is different to the use of the word “warrant” in some of the mathematics education literature. Toulmin accepted the use of a warrant and a modal qualifier in an argument can reduce uncertainty on the part of the reader. Slightly in contrast however, it has been suggested that a warrant should remove uncertainty (Rodd, 2000) and furthermore, removing uncertainty is “considered one of the hallmarks of proof” (Inglis et al., 2007, p.4). Without the warrants, one may not be convinced of a step within the proof or argument and fail to draw the same conclusion intended by the creator. Indeed, it has been suggested that the warrant should explain any challenges by the reader regarding why the data leads to the conclusion, with the backing refuting any challenge to the validity of the warrant (Stephan and Rasmussen, 2002). Since all proofs used for the purposes of this thesis are assumed to be valid, it is clear that understanding warrants is important to understanding the proof.
As briefly stated earlier, 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 recall that many of the steps \( A_i \rightarrow A_{i+1} \) require a Toulminian warrant (Toulmin, 1958) which can be left for the reader to infer rather than explicitly appearing in the text. Within a proof, the author can decide whether or not to leave the reader to infer the warrants required to understand the logic of the proof. An author may leave the reader to infer a warrant for themselves in a proof due to either the impractical nature of strict formal derivations (Fallis, 2003) or if the author believes the warrant is simple enough for the reader to infer (Inglis and Alcock, 2012). When a reader is left to infer the logical steps between parts of a proof, the warrants do not appear explicitly as part of the proof text. These warrants can therefore be considered as implicit warrants. This is in contrast to an explicit warrant, which also explains the logical step from one part of the proof to another but is explicitly written within the proof text. A very simple example of this comes from a proof taken from Alcock and Weber (2005) which is actually invalid.

Theorem:
\[ (\sqrt{n}) \rightarrow \infty \text{ as } n \rightarrow \infty. \]

Proof:

1. We know that \( a < b \Rightarrow a^m < b^m \).
2. So \( a < b \Rightarrow \sqrt[n]{a} < \sqrt[n]{b} \).
3. \( n < n+1 \), so \( \sqrt[n]{n} < \sqrt[n+1]{n+1} \) for all \( n \).
4. So \( (\sqrt{n}) \rightarrow \infty \text{ as } n \rightarrow \infty. \)

The major flaw in the proof comes in line 4. The line relies on the false “warrant” that since \((\sqrt{n})\) is an increasing sequence, it diverges to infinity. This is not true for all increasing sequences. Although this is an example of an invalid proof, it is also an example of an implicit warrant as the explanation does not appear within the proof text itself. Furthermore, for the purposes of proof comprehension, one also needs to be aware that there is a distinction between the warrants intended by the author and the warrants inferred by the reader. It is possible that there could be a difference between the two or the reader may determine a warrant is not required at that point in the proof at all since some lines may not require a warrant. For example, if one was to add a line at the beginning of the proof shown above stating: “Let \( a > 0, b > 0 \)”, this would be an example of a line not requiring a warrant as one is simply allowed to do this as part of
the set up of the proof. This suggests that inferring warrants within proofs is perhaps not a trivial task (Weber and Alcock, 2005). Within the studies reported for this thesis, the aim is for the student to infer the warrants intended by the author by helping them to become better at identifying the logical structure of proofs.

Many warrants are left as implicit because the majority of people interested in understanding them are fellow mathematicians and it would make the proof far too long if every warrant was made explicit (cf. Renz, 1981). Mathematicians do not need the extra information explicit warrants provide due to their prior knowledge and experiences so the mathematical community accepts proofs with implicit warrants, keeping the proofs shorter than they would otherwise be. Students, however, may not be as skilled as mathematicians. The added information an explicit warrant provides could be a useful tool in helping students understand a proof (Inglis and Alcock, 2012). Furthermore, students may not even consider locating or thinking about the warrants if they are not made explicit because recall it has been suggested they tend to focus on the “surface features” of a proof rather than the detailed logic (Selden and Selden, 2003), which was also confirmed by Inglis and Alcock (2012). Hence, Inglis and Alcock (2012) proposed that if implicit warrants were made explicit it could help reduce the burden when trying to validate a proof. On the other hand, Inglis and Alcock also warned that if too many warrants are made explicit in a proof, the level of cognitive engagement experienced by the reader may not be enough for them to fully comprehend the proof. Therefore a balance needs to be found such that students find it easier to think about the warrants contained within a proof without providing all the information.

Inglis and Alcock (2012) suggested that after one has accurately determined when a warrant is required, one needs to correctly infer the implicit warrant intended by the author. Finally, one needs to evaluate the warrant’s mathematical validity. Although this three-step process was suggested for validation tasks, the first two steps are useful for investigating the comprehension process. If the suggestion by Inglis and Alcock were correct then it would suggest that students need to be able to complete this triple task successfully in order to increase their understanding. If students were able to complete the first step then the issue with their understanding would perhaps lie with inferring the warrants. There would then be two options: make the warrants explicit, as suggested by Inglis and Alcock (2012), since they know where they appear in the proof anyway, or improve the students’ engagement with the proof by helping them direct their attention to focus more on the logical connections. The research reported in this thesis focuses on the latter.
2.5 Assessing proof comprehension and the reading processes of students in proof comprehension tasks

The opening to this literature review has discussed some of the theoretical issues surrounding the understanding of mathematical proof texts as well as the knowledge and skills required to successfully understand mathematical proofs. Considering these issues when researching methods of improving proof comprehension is important so that results can be reported clearly and so the work undertaken can be located within general teaching and learning research. Investigating potential methods of improving students’ understanding, however, requires one to also discuss the methods for measuring understanding of proofs. This next section therefore considers several methods that claim to measure proof comprehension and assess students’ proof comprehension abilities, arguing that one framework suggested by the literature is the superior framework for the work undertaken in this thesis.

To analyse proof comprehension in undergraduate mathematics in this thesis, and also the reading processes of students whilst comprehending proofs, two distinct methods are used. The first is to use a framework suggested in the literature to help set questions that can be used to assess students’ understanding of different proofs. To begin, there is an explanation of why proof comprehension needs to be measured at the undergraduate level. Several models and frameworks that have been suggested will then be described in turn and it will be argued that, potentially, the most promising framework to use is that of Mejía-Ramos et al. (2012).

The second method, used mainly to determine the reading processes of students, is to use eye-tracking methodologies. Eye-trackers and eye-tracking methodologies have been effectively used to investigate the reading behaviour of students whilst working with proof. However, these methods have yet to be tested with students who are reading a proof for comprehension purposes and to determine any changes made to their proof reading behaviour after an intervention. The second half of this section of the literature review will therefore briefly discuss the development of eye-tracking research, explain why it is useful for gaining insight into the reading process of students, and how it has been used previously in proof research, explaining how the research reported for this thesis will expand upon existing research.
2.5.1 Frameworks for assessing undergraduate proof comprehension

In order for one to improve undergraduate proof comprehension, one needs a way of assessing students’ understanding of proof. Students’ comprehension of proof is rarely assessed (Mejía-Ramos et al., 2012; Weber, 2012), perhaps leading to students thinking they do not need to spend time studying proofs (Weber and Mejía-Ramos, 2014). It has been suggested that comprehension tests can combat this as they provide students with the motivation to understand the proof (Conradie and Frith, 2000). In order to create a comprehension test, one needs to consider the proof itself. Given the discussion on logic and argumentation previously, it can be suggested that in order to fully comprehend a proof, one must identify and understand logical inferences alongside understanding the holistic aspects of the proof. Several frameworks have been suggested that provide a method for creating proof comprehension questions that test such knowledge of the inferences and the holistic aspects of the proof.

Conradie and Frith (2000) first proposed the idea of a comprehension test as an alternative method of assessing mathematics students’ understanding. They set their paper in the context of proofs to explain the advantages and disadvantages of using such tests. Each question can be used to test understanding of different aspects of the proof and thus the lecturer can obtain a better understanding of what that specific student actually knows. Conradie and Frith (2000) suggested that, for pedagogical purposes, the best form of comprehension test is one that tests students’ understanding of an “unseen” proof. According to Conradie and Frith, a comprehension test on “unseen” proofs increases the effort a student devotes to understanding the proof. Doing so provides students with the “…experience in reading and trying to understand ‘new’ mathematics on their own” (Conradie and Frith, 2000, p.9).

The work of Conradie and Frith (2000) provided the foundations to produce an actual model of proof comprehension although the paper did not itself suggest such a model. Yang and Lin (2007) built upon the work of Conradie and Frith (2000) to propose a model for comprehension of geometry proof at the high school level. Yang and Lin hypothesised there were five “facets” to reading comprehension of geometry proof. The first, basic knowledge, measured understanding of terms, figures and symbols. Logical status measured recognition of the status of a premise, conclusion or applied property within the proof. Summarisation measured understanding of given claims or the critical idea in a proof. The fourth facet, generality, measured the recognition of what the proof tries to prove. The final facet, application, measured the ability to apply a proposition
To test the five facets, Yang and Lin developed 16 questions and used these questions to test 601 high school students. A multidimensional scaling (MDS) analysis, which is an alternative to factor analysis, was conducted on the data. MDS analysis provides a representation of overall similarities between factors through a Euclidean distance model but this often produces vague results (Kruskal, 1964), leaving them open to subjective interpretation. Nevertheless, the results of the analysis, according to Yang and Lin, suggested the five facets could be split into three different groups. Group one was basic knowledge, group two was logical status and summarisation and group three was generality and application. Yang and Lin conjectured that as a student acquires the skills in each group, they develop their reading comprehension of geometry proof and move through four hypothetical levels of comprehension: surface level, recognising elements, chaining elements and encapsulation. As a student acquires basic knowledge, they move from the surface level to recognising elements. As the student acquires logical status and summarisation skills, they move from recognising elements to chaining elements and as the student acquires generality and application skills they move from chaining elements to encapsulation. Figure 2.6 shows a visual representation of how the model of Yang and Lin is constructed.

![Figure 2.6: A visualisation of the model of Reading Comprehension of Geometry Proof (RCGP) suggested by Yang and Lin (2007)](image)

Essentially, this study proposed that a model of students’ proof comprehension skills is possible and the model can be used to tell the level of understanding of a proof a student has. Yang and Lin reiterated this by conducting a follow up study which used
their model and found similar results (Yang and Lin, 2008).

Later, the model suggested by Yang and Lin was expanded by Mejía-Ramos et al. (2012) to suggest a model of assessment for undergraduate proof comprehension. Mejía-Ramos and his colleagues suggested a seven-dimensional framework, arguing that the model proposed by Yang and Lin was too simplistic for modelling undergraduate proof comprehension. Indeed, Mejía-Ramos et al. claimed that Yang and Lin mainly focus on the first three levels (surface level, recognising elements and chaining elements), but stated that the model does not specify when, and if, a student reaches the encapsulation level. Furthermore, according to Mejía-Ramos et al., Yang and Lin did not include in their model whether a student understands a proof in terms of “higher-level ideas”. These are the methods involved in the proof or how the proof relates to specific examples. Within the paper, Mejía-Ramos et al. suggested ideas for multiple-choice questions for each dimension. The seven dimensions, along with an example of a question for each dimension, are described below.

The first dimension suggested was meaning of terms and statements, which amounts to understanding the meaning of symbols, terms and definitions. This first dimension is very similar to the surface level suggested by Yang and Lin. To measure this dimension, multiple choice questions on understanding the meaning of the theorem, individual statements and individual terms were suggested. The second dimension, justification of claims, assesses whether the student can see how new assertions in the proof follow from previous ones. Mejía-Ramos et al. claim this is analogous to the chaining elements level in Yang and Lin (2007). An example suggested to measure this dimension given was “In the proof, which justification best explains why. . . ?”. Logical status, the third dimension suggested and similar to logical status suggested by Yang and Lin, examines whether the student understands the framework of the proof (“What type of proof is this?”) or the purpose of a line within a proof (“What is the purpose of making this assumption?”). The fourth dimension, higher-level ideas, assesses whether a reader can identify a good summary of the proof as a whole or of a section of the proof (“Which of the following is the best summary of . . . ?”). This dimension was not considered by Yang and Lin according to Mejía-Ramos and his colleagues. The fifth dimension suggested was general method. This dimension looks at whether the reader can apply the methods in the proof to a different situation and is similar to the application facet suggested by Yang and Lin. A suggested form of assessment for this dimension is to ask which approaches used in the original proof can be used to prove a similar but different theorem. The sixth dimension, application to examples, determines whether the reader can use the ideas in the proof in terms of a specific example. For this dimension a typical
Table 2.1: Similarities between the suggestions of Yang and Lin (2007) and Mejía-Ramos et al. (2012)

<table>
<thead>
<tr>
<th>Mejía-Ramos et al. (2012)</th>
<th>Yang and Lin (2007a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meanings of Terms and Statements</td>
<td>Surface Level</td>
</tr>
<tr>
<td>Justification of Claims</td>
<td>Chaining Elements</td>
</tr>
<tr>
<td>Logical Structure</td>
<td>Logical Status</td>
</tr>
<tr>
<td>Higher Level Ideas</td>
<td>-</td>
</tr>
<tr>
<td>General Method</td>
<td>Application</td>
</tr>
<tr>
<td>Application to Examples</td>
<td>Application</td>
</tr>
</tbody>
</table>

question would be “Using the logic of the proof, which best exemplifies why \( x = 5 \) is not a solution to \( 4x^2 - x^4 + 2\sin x = 30 \)”. The final dimension, called *identifying modular structure*, measures students’ comprehension of a proof in terms of its main components and modules and the logical relationship between them. It was therefore suggested the questions used to assess this level are of the form “Which of the following best describes the logical relationship between...?”.

The suggestions of Yang and Lin and Mejía-Ramos et al. clearly have many similarities, with Table 2.1 showing how the two frameworks compare.

However, Yang and Lin’s model was created by working with high school students and only with geometric proofs. The work of Mejía-Ramos et al. is yet to be empirically verified but is a suggested framework that can be used to create proof comprehension questions with example questions provided to test the different dimensions of the theoretical model. Indeed, Mejía-Ramos et al. (2012) state “...we can imagine a student (or mathematician) who grasps the big picture of a proof without understanding the technical details of the proof and vice-versa” (p.5). The framework also provides a consistent method of testing proof comprehension which has perhaps been lacking in previous investigations. Thus the framework suggested by Mejía-Ramos et al. is used in the studies reported in this thesis.

### 2.5.2 Eye-tracking and reading comprehension

It has been shown that investigating the way a person reads a text can provide an insight into what they are thinking (Duchowski, 2007; Just and Carpenter, 1976; Poole and Ball, 2006; Rayner, 1977). Thus the second study reported in this thesis includes the use of eye-trackers. Eye-trackers measure a person’s eye movements which can then
be analysed. In this section of the literature review, the use of eye-tracking methods in previous research is discussed. To begin, there is a brief discussion on the history and development of eye-tracking in research before providing the reader with details of the types of eye movement often analysed to report eye-tracking data, known as fixations and saccades. There then follows a discussion on how eye-tracking data has been used in different fields, including mathematics and proofs. It will be shown that despite some attempts to use eye-tracking methods in proof research, there has been no study that investigates students’ reading habits whilst trying to understand a proof for the purposes of a comprehension test. Therefore this section will provide an argument for conducting Study 2 reported in this thesis which compares students’ reading processes before and after an intervention.

The development of eye movement research

Research into eye movements has been conducted for over a century with some of the first works being published in the 1800s. However, due to a lack of sufficient technology, the early research was limited and often crude, with participants often experiencing severe discomfort from the early forms of eye-tracker. Research was nevertheless still conducted because it was suggested that the eyes move to new stimuli in order to inspect new information for processing. Thus, eye movements can provide evidence of overt visual attention (Von Helmholtz, 1925). Later work adapted these ideas to suggest that visual attention is akin to imagination, anticipation and thought (James, 1980); measuring eye movements could therefore provide an insight into the information processed from visual stimuli. More recently, eye movement research has also become interdisciplinary (Duchowski, 2007) and takes into consideration various factors such as the language of the text and the native language of the reader (Koda, 2013), and word frequencies (Monsell, 2012) to name but two.

A breakthrough in eye movement research came with the eye-mind hypothesis suggested by Just and Carpenter (1980). The eye-mind hypothesis states that there is a close link between the direction of human gaze and the focus of attention. It is therefore assumed that people attend to and process visual information that is currently being looked at. The eye-mind hypothesis was used by Just and Carpenter as a basis for a model of reading comprehension. The model suggests that as the eyes move from word to word and sentence to sentence, the working memory extracts and processes the information contained within the text and associations are made with the long-term memory in order to understand the text. The model is also based upon two assumptions. Firstly, a reader
tries to interpret each content word of a text as it is encountered, even at the expense of making guesses that may turn out to be wrong. Secondly, the eye remains fixed on a word for as long as the word is being processed, so gaze durations reflect the time taken to execute comprehension processes (Just and Carpenter, 1980). This suggests that in order to report eye-tracking data and analyse it, one needs to consider the time spent fixating on a word, words, lines or areas of interest that are important to the goals of the research. These techniques of analysing eye-movement data are now described.

Analysing eye-tracking data in research

In order to analyse the data provided by eye-trackers, the majority of research, including the research reported in this thesis, analyses two types of eye-movements. The first is to look at fixations. Fixation durations correspond to the amount of time the eye fixates on a particular location in order for the reader to process that information (Irwin, 2004), with dwell time being the sum of all fixation durations in a prescribed area (Hauland, 2003). Mean fixation durations are used in research as a measure of cognitive effort: larger mean fixations are associated with more effortful cognitive processing (Duchowski, 2007; Just and Carpenter, 1976; Poole and Ball, 2006; Rayner, 1977) as this indicates difficulty in extracting information (Just and Carpenter, 1976). One can therefore assume that a greater mean fixation duration on an area of interest in the text suggests that the reader is trying harder to understand that particular area of interest.

The second approach is to analyse the saccades made by a participant between different areas of interest. A saccade is a rapid eye movement that repositions the eye to a new point of interest. Although no information from the text is actually processed during the saccade (e.g. Matin, 1974), several studies have shown that when a saccade is made to a location, cognitive processing occurs in an obligatory and involuntary fashion (e.g. Deubel and Schneider, 1996; Hoffman and Subramaniam, 1995; Kowler, Anderson, Dosher and Blaser, 1995). This seems intuitively obvious as one would expect a reader to move their attention between sections of the text to fully understand the overall structure and the information contained within the text. A large number of saccades between two areas of interest may therefore indicate an attempt to connect information between the two areas. For proofs, a large number of between-line saccades may indicate an attempt to connect information between those lines, or to find the warrant that connects them (Inglis and Alcock, 2012). Further research has indeed shown saccades to be an indication of cognitive processing with Vauras, Hyöniä and Niemi (1992) showing that looking back at previous sentences within a text is an integral component of
comprehension.

Previous eye-tracking research

Eye-tracking research has been used in many fields and for many different purposes. The aim for this thesis is to obtain an insight into students’ proof reading processes. Hence there now follows a discussion on what previous eye-tracking research tells us about reading comprehension.

In order to read a text efficiently, it is essential for a reader to select the most relevant information. Selecting the most relevant information usually leads to selecting new information over old, already processed information (Watson and Inglis, 2007). Furthermore, it has been shown that readers will only select the information relevant to the perspective they are reading the text for (Kaakinen, Johanna, Hyöniä and Janice, 2003). This reduces the information processed by the working memory and therefore would perhaps improve understanding of the text. However, in order to obtain a more thorough understanding, one needs to integrate new and existing knowledge in order to fully understand a text. To do so, a link between a reader’s working memory and long-term memory must be established. Accordingly, a reader with a larger working memory capacity can integrate more elements of a text at a time (Daneman and Carpenter, 1980) and will have a better understanding of the text. Integrating new and previously understood knowledge is important for good proof comprehension because recall that proofs often rely on the use of previously understood definitions and concepts.

There have been several eye-tracking studies conducted for the purposes of general mathematics education. Eye-tracking studies in mathematics improve our understanding because the way mathematical texts are presented (Duval, 1999) and the way in which students read mathematical texts are essential for the learning of the mathematics itself. Psychological papers that have conducted research into the way mathematical texts are read have mostly considered the use of symbols and words in representing numbers and how that affects reading and understanding. For example, Gielen, Brysbaert and Dhondt (1991) showed that the syllable length of number names affected the fixation times of their participants. These results have also been shown in a later study with university students where it was shown that as the magnitude of the number value increased, so did the overall reading time (Brysbaert, 1995). Further studies have also provided an insight into mathematical reading processes for those with dyscalculia.
(Moeller, Neuburger, Kaufmann, Landerl and Nuerk, 2009) and a different native language to that of the text, including whether a reader is bilingual (Chinacotta, Hyönnä and Underwood, 2007).

Other papers that focus on research into mathematical reading using eye-tracking methodologies have investigated how students of various ages solve simple mathematical problems whilst reading. From the eye-tracking research, there appears to be some agreement on how such problems are solved. Firstly, it has been shown students usually do a scan through of the problem initially (Hegarty, Mayer and Green, 1992), suggesting they analyse the problem first and quickly look through the text for potential solutions. Secondly, more complex problems require a higher overall dwell time (de Corte, Verschaffel and Pauwels, 1990). Finally, it has been suggested students have their own particular ways of solving problems and seem to have their own preferences for the layout and structure of a problem (Bethell-Fox, Lohman and Snow, 1984; Verschaffel, de Corte and Pauwels, 1992). These studies have focused on the reading processes whilst trying to understand general mathematics problems, although some recent studies have focused on the way students read mathematical proofs.

Lin, Wu and Sommers (2012) asked 50 undergraduate students to read two geometry proofs of differing levels of difficulty. The proofs also contained diagrams for reference but no comprehension test was asked of the students. The analysis revealed the “harder” proof required the students to make a greater number of saccades when compared to the “easier” proof. Secondly, there was a greater than average ratio of fixation time on the figures when compared with texts of a similar layout, such as adverts or scientific texts. It does perhaps seem natural that a student would spend a lot of reading time on a diagram if it is provided and these results are indeed consistent with previous findings (e.g. de Corte et al., 1990). This perhaps implies that a greater number of saccades can be considered as a greater attempt at connecting information, as suggested previously. It should however be questioned how the researchers defined one proof “harder” than the other since no comprehension test was administered and there was no basis provided for reasoning that one proof was “harder” than the other.

A different investigation by Inglis and Alcock (2012) measured dwell times on the formulae and non-formulae within the proof presented to the participants and showed that the undergraduate students in the study, when compared with the mathematicians in the study, spent “proportionally more time fixating on ‘surface features’ of arguments, suggesting they attend less to logical structure . . .” (p.358). Furthermore, students in
the study made significantly fewer between-line saccades than did the mathematicians, suggesting the mathematicians appeared to dedicate more effort in inferring the implicit warrants. Although these results suggest students read proofs significantly differently to mathematicians, no attempt has yet been made to investigate why this is the case or indeed see if an intervention can change students’ proof reading processes. Study 2 investigates the latter by determining the effect self-explanation training has on the proof reading processes of students.

