Evaluating novel pedagogy in higher education: A case study of e-Proofs

by

Somali Roy

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 Mathematics

May 2014

Declaration of Authorship

I, Somali Roy, declare that this thesis titled, 'Evaluating novel pedagogy in higher education: A case study of the effect of e-Proofs' 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: ______________________________

Abstract

Evaluating novel pedagogy in higher education: A case study of e-Proofs

This thesis is a single case study of the introduction and evaluation of new resources and new technologies in higher education; in which "e-Proof" was chosen as a single case. E-proofs are a multimedia representation of proofs, were created by Alcock (2009), and aimed to help undergraduates to read proofs for better proof comprehension. My thesis aimed to investigate whether the impact of reading such technology-based resource, e-Proofs, on undergraduates' proof comprehension was better compared to reading written textbook proofs and if so, then why (or why not).

To evaluate the effectiveness of e-Proofs, I used both qualitative and quantitative methods. First I measured undergraduates' satisfaction, which is a most common research practice in evaluation studies, by using self-reporting methods such as web-based survey and interviews. A web-based survey and focus-group interviews showed that undergraduates liked to have e-Proofs and they believed that e-Proofs had positive impact on their proof comprehension. However, their positive views on e-Proofs did not evidence the educational impact of e-Proofs. I conducted an interview with Alcock for better understanding of her intentions of creating e-Proof and her expectations from it.

Next, I conducted the first experiment which compared the impact of reading an e-Proof with a written textbook proof on undergraduates' proof comprehension. Their comprehension was measured with an open-ended comprehension test twice — immediately after reading the proof and after two weeks. I found that the immediate impact of reading an e-Proof and a textbook proof were essentially the same, however the long term impact of reading an e-Proof was worse than reading a textbook proof (for both high and low achieving undergraduates). This leads to the second experiment in which I investigated how undergraduates read e-Proofs and textbook proofs.

In the second experiment, participants' eye-movements were recorded while reading proofs, to explore their reading comprehension processes. This eye-tracking experiment showed that undergraduates had a sense of understanding of how to read a proof without any additional help. Rather, additional help allowed them to take a back seat and to devote less cognitive effort than they would otherwise. Moreover, e-Proofs altered undergraduates' reading behaviours in a way which can harm learning.

In sum, this thesis contributes knowledge into the area of reading and comprehending proofs at undergraduate level and presents a methodology for evaluation studies of new pedagogical tools.

Acknowledgements

My first and sincere appreciation goes to my supervisors Dr.Matthew Inglis and Professor Barbara Jaworski for their continuous help and support in all stages of this thesis. Their attitude to research inspired me and I cannot give enough thanks to them.

Special thanks to Dr.Lara Alcock for being so kind and wonderful to me. I am grateful to you for your help and encouragement, which has a significant contribution in the thesis.

A big thank you to Dr.Stephanie Thomas for being a great support to whom I could always talk about my problems and excitements. I would also like to thank all my colleagues for their help over the last four years.

Finally, special thanks to my husband for encouraging me to step into the field of mathematics education. My expression of thanks is not enough for my parents and sister, for their unconditional love and for being my strength always.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

This thesis is a case study of the evaluation of new resources and new technologies in higher education in order to improve learning. I selected the computer-based "e-Proof", created especially to help undergraduates to read and comprehend proofs, as a single case for the thesis to get an insight into how a new technology-based resource is used by learners at the university level.

Case study is "an empirical inquiry that investigates a contemporary phenomena within its real-life context" (Yin, 1994, p.13). A case study is not necessarily about studying a single person but it "is used in many situations to contribute our knowledge of individual group, organisation, social, political, and related phenomena" (Yin, 1994, p.1). In his book, Yin (1994) gave an example of Graham Allison's famous case study of governmental decision-making on international relationship in which the "Cuban missile crisis" was chosen as a single case. Case studies have been criticised on the issue of generalising a single case to the other situations; however as Yin argued a single case is "generalizable to theoretical propositions and not to the populations", that is, the goal of a case study is "to expand and generalize theories (analytical generalization) and not to enumerate frequencies (statical generalization)" (Yin, 1994, p.10).

Case study research is appropriate for *evaluation research* (Yin, 1992, 1994), especially when the aim is to explore how effective a new resource is in terms of learning compared to conventional existing resources and to understand its (lack of) effectiveness. These 'how' and 'why' questions lead to an exploratory as well as an explanatory case study approach. A thorough understanding of a technology's (or resource's) intended operations and outcomes is important in case study evaluation, which leads towards an exploratory study. An exploratory study allows researchers to set hypotheses as well as to explore what makes a technology (not) effective and why. One advantage of using a case study design is that it allows researchers to use a combination of qualitative and quantitative research evidence depending on the specific research goals. In this thesis, I used both qualitative and quantitative research evidence to address the effectiveness of e-Proofs.

1.2 Background

New technologies and new computer-based multimedia resources are emerging every day in all areas of education and promise to improve teaching and learning compared to conventional instruction. Often these technologies and resources are developed and implemented assuming that they are effective. However whether those technologies are fulfilling the purposes that they are supposed to is still questionable, because of the little research evidence on their effectiveness (Dynarski et al., 2007; Highfield and Goodwin, 2008; Hoyles and Noss, 2003). Technologies have been widely used in education during last three decades, and researchers have conducted meta-analyses to investigate what kind of impact technologies had on education. Those meta-analyses, however, reached different conclusions: technologies have positive impact, but with small effect size (Kulik, 1994, 2003), similar impact as the conventional instruction (Dynarski et al., 2007), negative impact (Schacter, 1999). Moreover, researchers often questioned the quality of the current state of effectiveness studies, especially their weak research design (Cuban, 2001; Dynarski et al., 2007; Murphy et al., 2001). Murphy et al. (2001, p.2) noted that "the poor overall quality of the current state of effectiveness research is severely restricting the field's ability to learn from the experiences of others and limiting the ability to develop a knowledge base that will help inform the work of decision-makers, practitioners, and designers". These authors claimed that for the meta-analysis "approximately two-thirds of all the effectiveness studies provided by software vendors were excluded for methodological inadequacies" (Murphy et al., 2001, p.2). So far, it is not yet well established that technologies are effective for teaching and learning in different areas of education, such as reading, mathematics, sciences; but researchers have admitted the importance of evaluating new technologies that aim to promote learning. As well as the evaluating their effectiveness, it is also crucial to consider how and why a technology succeeds or fails compared to a conventional teaching and learning method (Ross et al., 2010).

It is important to mention that these comments on effectiveness research are not intended to encompass all research approaches, but merely those which place a strong emphasis on self report. There are examples of research traditions which do use more sophisticated methods, for example design research. Design research, according to Wang and Hannafin (2005, p.6), "is a systematic but flexible methodology aimed to improve educational practices through iterative analysis, design, development, and implementation, based on collaboration among researchers and practitioners in real-world settings, and leading to contextually-sensitive design principles and theories". It utilises multiple mixed methods over time and researchers work closely with participants and continuously refine design interventions to make them more applicable to practice. That means design-based research goes through various evaluation processes as part of the research process to establish a new intervention.

At the start of my PhD, I was interesting in exploring how undergraduates read and comprehend mathematical proofs. At the undergraduate level, mathematical proofs are a major part of mathematics courses and undergraduates spend a significant amount of time in proof related activities, such as reading and comprehending proofs. It is a well established fact in mathematics education that undergraduates have great difficulties with proofs (Harel and Sowder, 1998). Plenty of research has been conducted to understand what makes learning of proofs difficult (Moore, 1994). One of the reasons identified is undergraduates' lack of understanding how a proof works (Selden and Selden, 1995; Weber and Alcock, 2005). To overcome this difficulty, many educators suggested different alternatives, such as giving more stress to understanding the definitions used in proofs (Moore, 1994), writing proofs in an elaborate manner (Lamport, 1995), breaking down proofs into components (Leron, 1983) etc. To make proofs more comprehensible to undergraduates, Lara Alcock took the initiative by creating computer-based *e-Proofs*[1]. Alcock's e-Proofs are multimedia representations of proofs in which explicit audio explanations are provided for each line, that can directly guide its readers in their comprehension;