This section of the literature review on measuring proof comprehension and proof comprehension processes has provided evidence for using two distinct methods for investigating undergraduate proof comprehension. It is clear from the previous research conducted that using eye-trackers and eye-tracking methods can provide one with a good insight into the reading processes of students. Furthermore, the previous eye-tracking research provides a framework for analysing the data: looking at mean fixation durations, to determine cognitive effort, and saccades, to determine any attempts at connecting two or more lines of the proof together. However, eye-tracking data does not provide information about the reasons why a reader was attending to a piece of information or about the success or failure of their attempt to process the information (Inglis and Alcock, 2012). Therefore eye-tracking data should be complemented with other performance measures (Hyönä, 2010) and it would be useful to obtain a framework for creating such comprehension questions; the model suggested by Mejía-Ramos et al. (2012) has the potential to do this. It should now be clear to the reader that using eye-tracking methods and using a framework to produce comprehension tests can help measure students’ proof comprehension. Furthermore, it should be clear from the first section of this literature review why investigating proof comprehension is important and what to consider when conducting research into improving proof comprehension of undergraduates. What follows now is how proof comprehension could be improved, reviewing some previous methods and also exploring some potential new ideas using work from other fields of research.

2.6 Methods of improving proof comprehension

Having established methods for measuring students’ understanding of proofs, this review now moves on to discuss some of the potential methods, and some of the methods already tested, for improving students’ understanding of mathematical proofs. Indeed, how the proposed alternatives to formal proofs present mathematical information to students is an important, open research question (Lai, Weber and Mejía-Ramos, 2012).
This section is split under two distinct headings that are determined by the methods of improvement that have been suggested previously in the literature. The first subsection discusses ways of changing the presentation of proofs. These methods include changing the structure of the proof, providing examples within the proof to aid the reader and, finally, providing added audio and visual commentaries with the text. It will be shown that these previous methods of changing the presentation of the proof text to potentially improve proof comprehension did not obtain the results we perhaps would have liked. Therefore, the second subsection considers a different approach to improving proof comprehension, discussing methods of changing students’ engagement with mathematical proofs by focusing mainly on the use of a technique called *self-explanation training*. To begin though, there is a very brief discussion on why changing the presentation of a proof has potential, given that proof texts are a special form of text.

**Proof texts: A special form of text**

Within this subsection there are many different suggestions on how we could improve proof comprehension by changing the presentation of the proof texts. These suggestions are taken both from the proof comprehension literature and from literature in other fields. However, the literature from other fields has mainly focused on researching ways of changing the presentation of general texts such as those you would find in newspapers, books, and articles. Proof texts are different to these texts as they are a series of statements which are logically connected, through warrants, which show a theorem is true. Although general texts may constitute arguments, the order of the argument may be non-essential to the argument itself. On the other hand, proof texts could be considered as mathematical arguments that are usually written in a logical order and require the reader to apply logical reasoning to understand. Clearly any research on general texts cannot automatically be applied to proof texts. However such research does provide possible areas for exploration with proof texts and the first text-based manipulation that is discussed below is *structured proofs*.

**Structured proofs**

Generally, proof texts are presented to students in a very linear fashion, usually proceeding unidirectionally from the hypothesis to the conclusion. Leron (1983a) claimed that this traditional format is unsuitable for communicating mathematics. He suggested a format for proof texts that is arranged into different “levels”. Within each of the levels
there exists short “modules” which embody one of the major ideas of the proof. At the
top level, the main line of the proof is described and as one descends through the levels
of the proof, more elaborations and details are added until the proof is “watertight”.
Figure 2.7 shows a diagram taken from Leron (1983b) comparing the linear method to
the structured method.

Figure 2.7: Leron (1983b, p.175) comparison of linear proofs and structured proofs

According to Leron, the way proofs are currently taught is similar to what his alter-
native presentation of proof texts provide. He claims that currently it is the teachers
who provide the structure to the proof through their presentations and explanations.
Even so, the traditional approach still provides students with difficulty in understanding
proofs. Transferring the explanations into the proof presentation itself however may
help students. Despite this, Leron himself stated “I do not know of any way to prove
(or disprove) my claim on the merits of the structural method” (p.176). By using case
studies as a “powerful...subjective tool...” (p.176) to test the claims, Leron stated the
benefits of using such a model of presenting proofs are that, firstly, structured proofs are
more communicative proofs and, secondly, new learning activities are provided, such as
asking students to discover the lower levels when given the higher levels.

As one can see from Figure 2.7, the structured proof format suggested by Leron has
many levels. The reader is required to jump around the proof as they descend through
the levels, which is slightly removed from the traditional approach to presenting proofs.
More recently however, Lamport (2012) suggested a similar form of structured proof,
but one that kept the more traditional linear presentation, stating: “making proofs
easier to understand is easy. It requires only the simple application of two principles:
structure and naming” (p.43). The suggested structure by Lamport keeps the formal
structure of the proof but adds extra detail for each line in the form of a sub-proof,
giving a hierarchical structure. Each step in the hierarchical structure proves the main
step from the formal proof. Below is an example that shows the hierarchical structure, which Lamport called “fixing the proof”.

**Corollary** If $f'(x) > 0$ for all $x$ in an interval $I$, then $f$ is increasing on $I$.

1. It suffices to assume
   
   (a) $a$ and $b$ are points in $I$
   (b) $a < b$

   and prove $f(b) > f(a)$.
   
   **PROOF:** By definition of an increasing function.

2. There is some $x$ in $(a, b)$ with $f'(x) = \frac{f(b) - f(a)}{b - a}$.
   
   **PROOF:** By assumptions 1a and 1b, the hypothesis that $f$ is differentiable on $I$, and the Mean Value Theorem.

3. $f'(x) > 0$ for all $x$ in $(a, b)$.
   
   **PROOF:** By the hypothesis of the corollary and assumption 1a.

4. $\frac{f(b) - f(a)}{b - a} > 0$.
   
   **PROOF:** By 2 and 3.

5. Q.E.D.
   
   **PROOF:** Assumption 1b implies $b - a > 0$, so 4 implies $f(b) - f(a) > 0$, which implies $f(b) > f(a)$. By 1, this proves the corollary.

In considering structured proofs as a method of improving understanding, there are two issues that could be raised. Firstly, although structured proofs add more information to the proof with the aim of making it easier for the reader to understand, there is clearly a potential for an expertise reversal effect to occur for both suggested layouts described above. In the expertise reversal effect, external information becomes redundant relative to a particular learner’s internal knowledge structures (Kalyuga, 2007). Kalyuga, Ayres, Chandler and Sweller (2003) explain that “when an instructional design that includes guidance is beneficial for novices (resulting in better performance when compared with performance of novices who receive a format wherein such guidance is omitted) but disadvantageous for expert learners (resulting in poorer performance when compared with performance of experts who receive a format wherein such guidance is omitted), we have an example of the expertise reversal effect” (p.24). In other words, the instructional
 technique that is better for one group is actually less effective for the other. Structured proofs may provide benefit through the extra information for students when they are inexperienced but they may find the extra information redundant as they gain more experience.

Secondly, it has been shown that students experience difficulty with this format of proof presentation. Fuller, Mejía-Ramos, Weber, Samkoff, Rhoads, Doongaji and Lew (2011) found that students had difficulty with the format of structured proofs because they were unfamiliar with their layout and structure. Furthermore, students in the study found that they were required to make large jumps around the proof shown to them and between different sections of the proof also, making it more difficult for them to comprehend. Indeed, the results showed that students who worked with the structured proof performed slightly worse, in comparison to those students reading a more traditional linear proof, on comprehension questions relating to justifications, transferring ideas to another proof and illustrating ideas of the proof using examples. Students reading the structured proof did however perform better at summarising the proof. These results suggest that although there are some promising signs for improving proof comprehension, the benefits do appear to be minimal and sometimes understanding may even be reduced when compared with traditional proof formats. Perhaps, therefore, an alternative format should be considered.

**Generic proofs**

Recall that it has been shown that some students use examples to show that a theorem holds (e.g. Chazan, 1993). It has been suggested that instead of trying to change how these students approach proofs, their approach should be developed in order to improve their understanding. This subsection describes the reasons for considering teaching proofs by example, or *generic proofs*, and discusses the research already conducted. To begin, there is an explanation of what a generic proof is, followed by a discussion on why presenting generic proofs has been suggested as a method of improving proof comprehension. To conclude, the theoretical arguments and the limited research investigating the use of generic proofs as a method of presenting proof texts is discussed. It will be shown that the use of generic proofs to present proof texts are potentially an effective method of presenting proofs, but alone they appear to not be as effective in improving students’ overall proof comprehension skills.
A generic proof is a proof or argument “...based on the use of examples offer[ing] confirming, yet incomplete, evidence, that a mathematical claim is true” (Stylianides, 2007, p.298). The examples illustrate the argument in a particular case which are seen as a prototype (Tall, 1999). Furthermore, “a generic proof aims to exhibit a complete chain of reasoning from assumption to conclusion, just as in general proof, . . . making the chain of reasoning accessible to students by reducing its level of abstraction” (Dreyfus, Nardi and Leikin, 2012, p.204).

The example used to achieve any reduction in the level of abstraction must not rely on any properties of the specific examples being tested. Indeed, Mason and Pimm (1984) state that when considering a generic proof in number theory, the generic proof must nowhere rely upon any specific properties of the number that the proof is given in terms of. The authors go on to suggest a generic example is an actual example, but one presented in such a way as to bring out its intended role as the carrier of the general proof.

There are several reasons for considering the presentation of generic proofs as a potential method for improving proof comprehension. Firstly, one often provides counterexamples to show that a mathematical definition or claim does not hold for certain mathematical objects. The use of counterexamples is therefore common in advanced level mathematics and so, generally, the use of examples is unavoidable (Alcock and Inglis, 2008).

Secondly, Alcock and Inglis (2008) suggested that using examples might be used to explore extensions of the concept definition, which is “...a form of words used to specify [a] concept [that] may be learnt by an individual in a rote fashion or more meaningfully learnt and related to a greater or lesser degree to the concept as a whole” (Tall and Vinner, 1981, p.152), in order to obtain a better concept image, which is the “...the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes ...built up over the years through experiences of all kinds, changing as the individual meets new stimuli and matures” (Tall and Vinner, 1981, p.152). Ensuring that the concept image is consistent with the definition is important for mathematical development. Dahlberg and Housman (1997) also argued that generating examples for understanding formal definitions can facilitate further work with the concept. However, more recent work comparing participants who were asked to generate examples to participants who were asked to read examples provided for them has shown that “...simply asking students to generate examples about a concept may not substantially improve their abilities to write proofs about that concept, at least not more so than providing students with examples to read.” (Iannone,
Inglis, Mejía-Ramos, Simpson and Weber, 2011, p.11). Conversely, it has been suggested that generating examples are useful for working with proof when students are aware of the examples and have the choice to construct them for themselves (Sandefur, Mason, Stylianides and Watson, 2013). Although it seems the utility of example generation has caused some disagreement, it appears all authors consider examples important for learning.

Finally, examples are used by students of all levels at university. At the undergraduate level, students who are beginning to work with formal definitions have reported that they actively use examples (Alcock, 2008; Weber, Alcock and Radu, 2005). At the postgraduate level, a case study of two doctoral students showed that one often used examples although the other almost never used examples (Alcock and Inglis, 2008). Furthermore, some mathematicians have reported that they use examples to gain understanding when approaching a proof comprehension activity (Alcock, 2008). Weber (2008) showed examples of cases whereby mathematicians used generic examples when evaluating the validity of possibly incorrect proofs. Although the above does show that successful mathematicians do not avoid using examples completely, case studies have shown they appropriately qualify their conclusions when using empirical reasoning (Inglis et al., 2007). Perhaps it is the appropriately qualified conclusions that separate mathematicians’ understanding of proofs to the understanding of students who use empiricism. However it does appear that example generation can support understanding of general statements for both students and experts alike, since examples appear to be used at all stages of mathematical development.

From the literature discussed above, it seems reasonable to suggest that there are good reasons to believe that generic proof presentations may improve proof comprehension. Currently however, the majority of research exploring the use of generic proofs has been non-empirical, with examples of the potential uses of generic proofs proposed to suggest and support the theoretical work (Dreyfus et al., 2012). Balacheff (1988), for instance, suggested that generic proofs can assist the transition from pragmatic, example-based proof to an intellectual, general proof. He suggests the generic character of the situation provided by generic proofs aids transition because a generic proof makes the reasons for the truth of the general assertion explicit.

The theoretical work of Rowland (2002) does however provide some general principles when producing generic proofs. Firstly, the particular case should be neither trivial nor
too complicated. The example should also be “tracked” through the proof and emphasise the major aspects of the proof. The reasoning employed should be constructive and, finally, one should consider extra “scaffolding” for novice students as necessary. Although the general principles were created from work in number theory and were not backed with a systematic empirical study of comprehension, Rowland stated there is grounds for considerable optimism. He claimed generic proofs provide the possibility that students might “see” the generality intended for them in arguments that educators want their students to see.

Although the majority of research into the use of generic proofs has not been empirical, there have been two small-scale empirical studies conducted. Malek and Movshoutz-Hadar (2011) presented ten first-year engineering students with either a formal proof or a generic proof, called a “transparent pseudo-proof” by the authors, in each of the five one-to-one study sessions held with the researchers. Results showed exposure to the transparent pseudo-proofs supported students, on the whole, in their articulations of the main ideas of the proof.

Weber, Fuller, Mejía-Ramos, Lew, Benjamin and Samkoff (2012) found also that generic proofs did seem to slightly improve students’ proof comprehension. The students reported they understood and were convinced by the generic proofs presented in the study and did reasonably well in the comprehension tests set for the proofs. However, this was a small scale study with only 10 participants and there was no control condition to compare the intervention with as this was just a preliminary study.

Despite some promising signs, when considering using generic proofs as a form of proof presentation, there are several issues that need to be addressed. Firstly, recall that using a set of examples to check a theorem holds true for an infinite set has been defined as naïve empiricism (Balacheff, 1988). Proof by example is not a formal method of proof in the conventional sense and within the mathematical community it has been suggested there is a commonly held view that generic proofs are formally inadequate (Rowland, 2002). Indeed Rowland (2002) also suggests that generic proofs are just a step on the road to formal proof, offering explanation and illumination, and should perhaps be complemented with further instructional techniques.

Secondly, case studies have shown generic proofs can provide enlightenment (e.g. Kidron and Dreyfus, 2009; Rowland, 1998: Tall, 1979) but the empirical research is extremely limited (Dreyfus et al., 2012). Although it could be argued that generic proofs facilitate
understanding of logical relationships that relate to the particular examples provided in
the generic proof, they do leave the reader to determine the application and generalisabil-
ity of those relationships to other proofs. Therefore generic proofs alone may arguably
not improve the proof comprehension of students in the way educators and mathematici-
cians would like. An alternative is to, as suggested above, complement generic proofs
with an alternative instructional technique or to look at changing the way students en-
gage with proofs. More evidence for this suggestion comes from the literature on the
final form of proof presentation considered by this literature review: e-Proofs.

**e-Proofs**

The final method of presentation considered by this literature review is the use of e-
Proofs. In this subsection there is a detailed description of e-Proofs and how they have
been considered in previous research. Despite early arguments for the use of e-Proofs
being promising, the subsection concludes by suggesting that e-Proofs could be consid-
ered slightly inferior to traditional learning methods.

Technology has advanced dramatically over the last decade which has led to a change in
how students learn at university. Many more resources are now made available online
for students to download, for example. It therefore seems sensible to consider a method
of proof presentation that embraces this change in technology. One such method has
been called an e-Proof.

An e-Proof is a computer-based presentation of a proof which consists of the proof
and theorem together with annotations and an audio explanation. The presentation is
usually a series of slides which highlight one section or one line at a time and provide
added explanations on the proof structure (e.g. logical relationships between lines and
sections). The audio commentary is replayable and attempts to capture the extra audio
explanations that a lecturer might give when presenting a proof (Alcock and Wilkinson,
2011).

Alcock (2009) suggested that e-Proofs are designed to help students understand proofs
by explicitly stating the proof structure and reasoning used without providing too much
information on the screen at a given time. The warrants contained in the logical rea-
soning of the proof that have been left for the reader to infer are essentially exposed by
the added commentaries. It could therefore be suggested that e-Proofs make implicit
warrants explicit and add further information through the replayable audio and minor annotations.

In principle, e-Proofs seem a sensible method of improving students’ understanding of proofs. Indeed, an e-Proof directs the students’ attention more precisely, highlights logical relationships, highlights the straightforward annotations, and the audio adds clarity and navigation through the proof, or to a specific point in the proof (Alcock, 2009). However, since the reasoning one would need to understand the proof is made explicit, one would need to also consider the amount of work the student needs to do themselves (recall the warning given by Inglis and Alcock (2012) regarding this issue).

Roy et al. (2010) investigated the benefits of e-Proofs, compared to reading a textbook proof or a lecture, both in the short term and after a two week delay. 80 students studying for a single or joint honours mathematics degree at Loughborough University took part in the study with 28 students in the reading group, 21 students in the e-Proof group and 31 students in the lecture group. All students were given their respective versions of the generalised mean value theorem and proof, which had previously been unseen. The comprehension test framework suggested by Conradie and Frith (2000) was used to create and ask nine comprehension questions immediately after the study and in the post-test. There was a methodological issue with the study, however, since the lecturer took longer than the allotted time to teach. This meant the lecture group spent longer on the task than the students in the e-Proof and reading groups. Nevertheless, results showed that the lecture group significantly outperformed the other two groups, but there was no significant difference between the e-Proof group and the reading group in the immediate post-test. However, the e-Proof group had the greatest drop-off in scores of the three groups between the immediate post-test and the delayed post-test.

There are several possible reasons for the results found in Roy et al. (2010). Firstly, an e-Proof is just a different form of explanation. Although, arguably, a lecturer could articulate their own understanding of a proof better than they would in a lecture situation (due to the added annotations directing the student’s attention), the explanation still relies on students’ current understanding and the effort they put in themselves to understanding the proof (Alcock, 2009).

Secondly, Roy et al. (2010) suggested that the e-Proofs do not consider an individual’s reading strategies or experience in the field of mathematical proof. For example, an expertise reversal effect may occur if students’ levels of expertise are not considered. The
presentation of an e-Proof does not change based on a student’s understanding and, therefore, since e-Proofs offer extra information aimed at guiding students’ understanding, perhaps only novice learners benefit from e-Proof presentations. Ideally, we would like a method of improving proof comprehension for students of all abilities.

Roy et al. (2010) concluded by stating that the lecture group may have performed best because they had to make more logical connections for themselves. The authors therefore conjectured that if students were forced to make their own logical connections when trying to understand a proof then they would improve their understanding. As stated in the previous subsection, this suggestion by Roy et al. (2010) adds further evidence for considering a change in student engagement rather than changing the presentation of the proof text itself. One such method is through self-explanation training. Self-explanation training and changing student engagement with proofs is the main focus of the next section of the literature review.

Changing students’ engagement

All of the approaches discussed in the previous section 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 replayable annotations and commentary. Changing the presentation in such ways would also require a rather substantial effort on the part of the instructor. Given the evidence presented in the previous section that changing the presentation of the proof text appears to be, in the most part, not as successful as we would like at improving students’ proof comprehension, it would suggest that the time spent by the instructor in changing all of the materials may not be time well spent. In this next section, an alternative approach is explored in which the extra or different explanations are generated not by an instructor, but instead by the reader. This alternative approach is called self-explanation training.

Self-explanations and self-explanation training

Self-explanation training is a pedagogical tool that can be used to enhance learning, which can be defined as generating explanations to oneself in an attempt to make sense of new information (Chi, 2000). When using self-explanations, a learner will use their existing knowledge on a topic to explain to themselves what any new content in a text
is telling them. Research into self-explanation training suggests that using previous knowledge of a subject should provide learners with a platform to produce explanations and draw conclusions from a text which contains new content. Thus, producing self-explanations creates the potential to produce a greater understanding of the new content. In this section, it will be shown that self-explanations can be considered as a form of constructive learning, which is assumed by some to be a superior form of learning (e.g. Chi, 2009). There will then be a discussion on the development of self-explanations and self-explanation training through the previous research conducted in various fields. The effects of self-explanations have also been tested in the field of mathematics, with differing results, and these are discussed in detail. Some potential drawbacks to the use of self-explanation training are also considered before concluding that self-explanation training appears to be a sensible method that one can consider for improving undergraduate proof comprehension.

It has been suggested that self-explanations can be seen as a type of constructive learning (Chi, 2009). In constructive learning, learners produce an output (either written or oral) which often, but not always, contains new content-related ideas that go beyond the information given. The other two types of learning are active learning and interactive learning. In active learning, a learner will be actively doing something such as underlining or highlighting, but will only be producing outputs that are part of the original material. In interactive learning, the learner will discuss a topic with someone else and interact with that person’s contributions and ideas. Interactive learning can therefore be useful if one is in a group situation. However, when one is learning on an individual basis, active learning and constructive learning techniques are available to the learner. Chi (2009) suggested constructive learning is superior to active learning and since self-explanation training could be considered a form of constructive learning, it may have the potential to be a useful tool for learning on an individual basis.

There have been several studies that look into the use of self-explanations and providing self-explanation training to students outside the domain of mathematics. Chi et al. (1989) were some of the first researchers to use the term “self-explanations”. They prompted undergraduate students in their study to self-explain and the students were asked to read chapters in a book on Newtonian mechanics. The researchers encouraged the students to create their own explanations, and subsequently gave them 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 than did the “poor” students (who had a mean success of 46% on the problems). Self-explanations in the study were defined to be interpretations of what had been read that involved information
and relationships beyond those literally contained in the text. The researchers raised the possibility that constructing self-explanations is one of the signatures of effective reading comprehension. However, it is interesting to note that there was no time limit on the study and time was not considered as a potential confounding factor, despite the “good” students having more time on task. Chi et al. claim that the “good” students spent longer because it was their choice to do so, “…rather than a tendency to dwell unnecessarily on the examples…” (p.160). Despite this claim however, it could be argued that the “good” students performed better simply because they took more time to understand the content.

In their later work Chi et al. (1994) investigated whether students could be taught to self-explain. The students were split into two groups: the self-explanation group and the reading group, and each group read the same passage containing information on the path the blood takes around the body. Each of the students in the self-explanation group were told to read the 101-sentence passage, where each sentence was written on a separate page, and invited to explain what each sentence meant. The reading group were told to read the passage twice. All students then took a post-test that involved answering 23 comprehension questions and drawing the blood-path diagram. The results showed that students in the self-explanation group understood significantly more than the control group. However students who were able to produce more self-explanations had an improved comprehension, regardless of the experimental group they were in.

Ainsworth and Burcham (2007) investigated the nature and quality of self-explanations produced by students. A training booklet was produced that instructed students how to self-explain and the reasons why self-explanations are beneficial. The students were also provided with examples in the booklet of other students’ self-explanations in order for them to see how the self-explanations should be done. All students in the study were asked a series of comprehension questions to test their knowledge and understanding of a text on the circulatory system. The results showed that not only did those students who produced more self-explanations understand the material the best, those students also produced higher quality explanations. In order to judge and analyse students’ explanations, a coding scheme using eight headings was used. The eight headings were principle-based explanation, goal-driven explanation (an explanation that inferred a goal to a particular structure or sentence (Ainsworth and Burcham, 2007, p.9)), elaborative explanation, noticing coherence (noticing any connections between previous and current material), monitoring negative/positive (stating they did not/did understand the material), paraphrasing, and false self-explanation. Students who produced more “false” self-explanations scored lower on the comprehension test whereas students who
produced more positive monitoring and principle-based statements performed better. One would perhaps expect these results since false explanations clearly reveal a lack of understanding and producing positive monitoring statements reveal confidence in one’s ability, albeit that these statements are perhaps not as beneficial to understanding.

Self-explanations have also been shown to have positive effects in other domains including history (Leinhardt, 1993), programming, and multimedia (Roy and Chi, 2005). Furthermore, there have been several studies conducted that investigate the effects of self-explanation training in a mathematical domain. Wong et al. (2002) used 47 grade nine high school students (who were in the top classes for mathematics in their year group) to investigate the effect of self-explanation training on understanding geometry texts. Students were split into two groups and participated in six training sessions, with the sessions for each group being the same except for one. In this exceptional training session, the explainers heard examples of self-explanations given on a tape whilst the control group had think-aloud training before applying it to textbook problems. It could be argued that being given training in think-aloud protocols is much the same as being given training in self-explanations. Nevertheless, all students then took a post test where they were asked open ended questions. The analysis conducted showed that the self-explanation group had a “meaningful advantage” (Wong et al., 2002, p.453) and performed significantly better at the post-test questions. The results from Wong et al. (2002) therefore suggest that the use of self-explanations seem to improve high school students’ problem solving and general mathematical skills.

Rittle-Johnson (2006) similarly found that prompts to self-explain led to greater learning in the context of instruction about mathematical equivalence. The students involved in this study were of third to fifth grade level and those students in the self-explanation group were given an intervention of self-explanation training through practice problems. When students had read through a problem they would be shown two answers, one being correct and the other incorrect. The experimenter would then verbally ask how and why the answer given on screen was created. Students in the control group just saw the correct answer and had no explanation asked of them. A post-test was taken and students in the self-explanation group were shown to once again outperform those in the control group.

Other studies that have investigated the effect of self-explanations on mathematical understanding have also found positive results (e.g. Mitrovic, 2005; Renkl, 2002) and some
studies have also considered how self-explanation training can be implemented successfully in the mathematics classroom using computer software (Aleven and Koedinger, 2002). However, some studies have shown no effect of self-explanation on mathematical understanding. For example, Matthews and Rittle-Johnson (2009) showed that children who were given self-explanation prompts whilst receiving instruction about mathematical equivalence did not significantly outperform children who did not receive the prompts. Matthews and Rittle-Johnson claimed these results show that the benefits of self-explanation may vary with the type of instruction provided. They suggested that conceptual instruction alone may be sufficient enough to promote understanding.

Some other studies investigating the effect of self-explanations on mathematical understanding have even shown negative effects (e.g. Berthold and Renkl, 2009). These negative effects are possible because there are several issues with self-explanations that one needs to address. Firstly, a learner may rely on self-explanations alone. Relying solely on self-explanations has its disadvantages because the learner may not be able to self-explain a specific solution step or the learner may give an incorrect self-explanation. Indeed, “if the self-explanations produced by the learner are nonsensical, irrelevant or [include] verbatim utterances, then a learner is merely being active and not constructive” (Chi, 2009, p.78). However, if students are prompted to do so, they may be able to produce good explanations. It is therefore possible that some previously successful studies into self-explanations may have only been successful because of the prompts to produce explanations given by instructors to students that were beyond the initial self-explanation training. For the purposes of this thesis, self-explanation training is provided to students through a training booklet that is either on screen or on paper. Students are not prompted to produce verbal or written explanations but are asked to use the training they have been given to understand the proof. By doing so, this perhaps provides a fairer reflection of the effects of self-explanation training on proof comprehension.

However, Schworm and Renkl (2006) suggested that “self-explanations must be prompted or elicited in order to counteract the passivity of many learners” (p.427). They went on to suggest that it may be useful for these passive learners to use worked out examples with the prompted self-explanations to ensure they gain the most benefit. Worked-out examples consist of a problem formulation, solution steps and the final solution itself. This is typically employed in mathematics textbooks in the form of a theorem being introduced followed by a worked-out example and then one or more to-be-solved problems (Schworm and Renkl, 2006, p.427-8). Learners who use worked-out examples, it is suggested, will then be able to use prompted self-explanations to enhance their learning experience. Although it is true that the learner is required to produce the explanations
for themselves, it is questionable whether worked examples could be used with proofs in the same way; proofs generally relate to a specific theorem or premise.

Similar suggestions have been made regarding novices. It has been shown that, in the domain of statistics, students with low prior knowledge learn best from worked-out examples whereas as high prior knowledge students learn best from producing self-explanations; a form of expertise-reversal (Leppink, Broers, Imbos, van der Vleuten and Berger, 2012). Indeed, in the expertise-reversal literature it has been suggested that learners “...with considerable prior knowledge [who are given] strong guidance while learning...[have] most often found [it] to be equally effective as unguided approaches” (Kirschner, Sweller and Clark, 2006, p.84). Hence, low prior knowledge students should perhaps therefore begin by learning from worked-out examples before proceeding to use self-explanations.

From the discussion above, there is good reason to believe that there would be a positive effect on proof comprehension when students are provided with self-explanation training. Indeed, it has been shown through a small-scale meta-analytic study that, in general, there is a small, positive effect for students who self-explain when learning mathematics (Durkin, 2011). Furthermore, the wider-scale studies investigating the effects on text comprehension in other fields have also generally shown positive results. However, only studies outside the field of mathematics (e.g. O’Reilly, Best, and McNamara, 2004) appear to have considered the effect of self-explanations beyond an immediate post-test. Since there had previously been no study that considered the effect of self-explanation training on proof comprehension, this provided the motivation for conducting the three studies reported in this thesis.

2.7 Summary

Before moving on to describe the general methodological issues associated with the studies reported in this thesis, a list of the main points arising from the literature review are provided for the reader.

- There is currently no formal definition of proof (Selden and Selden, 2003) and there may never be a formal definition. Furthermore, proofs can come in different forms and a person’s proof scheme defines for them what proofs convince them
that a theorem is true (Harel and Sowder, 2007).

- Students encounter “textbook proofs” everyday which philosophers would typically describe as informal proofs, whilst mathematicians would describe them as formal proofs. “Textbook proofs” will be used for the purposes of the studies reported in this thesis.

- Students have their own preferences and conceptions about proof (Healy and Hoyles, 2000) and have many different beliefs and approaches to proof. For example, some students believe diagrams can constitute a proof (Weber, 2010) and some use examples to test whether a proof holds (Chazan, 1993).

- The Toulmin (1958) scheme of argumentation can provide an insight into the logical relationships of a proof and has been used throughout the mathematics education literature. This model will be used throughout this thesis to describe the logical steps in proof that prove a theorem.

- Understanding warrants, which are connections between some data and a conclusion, is important to understanding proofs as a whole and it has been suggested that students need to undertake a three-step process involving the warrants in order to validate proofs (Inglis and Alcock, 2012). The three-step process is: determining when a warrant is required, inferring the warrant, and judging the validity of the warrant.

- Methods of measuring and modelling proof comprehension have been suggested in the literature. These methods have been suggested at the high school level (Yang and Lin, 2007) and the undergraduate level (e.g. Mejía-Ramos et al., 2012). The model suggested by Mejía-Ramos et al. (2012) will be used in the studies as a framework for producing the comprehension questions that test students’ understanding of the proofs used.

- Eye-tracking studies have been useful in previous proof comprehension studies to obtain insight into the reading processes of students (e.g. Inglis and Alcock, 2012; Lin et al., 2012). In this thesis, eye tracking is employed in Study 2 to determine the effects of self-explanation training on reading strategies.
• Previous attempts at improving proof comprehension have not been as successful as we would like. They have focused on manipulations of the proof texts by creating modules (Leron, 1983b), providing generic proofs that use examples to illustrate the logical relationships in the proof (Rowland, 2002), or adding extra verbal or written instructions with the proof text, known as e-proofs (Alcock and Wilkinson, 2011).

• An alternative is to consider trying to change the way students engage with proofs. A method of doing this is through self-explanation training whereby the student combines their own knowledge with the information provided in the text to create their own explanations. This method has been shown to be successful in the domains of biology (e.g. Chi et al., 1989), history (Leinhardt, 1993), programming (Roy and Chi, 2005), and with more general mathematics texts (Wong et al., 2002).

• No study had previously investigated the effect of self-explanation training on proof comprehension at the undergraduate level and this provided the motivation for the studies reported in this thesis.
Chapter 3

Methodology

When undertaking any research project, it is of course vital to ensure that appropriate and well thought out methodologies are used. Therefore, before discussing the research studies undertaken for this thesis, general methodological issues are discussed. These general methodological issues relate to all of the studies reported in this thesis and scientific investigations more generally. The chapter begins by looking at the philosophies underlying science, leading to the reasons for using an experimental design in all of the work reported in this thesis. This is followed by a discussion on the quantitative methodologies used whilst undertaking the research studies described later. Also, since eye-tracking featured heavily in Study 2, general eye-tracking methodological issues are discussed. There is then a brief discussion relating to ethics, and to conclude, there is a brief introduction to the studies designed to investigate undergraduate proof comprehension and the ideas that pin those studies together.

3.1 The philosophy of science and experimental design

To conduct a research study there are fundamentally two different approaches a researcher can adopt in order to answer their research question. The first approach is through observation where things that happen naturally in the world are observed without interference or manipulation. This approach is called the correlational or observational method (Field and Hole, 2003). The second approach is to change something in an environment and report the effects such a change causes. This approach is called the experimental method (Field and Hole, 2003). The observational method does have some merit because the subjects in such studies are seen in a naturally occurring environment and therefore one may expect the observations made could be more easily applied to
other situations. However, a lot of modern scientific enquiry is based upon experimental methods because of the framework for scientific discovery and the philosophies underlying science.

According to Popper (2002) there is a difference between scientific and non-scientific discovery. Those statements that have empirical evidence for verification are scientific statements. Non-scientific statements are statements that cannot be empirically verified. So, for instance, the statement: “Drinking tea before doing mathematics improves your understanding,” is a scientific statement because one can get empirical evidence for such a statement. On the other hand, the statement: “Mathematics is better than Physics,” is non-scientific because one cannot measure how much more mathematics is better than physics. Although observational methods can provide empirical evidence (one can watch people do mathematics after they have had a cup of tea), these observations may not provide a fully scientific discovery. In order for a fully scientific discovery to be made, causality may also need to be considered.

The data obtained from research studies can often be broken down into a cause and effect, known as the variables, and the goal is to determine a causal relationship between these two variables. In order to do so, cause has to precede effect. Also, cause and effect should correlate and finally, all explanations, other than the one suggested, of the cause-effect relationship must be ruled out (Field and Hole, 2003, p.13). Crucially, observational designs violate these conditions because they do not take into account the external factors that may determine the cause-effect relationship.

Popper (2002) also noticed the fundamental importance of controlling for external factors. He suggested that increasing the amount of empirical evidence for a scientific discovery is not enough to show the truth of such a discovery because external factors cannot be completely ruled out. Furthermore, Popper believed it was more powerful to disprove a theory through showing one instance where the hypothesis breaks down rather than to have many instances that confirm a hypothesis. Indeed he stated: “the game of science is, in principle, without end. He who decides one day that scientific statements do not call for any further test, and that they can be regarded as finally verified, retires from the game” (Popper, 2002, p.32). However, in order to disprove a theory one needs to ensure that the theory and the opposing hypothesis being compared only differ in one respect. Hence, all scientific studies must control for external factors. Using the mathematics and tea analogy again, it is clear that observational research does not control for external factors. If one was to ask the participant what else they
had eaten, drunk or indeed anything else they had done before doing the mathematics, then there would be a violation of the assumption of a natural environment. Therefore by not discovering what else the subject had done, one cannot rule out external factors when using observational methods.

All of the philosophical evidence above suggests that an effective method to approach conducting research studies is to use the experimental method. However it has been suggested that even this method has its drawbacks. One obvious drawback is that it only tests the causal relationships that we, as researchers, manipulate. One cannot manipulate age or sex for example. It is possible to test non-manipulable events such as age or sex through quasi-experimental designs but identifying the causal relationship is much more difficult due to “the difficulty in making an unambiguous connection between the event and the presumed cause” (Christensen, 2004, p.76).

Linked to the drawback cited above is the problem of artificiality. The claim is that the experimental approach obtains results through artificial situations where subjects are put in “...situations in which [they] are totally controlled, manipulated and measured” (Bannister, 1966, p.24). However it has been suggested that this criticism is unjustified because as long as one is careful about the conclusions drawn from such studies, artificiality need not be a problem. As Underwood stated (quoted in Christensen, 2004, p.76): “...it [is] reasonable to suspect these factors would also be important in the [everyday world]. But one would not automatically conclude such; rather, one would make field tests...to deny or confirm the inference...”.

In summary, experimental methods test different hypotheses and look for causal relationships by randomising participants to conditions, which aim to remove the effect of external factors. Generally, observational methods do not allow causal statements to be made. This is not to say however that observational methods should be rejected, as good research in mathematics education has been conducted through non-experimental approaches. For the research conducted for this thesis however, non-experimental approaches were considered not appropriate and an experimental approach has been taken in all of the studies.
3.2 Quantitative methodologies

Having chosen an experimental approach to research, all studies conducted in this thesis are quantitative in nature. Quantitative research studies establish statistically significant conclusions about a population by studying a representative sample of the population and controlling certain variables (Cresswell, 2003). This is in contrast to qualitative research which is a subjective way to attempt to explain the studied behaviour (Walsh, 2003). Although one of the studies reported in this thesis, Study 1, has what would be generally classed as having a qualitative element to it (due to the coding of the verbal protocols), the design is in an experimental setting and the codings were analysed statistically. Hence all studies reported can be considered quantitative. This section is therefore used to describe the issues surrounding quantitatively based experimental studies and how the general designs of the studies have considered and maintained the three aims of research (reliability, validity and importance).

Reliability

Reliability refers to “...the extent to which a measurement procedure yields the same answer however and whenever it is carried out” (Kirk and Miller, 1986, p.19). According to Field and Hole (2003), a study cannot possibly be important if the findings are unreliable. Therefore all possible steps should be taken to ensure the research conducted is reliable. The main threat to the reliability of the research reported in this thesis that needed to be considered was the precise measurement of the variables. If participants are unsure as to what they are being asked to do, the results may be unreliable (as well as providing a threat to the validity of the results). Furthermore, if what is being measured is not clearly defined then there would also be ambiguity in the results. Therefore all definitions and instructions are clearly defined in each study for the participants so they should be clear about what it is they are being asked to do.

Validity

Validity generally refers to “...the extent to which [the experiment] gives the correct answer” (Kirk and Miller, 1986, p.19). There are two types of validity. The first is internal validity. One obtains internal validity in experiments when the measurements obtained are caused by the experimental manipulations and not by other factors. The second type of validity is external validity. One obtains external validity in experiments
when the findings of a study are representative of a global population and not just the population under the influence of the experimental conditions (Field and Hole, 2003). It is important to therefore consider both internal and external validity threats when designing an experiment. This is now discussed, beginning with some of the more common threats to internal validity that would affect the research reported in this thesis and how the methods incorporated to the designs of the studies consider and address the threats to validity.

Threats to internal validity

1. **Group threats:** If the participants chosen are put into experimental groups and everyone in one group has the same background and everyone in the other group has the same background, but a different background to the first group, then there is a group threat to validity. For example, if in the research reported in this thesis contained only first year undergraduates in one group and only third year undergraduates in the other group, this would clearly constitute a threat to validity. To combat this, all assignments of participants to groups in the studies reported are randomised, reducing the chances of obtaining a group threat to internal validity.

2. **Maturation:** Maturation occurs when participants improve, or deteriorate, as a consequence of development and not the experimental design. Reasons for maturation can include being taught something new between tests, fatigue, practice or being alerted to the study's purpose. To control for this, the content of the modules taken by the students was thoroughly checked to ensure they would not be familiar with the material. Furthermore, materials were only released to students after the completion of the study.

3. **Differential mortality:** This threat occurs when participants drop out of a study for various reasons before completing the experiment. Clearly these participants dropping out can have a major impact on the results of the study as those participants who drop out may increase the chances of group threats, for example. Of course, no researcher can force participants to stay for the duration of an experiment and thankfully no participants dropped out during any of the studies reported. However, due to issues with data collection from the eye-tracking research conducted in Study 2, some participants’ data had to removed from the analysis of that study. This is discussed in more detail in the methods and discussion sections of Study
2.

4. **Reactivity:** When compared to non-experimental conditions, a participant may react in a different way under experimental conditions due to the pressure of doing an experiment. This is called a reactivity threat to validity. Participants may change the way they would act in a normal situation because they want to please the experimenter for example. To prevent reactivity threatening the internal validity of the reported experiments, studies were undertaken in natural learning environments, such as lecture theatres and mathematics learning support centres\(^1\). It was hoped that students would then not feel under pressure and would perform as they would otherwise. Clearly it is impossible to state categorically that the threat of reactivity was removed completely, but measures were taken to reduce the threat of reactivity.

5. **Experimenter effects:** The way an experimenter acts or appears can also effect how a participant performs. Often experimenters do not realise they are subtly and unconsciously affecting the results. Age, race, and sex may also affect results (Rosenthal, 1966; Rosenthal and Rosnow, 1969). To control for this in each of the studies, students were put under exam conditions, which they were used to, and there was no contact between experimenter and participant in the time between giving the instructions and the end of the experiments.

**Threats to external validity**

1. **Over-use of specialist participant groups:** Since the work undertaken for this thesis is to improve undergraduate proof comprehension, it is acceptable to use only undergraduate students in the studies reported. However, if say only first year students were used in the studies, then it would be difficult to conclude anything about the wider student population of mathematics students that incorporates all years of study. It has also been shown that using volunteers can cause problems. Rosenthal and Rosnow (1975) showed that volunteers recruited for studies tended to be more intelligent, better educated, more sociable and had more respect for science and scientists. Again, this would suggest difficulties in relating results to the wider population. These issues are known as the over-use of specialist participant groups threat to external validity. Although all participants in the studies

\(^1\)The only exception to this was with Study 2 which were undertaken in front of an eye-tracker in an eye-tracking laboratory. The issues surrounding this, and eye-tracking methodology in general, are discussed in detail later.
reported were from one institution, the institution is fairly typical of the institutions one would find in the UK. This is discussed further in the methodology of Study 1 and the conclusions chapter. Furthermore, a wide range of students were recruited in the studies using first year undergraduates to finalist undergraduates.

2. **Restricted numbers of participants:** It is of course vitally important that the data obtained from doing research provides enough statistical power, which is defined as “...the probability of rejecting the null hypothesis” (Christensen, 2004, p.359), to convince other researchers that a study is valid. Without enough statistical power, conclusions cannot be drawn and wider inferences cannot be made. Making a claim based upon the results of a study in which two participants took part, for example, clearly would not have enough statistical power to be externally valid. Cohen (1988) stated that the majority of experiments in the field of psychology use too few participants, reducing the statistical power and questioning the claims made by the researchers about the general population. The same could be said of the majority of other fields as well. In the same book, Cohen also provided tables that can be used for calculating power-to-participant ratios. Using these tables and conducting power analyses before running the studies reported meant that the statistical power was high enough in the studies reported because an estimation of the number of participants required could always be calculated.

**Generality**

Generality is very similar to external validity and relates to whether the findings of research studies can be generalised to other people taking part in the same, or similar, studies at different times and in different places. The studies reported use proofs taken from textbooks used in everyday teaching, with some being used for the purposes of previous studies also. Therefore, as Field and Hole (2003) state: “All other things being equal, the results from a study on basic cognitive processing [(which one could argue the studies reported are about)] are more likely to be generalisable...” (p.63).

Clearly it is difficult to completely eradicate all threats to any piece of research. However, details in each of the methods sections of the individual studies reported will show that threats to the research carried out in this thesis have been well considered and sensible steps and precautions have been taken to reduce these threats.
Quantitative study designs

Having discussed the issues surrounding the threats to quantitative research and how the methods used combat such threats, this section now moves on to discuss the general quantitative study designs used in this thesis. The studies reported have between-groups designs of various forms and so the issues surrounding using such designs are now discussed.

Between-groups design

A between-groups study randomly allocates participants into each of the different conditions and those participants are often only tested once. Each participant therefore only experiences one of the conditions. As with any study design, there are advantages and disadvantages to using such a design. One advantage is the simplicity that a between-groups design brings. Participants are randomly allocated to one of the groups and only experience one of the conditions. Issues such as fatigue and practice effects therefore do not affect between-subjects designed studies. A between-groups design also accounts for unchangeable characteristics such as sex and age.

One drawback to a between-groups design is the number of participants needed. Of course, when each participant is only experiencing one condition, one needs to double the number of participants of an experiment that exposes participants to two of the conditions. This can be seen in the studies reported in this thesis with a between-groups design. These studies have a greater number of participants than studies reported in the literature with a repeated-measures design, which exposes participants to all levels of the experimental manipulation. A second, more serious drawback is that of insensitivity. In an ideal world, putting participants into different groups and exposing them to only one condition would show that any difference between the groups is purely down to the experimental manipulations we place on them. In practice, this is not the case. As discussed with the treats to validity, there can be within-group variations causing variance or “noise” to the results. However, as stated previously, sensible precautions have been taken to reduce this factor in the study designs and the conclusions drawn from between-groups designed studies are carefully considered.
**Types of between-groups designs used in this research**

There are two types of between-groups designs used in this thesis. The first is a post-test/control group design and the second is a pre-test/post-test design.

**Post-test/control group design**

The most simple version of the between-groups design used in the studies reported is the post-test/control group design. In this design, participants are randomly allocated to either the experimental group, who experience the treatment that is being researched, or the control group, who do not experience any form of the treatment being researched. As the name suggests, after receiving the treatment or the control, there is a test that determines the effects of the treatment compared to the control.

The design described above does suffer from one weakness. The randomisation technique used to assign participants to the different groups means the experimenter can only assume that the groups are equivalent before any experimental manipulations take place. However, it is probabilistically true by definition that nothing can systematically correlate with a random variable. Of course, post-hoc tests can further reveal some answers to the equivalency question but an alternative design, also used in the research of this thesis, is described below and is called the pre-test/post-test design.

**Pre-test/post-test design**

In the pre-test/post-test design, participants are again randomly allocated to the different groups. All participants are then given a pre-test before undergoing the different treatments. Participants are then all given the same post-test to determine the effects of the different treatments.

The advantage of this design over the post-test/control group design is clear. All participants, regardless of the group they are in, are tested before any experimental manipulations take place. It can therefore be shown that the groups were comparable before being subjected to the experimental conditions so one obtains increased statistical power. However, because participants then do a post-test, it has been suggested that practice effects can affect their subsequent performance (Field and Hole, 2003). This is discussed in the methods and discussion sections of Studies 2 and 3, which is where this problem may have affected the results if it had not been considered.


3.3 Comprehension test design

Within the studies reported in this thesis there are two types of comprehension test used. The first is a multiple choice test whereby the participant has a choice of answers with only one definitive correct answer. The second is to use an open-ended comprehension test whereby the participant has to provide their own answers. This brief section will discuss the reasons for using both types of test for the studies reported in this thesis.

Multiple choice tests

There are several distinct advantages to using multiple choice tests in research studies, and for exams testing understanding of a subject more generally; multiple choice tests are objective, easy to mark, efficient, and can cover a wider range of the subject being tested (Higgins and Tatham, 2003). Clearly, there are also some disadvantages associated with using multiple choice tests, such as a greater potential for guessing and potentially encouraging surface, or rote, learning, but these can be overcome through well-designed questions (Higgins and Tatham, 2003).

For the purposes of this thesis, the main reason for choosing multiple choice questions in the majority of the studies reported is due to the framework of Mejía-Ramos et al. (2012). In their earlier paper they suggest that “…[a multiple choice test] lends itself to large scale studies on students’ understanding of proofs given in different formats” (Mejía-Ramos, Weber, Fuller, Samkoff, Search and Rhoads, 2010, p.5). The authors go on to suggest a framework for assessment of proof comprehension and provide multiple-choice answers to the different “dimensions” of their framework. It therefore seemed appropriate to use this model to create questions in each of the studies, as discussed previously.

The format of the multiple choice tests also needed to be considered. For example, how many answers should there be in each question? Indeed, studies of the type reported in Study 2 and Study 3 had not been conducted before so there was no precedent for the number of questions and the number of possible answers for each question. Originally, five possible answers for each question, including “all of the above” and “none of the above” options, as suggested in Mejía-Ramos et al. (2010), were considered. However, research by Haladyna and Dowing (1993) into multiple choice tests suggested that three
response items per question should be used for multiple choice tests, with two distractors and one correct answer. Their analysis on various multiple choice tests showed most items had only one or two working distractors no matter how many choices were offered, suggesting three items is a natural choice. Haladyna (2004) also argued that having an “all of the above” or “none of the above” option places more emphasis on test taking skills rather than testing knowledge. For example, if a student has the option of “all of the above” but is sure at least one of the possible options is wrong, they can automatically eliminate the “all of the above” option as a candidate for the right answer. Based on this argument, it was therefore decided to reduce the number of choices to three and to remove the “all of the above” and “none of the above” options for all of the studies containing multiple choice tests.

Open-ended questions

Open-ended test items provide a method for gaining insight into the knowledge of a student since the students have to provide the answers to a question themselves, unlike multiple choice tests which provide students with possible solutions. Indeed, open-ended questions continue to be favoured by test developers for mathematics tests in schools (Cooper and Dunne, 1999). However, the major disadvantage open-ended questions have are the opportunity for subjectivity in the marking (e.g. Laming, 2003). To combat this, a good mark scheme should be produced and independent checks should be carried out on the markers’ scores for each participant.

It was decided to use open-ended questions to test students’ proof comprehension in Study 1 reported in this thesis and the reasons for this are discussed in Chapter 4. The framework of Mejía-Ramos et al. (2012) was again used to help create the questions and provide some of the answers that the markers were looking for in order to give marks to participants’ answers. Furthermore, tests that students face during the course of their degrees are usually open-ended but can also be multiple-choice, hence the decision for using both test types in the studies reported.
3.4 Eye-tracking methodologies

Now that it is clear what study designs have been used and the reasons for using them, this section concerns the methodological issues of using eye-tracking in research. Eye-tracking features heavily in the work of Study 2 reported in this thesis so the general issues surrounding the use of eye-tracker and eye-tracking methodology are discussed below.

General eye-tracking methodology and this research

Eye-tracking studies generally involve changing some form of display as the treatment or manipulation being considered. Examples include varying the Graphical User Interface (GUI), which is the way a user interacts with the computer (e.g. through a mouse, keyboard or touch screen), or varying forms of visual stimulus, such as two different images. A usual method of testing participants is to ask them to execute a specific task on screen. Reaction times, error rates and measurements relating to participants’ eye movements, such as fixations and fixation durations are then analysed to determine any effects of the experimental manipulations (Duchowski, 2007, p.158).

For the research reported in Study 2 of this thesis, the analysis undertaken involved using a method including analysing participants’ fixations which, as described in the literature review, is a method used in many previous eye-tracking studies. However, the treatment that differs is the form of training each group gets rather than varying the visual display. The analysis described above then investigates the differences in reading patterns caused by this training. This is discussed in detail in Chapter 5.

There are two main options when determining where to do an eye-tracking study. The first is to do the study in a laboratory. According to Duchowski (2007), “Control is probably the chief reason for holding experiments in the laboratory” (p.160). One can often observe participants without the risk of experimenter effects and the eye-tracker used can be calibrated and set up so the desired results are achieved. Furthermore, laboratory experiments can often be a more comfortable setting for the participant as the eye-tracking hardware used doesn’t necessarily need to move around with the participant. The obvious issue with laboratory experiments is a generalisability issue.
Laboratories are artificial settings and therefore “...suffer from a lack of ecological validity...” (Duchowski, 2007, p.160).

The second option is to do the study “in the field”. This of course increases the ecological validity of the experiment but there are often constraints due to a lack of equipment, which rather dictate whether field research is possible. Although head-mounted equipment is becoming increasingly less cumbersome and more affordable (Li, Babcock and Parkhurst, 2006), it is questionable as to whether wearing a head-mounted eye-tracker doesn’t affect the way a participant performs a task.

The eye-tracking study reported in this thesis was restricted due to the equipment available. A static, Tobii studios eye-tracker was used in the study and therefore all studies were undertaken in a laboratory setting. Despite the drawbacks to laboratory based work described above, being able to eye-track students in the field would not be too distant from the conditions they experienced in the laboratory settings since students are expected to study outside the lecture theatre. Again, the methods section of the eye-tracking study conducted for this thesis discusses in detail the methods used to further combat the issues described above with using a laboratory.

Conducting eye-tracking studies

Recall that the major reason for using eye-tracking data is from the eye-mind hypothesis, suggested by Just and Carpenter (1980), which states there is a close link between the direction of human gaze and the focus of attention. One can obtain the direction of gaze through the point of regard, which is discussed shortly. However, although the research cited in the literature review does confirm the eye-mind hypothesis is positive in reading tasks, it does have some problems that need to be considered. For example, one may gaze out of the window whilst formulating thoughts. In this case, gaze behaviour is not what is being attended to. As Hyönä (2010, p.173) stated: “…[gaze behaviour] can serve as an index of current attentional processes only as the available visual environment in front of our eyes is pertinent to the task we would like to study.” Therefore, one needs to be careful when considering eye-tracking data to ensure it is fixation durations on the text that are analysed as it is there where the intake of visual information takes place (Hyönä, 2010).
When conducting eye-tracking studies, there are two types of measurement available to the researcher. The first is measuring eye movements relative to the participants head position and the second is measuring the orientation of the eye in space, also called the “point of regard” (Young and Sheena, 1975). Point of regard measurements are best when the concern of the researcher is the identification of elements in a visual scene (Duchowski, 2007) which is therefore why all measurements for eye-tracking studies in this thesis use the point of regard.

In order to record eye movements to obtain measurements, there are four broad categories of methodologies available according to Duchowski (2007). The first is called electro-oculography (EOG) (see Duchowski, 2007, p.52). EOG was most widely applied in studies conducted in the 1960s and 70s but it is still used today. It measures the skins’ electric potential differences by placing electrodes around the eye. The measurements obtained are very useful for providing eye movement to relative head position measurements but the technique can be uncomfortable and slightly invasive for the participant. Furthermore, EOG is not useful for obtaining point of regard measurements.

A second method available is called the scleral contact lens search coil (see Duchowski, 2007, p.52). This method involves attaching a mechanical or optical reference object to a contact lens worn directly on the eye, usually a wire coil. The coil is embedded in the scleral contact lens and is measured moving through an electromagnetic field to produce the results. Clearly the results obtained from this method are extremely precise as the coil moves exactly with the movements of the eye. However this is the most invasive method available and can be extremely uncomfortable for the participant. Great care and skill is required also needed when inserting the lens into the participant’s eye. Again, this method is good for measuring eye position relative to head position but it is not suitable for measuring the orientation of the eye in space.

The third method is called photo-oculography (POG), also known as video-oculography (VOG) (see Duchowski, 2007, p.53), and groups together a wide range of eye movement recording techniques that do not fit into one of the other three groups. POG measures distinguishing features of the eyes when given the different experimental conditions such as the apparent shape of the pupil, the positions of the boundary of the iris, and corneal reflections. The measurements are usually assessed manually or through watching a video playback of a recording of the eyes. The problem with this technique is that it is very prone to error as the majority of analysis is done manually. The researcher can only get data as good as the sampling rate of the video device used to record the experiment.
The final, most commonly used method is video-based pupil or corneal reflection and measures the point of regard (see Duchowski, 2007, p.54). To do so, corneal reflections, known as the Purkinje reflections or images (Crane, 1994) are measured relative to the location of the centre of the pupil, with the point of regard being measured in real time. An infra-red beam is usually shone into the eyes of the participant to obtain the reflections. Although the idea of an infra-red beam shone into ones eyes may sound painful, the participant feels no pain and does not realise that the beam is being shone into their eyes. This method is therefore the most practical and best non-invasive method available for eye-tracking studies. It is this method that is used in the eye-tracking study reported in this thesis.

To conduct experiments using a video-based pupil or corneal reflection, two options are available. The first is through head-mounted equipment. As the name suggests, this is an eye-tracker mounted to the participant’s head. The tracker is calibrated to show what the participant is seeing whilst being subjected to the different experimental conditions. This equipment is usually used in observational studies rather than experimental studies as the researcher can leave the participant to act naturally. However, head mounted gear can be used in an experimental study to fix the participants head in place, thus obtaining more accurate results (Duchowski, 2007).

The second type of tracker is a table-mounted eye-tracker. Again, as the name suggests, the eye-tracker is mounted on a table, usually in the form of a computer monitor. Participants sit in front of the monitor and follow on-screen instructions. The eye-tracker is calibrated before the study to follow the participant’s eyes but also to ensure there are consistent results with all participants in the study. This eye-tracker is arguably less intrusive than the head mounted hardware because participants are not restricted in their head movements and can therefore act more naturally. A problem that follows from the ability to move their heads freely is that if participants move so much that their eye movements fall away from the screen, the calibration can be lost and results can be misleading. Therefore it is important to ask the participant to remain focused on the screen and not to take their eyes off it during the study.

Study 2 reported in this thesis use a table-mounted eye-tracker for two reasons. Firstly, and crucially, this was the only option available at the time of conducting the research. Secondly, since participants were reading proofs, their eyes only needed to be followed for playback and it can get very tedious.
for what they were reading and not anything else. Therefore even if head-mounted eye-tracking systems were available, they were not needed for the studies reported. Study 2 discusses the set-up and use of a table-mounted eye-tracker in more detail.

### 3.5 Ethics

In order to conduct a study, participants have to be subjected to a series of different manipulations. These are necessary to further scientific knowledge by determining the effects and influences of the manipulations. But it is important to remember that these manipulations can cause negative effects for participants and that it is crucial to obtain their consent in order for them to take part in research studies. Such considerations are known as ethics and it is integral to the conduct and development of research that researchers consider all ethical issues when designing studies (Sieber and Stanley, 1978). This section therefore discusses general ethical issues that affected the studies reported in this thesis and how they were considered in the design of the studies.

According to Christensen (2004), research ethics are “a set of guidelines to assist the experimenter in conducting ethical research studies” (p.115). There are three main areas that cause the consideration of ethics in research studies. These areas are the relationship between society and science, professional issues and the treatment of research participants (Diener and Crandall, 1978).

The first area, the relationship between society and science, relates to “the extent to which societal concerns and cultural values should direct the case of scientific investigation” (Christensen, 2004, p.115). Researchers tend to adapt their research proposals towards priorities that will get them money or are determined by their own cultural beliefs, which can clearly be considered as an ethical issue.

The second area relating to professional issues concerns the issue of presenting fraudulent work. “A scientist is trained to ask questions, to be sceptical, and to use the scientific value in searching for truth...,” (Christensen, 2004, p.116) but some scientists may manipulate and change their results. This is clearly unethical and is completely against the idea of scientific enquiry and furthering knowledge. It is also seen as unacceptable and unethical for researchers to produce piecemeal (several articles based on one set of data) and duplicate publications as this can confuse the literature and make understanding
The third area relates to the treatment of research participants. Research should be conducted in a way that does not cause participants distress, pain or psychological or physical harm. In the pursuit of furthering knowledge, some researchers may not consider this as they only see the goal of improving scientific knowledge. A researcher therefore must consider the ethical dilemma of determining whether the knowledge gained from participants outweighs the cost to the research participant (Christensen, 2004, p122).

The ethical issues described above must be considered in research designs. One method of doing so is to ask the recommendations of others, such as scientists in related fields, students, or lay individuals rather than just the researchers themselves or their research colleagues. The reason for this is because the researcher and their colleagues may be so involved in the study that the potential scientific method may be exaggerated. By doing so, ethical issues can be neutrally discussed and problems can be removed from the study design.

Further to the method described above, the American Psychological Association (APA) sets out guidelines for research and publication known as the code of ethics (see American Psychological Association, 2013). Below, some of the guidelines that apply to the work conducted for this thesis are discussed.

**APA code of ethics**

_Institutional Approval:_ There is a requirement from most, if not all, institutions that doing research that involves human subjects must have such research reviewed by an independent review board. Loughborough University, which is where all of the studies reported in this thesis took place, conforms to this requirement. Thus, the studies conducted were passed and approved by the University’s ethics committee.

_Informed Consent:_ The APA code of ethics states that research participants must be fully informed about all aspects of the study because they can then give their informed consent. It is considered vital to gain the consent of a participant “…because of the sacredness of the principle that individuals have a fundamental right to determine what is done to their minds and bodies” (Christensen, 2004, p.137). Once the information...
about the study is passed on to the participant, it is assumed they can make an informed
choice as to whether to participate in the study or withdraw. However, the APA code
of ethics does understand that occasionally informed consent may not be appropriate.
For example, Resnick and Schwartz (1973) found that completely informing participants
about all parts of their study reduced the participants’ incentive to take part. Furthermore,
Rind and Bordia (1996) did not get the consent of the customers of the restaurant
used in their study. Waiters were asked to put happy faces or unhappy faces on the backs
of receipts to see which gained more tips and whether there was a difference between the
male and female waiters. Clearly, informing the customers of the purpose of the study
would have been detrimental to the study. Therefore, the APA code of ethics states that
informed consent may be dispensed with under specific conditions and only if the study
will not cause harm or distress to the participant. All studies reported in this thesis
conform to these guidelines. An informed consent form was provided for all participants
in all studies and therefore participants provided active consent in all studies reported.

Passive Deception: This refers to withholding some information from participants de-
liberately in a study for experimental reasons. Clearly this goes against the idea that
all information should be provided to the participant so they can make an informed de-
cision. But, as described above, the code of ethics states that researchers do not have to
provide all information to participants in an experiment so long as it is not expected to
cause pain or severe emotional distress. Therefore, if no pain or severe emotional distress
is expected from the study and as long as it is done for the right, scientific reasons, we
do not have to inform our participants about all parts of the study. If our scientific un-
derstanding is improved by not informing participants about all parts of the study then,
assuming the researcher conducts themselves with scientific integrity, the code of ethics
allows for passive deception. It is clear that no study reported in this thesis was expected
to cause pain or severe emotional distress, therefore some passive deception was allowed.

Debriefing: APA code of ethics standard 8.08 specifies that psychologists must debrief
participants as soon as possible after the completion of a study. This is clearly particu-
larly important if a participant has been subjected to any physical or psychological
effects. Debriefing has two main objectives which are to desensitize or de-hoax through
providing information about any deceptions used in the study. Furthermore, debriefing
can actually be beneficial to the researcher as further data can be gained. In all of
the studies in this thesis, a debrief took place with participants, either individually or
through an e-mail, straight after the completion of the study.
Confidentiality: Regardless of the research study being conducted, participants have the right to confidentiality. Each individual should only be identified in the study by, preferably, a code and real names should never be used (Christensen, 2004). Data protection must also be followed, with the laws of the Data Protection Act (1998) determining the use, storage and the amount of time before data must be destroyed in the United Kingdom where these studies took place. The present research conducted conforms to all of the above and by signing an informed consent form for each of the studies, participants were told of their confidentiality rights and also allowed the researcher to view their exam results to aid with the analysis.

It has therefore been shown that the highly important consideration of ethics has been addressed and well thought about in each of the study designs. Hence, results reported can be considered as not only scientific, but also as ethical. The studies which these results come from are introduced in the next section where there is a discussion on how each study links together to provide a picture of undergraduate proof comprehension and how this is used to provide different tests of one method of improving undergraduate proof comprehension.

3.6 Studies reported in this thesis

The aim of the research reported in this thesis is to build upon our existing understanding of undergraduate proof comprehension and to offer an insight to potential methods of helping students to improve their proof comprehension skills. Each study reported in this thesis builds upon the last by using the results and conclusions drawn to drive the development of the next study. The three studies are now introduced in brief below.

The first study investigates a method of changing the way a student reads a proof text and the processes undertaken by the student whilst reading the proof text. The method chosen was to use self-explanation training, which has been shown to be successful in other fields. It is shown that the same effects occur for proof comprehension and self-explanation training appears to significantly improve students' abilities to understand proofs. Furthermore, the study shows that self-explanation training promotes higher quality verbal explanations of the structure of the proof.
The second study reported in this thesis looks at the effects self-explanation training has on students’ reading patterns. An eye-tracker was used to compare the reading patterns of students before and after being given self-explanation training and also the difference between being given self-explanation training and not. This study revealed that self-explanation training increased levels of concentration and also the number of times students in the study moved between lines within the proofs provided.

The final study looks at the longer term effects of self-explanation training in a genuine pedagogical setting. Students were tested over a period of three weeks through one post-test and one delayed post-test during their scheduled lectures. Half of the group received self-explanation training whilst the other half received a control activity on study habits. The results of this study confirmed the findings of the previous two, suggesting that self-explanation training is a powerful pedagogical tool that appears to help students proof comprehension improve in both the short term and over the longer term also.

All of the studies described in brief above are now reported in detail in the following chapters.
Chapter 4

Study 1: Using self-explanation training to improve proof comprehension

In the literature review, it was argued that previous attempts to improve proof comprehension in undergraduate mathematics have had somewhat disappointing results. These previous investigations have taken an approach whereby the layout or presentation of the proof text was changed in some way. However, an alternative would be to change the engagement a student has with a proof text. One such way is through the provision of self-explanation training. Recall that research into self-explanation training has shown that students understand a text significantly better if they are taught to generate self-explanations whilst reading the text. A natural question is therefore this: does self-explanation training have any effect on proof comprehension? The following study reported in this chapter addresses this question.

4.1 Methods

The purpose of the study

The purpose of conducting this study was to determine whether self-explanation training can improve undergraduate proof comprehension. Using an adapted training booklet taken from Ainsworth and Burcham (2007), one could see whether the mainly positive effects of self-explanation on text comprehension described in the literature apply to mathematical proofs.
Task procedure

76 undergraduates currently studying for a maths degree at Loughborough University were tested in this study over a six week period from February to March 2012. Figure 4.1 above shows the design of the study. The participants were tested on an individual basis and were randomly assigned to either the control group (38 participants) or the self-explanation group (38 participants). These students were taken from all three years of study (26 from Year 1, 24 from Year 2 and 26 from Year 3) and from a range of mathematics courses such as Mathematics MSc, Mathematics, Accounting and Financial Management, Mathematics and Physics, and Mathematics and Sports Science. These courses all contain core mathematics modules (Calculus and Linear Algebra) as well as pure and applied modules and the students attend a mixture of lectures and small group tutorials whilst also being expected to do their own independent study. This means their academic experience is similar to that of undergraduates studying mathematics at other UK universities.

The self-explanation group were given self-explanation training through PowerPoint slides adapted from the training given to participants in the study by Ainsworth and Burcham (2007). Students were then asked to try and understand the proof chosen for the study. Participants in the control group were given a passage to read and answered some short questions on the history of right-angled triangles that was of approximately the same length as the training slides given to the self-explanation group. This ensured students in both groups spent approximately the same amount of time on task. There was no time limit to the time spent on reading, studying the proof or answering the questions; however there was a total time limit of one hour.

All participants were audio recorded throughout the study so that an accurate measure of their time in the experimental situation was obtained but, more importantly, so their explanations could be coded and analysed. Participants read through the information at their own pace, pressing the spacebar to proceed to the next slide. Going back to any
slide was not allowed. Once participants had completed the training or read the passage on right-angled triangles and answered the related questions (according to the group they were randomly assigned to), the slides told them to study the proof chosen for the study without saying anything at first. Both groups were then given the same proof again but this time with one line highlighted in red at a time. Participants in the control group were told to make any comments on the highlighted line that helped them to understand the proof whilst participants in the self-explanation group were told to try and use the training given to help them understand the proof. Pressing the spacebar moved the highlight on to the next line and a sound indicated a successful move; the participants were made aware this would happen. The sound was so that when analysing the recordings, the experimenter could tell when a participant had moved on to the next line.

The experimenter sat in the study room with each participant. Participants were told the experimenter was in the room to rectify any technical issues only and told to otherwise ignore the fact the experimenter was there. After participants had completed all parts of the study they were given £8 inconvenience allowance and once all the data had been analysed, all participants were e-mailed the results along with both the PowerPoint presentations.

**Task Construction**

To conduct a study that would answer this research question, a proof was chosen so that students from all years of study would be able to take part and understand. Finding a proof that students of all years of study would be capable of understanding meant that there was a larger pool of students to recruit participants from and also meant the results would show the effects on a larger cross-section of mathematics students. The proof also needed to require the students to use prior knowledge since one of the aims of self-explanation training is to help integrate new knowledge with existing knowledge. Furthermore, the proof needed to be of a standard that was typically found during the students’ studies and one that they had not seen before. It was agreed to use a proof from number theory that had been used by Mejía-Ramos et al. (2012) to give examples of their seven types of question for assessing proof comprehension. It was decided that since students may not have seen the terms monadic and triadic before, the definitions of these terms would be provided to the students on their first viewing of the proof. The proof is given below.
Note: We say a number is monadic if it can be written in the form $4k + 1$ and triadic if it can be written in the form $4k + 3$.

Theorem
There are infinitely many triadic primes.

Proof

(L1) 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.

(L2) Similarly, the product of any number of monadic numbers is monadic.

(L3) Now, assume the theorem is false, so there are only finitely many triadic primes, say $p_1, p_2, \ldots, p_n$.

(L4) Let $M = 4p_2 \ldots p_n + 3$, where $p_1 = 3$.

(L5) $p_2, p_3, \ldots, p_n$ do not divide $M$ as they leave a remainder of 3, and 3 does not divide $M$ as it does not divide $4p_2 \ldots p_n$.

(L6) We conclude that no triadic prime divides $M$.

(L7) Also, 2 does not divide $M$ since $M$ is odd.

(L8) Thus, all of $M$’s prime factors are monadic, hence $M$ itself must be monadic.

(L9) But $M$ is clearly triadic, a contradiction. □

To provide the self-explanation training and the proof to the students, an adapted version of the training slides created originally by Ainsworth and Burcham (2007) was produced in a PowerPoint presentation. The original presentation had been previously used in a successful study on self-explanation training and was therefore deemed appropriate to use in this study. Also, Ainsworth and Burcham tested their participants on an individual basis using the original presentation and it was therefore decided to do the same with the present study. The slides used in this study are shown below with the adaptations from the original presentation discussed after.

Slide 1

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 in the proof.
- Attempt to explain each line presented to you in terms of previous ideas. These may be ideas from the information in the proof, examples from previous theorems/proofs or ideas from your own prior 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?

You will now 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:

No odd integer can be expressed as the sum of three even integers.

Proof:

(L1) Assume, to the contrary, that there is an odd integer $x$, such that $x = a + b + c$, where $a$, $b$, and $c$ are even integers.
(L2) Then, $a = 2k$, $b = 2l$ and $c = 2p$, for some $k$, $l$, $p$ integers.

(L3) Thus $x = a + b + c = 2k + 2l + 2p = 2(k + l + p)$.

(L4) It follows that $x$ is even; a contradiction.

(L5) Thus no odd integer can be expressed as the sum of three integers. \hfill \square

After reading this proof, one student made the following self-explanations:

- The proof uses the technique of proof by contradiction.
- Since $a$, $b$, and $c$ are even integers, we have to use the definition of an even integer, which is used in line 2.
- The proof then replaces $a$, $b$, and $c$ with their respective definitions in the formula for $x$.
- The formula for $x$ is then simplified and is shown to satisfy the definition of an even integer also; a contradiction.
- Therefore, no odd integer can be expressed as the sum of three even integers.

You must also 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.

*Slide 4*

**Paraphrasing**

“$a$, $b$, and $c$ have to be positive or negative, even 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 what is already represented in the text by the words “even integers”. 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 that $2(k + l + p)$ is an even integer.”

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, $2(k + l + p)$ is an even integer because the sum of three integers in an integer and two times an integer is an even integer.

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, either in your head or by making notes in the space below the proof, using the training you have been given.

**Theorem**

$(0, \infty)$ is not bounded

**Proof**

(L1) Assume the theorem is false and that $(0, \infty)$ is bounded.

(L2) Therefore, by assumption, there exists a constant $C > 0$ such that $(0, \infty) \subset [-C, C]$.

(L3) Note, $C + 1 > 0$, thus $C + 1 \in (0, \infty)$.

(L4) However, $C + 1 > C$ so $C + 1 \notin [-C, C]$. 
(L5) This contradicts the assumption that \((0, \infty) \subset [-C, C]\).

(L6) Thus, \((0, \infty)\) is not bounded. \(\square\)

Clearly, some of the slides had to be slightly altered from those used by Ainsworth and Burcham (2007). There were three main differences between the slides used in the present study and the study of Ainsworth and Burcham. Firstly, the obvious difference was the examples of self-explanations given by students in the slides were on mathematical proofs rather than biology. Secondly, instead of the slides training students to use self-explanations on paragraphs, the training asks students to consider self-explanations on a line-by-line basis. Finally, the major difference in this study is instead of having to study various paragraphs on different subjects, the participants only studied one proof and not several proofs. It seemed reasonable to consider only one proof in a study such as this one but future studies could consider more than one proof. Other than these alternatives described, the slides followed the same structure as those provided in the study of Ainsworth and Burcham (2007) including relevant pieces of text (proofs, in this case), example self-explanations, and providing practice.

An alternative presentation of the training materials was also considered but dismissed. The alternative option would have provided the training to the students verbally, either as a group or individually; the students would then write their explanations down on separate pieces of paper. Each piece of paper would contain one line of the proof. This method would have been consistent with the work of Chi et al. (1994). However, in the study by Chi et al., the experimenter went round and prompted the students if their self-explanations were sketchy or incorrect. Students in this study were not prompted by the experimenter at all. Also, in order to simulate learning conditions familiar to the students, the participants in this current study needed to see the whole proof as it was constructed, rather than an individual line at a time. This was decided because recall that it has been shown previously that students struggle with proof formats that they are unfamiliar with (Fuller et al., 2011). Both of the above methods therefore kept conditions familiar for students in the hope that the issues found previously would not occur. Furthermore, seeing the whole proof as it was originally constructed by the author of the proof meant better comprehension questions could be asked to test participants’ understanding of the whole proof and so the PowerPoint presentation was provided to the students instead.
As in Ainsworth and Burcham (2007), an example was provided for the participants to practice with before they saw the main proof. Providing this example gave the participants a chance to apply the training before tackling the main proof. This method is in line with previous studies. However, some of those studies (e.g. Chi et al., 1994) told participants to stop making monitoring statements and paraphrasing during the practice tests and during the main study. The experimenter also often asked leading questions that would ensure participants made explanation statements. Students in this study were left to complete the study with no interaction with the experimenter (except for administration purposes). The purpose of this method was to hopefully provide a fairer reflection of the effects of self-explanation training on proof comprehension.

Participants in the control group were given a passage to read on the history of right-angled triangles that was of a similar length to the training given to the self-explanation group. Ideally, these students would have had a task that offered a different form of training that was shown previously to improve proof comprehension. However, recall that in the literature review it was shown that there was no previously accepted successful intervention that improved proof comprehension for undergraduate students. It was therefore decided to give the students a task that was mathematically based. Furthermore, since students were brought into a somewhat unfamiliar lab situation, it was considered desirable for them all to do something of equivalent length before beginning the proof comprehension task. Hence, time spent in the experimental situation could also not be considered as a confounding factor. Participants in the control group also had to answer five questions on the passage they had just read to ensure they did not just click straight through to the proof. The passage given to participants in the control group can be found in Appendix A.

Another consideration was whether the experimenter should sit in the room with the participants whilst they were doing the study. Research has shown that sometimes when an experimenter, or someone of authority, sits in a room whilst a study is conducted, the participant performs significantly differently (see Rosenthal, 1966). This is a confounding factor known as the experimenter effect. Careful consideration was given as to whether this would cause problems for this study. It was decided that the problems created if the experimenter was out of the room, such as the participant not doing as they were instructed, going back through the slides or not staying on task, outweighed the possible problem of experimenter effects. Therefore, the experimenter sat in the room with the participants throughout the study.
Creating the questions

Recall that the proof used for this study was taken from Mejía-Ramos et al. (2012). In their paper, Mejía-Ramos et al. provided example questions using this proof to show how understanding of a proof could be assessed. A selection of these questions were therefore chosen alongside new questions created based upon suggestions made by Mejía-Ramos et al. In total, 14 questions, two from each of the seven sections suggested for assessing proof comprehension, were given to the participants to answer. Questions were left open ended rather than providing multiple choice options. Multiple choice questions do provide a simpler and more secure method of marking. However, giving open ended questions is in line with previous studies that looked into self-explanation training and also meant the students were able to express their ideas freely, providing a greater depth and insight into their understanding of the proof. Each question had a maximum mark of either 1, 2 or 3 according to the complexity and there was a maximum mark of 28. The questions, along with the number of marks available for each question, are provided below.

Proof A - Open Ended Questions

1) What does it mean for a number to be triadic? [2 marks]

2) What does it mean for a number to be a prime? [2 marks]

3) What type of proof is this? [1 mark]

4) In line 3, what is the purpose of assuming that the theorem is false and that there are only finitely many triadic primes? [3 marks]

5) In line 5, why does the fact that 3 does not divide $4p_2 \cdots p_n$ imply that 3 does not divide M? [2 marks]

6) Which claim(s) in the proof logically depend on line 2 of the proof? [2 marks]
7) Which of the following summaries best capture the ideas of the proof? (Please circle the letter of your choice) [1 mark]:

A. The proof assumes there are finitely 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 $M = 4p_2 \cdots p_n + 3$, 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 $M = 4p_2 \cdots p_n + 3$ 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.

8) Summarise in your own words how the proof arrives at the conclusion that $M$ itself must be monadic. [3 marks]

9) Why does the proof include the subproof that the product of monadic numbers is monadic? [2 marks]

10) 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. [2 marks]

11) 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 marks]

12) 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? [2 marks]
13) If 3, 7, 11 and 19 were the only triadic primes, what would the value of M be? [2 marks]

14) 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. [2 marks]

Marking of question papers

In order to prevent biased marking, question papers were given to two mathematics PhD students at Loughborough University to mark. The markers did not know whether the paper had come from the control group or the self-explanation group. Markers were provided with a mark scheme to use which had been created by the experimenter in discussion with mathematics lecturers at Loughborough University. Once one of the markers had marked a paper, they gave it to the other to check. If there were any discrepancies, they were discussed with the experimenter until an agreement was reached. This ensured fair and consistent marking throughout the study. Although this protocol was put in place, the experimenter did not have to discuss any marking decisions with the markers. The mark scheme is provided in Appendix B.

Coding of verbal protocols

The verbal explanations produced by all participants were transcribed and coded according to a scheme based on Ainsworth and Burcham (2007) (see p.47). The work of Ainsworth and Burcham provided a qualitative method of gaining insight into students’ possible self-explanations for a proof. Seven of the eight categories used by Ainsworth and Burcham were used in the coding of the data in the present study; it was found that the “elaborative explanations” category was unnecessary. Some slight changes were made to the descriptions of the other seven categories to ensure they were suitable for coding verbal protocols on mathematical proofs. Each category was assigned by the investigator as either a non-explanation category or an explanation category. The seven categories used are as follows:

Non-explanation categories
False Explanations: This category was scored if a participant gave a wrong explanation.¹

Negative Monitoring: This category was scored if a participant stated “I don’t understand this”, “How is this true?” or words to that effect.

Positive Monitoring: This category was scored if a participant stated “I understand this”, “OK, this makes sense” or words to that effect.

Paraphrasing: This category was scored if a participant simply repeated the line or part of the line using similar words or the same words as used in the line.

Explanations categories

Goal Driven Explanations: This category was scored if a participant gave any explanation that related to the structure of the proof (how it is used in order to reach the goal of proving the theorem). For example, if a participant said: “OK, we’re doing this because we are going to use it later on in the proof.”

Noting Coherence Explanations: This category was scored if a participant gave any explanation that related back to a previous idea used in the proof. For example, if a participant said: “...this is because in line 5 we introduced...”

Principle-Based Explanations: This category was scored if a participant gave any explanation that was based upon definitions, theorems or ideas not explicitly written in the proof and can be considered as inferring a warrant. For example, if a participant said: “...this is because by the definition of triadic...”

False explanation, negative monitoring, positive monitoring, paraphrasing have meanings that are analogous to those used in Ainsworth and Burcham (2007). The explanation category statements differ slightly and are illustrated by the following examples taken from the study:

Goal Driven Explanations

(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.”

¹If a student made no remarks on that line it was not classed as a false explanation, it was just removed from the analysis.
This comment explains why setting up \( M \) as a triadic early on in the proof has helped to reach a conclusion later on, leading to a contradiction and the goal of the proof.

**Noticing Coherence Explanations**

(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.”

This comment refers back to previous lines in the proof to explain what the current line is stating. This is different to the goal driven explanations because it does not explicitly explain how the line relates to getting to the goal of the overall proof.

**Principle-Based Explanations**

(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...”

These comments were classed as principle-based because they involve the use of definitions that are not explicitly written in the proof.

**Calculating the number of participants required**

In order to calculate the number of participants required for this study, a power analysis using effect sizes from previous studies into self-explanation training. Cohen (1969) states the power of a statistical test is “…the probability that...[the test] would lead to statistically significant results...” (p.vii). Power is linked to the sample size of the test, the effect size, and the required probability level for any differences between two groups (for example) to be judged significant (Field and Hole, 2003). For this particular study, the standard probability level that a test is significant is set at 0.05, the
recommended power is set at 0.8 (Cohen, 1992) and evaluation of previous effect sizes found in self-explanation studies suggested an effect size of 0.58 should be expected. By using a program called “G-Power” which completes the necessary calculations for you, the power analysis suggested a sample size of 76 students.

4.2 Results

Effects of self-explanation training on proof comprehension

Firstly, the comprehension test for this study, created by using the framework of Mejía-Ramos et al. (2012), was subjected to a test of internal reliability. Internal reliability is measured by using Chronbach’s alpha and the overall Cronbach’s alpha for this comprehension test was 0.672, which can be considered as an acceptable level of internal reliability. Therefore, the analysis proceeded by using the scores from the comprehension test.

To investigate the effects of self-explanation training on students’ proof comprehension, the participants’ scores on the comprehension test were subjected to an analysis of covariance (ANCOVA) with one between subjects factor (condition: self-explanation training or control) and one covariate (proof study time). Participants’ proof study time was included as a covariate because it was found this correlated with their scores on the comprehension test, $r = 0.439, p < 0.001$. The analysis revealed a significant effect of condition, $F(1, 76) = 13.315, p < 0.001, \eta^2_p = 0.154$. This suggested that self-explanation does appear to have a positive effect on students’ proof comprehension. Indeed, participants in the self-explanation condition scored an average of $18.2/28$ (SD = 4.2) for the comprehension test compared to students in the control condition who scored an average of $14.2/28$ (SD = 4.0). The difference in scores corresponds to a large effect size, $d = 0.950$, suggesting the effect self-explanation training has on proof comprehension is both large and positive.

Critical consideration of the relationship between the self-explanation training and the proof comprehension test led to a concern that the examples in the training provided relevant practice and, in particular, an answer to one of the questions. Therefore, an added analysis was conducted with the question “What type of proof is this?” removed. This further analysis yielded the same results, showing a significant effect of condition,
In order to determine the effect of self-explanation training on proof comprehension by year of study, a 3 (year of study) × 2 (condition) ANCOVA with one covariate (proof study time) was conducted. This analysis revealed a main effect of year of study, $F(2, 69) = 3.456, p = 0.037$, which is perhaps unsurprising. Indeed, the participants in their third year of study scored an average of 17.8/28 (SD = 4.2) on the comprehension test, out-performing their colleagues in years two (M = 15.8, SD = 5.2) and one (M = 14.9, SD = 3.9). These results are consistent with the idea that students improve their mathematical ability as they go through university with Figure 4.2 showing the average scores for each of the two groups split into the participants’ year of study graphically.

![Figure 4.2: Study 1: The average score for participants in each group by year of study](image)

More importantly however was the analysis revealed no significant condition by year group interaction, $p > 0.2$ and so this shows that there is no substantial difference between the effect self-explanation training has on proof comprehension for a first year student or a final year student.

**Effects of self-explanation training on the verbal explanations given**

Participants’ verbal comments were analysed and coded according to the scheme explained in the methods section of this chapter. The median number of each type of comment given by participants in each group, as shown in Figure 4.3 above, were
Figure 4.3: Study 1: The median number of verbal “explanations” and “non-explanations” for each group compared using a series of Bonferroni-corrected Mann Whitney $U$ tests. Bonferroni-corrected Mann Whitney $U$ tests were conducted to reduce the chance of false-positive results, given that seven different verbal explanations were possible (see e.g. Cabin and Mitchell, 2000). The analysis revealed that participants in the self-explanation group produced significantly more statements classed as “explanations”; these were more comments classed as principle-based, $U = 386$, $p < 0.001$, noticing coherence, $U = 399$, $p = 0.001$, and goal-driven, $U = 407$, $p = 0.001$. Participants in the self-explanation group also produced significantly fewer comments classed as “non-explanations”; these were comments classed as negative monitoring, $U = 440$, $p = 0.002$, and positive monitoring, $U = 400$, $p < 0.001$. Analysis of the difference between the number of comments classed in the categories “false explanation” and “paraphrasing” did not reach significance. Overall, students in the self-explanation group provided an average of 11.5 comments classed as explanation-type statements which is significantly more than the students in the control group, who gave on average 5.8 comments of this type, $t(52.5) = 4.434$, $p < 0.001$, $d = 1.017$. These results therefore suggest that students who receive self-explanation training also provide higher quality explanations whilst trying to understand a proof.
4.3 Discussion

This study was the first to look at self-explanation training in the context of undergraduate mathematical proofs. The findings are consistent with the majority of the previous research into self-explanation training: self-explanation training appears to improve comprehension of texts, including proof texts. Participants who had been trained to self-explain performed significantly better on the proof comprehension test than participants who had not been trained to self-explain.

Again, in line with previous studies, it was naturally expected that this study would show that self-explanation training improves the explanations given by students. Indeed, the findings show participants in the training group produced significantly more explanations than those that did not receive the training. Furthermore, no participant in the self-explanation group was prompted to make explanation statements by the experimenter. Therefore the findings suggest that, under experimental conditions, undergraduate students can be trained to produce self-explanations that improve their understanding of a mathematical proof without prompting beyond the PowerPoint-based training. Moreover, there is a highly significant difference between the number of noticing coherence and principle based statements produced by the two groups. Successfully understanding a proof requires logical maturity (Weber, 2001) which goes hand-in-hand with understanding the structure of the proof. Increased numbers of statements recognising coherence between lines and the principles used to create the proof would imply that students are being encouraged to make better use of their existing understanding of logical reasoning. Indeed, the findings of this study suggest this.

Recall that Inglis and Alcock (2012, p.9) suggested that when considering a proof line-by-line, the reader is faced with a triple task: (i) accurately determine when a warrant is required, (ii) correctly infer the implicit warrant that the author of the purported proof intended, and (iii) evaluate the warrant’s mathematical validity. Although Alcock and Weber (2005) showed students are not particularly good at evaluating a warrant’s mathematical validity, this study suggests that with self-explanation training, students could perhaps do the first two tasks. The participants in the self-explanation group produced on average 6.0 principle-based explanations compared to 3.2 for those in the control group which perhaps suggests that they could infer and explain twice as many warrants after being given self-explanation training. Therefore, although this study did not set out to investigate whether self-explanation training can improve students’ ability to complete the triple task, it does suggest they have the capability to explain the warrants required for the proof used in this study and, furthermore, can perhaps infer
them correctly. Clearly this study does not show self-explanation training helps students improve their ability to evaluate the mathematical validity of a warrant so this is an area for future study.

The results also show that there was no significant difference between the number of false explanations or paraphrasing statements given by the two groups. A possible reason for this is while some students were thinking about their explanations they would often just repeat the line or say something before realising it was wrong. Another reason could simply be due to the experimental conditions faced by the students. Thinking out loud is quite unnatural and talking to a computer whilst someone is in the room can be unnerving. Several students commented that it was “a weird experience” once they had given their explanations on the proof. Also, two students in the self-explanation training group and three students in the control group did not say anything at all about some lines in the proof. A possible reason for this is because they were nervous. However, this should not be considered a major issue in the analysis since the differences between the groups’ scores are highly significant and also because both groups were under the same conditions.

One possible limitation with this study is the nature of the of the control study and the nature of the training. The control task has no significant relationship with proof although it does contain mathematical content. Furthermore, within the training booklet there are examples of a student’s self-explanations on a proof which includes stating what type of proof it was. Indeed, the self-explanation training gives students the opportunity to read a proof whereas the control task does not, giving the self-explanation group extra practice before the main study task and, arguably, an unfair advantage. Therefore it could be argued that the results seen in this study were down to practice or memory. For example, one question asks the student to determine the type of proof. The proof used in this study is a proof by contradiction and in the self-explanation training presentation the proof used is also a proof by contradiction. Therefore, for this question, a student in the self-explanation group has received relevant practice and may have an advantage over a student in the control group. However, recall that further analysis of the results when the question “What type of proof is this?” was removed revealed the same results. Furthermore, it should be made clear that providing students with examples to practice with and read through is an integral part of the self-explanation training. This is consistent with previous studies that have investigated the effects of self-explanation training on text comprehension (e.g. Ainsworth and Burcham, 2007). Although this suggests that it is perhaps unlikely that practice or memory is the main reason why the self-explanation group outperformed the control group, effects of practice
are still a plausible explanation. Therefore such alternative explanations are considered and investigated through the study design of the later studies reported in this thesis.

Also, although the proof was chosen for its accessibility for students of all years of study, it does not show that self-explanation training improves understanding of all proof texts. The question remains whether a result for one proof is analogous for all proofs in mathematics. Whilst the proof used in this study is typical of the proofs that might be encountered by students in an undergraduate degree, it remains possible that it has some specific features that interact with the self-explanation training in such a way as to make the training look more effective than it actually is. This study should however be considered as evidence that the training can be effective.

One should also consider who these results may apply to. The study was conducted at one university with one group of students so one cannot say for sure whether the effects seen in this study would apply to other groups of students. However, Loughborough University is a typical UK university offering a typical mathematics degree for the UK. Indeed, the university offers a wide range of mathematics degrees, including single and joint honours at both the bachelors and masters levels that take between three and five years to complete and require a grade A in mathematics at A-Level (or equivalent) on entry. The first year requires only students to obtain 40% or more to pass the year and does not count towards the final degree classification. From the second year onwards, students’ results in each of the modules taken are weighted with final year modules worth more towards the final classification than any of the previous years. Each module is worth either 10 or 20 credits and students require 100 credits to pass the year or 120 credits to obtain an honours degree. Furthermore, students on mathematics courses cover a range of core, pure, and applied mathematics modules during their degree, taught through a combinations of lectures, small group tutorials, and individual study. This set-up is fairly typical of a UK university. Therefore, in the absence of alternative research, one can argue that the results seen in this study would be likely to be seen for students in a similar university in the UK.

A final issue when considering the results of this study was that it was undertaken under experimental conditions, so it cannot be said for certain whether these effects would be seen in normal learning conditions. It would be interesting to test whether these effects do indeed repeat under normal learning conditions as the results could suggest that self-explanation training should be given to undergraduate students when they start their
Discussion, Study 1

mathematics degrees. Indeed, this issue is investigated later in Study 3.
Chapter 5

Study 2: Does self-explanation training change the way students read proofs?

The results of Study 1 into the effects of self-explanation training on undergraduate proof comprehension are a promising indicator that it is possible to improve undergraduates’ proof comprehension skills. The question remains however as to how self-explanation training actually does this. It would be beneficial to see in what ways self-explanation training changes the behaviour of the students when they are trying to understand a proof. By investigating different ways in which self-explanation training changes students’ behaviour, future studies, and indeed pedagogical instruction, can be tailored so that students gain the maximum benefit.

An obvious candidate for first investigation is to look at the changes self-explanation training the way students read proofs. Therefore Study 2 asks the following research question: does self-explanation training change the way students read mathematical proofs? Recall that larger mean fixation durations are associated with greater cognitive processing (e.g. Duchowski, 2007; Just and Carpenter, 2003) so by measuring mean fixation durations in this study, one can investigate the impact of self-explanation training on a student’s effort whilst reading a proof.

Also, recall that Inglis and Alcock (2012) showed that when students have only experienced traditional learning methods, such as lectures and tutorials, they appear to read proofs significantly differently to mathematicians. Given these results, it would be interesting to see whether self-explanation training can aid students in producing more
between-line saccades, like mathematicians appear to do, as this has been argued by researchers to show greater attention to the logical relationships (e.g. Inglis and Alcock, 2012). A subsidiary question therefore asked in this study is: does self-explanation training increase attention to the logical relationships between the lines in a proof? Since the results of Study 1 show self-explanation training has a strong, positive effect on proof comprehension, it would be expected that self-explanation training would indeed change the way students read proofs and would move their reading of proofs to perhaps become more in line with the way mathematicians read proofs.

5.1 Methods

The purpose of the study

Given the significant and positive effects of self-explanation training on proof comprehension reported in Study 1, the purpose of this study was to determine how self-explanation training improves proof comprehension. The hypothesis tested is that self-explanation training changes the way students read proofs, helping them to focus harder on the proof and increase the number of times they shift their attention around the proof. The study uses eye-tracking methodologies to measure fixation durations and saccades which, when analysed, determine the general reading behaviour of the students through assessing how hard they concentrated and how much they moved their attention around the proof.

Task procedure

The participants were tested on an individual basis in a small computer lab and were randomly assigned to one of four experimental groups (eight participants per group) with groups one and two as the self-explanation condition group and groups three and four as the control condition group. Group one was given a proof (Proof B)\(^1\) to read before having to answer 10 multiple choice questions. Students in group one were then given self-explanation training before being given a second, different proof (Proof C). Again, there were then 10 multiple choice questions to answer relating to Proof C. Group two did the same as group one except they read Proof C first before reading Proof B second. Group three read Proof B first before answering the multiple choice questions on that proof.
proof. Students in group three were then given a passage on the history of right-angled triangles to read (the same passage as used in Study 1 for the control group) before being given Proof C and the related multiple choice questions. Group four did the same as group three except they read Proof C first and Proof B second. Figure 5.1 shows a diagram of the study design.

All students in the study were given the same proofs and the same multiple choice questions related to each proof. There was no time limit to the time spent on reading, studying the proof or answering the questions. There was however a total time limit of one hour. All participants’ eye movements were recorded using a Tobii T120 eye tracker. Participants went through the study at their own pace, clicking a mouse button to proceed to the next slide. Going back to any slide was not allowed. The experimenter sat in the room with the participants who were told the experimenter was in the room to rectify any technical issues and to administer the question papers. Participants were told to otherwise ignore the fact the experimenter was there. Once participants had completed all parts of the study, they were given £8 inconvenience allowance, thanked and then dismissed.

Task Construction

The first consideration when constructing this task was the design of the study. It was decided to use a mixed design whereby all students studied one of two proofs then half of them experienced self-explanation training before studying the second proof. Using this study design meant that fewer students were required to take part in the study since the statistical power of a mixed design is much greater (Field and Hole, 2003).
Methods, Study 2

97

given the fact that over 70 students were recruited for the previous study, the pool of students to target was much smaller as no student who had taken part in the previous study was allowed to take part in this study. Hence, reducing the number of students required whilst keeping the high statistical power was useful.

Although repeated measures designs do provide greater statistical power, a criticism of using such a design in studies is there could be a “learning effect” which can confound the results (Field and Hole, 2003). The “learning effect” is created because all participants experience, in this case, both of the proofs. Therefore, after experiencing one proof, participants get used to the task and automatically improve. Putting participants into groups randomly and providing them with the proofs in a random order is a technique called counterbalancing. Counterbalancing equalises the amount of practice across the study, thus reducing the effects of practice on the results.

A further reason for using a repeated measures design is that it allows for direct comparisons to be made between both pre-intervention and post-intervention reading behaviours (in this case). Thus it allows for one to compare students’ “natural” eye movements before the intervention to their eye movements after. The pre-test therefore allows better control for methodological threats to the results and greater overall statistical power. In the previous study, a repeated measures design could have been used but it did not have to be used; randomisation alone works well to control for methodological threats when there is a larger sample size.

The second consideration in constructing this study was in deciding which proofs to use. As with Study 1, the proofs needed to be chosen so that students from all years of study could understand them. By doing so, there would be the potential for a greater number of students from different years of study and different mathematical experiences.

As stated previously, it was also decided that no student who took part in the previous study should take part in this study because their experiences in the previous study may have affected their performance in this study. Making the proofs accessible to students from all years of study therefore made the recruitment process easier. The two proofs chosen were taken from number theory but were chosen such that it was unlikely that students could have seen them before. The two proofs are shown below.

---

2see Chapter 3.2

3Courses that involve number theory at Loughborough University were checked to ensure these proofs did not appear in the course notes.
Theorem B

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

Proof B

(L1) Let \( n \in \mathbb{Z} \) and \( n > 0 \).

(L2) By definition, if \( n \) is even then \( \exists k \in \mathbb{Z} \) such that \( n = 2k \).

(L3) Then, \( 3n^2 + 8 = 3(2k)^2 + 8 = 12k^2 + 8 = 2(6k^2 + 4) \).

(L4) Therefore, \( 3n^2 + 8 \) is even.

(L5) Now, assume \( n \) is odd and we will show that \( 3n^2 + 8 \) is odd.

(L6) By definition, if \( n \) is odd \( \exists a \in \mathbb{Z} \) such that \( n = 2a + 1 \).

(L7) Then, \( 3n^2 + 8 = 3(2a + 1)^2 + 8 = 3(4a^2 + 4a + 1) + 8 = 2(6a^2 + 6a + 5) + 1 \) which is odd.

(L8) Therefore, \( n \) is even. \( \square \)

Theorem C

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

Proof C

(L1) Clearly, \( p \) cannot be 2, so we need only show that \( p \not\equiv 3 \pmod{4} \).

(L2) Suppose \( p = 4k + 3 \) for some \( k \in \mathbb{Z} \).

(L3) Let \( y = 2n \).
Then, by Fermat’s Little Theorem, $y^{p-1} \equiv 1 (mod p)$ since $p$ is not a divisor of $n$.

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

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

But this cannot be the case.

Therefore, $p \equiv 1 (mod 4)$. □

It should be noted that for Proof B the conclusion in line L8 does not appear to directly follow from line L7 and appears to be the wrong conclusion. However, the original proof had “Therefore, by the contrapositive, if $3n^2$ is even, then $n$ is even” as line L8. Since the students are asked to determine what the type of proof is for one of the questions, it was decided to remove part of this line and indeed references in other lines to the proof being a proof by contraposition. Thus these references became implicit and students had to use their logical reasoning skills to determine not only the type of proof being used but also how the author has arrived at that particular conclusion. Also, this partially addresses the concern raised in the previous study that the specific types of example proofs in the self-explanation training give an unfair advantage.

When choosing these two proofs, there was some concern that Proof C was more difficult than Proof B. However, even if one proof was considered more advanced than the other, the study is interested in looking at whether self-explanation training changes the way students read proofs. Therefore, it is only interesting to note any differences in the reading patterns between the students who were given the training and those who were not. Indeed, effects of proof difficulty are further considered by the analysis of the data (as discussed in the results section of this chapter).

In order to provide the training to the students, the same PowerPoint presentation used in Study 1 was used in this study. Also, the same passage on the history of right angled triangles used in Study 1 was shown to those students in the control groups. Since both presentations had been used in the previous study a direct comparison could be made between this study and the previous one, providing continuity and consistency whilst also recognising the concerns regarding the control group task discussed in the previous
study. Furthermore, since the training given to the students in Study 1 was so successful, it made sense to continue with that same training in order to see whether it does have an effect on the way students read proofs. If the method of training students in this study was different to that used in Study 1 there would be less confidence that the training used led to the same level of improvement.

By using two proofs that were different to the proof used in Study 1, one could also gain more evidence to suggest that self-explanation training does indeed improve proof comprehension, particularly with number theory proofs. Therefore, although this study does not address whether self-explanation training can improve proof comprehension for a wider range of proofs, it does provide consistency and further evidence for the findings of Study 1. Indeed, since it has been shown this training does appear to have a positive effect on proof comprehension, the results of this study can offer some evidence regarding whether or not self-explanation training improves proof comprehension through changing the way students read.

Recall that in Study 1, students were asked to give their explanations out loud in order for them to be coded and analysed. In this study it was decided that students should not be asked to given their explanations out loud. However they were told they could do so if they felt inclined. This decision was mainly based upon observations made in Study 1 but also from the literature. For example, students in Study 1 stated that talking out loud to a computer in front of a person they didn’t know was a strange experience. Hence, it is possible that this could lead the students to change the way they read without any effects of training. Bojko (2005) also noted this phenomenon and suggested that although thinking out loud is certainly very useful for gaining insights, it can also change how the participant acts during the task and, more importantly, may change where they are looking. Furthermore, Duchowski (2007) noted that eye movements can differ somewhat when reading silently from reading aloud; the mean fixation durations are larger when reading aloud (or whilst listening to a voice reading the same text) than in silent reading. Similarly, Gielen et al. (1991) found their results were directly caused by participants speaking out loud whilst reading, adding further evidence that talking whilst undertaking an eye-tracking study can affect eye movements. Therefore, one should carefully consider asking participants to talk out loud when designing such experiments. By not forcing participants to talk out loud, this phenomenon of changing gaze patterns involuntarily is reduced and the present study also represents similar conditions to when students study a proof for the first time on their own. Therefore it could be suggested with more confidence that any differences between the eye movements of the participants in the control group and the self-explanation groups were caused by the
effects of the training.

Students in this task were allowed to have the proof in front of them whilst they answered the comprehension questions, similar to Study 1. Again, the aim of this study was not to test students’ memory skills but to test their reading patterns and, as a secondary result, their understanding of the proofs. Also, by letting students have the proofs in front of them whilst they answered the questions, students were not worrying about trying to remember all the parts of the proof. If they were worrying about remembering the proof, this could again also affect their reading patterns and hence confound the results. Furthermore, using this method also allowed more challenging questions to be asked. The methods used to create these questions are now discussed below.

Creating the proof comprehension questions

In order to create the proof comprehension questions, the framework suggested by Mejía-Ramos et al. (2012) was once again used. A total of 10 questions for each proof was created based on the framework with each question having three candidates for an answer to choose from. It was decided to proceed with 10 multiple choice questions because, firstly, the effect size in Study 1 was so large that using a subset of questions was considered appropriate to test the proof comprehension of the students. Secondly, this study was mainly interested in answering the question: does self-explanation training change the way students read proofs? Therefore, the difference between the comprehension levels of the control group and the self-explanation group was a useful but a secondary result. Furthermore, making the proof comprehension task only 10 questions long meant students spent a reduced amount of time being tested, reducing the effects of fatigue, but it also meant the students had a reason to concentrate on the training and the proof (the students were told upfront there would be a comprehension task after reading each proof). It is also true that multiple choice questions take less time for the participant to complete (Higgins and Tatham, 2003) and provide cleaner results, adding to the argument for using the methods described above. The multiple choice questions for both proofs are provided in Appendix C.

Using the eye tracker to determine reading patterns

As stated in the literature review and methodology chapters, using an eye tracker is an useful, non-intrusive method for gaining insight into participants’ reading patterns. It
was therefore decided to record all students in this study using an eye tracker. For this study, eye movement data was analysed using Tobii Studio 2.1 configured to use the Tobii Fixation Filter (Tobii Technology, 2010). This eye tracker consists of two binocular infrared cameras underneath a 17” TFT monitor. The information participants are expected to read appears on the monitor screen and participants view this information without head restriction.

In order to use the eye tracker for this study, the proofs were written using the program “\textup{\LaTeX}” before being turned into “pdf” files. This meant that for each part of the study, the text size and the text font remained the same which is crucial because even the slightest change in the text can alter the reading patterns of a participant (Duchowski, 2007).

Participants’ eye movements were recorded throughout the study but only the movements recorded for the first viewing of each of the proofs were analysed, as this was the objective of the study. During the comprehension test phases of the study, it was expected that participants’ eye positions in relation to the screen would change because they were looking at both the screen and a piece of paper containing the comprehension questions. Therefore, to ensure a consistent measurement throughout the study, participants were calibrated at 60cm away from the screen before reading the first proof and again before reading the second proof. In order to calibrate each participant’s eye movements with the eye tracker, the screen showed a nine-point display consisting of a red spot that moved around the screen. The accuracy of such calibration is typically 0.5°.

The data one receives from an eye-tracker is a large table of numbers. This table contains the participant number, the fixation duration\(^4\) in milliseconds on a particular area of the screen and the location on the screen of each fixation. A further output contained within the table is which “area of interest” a particular fixation is in. One can use the eye-tracking software to pre-define any areas of interest one particularly wishes to investigate. For this study, the areas of interest are the lines of the proof, as shown in Figures 5.2 and 5.3. Each line of each proof was individually assigned an area of interest. The fixation durations that were recorded in the table as being in an individual area of interest were summed for each participant to give dwell times. The mean fixation durations and total dwell times for each line of the proofs, including the theorem, were then statistically analysed using SPSS.

\(^4\)see section 2.5.2 for information on fixation duration and saccades
In order to calculate the number of between-line saccades made by each participant, the number of times a participant made a fixation on a line, say $X$, followed by a fixation on a line $Y$, was counted. This method was used for both proofs, and the total number of between-line saccades for both conditions was compared.

Participants

Since the results of the self-explanation training study reported in Study 1 of this thesis had such a large effect size with small $p$-values, far fewer participants were needed for this study. A power analysis was once again conducted in order to reveal the exact number of participants needed for this study and this was shown to be 32. Therefore, 32 undergraduate students who were studying for a mathematics degree at Loughborough University in May 2012 took part and none of these students had participated in
Study 1. However, it should be noted that when analysing the data, four participants had to be removed from the analysis due to recording problems. Their data from the comprehension tests were also removed to keep the analysis consistent. Fortunately, one participant from each group was lost through these problems meaning there were seven participants per group. Therefore, 28 participants were analysed in the results. Despite the power analysis suggesting 32 participants were needed, this loss of data appeared to have no major impact on the final results of the study.
5.2 Results

This section reports the results of the above study investigating the effects of self-explanation training on reading mathematical proofs. Firstly, there is a report on the results of the comprehension test for all students in the study. Secondly, the results of the effect of self-explanation training on cognitive effort is reported. Finally, the effect of self-explanation training on the number of between-line transitions made by the participants in the study is reported. Again all results were produced using SPSS.

The results of the proof comprehension tests

Although it is not the main result for this study, the results of the effect of self-explanation training on proof comprehension are reported since they add to the results of the previous study reported in this thesis. To report the effects of self-explanation training on proof comprehension, an ANCOVA with two between-subjects factors (condition: self-explanation, control; proof read second: Proof B, Proof C) and one covariate (proof comprehension score from the first reading attempt) was conducted. The covariate of proof comprehension score from the first reading attempt is considered because one would expect a participant to improve on their score from any first attempt if completing a second attempt straight after, even if the second attempt is with a different proof. Therefore, by including a covariate, there is an element of control for any effects of the first attempt on the second. Furthermore, the covariate controls for pre-existing differences in individual proof-reading ability, so the results show the change in response to the training, or otherwise.

The results showed that students who had received self-explanation training achieved a higher mean comprehension score, 7.56 (SD = 1.9), compared to those in the control group, 5.56 (SD = 1.5). Indeed, the analysis revealed a significant effect of condition, $F(1, 27) = 8.850, \ p = 0.006, \ \eta^2_p = 0.247$, but no significant effect of the order in which participants read the proofs nor a significant condition-by-proof-order interaction, both $F < 1$. Furthermore, there was no significant difference between the two groups for the mean time spent reading the proof, $p > 0.1$ for both proofs. These results therefore suggest that students who receive self-explanation training perform better on a subsequent comprehension task, and are analogous to the results in found in the previous study.
The effects of self-explanation training on cognitive effort

To investigate the effects of self-explanation training on cognitive effort, the mean fixation durations of participants for the second proof reading attempt were subjected to an ANCOVA with two-between subjects factors (condition: self-explanation, control; proof read second: Proof B, Proof C). This analysis was conducted for several reasons. Firstly, it has been suggested that longer fixation durations can be associated with an endeavour to process information more thoughtfully (see e.g. Duchowski, 2007; Inglis and Alcock, 2012; Just and Carpenter, 1976; Poole and Ball, 2006; Rayner, 1977). Secondly, the mean fixation durations for the proof read first were included as a covariate because one would expect that the effort a participant puts in to the second reading attempt would be related to the effort they put into reading the first; it would perhaps be expected that there would be large individual differences. It has been shown that the individual differences in eye-movement behaviour can be controlled using this method of analysis but the statistical power of the results is also maximised (van Breukelen, 2006).

The results showed that there was a significant effect of condition, $F(1,23) = 14.234$, $p = 0.001$, $\eta^2_p = 0.382$, but also revealed there was no significant effect of proof order, $p > 0.3$, and no significant interaction between condition and proof order, $p > 0.3$. On average, those who received the self-explanation training had mean fixation durations of 301ms (SD = 33.5) for the second proof read whilst the control group had a mean fixation duration of 276ms (SD = 30.0). These results suggest that self-explanation training produces deeper engagement with mathematical proofs and show that the training does appear to change the way students read mathematical proofs.

The effects of self-explanation training on the number of between-line transitions

The final analysis reported in this results section concerns the effects of self-explanation training on the number of between-line transitions and therefore how self-explanation training changes the way participants move their attention around the proof text. To do so, the number of between-line saccades produced by participants on their second reading attempt was subjected to an ANCOVA with two between-subjects factors (condition: self-explanation, control; proof read second: Proof B, Proof C) and two covariates (number of between-line saccades made during the first reading attempt, and the overall duration of the second reading attempt). These covariates were included because one would expect the number of between-line saccades in the first reading attempt to be
closely related to the number in the second reading attempt and also one would expect the longer a participant spends reading the second proof, the more between-line saccades they are likely to make. The analysis revealed a significant effect of condition, $F(1, 22) = 10.394, p = 0.004, \eta^2_p = 0.321$, and a significant effect of proof order, $F(1, 22) = 8.449, p = 0.008, \eta^2_p = 0.277$, but there was no significant interaction between condition and proof order, $p = 0.742$. Figure 5.4 below shows the mean total number of between-line transitions for the proof read second, split by condition.

![Figure 5.4](image)

**Figure 5.4:** Study 2: The mean total number of between-line transitions for the proof read second, split by condition. Error bars show ±1SE of the mean.

These results suggest, firstly, that self-explanation training has a significant effect on the number of between-line saccades conducted by the reader and thus increases the number of changes in attention around the proof of a participant. Secondly, and possibly somewhat unsurprisingly, the results show that the proof that is being read can have a significant effect on the number of between-line transitions made. This suggests the nature of the proof itself can affect reading behaviour. Finally, and crucially, the results show there is no significant difference in the effect self-explanation training has on the number of between-line saccades for different proofs, suggesting that self-explanation training has the same effect on the reading patterns of participants for different proofs.
5.3 Discussion

Building upon the findings of the previous study reported in this thesis that looked into the effects of self-explanation training, this study set out to determine whether or not self-explanation training has any effect on the way students read mathematical proofs. The results not only confirm the findings reported in Study 1 but also show that self-explanation training does have an effect on the way students read proofs, both in terms of the amount of effort put into reading and understanding the proof and the number of attempts to connect the information contained within each line of the proof. These results, combined with the results of Study 1, therefore suggest that students who are given self-explanation training perhaps improve their proof comprehension by changing the way they read proofs.

Another important finding from this study is that the effect self-explanation training had on reading was the same for both the proofs used in the study. The results showed that for both fixation duration and between-line saccades there was no significant interaction for proof and condition. Therefore, it can be suggested that giving self-explanation training to students could change the way students read proofs and improve their understanding.

As reported in the discussion section of the previous chapter, some of the limitations of the study design of Study 1 are at least partially addressed by this study whilst others are not. Firstly, one of the limitations of the study design in Study 1 was that proof used in the study was similar in structure to that of the proof used in the self-explanation training. Thus the effects seen in the results may be accounted for by the practice element of the training alone rather than the specific instructions about self-explanations. Proof B used in this study is a proof by contraposition and is structured in a different way to a proof by contradiction. However, Proof C is a proof by contradiction. Therefore, Proof B does not suffer from the same limitation that was reported in Study 1 although Proof C does. Hence, one can only conclude that this study partially addresses the issue of the proofs being similar to that given in the training.

Secondly, it was suggested that in Study 1 students who received the self-explanation training had an advantage because they were able to practice with a proof in their task whereas the control group did not. Although the control activity and the training in this study are the same as the previous study, all students in this study were asked to read a proof and answer comprehension questions on that proof before being given one of the
activities. Therefore, all students would have gained some practice at reading a proof then answering related comprehension questions before completing one of the activities and then reading the second proof. Again, although this does not totally exclude the suggestion that the effects seen are down to practice, it does at least reduce this possibility.

One limitation that this study does not address is whether self-explanation training improves proof comprehension for proofs outside of number theory. However, as with the previous study, it was important to open the study up to as many students as possible by making the proofs accessible to all years. Number theory proofs offer such benefits; one does not necessarily need specific knowledge gained in later years of study to understand the proof that would be needed for a specific calculus or analysis proof for example. Furthermore, by using a number theory proof, one obtains more consistency and continuity between the studies.

Despite the limitations considered above, the main aim of this study was to consider whether or not self-explanation training changes the way students read proofs. The study design allows for this by getting students to read a proof before the intervention so “natural” eye movements can be compared. Therefore the limitations discussed previously are less of an issue for this study than they are for Study 1.

It must also be noted that these results have been produced using students who were in a laboratory situation rather than a lecture theatre or tutorial situation and the results also show the short-term effects of self-explanation training on proof reading only. Further studies therefore should consider conducting experiments under traditional learning situations and should also be longitudinal. For example, students’ reading techniques could be tracked before being given a lecture on self-explanation training in their first week of a course. Later in the course, the same students could then be tracked again to make comparisons. In order to remove the possibility that any improvements from general learning from the course would cause the change in reading, a control group of students who did not receive the training could also be measured at the start of the course and measured again later. The difference in improvement between both groups could then be compared. A more simple study design that only compares students’ proof comprehension in the short and longer term without the use of eye-tracking data is reported in the next chapter.
Chapter 6

Study 3: Does self-explanation training work in a pedagogical setting and over the longer term?

Study 1 and Study 2 showed that self-explanation training has positive effects on both reading processes and learning outcomes, but they did so under laboratory conditions, and only for proofs read immediately after being given self-explanation training. In this final study, self-explanation training is shown to improve proof comprehension in a genuine pedagogical setting and over the longer term.

6.1 Methods

The purpose of the study

The purpose of this final study was to determine if self-explanation training could be implemented in a genuine pedagogical setting. Clearly, for self-explanation training to truly improve proof comprehension, it should be a method that can be taught to students in a traditional learning environment (e.g. lectures and tutorials). Therefore, this study was conducted during a live lecture with first year students, as primarily this would be the intended target audience for the training. Furthermore, the lasting effects of self-explanation training are unclear. There have been very few studies that have considered the long-term effects beyond a week or so. This study considers the effects of self-explanation training on proof comprehension over roughly three weeks by asking
students to complete a delayed post-test. The study design is shown below in Figure 6.1

**Figure 6.1: Study 3: Diagram taken from Hodds et al. (2014), Fig. 1, p.75**

### Task Construction

The self-explanation training was provided in a paper booklet format so that it could be administered during the course of a normal lecture. The training booklet contained the same materials and instructions as the electronic version used in Studies 1 and 2. However, the practice proof was changed so that the self-explanation group did not spend extra time on a calculus proof, possibly giving them an unfair advantage over the control group both in the study and in their course. The replacement practice proof is provided in Appendix D.

The control group also received a paper booklet but it contained information on time management and how they could improve their time management skills at university. The booklet was again of similar length to the self-explanation training and was structured so that students were asked to provide written answers to questions on their current time management (e.g. How many tutorial sheets do you complete at home 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. Since both groups would eventually receive both conditions at the end of the delayed post-test, this control booklet seemed more appropriate as it ensured both groups spent lecture time studying material on study skills, something that would be beneficial to them considering that
lecture time was being taken up by the study.

Creating the proof comprehension questions

This study used Proof B and the associated multiple-choice questions from Study 2, and the triadic primes proof, Proof A, from Study 1 along with a newly-constructed 10-item multiple-choice proof comprehension test. Again, the order of the comprehension test questions was randomised for each participant, and each item was allocated one point to give a maximum total score of 10 for each proof. The multiple-choice test for Proof A is provided in Appendix E.

Procedure

The study took place during two scheduled lectures during the first semester of the academic year. This was chosen so that the students, who were all first years, had not taken part in either of the previous self-explanation studies, and it simulated providing the training to students in their first few weeks at university. 107 first-year undergraduate mathematics students undertaking a calculus course at Loughborough University participated in the study. A total of 139 students took part in the first session and 122 in the other; data was only used from those 107 who were present in both. 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 were assigned to the control group and odd identification numbers were assigned to the self-explanation group.

In the first lecture, which was week 2 of the course, the self-explanation group were given 15 minutes to read the self-explanation training booklet whilst those in the control group were given the same time to read the time management booklet. Both groups then read Proof B and answered the related multiple-choice questions. Again, participants were given 15 minutes to complete this task due to the timings of the lecture. For the duration of the study, participants were asked to work individually and in silence. Furthermore, participants were asked to sit on the right hand side of the lecture theatre if they had an even identification number and on the left hand side of the lecture theatre if they had an odd identification number. This ensured that students from both groups were separate and could not see the other group’s materials.
Just under three weeks later, in week 4 of the course, both groups were asked to read the triadic primes proof from Study 1 and answer the related multiple-choice questions. The groups were split in the room according to their identification numbers, were asked to work individually and in silence and had 15 minutes to do the task. Once all of the papers had been collected, participants in the self-explanation group were given the control materials and those in the control group were given the self-explanation training booklet. This ensured all participants were given the same access to the self-explanation technique, and the time-management materials, so no group was left disadvantaged. Again participants worked individually in silence, and had 15 minutes for each task.

Once all of the analysis had been completed, all of the materials from the study were provided on the university’s virtual learning environment. This enabled students to access both of the booklets, the answers to the multiple-choice tests and feedback, which contained information on the purpose of the study and the results including the benefits of having good time-management skills and the positive effect self-explanation training can have on one’s proof comprehension.

### 6.2 Results

This section reports the results of the above study investigating the effects of providing self-explanation training, in a genuine pedagogical environment, on reading mathematical proofs, both in the short term and the longer term.

**The results of self-explanation training on proof comprehension in a pedagogical environment over the longer-term**

To investigate the immediate and longer-term effects of self-explanation training as administered in a lecture, an ANOVA was conducted on the proof comprehension scores with one within-subjects factor (time: post-test, delayed post-test) and one between-subjects factor (condition: self-explanation, control). The results showed the self-explanation group scored significantly higher on the immediate post-test ($M = 5.94$, $SD = 1.91$) than the control group ($M = 5.19$, $SD = 1.79$) and also on the delayed post-test (self-explanation group: $M = 6.04$, $SD = 1.59$; control group: $M = 5.44$, $SD = 1.79$). Indeed, a main effect of condition was found, $F(1, 105) = 6.024$, $p = 0.016$, $\eta^2_p = 0.054$, but no significant effect of time or interaction between condition and time was reported, $F < 1$ in both cases. Thus, these results show that self-explanation training in a typical
pedagogical environment improves proof comprehension significantly in the short term and in the longer term also. Figure 6.2 below shows the mean comprehension tests scores split by condition at post-test and delayed post-test; the differences at both times correspond to effect sizes of $d = 0.410$ at post-test and $d = 0.350$ at delayed post-test.

![Figure 6.2: Study 3: Mean comprehension test scores by condition for post-test and delayed post-test](image)

6.3 Discussion

This final study looked at the effects of self-explanation training on proof comprehension in a genuine pedagogical setting and over a longer period of three weeks. The results have confirmed the effects seen in the previous two studies, showing that self-explanation training is effective at improving proof comprehension in the immediate term. The results also show that the training works in a genuine pedagogical setting and the effects are still shown three weeks later. These results are therefore positive and suggest that lecturers might wish to consider incorporating self-explanation training into lectures that involve working with mathematical proofs.
Again, this is the first study in the field of mathematics education to look at the effects of self-explanation training on proof comprehension beyond one week in a pedagogical setting. Previous studies (e.g. O’Reilly et al., 2004) have shown the effects of self-explanation training in other subjects after one week but they also provided prompts to remind participants of the training. Students in Study 3 were not reminded of the training they received and so it could be argued that the results of Study 3 are perhaps stronger than the previous studies into longer-term effects of self-explanation training.

These results are also positive for general teaching practices at the post-compulsory level. The self-explanation training was provided as a four page booklet; the same that was provided electronically in the previous two studies reported in this thesis. It was therefore easy to administer and can be considered as a resource-light method of providing the training to the students. If such effects can be found with students studying a proof, it is possible that these effects could be seen in other areas of mathematics and indeed in other subjects. Lecturers would not need to change their style of teaching or spend time working on the materials which is a positive aspect of the training booklet. The training could therefore easily be implemented in the lecture theatre.

Furthermore, the results of this study go some way towards demonstrating that the results of the earlier studies were not simply due to decisions made regarding the methodology of those studies. Indeed, one of the main concerns was the effects shown in the previous two studies were perhaps down to relevant practice with specific types of proof. In this study, Proof B is a different type of proof to that shown in the training booklet and is the first proof that the students are tested on. Therefore, the effects of practice are reduced. Although the second proof used in the study is the same type of proof used in the training, it was given to the students in a lecture around three weeks later. Since students were not allowed to take the booklet away with them until after the second lecture and students were not prompted about the training in the second lecture, it seems highly unlikely that the results seen in this study are due to the effects of practice rather than the self-explanation training.

Although these are results are encouraging, they only show the effects of the training over a period of three weeks. Although three weeks is a longer period than asking students to study a proof immediately after being given the training, it does not show the effects of the training last over an extended period. Ideally, as educators, we would like our students to be able to apply the training for the whole time that they are at university. Therefore, it would be useful to determine whether the effects seen in this
study last over the course of a student’s study at university. Due to time constraints this was not possible for the purposes of this thesis, but would be a worthwhile future study.

The results of the study reported are promising and add further weight to the argument that self-explanation training could be incorporated as a pedagogical tool to improve proof comprehension in undergraduate mathematics. Naturally, further work needs to be done before self-explanation training can be implemented. For example, since this study uses the same proofs as used in the previous two studies, one can still not state whether self-explanation training improves proof comprehension for a wider range of proofs. Therefore if further work shows that self-explanation training does indeed improve proof comprehension for a wider range of proofs with different students (such as students from different countries and at different institutions) and over an even greater period of time then there would be good evidence to suggest self-explanation training should be incorporated into undergraduate mathematics courses.
Chapter 7

Conclusions

The final chapter of this thesis provides some conclusions that one can draw from the research undertaken and how these conclusions can be taken forward in future research. First, the three studies are summarised, detailing the main results of each. Next some general limitations of the studies are reported. This is followed by some suggestions on where to direct future study that may address some of those limitations and extend our understanding of the phenomena studied in this thesis. The chapter concludes by offering some suggested pedagogical implications, not just for the teaching of proof, but the teaching of mathematics more generally.

7.1 Summary of the studies reported

Study 1 took the suggestion from the general reading literature that a short course of self-explanation training can improve one’s understanding of a text. Students from all three years of study at one university were invited to take part, with half of the participants receiving self-explanation training and the other half receiving a control activity on the history of right-angled triangles. Both groups were then invited to explain out loud their thoughts on a proof of the theorem that states there exist infinitely many tridiadic primes. A comprehension test, using the framework of Mejía-Ramos et al. (2012), was then administered. The analysis of the data obtained from the comprehension test revealed the self-explanation group significantly out-performed the control group and the effect was consistent across all years of study. Furthermore, using a framework from Ainsworth and Burcham (2007) as a guide to analyse the participants’ verbal responses, it was found that students in the self-explanation group produced significantly more
higher quality explanations than did the control group.

The second study reported in this thesis investigated whether self-explanation training changes the way a student reads a proof text by using eye-tracking technology to analyse eye movements. Again, students from all years of study (who had not participated in Study 1) were invited to take part and were first asked to read and answer comprehension questions on a proof they had not seen before. Half of the students in the study were then given self-explanation training with the other half receiving the same control activity used in Study 1. A second proof was then administered to the students in both groups with a related comprehension test. By analysing the eye-movements of students in both groups, the results showed self-explanation training does indeed change the way a student reads a proof, both in terms of the number of between-line transitions and the length of the student’s mean fixation durations.

The final study investigated whether the effects of self-explanation training found in both Study 1 and Study 2 would be found in a genuine pedagogical setting and over an extended period of three weeks. Students in a first year calculus class took part in the study and were split based on their individual identification numbers. Those with odd identification numbers received self-explanation training and those with even identification numbers received an activity that asked them to reflect on their own study habits. Both groups were then given a proof and a related comprehension test to complete. Three weeks later, all students received a different proof and completed a comprehension test related to that proof. Results revealed again that there was a significant effect of self-explanation training on proof comprehension, both in the short term and in the longer term.

Having summarised the studies reported in this thesis, this section now moves on to discuss some of the general limitations of the research reported followed by some implications of the results for the learning and understanding of mathematical proofs.

### 7.2 Some general limitations of the research reported

Any research conducted in any field will have issues that need to be addressed and will also have certain limitations. One possible limitation of the research reported in this thesis is due to the wide range of proofs used in mathematics. Clearly, in an ideal world,
more proofs would be tested with a larger number and broader range of students for each research question. However, given the circumstances and the constraints, such as time and money, only one or two proofs are used in each study reported. The individual studies are therefore necessarily limited in the number of proofs they can consider, so not all domain areas or proof types are covered. Indeed, recall that the majority of proofs used in the studies were taken from number theory. This was to ensure that students from all three years of study were able to participate in the studies. Furthermore, Proof A and Proof C used in the studies were proofs by contradiction, matching that of the example proof given in the training and leading to the suggestion that the effects found were down to practice alone and not due to the training. Recall however that Proof B, used in Study 2 and Study 3, was a proof by contraposition and that further analysis conducted in Study 1, where a question was removed from the results, suggested that it was unlikely that the effects seen were due to merely remembering the content of the training (although that cannot be totally discounted). Moreover, the proofs used for the purposes of this thesis are fairly typical of those that undergraduate students encounter every day in terms of length and logical complexity. Indeed, the majority of proofs used in the studies reported were taken from textbooks or lecture notes from undergraduate courses. Finally, the results of the delayed post-test in Study 3 further reduced the likelihood that practice effects were the cause of the difference in comprehension performance.

Secondly, the studies reported in this thesis took place in one university in the UK. Therefore, conclusions cannot be drawn for all students as students in different universities in different countries will have different mathematical backgrounds. The self-explanation training that worked for students at the university used in this thesis may not necessarily work for others. One should therefore consider that the term “student” used in the conclusions drawn from the research in this thesis could perhaps only refer to those at a university similar to that used in the studies. Indeed, the university used is fairly typical for one in the UK. Students are currently, at the time of writing, expected to achieve AAB to obtain a place on a mathematics course, including an A in mathematics at A-Level (or equivalent). This entry requirement is high but not the highest grades that would be expected of students to obtain a place at Oxford or Cambridge for example. The subjects taught are a mix of pure and applied mathematics and involve a mix of lectures and small group tutorials. Therefore, although one should be cautious in stating which students the effects seen in the studies reported apply to, one should be fairly confident that self-explanation training could improve proof comprehension for many students in at least the UK. Indeed, the large effect sizes seen in the studies support this further.
Finally, in order to allow one’s own students to gain the most benefit from self-explanation training, one may need to consider individual students’ circumstances. For example, students’ background mathematical knowledge and prior teaching experiences could affect how self-explanation training changes the way students understand proofs. Results from Study 1 suggest however that the effect self-explanation training has on proof comprehension is consistent across all three years of study. Nevertheless, self-explanation training may not work for all students and educators should perhaps be aware of their students’ strengths and weaknesses before providing the training. A consideration for future research should perhaps therefore be to determine whether different groups of students might benefit differently from the training.

7.3 Directions for future study

Clearly the research reported in this thesis has offered some enlightenment as to how we, as educators and researchers, can improve undergraduate proof comprehension. However further research will always need to be conducted to ensure we are providing the best support for our students. Furthermore, as researchers we are also interested in what we have learned and might yet learn about mathematical reading itself and about its relationships with other skills. Indeed, the main issue that is not addressed by the research reported in this thesis is what groups of students benefit from self-explanation training and for what proofs. Therefore, this subsection offers some potential areas, and direction, for future study in proof comprehension, and the field of mathematics education more generally, that may address some of the limitations reported above whilst also extending the work on self-explanation training.

As stated in the limitations described above, the studies reported in this thesis only consider one group of students from one university. Other groups of students, ranging from school children to postgraduate mathematicians, were not included in the studies and therefore one cannot say the effects seen in the studies reported would be seen for these students. Firstly, it would be beneficial to consider the use of self-explanation training with school children as it would clearly be beneficial if students came to university with a better understanding of proof and how to understand proofs in general. Personal discussions with secondary level educators have shown that students at the secondary level and in higher education have a limited knowledge about proof. Research also backs this claim, suggesting students often see proof as just something to study rather than a way of understanding and communicating mathematics (Knuth, 2002). By investigating
whether self-explanation training improves understanding of proof texts and mathematical texts more generally at the secondary level, it maybe possible to produce better prospective mathematicians before they arrive at university. Secondly, it is not just the students in the lower levels of education that may benefit from self-explanation training. Perhaps even PhD and research students could benefit from self-explanation training for proof comprehension. Postgraduate students do generally have more mathematical knowledge and experience than undergraduates, or secondary level students, and experience and ability is something that may effect the impact of self-explanation training on a student’s proof comprehension skills. Although the results of Study 1 show that self-explanation training does not substantially benefit one year group over another, the mathematical ability, or problem solving ability, of the students in the study were not specifically considered in the analysis. Therefore it is possible that when mathematical ability is considered, the effect self-explanation training has on proof comprehension could be different to the results reported in this thesis. Further research would then need to consider tailoring the training based on ability by adding more examples or by complementing the training with additional tuition.

The studies reported in this thesis also do not take into account individual student’s circumstances, other than their year of study, and thus it is unclear which students benefit most from self-explanation training. One of the issues that would need to be addressed is whether self-explanation training is less susceptible to the expertise reversal effect when considering students work with (and showing understanding of) proofs. Recall that the expertise reversal effect is said to occur when an instructional technique that is better for one group is actually less effective for the other and results in a reduced performance. A future study could therefore investigate this. In order to do so, students with high and low knowledge could be identified by doing a pre-test. Students could then be given self-explanation training followed by a proof and related comprehension test to see whether or not an expertise reversal effect occurs.

Not only could we consider which groups of students self-explanation training works best for, we could also investigate the types of proof and the types of mathematical texts more generally self-explanation training works best for. For example, Study 2 revealed a significant effect of proof read first on reading behaviour suggesting that the proof itself has an effect on one’s reading behaviour. This raises the question as to which proofs does self-explanation training work best with. The results of the three studies reported in this thesis suggest that the answer to the question would perhaps be that the training works for all proof types. However, non-formal proofs and proofs with diagrams, for example, are very different to the proofs used in the studies reported. Hence
the self-explanation training may need to be adapted if we want students to be able to fully understand the whole range of mathematical proofs available.

For mathematical texts more generally, Wong et al. (2002) provided a promising start with their results. However, further research would increase the evidence for incorporating self-explanation training into the mathematics curriculum at the post-compulsory level. Indeed, one could invite students to take part over a sustained period of time and ask them to read a proof, or mathematical text, before providing a corresponding comprehension test. Some students would then receive self-explanation training whilst others would receive a control condition similar to that provided in Study 3. The students would return after several defined time periods, perhaps after 1 week, 1 month and 3 months, and would be tested on their understanding of different proofs, or mathematical texts, to make a comparison. Again, this could be conducted using an eye tracker to obtain even more data on students’ reading habits, although obtaining a large number of students for multiple visits to an eye-tracking laboratory may prove to be difficult. This study could instead be conducted during a lecture scenario, making the recruitment of large numbers of students easier to obtain.

Finally, although the results of Study 2 show that the self-explanation group made significantly more between-line transitions than the reading group, the results did not investigate whether self-explanation training increases one’s ability to focus on lines that require warrants. Furthermore, the study did not ask students to infer the warrants required in the proof. Perhaps, therefore, it is not only that self-explanation training changes the way students read a proof, but that it also improves their ability at inferring the required warrants within a proof. By conducting a future study that investigates this, one could determine whether students could complete the second step (which is to infer the warrants required) of the three-step process suggested by Inglis and Alcock (2012). Such a study could also add greater insight into the aspects that self-explanation training improves in order to improve undergraduate proof comprehension.

7.4 Pedagogical implications for undergraduate proof comprehension

The limitations of the studies notwithstanding, it can be argued that one of the main implications from this research is that students in the studies do appear to have at
least some of the knowledge and skills required for proof comprehension and perhaps they just need some guidance on how to use that knowledge and those skills. Self-explanation training appears to offer that guidance. This section therefore considers this, and other implications for undergraduate proof comprehension.

Recall that previous studies that investigated ways to improve undergraduate proof comprehension have mainly focused on changing the presentation of the proof text itself. Given the success of the self-explanation studies reported in this thesis and the limited success of the suggestions in the literature regarding the change of the presentation of the proof text (e.g. Alcock and Wilkinson, 2011; Leron, 1983a; Rowland, 2002) it seems reasonable to suggest that changing the way students engage with proof will have a better impact on students’ understanding of proofs. It is perhaps possible that the layout of proof texts are as good as they are ever going to be for the purposes of understanding. There may be some slight arrogance on our part as researchers and educators if we believe we can change the format of proofs, after centuries of evolution, and assume this will automatically help students. Indeed, recall it has been shown that changes to the layout of a proof text have caused issues for students because the format is unfamiliar (Fuller et al., 2011). Furthermore, the research reported in this thesis agrees with the suggestion made by Roy et al. (2010) that when students are forced to create their own understanding of proofs they will have a better general understanding of the proof text. Moreover, it has been suggested that proof comprehension can be improved through emphasising to students the importance of being able to communicate and understand mathematical proofs (Healy and Hoyles, 2000) which, arguably, self-explanations promote, rather than changing the layout of the proof text itself.

The majority of research has also discussed what students cannot do or, indeed, what they do differently to mathematicians when trying to understand a proof. For example, they may not have deductive reasoning skills (Selden and Selden, 2003), they may incorrectly evaluate inferred warrants within a proof (Weber and Alcock, 2005) and some gain conviction from considering just a few examples (Weber, 2001), to name just a few differences. Those previous studies have focused on reporting the negative aspects, rather than reporting what students can do and suggesting how to develop those aspects. Again, the results of the studies in this thesis show that students are perhaps able to understand certain aspects of proof and can improve their comprehension if they are given direction and encouragement through self-explanation training, which is a different line of enquiry to previous research.
However, the results of the studies reported are consistent with the suggestion in the literature that students struggle to understand proofs because they do not read the proofs in a thorough or strategic manner; students in the control group in Study 2 had shorter mean fixation durations, suggesting less cognitive effort, and made fewer between-line saccades, suggesting fewer attempts to make logical connections. Furthermore, students who did not receive self-explanation training in Study 1 made more false explanations and paraphrased more than students in the self-explanation group. Again though, these statements were not due to a serious lack of knowledge, as the comprehension scores suggest. It is perhaps they are misguided on how to use and apply the knowledge they have; as suggested previously, self-explanation training helps students to apply that knowledge and could be considered as a pedagogical tool to help students improve their comprehension.

If one is to offer guidance to students on how to apply their knowledge of proofs correctly in order to improve their proof comprehension, one must make sure the training they are given works for them. Therefore, the expertise reversal effect must be considered. Study 1 showed that self-explanation training was effective across all three years of study and this is perhaps because self-explanation training is relatively invulnerable to the expertise reversal effect. Self-explanation training does not add information to the proof text that the student may already know. Study 2 also suggested that self-explanation training increases the cognitive effort made by the student. It therefore implies that the training helps students to focus more of their efforts on understanding the information contained within the proof. Thus it can be argued that, in comparison to other interventions such as e-proofs, there is less chance an expertise reversal effect would be seen with successful students for whom any added explanations would be superfluous. However, this was not directly tested for within the studies reported and further research would be required in order to confirm this hypothesis. A future study is described previously.

If we are to offer self-explanation training to students in order to guide their understanding, it is important to also consider educators. Self-explanation training is very generic; contained within the booklet are instructions that potentially apply to all proofs and all proof types, and also some specific examples to guide students in the process of self-explaining. Since the training is resource-light, instructors do not need to change the way they teach mathematical proofs. Previous attempts to improve proof comprehension, such as generic proofs (Lamport, 2012; Rowland, 2002), modular proofs (Leron, 1983a), and e-Proofs (Alcock and Wilkinson, 2011; Roy et al., 2010) differ in this way as teaching materials would need to significantly change. Self-explanation training can
also be taught and administered in 15 – 20 minutes, as shown in Study 3.

The modern learning environment is also constantly changing, especially with the fast pace of improvements in technology. Students seem to spend more time learning using computers and less time interacting with instructors due to online learning becoming increasingly common (Allen and Seaman, 2011). Instead of changing the way an instructor teaches, one could adapt the learning materials; self-explanation training can be adapted in the ways that would fit with the changing learning techniques of the modern student. For example, the materials can be made available online as an easy to follow PowerPoint or integrated to a university’s own online learning system.

Self-explanation training also leaves the reader to do the understanding for themselves, reducing the effort required by the educator. Students who prefer to work on their own or in groups would therefore perhaps learn more effectively using self-explanation training. Indeed, the results of the studies reported suggest that perhaps self-explanation training could also improve students understanding of new proofs, given that the students who participated in the studies had not seen the proofs before. It could therefore be suggested that self-explanation training could potentially be a useful pedagogical tool, not only in teaching proofs or mathematics generally, but across a wider range of subjects, as previous studies have shown (e.g. Chi et al., 1994; Leinhardt, 1993), and also in new situations.

7.5 Concluding remarks

By conducting the research reported in this thesis, we have set out on the road to improving and developing undergraduate proof comprehension. Some theoretical suggestions have been empirically tested, with self-explanation training being shown to be an effective pedagogical tool. The conclusions drawn from the three studies and the potential future studies suggested will hopefully enable us, as researchers and educators, to improve the way we teach and learn mathematical proofs. It has been suggested that “...proof is math and math is proof” (Hersh, 1993, p.396) so it is important to improve students’ understanding of mathematical proofs. By providing a better service to our students so that they are able to understand proofs, we can perhaps produce better mathematicians and mathematically well-educated people generally for the future.
Appendix A

The passage for the control group used in Studies 1 and 2 taken from Gonczy (2003)

• In the following study you will be presented with a proof to study and asked to say out loud any thoughts you have that help you to understand that proof.

• You will now read a passage about the history of mathematics and in particular, right-angled triangles. There will be a few short questions on what you have read at the end of the passage. Please press the spacebar to move on to the next slide when you are happy to move on.

Right-angled triangles and their applications

Mathematics is often thought of as a purely European development. History of mathematics books often focus on Greece as the epicenter for early mathematical discoveries. For example, the main theorem about right triangles (the sum of the squares of the two smaller sides of a right triangle equals the square of the hypotenuse) is attributed to and now bears the name of a Greek, Pythagoras of Samos, born around 570-560 BC. Pythagoras is often credited with the first proof of the theorem; however his actual written proof has not been found. Earlier civilizations definitely knew about this geometric fact, and perhaps after his travels, Pythagoras took this information back to Greece.
Around the same time as Pythagoras, the Sulvasutras by Apastamba demonstrates that the Indian civilization was familiar the Pythagorean Theorem and Pythagorean Triples (sets of three numbers that satisfy the Pythagorean Theorem). The Babylonians discovered the Triples much earlier. In the Babylonian tablet known as Plimpton 322, Pythagorean Triples are arranged so that the first row corresponds to the ratio $\frac{c^2}{b^2}$, the second to the number $a$, and the third to the number $c$, such that $a^2 + b^2 = c^2$. This tablet is thought to come from approximately 1900-1600 BC, long before Pythagoras time.

While the Babylonians certainly knew about Triples very early on, the Chinese may have proven the Pythagorean Theorem the earliest; some estimates are as early as 1100 BC, although sixth century BC is more generally accepted. The Pythagorean Theorem and right-angled triangles were very prominent in Chinese writings, both in mathematical treatises and in more practical science books. The Chinese grasped many right-angled triangle principles early on, and applied them to practical problems. China is an ancient civilization, akin to Babylon and Egypt. Like the notched wolf bone in Europe, the first signs of math in China are the markings on tortoise shells and cattle bones, known as oracle bones, from the Shang dynasty, approximately 1200 BC.

Later, the Chinese did calculations using small bamboo counting rods, which lead to the use of rod numerals and a positional system for writing numbers. Unlike the Babylonians and their “sexagesimal” system, the Chinese tended to decimalize fractions, and unlike the Egyptians and their unit fractions, the Chinese used common fractions and were able to find the lowest common denominator in order to add different fractions. The Chinese were also comfortable with negative numbers, using a red set of counting rods for positive numbers and a black set of counting rods for negative numbers, similar to modern accounting except with the colours are reversed.

Like early writings in Egypt and Babylon, ancient Chinese mathematic books tended to be collections of practical problems, giving the problem first, then the answer, and sometimes the solution method. The earliest books on mathematics seem to have been written in the Han dynasty, although dating the books is very difficult as the emperor Shih Huang-ti of the Chin dynasty ordered a burning of books in 213 BC.

In the following Han dynasty, mathematicians had to rewrite all of the old books from memory or from hidden scrolls. The early books are thus thought not to be the work of any one mathematician, but rather a collection of the mathematical knowledge up to
that age. The main early writings include the Zhou bi suan jing (The Arithmetic Classic of the Gnomon and the Circular Paths of Heaven), which is sometimes written as Chou Pei Suan Ching; and the Chiu chang suan shi (The Nine Chapters on the Mathematical Art), which is sometimes written as Jiu zhang suan shu. Both books were commented on by many later mathematicians with Liu Hui, an official in the Wei kingdom, being the first to comment on the Nine Chapters.

In addition to remarking on the existing work, Liu Hui also added nine problems to the end of the Nine Chapters; this addition was later published as the Haidao Suanjing (The Sea Island Mathematical Manual). The added nine problems, as well as the last chapter of the Nine Chapters and several discussions in the Zhou bi, all involved the application of right-angled triangles and demonstrated the Chinese knowledge of the Pythagorean Theorem.

In ancient China, the base of a right-angled triangle was called kou or gou (meaning leg), the altitude was ku or gu (meaning 'thigh'), and the hypotenuse was hsian or xian (meaning bowstring). The problems in the ninth chapter, called the Kou-Ku chapter, progress from easy to hard. The first three problems only require simple knowledge of the 3-4-5 right triangle. In each problem, two of the three sides are given, and the reader is asked to find the third side. In Liu Hui’s commentary, he refers to a diagram similar to one used by Euclid in his Elements proof, showing the squares of the lengths on the sides of the triangle.

The rest of the problems in the Kou-Ku involve figuring out where the triangle is in the practical situation given and then finding a side of the triangle, given different lengths or ratios. The methods and explanations given clearly make use of the Pythagorean Theorem. In Problem 6, the kou and the difference of hsien and ku are given in the context of a reed plant in the middle of a square pond. The given method to find ku and hsien is to “find the square of [kou], and from it subtract the square of [hsien minus ku]. [Ku] will be equal to the difference divided by twice [hsien minus ku]. To find [hsien] we add [hsien minus ku] to [ku]”. When using our modern notation of “$a, b$ and $c$”, this can be shown to form the theorem of $a^2 + b^2 = c^2$.

Questions
Please now answer these questions, on what you have just read, out loud.

1) The theorem for right-angled triangles is named after Pythagoras of where?
2) Which 3 groups of people discovered the Pythagorean triples much earlier than Pythagoras?
3) Chinese calculations were done on what kind of rods?
4) How many chapters are there in the Chinese book on the Mathematical art?
5) In problem 6, the kou and the difference of hsien and ku are given in the context of a what in the middle of a what?
Appendix B

The mark scheme used in Study 1

Mark Scheme

Note: Answers do not have to be exact to the ones provided but should be close - use discretion.

1. What does it mean for a number to be triadic?
A number is triadic if it can be represented in the form $4k + 3$ (1 mark) for some integer $k$ (1 mark).
Do not accept: A number is triadic if it can be represented in the form $4j + 1$ for some integer $j$.

2. What does it mean for a number to be a prime?
A number is prime if it is greater than 1 (1 mark) and is divisible by 1 and itself (1 mark).
An answer of the form: A number is prime if it is divisible by 1 and itself will score one mark only.
Do not accept: A number is prime if it is 1, 2, 3, 5, 7, 11... 

3. What kind of proof is this (e.g. Proof by induction)?
This is a proof by contradiction (1 mark).
Do not accept: This is a proof by induction, This is a direct proof etc.

4. In line 3, what is the purpose of assuming that the theorem is false and that there are only finitely many triadic primes?
The purpose is to set up a proof by contradiction. (1 mark) We assume that the theorem is false (1 mark) and prove this cannot possibly be the case thus showing the theorem has to be true (1 mark).
Do not accept: The purpose is so we can use monadic primes rather than triadic primes as they are easier to work with in a proof.
5. In line 5, why does the fact that 3 does not divide \( 4p_2 \cdots p_n \) imply that 3 does not divide \( M \)?
Because, by rules of division, we know that if 3 does not divide \( 4p_2 \cdots p_n \), it cannot divide \( M \) since \( M = 4p_2 \cdots p_n + 3 \) (2 marks)

6. Which claim(s) in the proof logically depend on line 2 of the proof, the claim that the product of monadic numbers is monadic?
The claim that \( M \) itself is monadic because \( M \) is made up of a product of monadic numbers. (2 marks). Also accept Line 8.
Do not accept: The claim that \( p_2 \cdots p_n \) do not divide \( M \) and hence 3 does not divide \( M \).

7. Which of the following summaries best captures the ideas of the proof:
a. It assumes there are finitely 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. This cannot be true as \( M \) is triadic and thus the theorem is proved.
b. It lets \( M = 4p_2 \cdots p_n + 3 \), where \( p_i \) are prime numbers and \( p_i \neq 3 \). 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 \( M = 4p_2 \cdots p_n + 3 \) and shows 2 does not divide \( M \), since 2 is even and \( M \) is odd. It then uses monadic primes to create an infinite triadic prime.
A is the correct selection (1 mark)

8. Summarise in your own words how the proof arrives at the conclusion that \( M \) itself must be monadic.
No triadic prime can divide \( M \). (1 mark)
Shows 2 cannot divide \( M \). (1 mark)
Uses the subproof to show \( M \) is monadic. (1 mark)

9. Why is the sub-proof that the product of monadic numbers is monadic included in the proof?
So that we can show \( M \) itself is a monadic number (1 mark).
This provides our contradiction (1 mark).
Do not accept: It is just there as extra information, it is not actually needed in the proof or There is no proof, it is just a statement or So we can conclude no triadic prime divides \( M \).

10. Do lines 3 – 7, which establish that \( M \) is not divisible by a triadic prime, depend on the statements in lines 1 – 2, which establish that the product of monadic primes is monadic? Explain your answer.
The lines are logically independent (1 mark). Student provides good reasoning (1 mark).
Do not accept: Lines 3–7 rely (logically depend) on statements made in lines 1–2 or Lines 1–2 rely (logically depend) on statements made in lines 3–7.

11. Using the method of the proof you have been working with, what would be an appropriate value for $M$ if you were writing the proof for the theorem that there are infinitely many primes of the form $6k + 5$?
$M = 6p_2 \ldots p_n + 5$ (1 mark) where $p_1 = 5$ (1 mark)

12. 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?
No it is not (1 mark)
Because the product of triadic numbers need not be triadic you cannot use this in the way this proof uses the product of monadic numbers is monadic (1 mark)
Do not accept: Because the proof uses a contradiction of monadic numbers and since we are trying to prove there are infinite monadic numbers, we can’t contradict what we are trying to prove.

13. If 3, 7, 11, 19 were the only triadic primes, what would the value of $M$ be?
$M = 4p_2 \ldots p_n = 4 \times 7 \times 11 \times 19 + 3 = 5852 + 3 = 5855$. (2 marks)

14. 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.
No (1 mark) because $p_1 = 3$ so $87 = 4p_2 + 3$ so $\frac{84}{4} = p_2$ so $21 = p_2$ which is not prime. (1 mark).
Appendix C

The multiple choice questions used in Study 2

Correct answers indicated with a tick.

Proof B

1 According to the theorem, which of the following is the most appropriate definition of \( n \)?

a) \( n \) belongs to the integers.

b) \( n \) belongs to the positive integers. \( \checkmark \)

c) \( n \) belongs to the negative integers.

2 Which justification best explains why \( 3n^2 + 8 = 2(6m^2 + 4) \) implies \( 3n^2 + 8 \) is even?

a) \( 6m^2 + 4 \) is even because of the plus 4, so \( 3n^2 + 8 \) must be even.

b) \( 6m^2 + 4 \) is even so \( 2(6m^2 + 4) \) is also even.

c) \( 6m^2 + 4 \) is just another integer, say \( k \), so by definition \( 3n^2 + 8 \) is even because \( 3n^2 + 8 \equiv 2k. \) \( \checkmark \)

3 Which justification best explains why showing that if \( n \) is odd then \( 3n^2 + 8 \) is odd helps to prove the theorem?

a) Because all numbers are either odd or even so if \( 3n^2 + 8 \) is odd because \( n \) is odd, then if \( 3n^2 + 8 \) is even, \( n \) has to be even.

b) Because we showed in the first half of the proof that if \( n \) is even then \( 3n^2 + 8 \) is even. Therefore, by showing if \( n \) is odd then \( 3n^2 + 8 \) is odd, we can conclude that \( n \) is even if and only if \( 3n^2 + 8 \) is even. \( \checkmark \)
c) We have shown this in the first half of the proof. The second half is a proof by contradiction which adds to the proof.

4 Which of the following best describes the logical relation between lines (L2) and (L6)?

a) The lines are logically independent. √

b) (L2) logically depends on statements made in line (L6).

c) (L6) logically depends on statements made in line (L2).

5 Which of the following best describes the logical relation between lines (L5), (L6) and (L8)?

a) The lines are logically independent.

b) (L8) logically depends on statements made in both lines (L5) and (L6). √

C) (L8) logically depends on statements made in line (L5) and is independent to statements made in line (L6).

6 Which of the following best explains why the proof does not stop at line (L4)?

a) Because the proof would be incomplete - we need to show both implications of the if and only if statement. √

b) Because the proof would be incomplete - we need to show if $3n^2 + 8$ is odd then $n$ is odd also.

c) The proof is complete at this point but the extra lines are additional pieces of information to help understanding.

7 Which of the following best describes the method of the second half of this proof?

a) Proof by contradiction.

b) Proof by contraposition. √

c) Proof by induction.

8 Which of the following best summarises the proof after line (L5)?

a) We assume $n$ is odd, so $n = 2a + 1$. We replace $n$ with $2a + 1$, re-arrange the terms and factorise. This gives $2(6a^2 + 6a + 5) + 1$. This is odd because of the plus 1. Therefore, by contradiction, $n$ is even if and only if $3n^2$ is even.
b) We assume $n$ is odd, so $n = 2a + 1$. We replace $n$ with $2a + 1$, re-arrange the terms and factorise. This gives $2(6a^2 + 6a + 5) + 1$. This is odd because we can replace $6a^2 + 6a + 5$ with, say $k$, as $6a^2 + 6a + 5$ is just an integer. Therefore, $2(6a^2 + 6a + 5) + 1 = 2k + 1$ which shows $3n^2 + 8$ is odd. Hence by contraposition, $n$ is even if $3n^2 + 8$ is even. \[\sqrt{\}

c) We assume $n$ is odd, so $n = 2a + 1$. We replace $n$ with $2a + 1$, re-arrange the terms and factorise. This gives $2(6a^2 + 6a + 5) + 1$. This is odd because $6a^2 + 6a + 5$ is just an integer. Therefore, $2(6a^2 + 6a + 5) + 1 = 2k + 1$ which shows $3n^2 + 8$ is odd. But this is a contradiction as we assumed $n$ was even in the first half of the proof. Hence by contradiction, $n$ is even if and only if $3n^2 + 8$ is even.

9 Which of the following best describes why we do not explicitly prove $n > 0$ and $n \in \mathbb{Z}$?

a) Because we are allowed to chose an arbitrary $n > 0$ where $n \in \mathbb{Z}$. \[\sqrt{\}

b) Because we are proving $n$ is even if and only if $3n^2 + 8$ is even and not $n > 0$ where $n \in \mathbb{Z}$.

c) Because it is clear $n > 0$ where $n \in \mathbb{Z}$. Working with $n > 0$ where $n \in \mathbb{R}$, say, would require a more complex proof.

10 According to the theorem, if $3n^2 + 8$ was an odd number, would this imply that $n$ was odd also?

a) No as the theorem talks about even numbers, not odd numbers.

b) Yes because this is the contrapositive of the theorem statement. \[\sqrt{\}

c) Yes, but not because of the theorem - it is because any odd number squared, times three and plus eight gives an odd number.

Proof C

1 Which of the following best defines the symbol $\equiv$ in this proof?

a) Equivalent to.

b) Congruent to. \[\sqrt{\}

c) Equal to.

2 Which justification best explains why $p$ cannot be 2?

a) Because 2 divides into 4 so you cannot have $p \equiv 2(\text{mod } 4)$. 
b) Because \(2 \mod 4 = (-1) \mod p\) which is shown later in the proof.

c) Because \(4n^2 + 1\) is odd so 2 does not divide into it. √

3 Which justification best explains why \(y^2 + 1 \equiv 0?\)

a) Because \(y^2 + 1\) is divisible by \(n\).

b) Because \(y^2 + 1 = (2n)^2 + 1 = 4n^2 + 1\). √

c) Because \(p\) does not divide \(n\) so \(y^2 + 1 \equiv 0 \mod (p)\).

4 Which of the following best describes the logical relation between lines (L1) and (L2)?

a) The lines are logically independent.

b) (L1) logically depends on statements made in line (L2).

c) (L2) logically depends on statements made in line (L1). √

5 Which of the following best describes the logical relation between lines (L4), (L5) and (L6)?

a) The lines are logically independent.

b) (L6) logically depends on statements made in both lines (L4) and (L5). √

c) (L6) logically depends on statements made in line (L5) and is independent to statements made in line (L4).

6 Which of the following best explains why showing that \(p \not\equiv 3 \mod 4\) proves the theorem?

a) 3 is the first odd prime number. Therefore, if \(p \not\equiv 3 \mod 4\), it had to be \(1 \mod 4\) because primes are only divisible by themselves and 1.

b) Prime numbers are either monadic (1 mod 4), triadic (3 mod 4) or 2. Since we are told that \(p\) cannot be 2, by showing it cannot be triadic it has to be monadic. √

c) \(3 \mod 4 = (-1) \mod 4\). Therefore, if \(p \not\equiv 3 \mod 4\), \(p \not\equiv (-1) \mod 4\). This means it must be 1 mod 4 by rules of modulo arithmetic.

7 Which of the following best describes the method of this proof?

a) Proof by contradiction. √

b) Proof by contraposition.

c) Proof by example.
8 Which of the following best summarises the proof after line (L3)?

a) We are told \( y^2 + 1 \equiv 0 \pmod{p} \). By doing some substitutions we show \( y^{p-1} \equiv (-1) \pmod{p} \). But this cannot be the case because we know \( p \) divides \( 4n^2 + 1 \) and by Fermat’s Little Theorem, \( y^{p-1} \equiv 1 \pmod{p} \). Therefore, we have shown \( p \not\equiv 3 \pmod{4} \) and proved the theorem. √

b) We are told \( y^2 + 1 \equiv 0 \pmod{p} \). We show \( y^{p-1} \equiv (-1) \pmod{p} \) by doing some substitutions. But this cannot be the case because we know \( y^2 + 1 \equiv 0 \pmod{4} \) so \( y \not\equiv (-1) \pmod{p} \). Therefore, we have proven \( p \not\equiv 3 \pmod{4} \) and proved the theorem.

c) We are told \( y^2 + 1 \equiv 0 \pmod{p} \). We show \( y^{p-1} \equiv (-1) \pmod{p} \) by using Fermat’s Little Theorem. But this cannot be the case because we know \( p \not\equiv 2 \) and \( y^{4k+2} \) divides 2. Therefore, we have proven \( p \not\equiv 3 \pmod{4} \) and proved the theorem.

9 Which of the following best explains why we set \( y = 2n \)?

a) Because we know \( p \not\equiv 2 \) so if \( y = 2n \), \( p \) divides \( y \) which cannot be the case. Therefore, setting \( y = 2n \) helps us to prove \( p \not\equiv 3 \pmod{4} \).

b) Because we can use Fermat’s Little Theorem to show \( y^{p-1} \equiv 1 \pmod{p} \). This is then used to show \( p \not\equiv 3 \pmod{4} \) because by modulo arithmetic, \( y^{p-1} \equiv 1 \pmod{p} \) implies \( p \equiv 1 \pmod{4} \).

c) Because we can use Fermat’s Little Theorem to show \( y^{p-1} \equiv 1 \pmod{p} \) and because \( y^2 + 1 = 4n^2 + 1 \), which is divisible by \( p \) by the theorem. This then sets up a contradiction which we use to prove \( p \not\equiv 3 \pmod{4} \). √

10 According to the theorem, is \( 133 \equiv 1 \pmod{4} \)?

a) Yes because \( p = 133 \) divided by 4 is 33.25 which is 1(mod 4).

b) No because \( p = 133 \) is not prime. √

c) No because \( p = 133 \) does not divide \( (4n^2 + 1) \forall n \in \mathbb{Z} \).
Appendix D

The alternative practice proof used in Study 3

Practice Theorem

There is no smallest positive real number.

Proof

(L1) Assume, to the contrary, that there exists a smallest positive real number.

(L2) Therefore, by assumption, there exists a real number \( r \) such that \( 0 < r < s \) where \( s \) is any other positive real number.

(L3) Consider \( m = \frac{r}{2} \).

(L4) Clearly, \( 0 < m < r \).

(L5) Therefore, this is a contradiction since \( m \) is a positive real number that is smaller than \( r \).

(L6) Thus there is no smallest positive real number. \( \square \)
Appendix E

The multiple choice questions used for Proof A in Study 3

Proof A - Multiple Choice Questions (Correct answers indicated by a tick)

1. According to the proof, which of the following would be the first possible value for $M$?
   a) $M = 87$.  
   b) $M = 135$.  
   c) $M = 311$.  √

2. In line (L7), why does the proof show that 2 does not divide $M$?
   a) Because 2 is neither monadic nor triadic but is a prime so it needs to be shown not to divide $M$ for $M$ to be monadic.  √
   b) Because 2 can also be considered as a triadic prime so for $M$ to be monadic we must show that all triadic primes do not divide $M$.  
   c) Because 2 is the only even prime number so if 2 does not divide $M$ then no even number will divide $M$.

3. Which of the following best defines a prime number?
   a) Any real number that greater than 0 and is only divisible by 1 and itself.  
   b) Any positive integer that is only divisible by 1 and itself.  
   c) Any positive integer that is greater than 1 that is only divisible by 1 and itself.  √

4. Which of the following best describes the logical relation between lines (L2) and (L8)?
   a) The lines are logically independent.  
   b) (L2) logically depends on statements made in line (L8).  
   c) (L8) logically depends on statements made in line (L2).  √
5. Which of the following best describes the logical relation between lines (L5) and (L6)?
   a) The lines are logically independent.
   \(\sqrt{\text{b)} (L5) \text{ logically depends on statements made in line (L6).}}\)
   c) (L6) logically depends on statements made in line (L5).

6. Using the method of the proof you have been working with, which of the following would be an appropriate \(M\) to use if you were trying to prove there were infinitely many primes of the form \(6k + 5\)?
   a) \(M = 4p_2...p_n + 5\) where \(p_1 = 5\).
   \(\sqrt{\text{b)} M = 6p_2...p_n + 5\) where \(p_1 = 6\).
   c) \(M = 6p_2...p_n + 5\) where \(p_1 = 5\).

7. What type of proof is this?
   a) Proof by contradiction. \(\sqrt{\text{b)} \text{ Proof by contraposition.}}\)
   c) Proof by induction.

8. Which of the following summaries best capture the ideas of the proof?
   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. \(\sqrt{\text{b)} \text{ The proof lets } M = 4p_2p_n + 3, \text{ where } p_i \text{ are prime numbers and } p_i \text{ does not equal 3.}
   \text{Thus, } 2 \text{ does not divide } M \text{ because } M \text{ is odd. Further, } p_i \text{ does not divide } M \text{ because it leaves a remainder of 3.}}\)
   c) The proof introduces monadic primes to be used later on in the proof. It lets \(M = 4p_2p_n + 3\) 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.

9. Can we conclude from this proof that the product of two triadic primes is itself triadic?
   a) No - the proof only shows the product of two monadic numbers is monadic. \(\sqrt{\text{b)} \text{ Yes - triadic and monadic primes are closely linked, as shown in the proof, so we are allowed to assume that the product of two triadic primes is triadic.}}\)
   c) Yes - this is used in the proof because \(M\) is a triadic number and this can only occur if the product of triadic primes is triadic.

10. Why does the proof include the sub-proof that the product of monadic numbers is monadic?
a) Because in line (L4) we have a product of monadic number so $M$ itself needs to be shown as monadic.

b) Because by showing that the product of monadic numbers is monadic we can then assume the product of triadic numbers is triadic.

c) Because the proof uses it in line (L8) to show that $M$ is in fact monadic leading to a contradiction. √
Bibliography