[1] A detail discussion on "e-Proof" is given in section 2.3.4 in Chapter 2

whereas a textbook presentation of the same proof might not provide such explicit explanations (Alcock, 2009).

I became interested in whether the multimedia representation of proof, the e-Proof, created precisely aiming to improve reading and comprehension of proofs at undergraduate level, is successful. So my PhD interest slightly diverted from reading comprehension of proofs in general to the effectiveness of e-Proofs. This serves three purposes:

- it allowed me to get an insight into the area of reading and comprehension of proofs at the undergraduate level.
- it also allowed me to explore the issues related to impact of the new resource called e-Proof.
- it allowed me to investigate and reflect upon different methods of evaluating novel educational resources.

1.3 Research goals

My thesis aimed to explore whether "e-Proof" is fulfilling its purpose of improving comprehension of proofs at the undergraduate level; and if so (or not), then why. More precisely:

- What was the purpose of creating e-Proofs?
- What do undergraduates think is the impact of using e-Proofs?
- What is the immediate and the long term impact of reading an e-Proof on undergraduates' proof comprehension compared to reading a written text-book proof?
- What are the reasons why reading an e-Proof is better (or worse) than reading a textbook proof?

  The first three research questions are exploratory in nature, whereas the last one is an explanatory question.

1.3.1 Methodological approaches of effectiveness research

I used both qualitative and quantitative research methods to address the above research questions. Methods commonly employed for effectiveness research are observations, expert's reviews, surveys and other qualitative and descriptive approaches. However, relying only on these qualitative research approaches is often criticised (Dynarski et al., 2007; Murphy et al., 2001; Ross et al., 2010) and one of the reasons is as Ross et al. (2010, p.27) mentioned, "receiving researchers' subjective (i.e., qualitative) impressions as the only evidence would not be sufficient". On the other side, experimental research methods allow researchers control over one or more manipulations to get conclusion about the effect of manipulating variables. Experiments compare directly between a control group (who are not exposed to a new technology) and a treatment group (who are exposed to a new technology) to examine whether that additional treatment (exposure to a new technology) has any impact on the achievement of the group (Christensen, 2003).

To understand the purpose of creating e-Proofs and what undergraduates' think of e-Proofs qualitative research methods were more suitable rather than an experimental method, whereas to understand the impact of reading an e-Proof compared to reading a textbook proof, direct comparison was essential and for that reason an experimental research approach was adopted. A detailed discussion on the methodologies adopted to address the above research questions is given later in Chapter 3, Chapter 4 and in Chapter 6.

1.4 Outline of the thesis

This thesis contains eight chapters in which three chapters report empirical studies and two chapters describe two methodologies — the experimental methodology and the eye-tracking methodology. This first chapter gives an introduction of the thesis which describes the background that inspires the research and the aims that this research wanted to achieve.

1.4.1 Chapter 2: Literature review

In the next chapter, Chapter 2, I review the relevant literature. The literature review chapter has three main sections — proof and proving, proof presentations and learning from texts and multimedia presentations. In the *proof and proving* section I discuss why proofs are important in learning of mathematics; students', especially undergraduates' notions of mathematical proofs; what makes learning proofs difficult for students; and how they engage with proving activities, such as proof construction, proof validation, proof comprehension. The next section is *proof presentation*, in which I review how proofs are presented in undergraduates classrooms and what the alternative presentations of proofs have been proposed by mathematics educators. The e-Proof is an alternative proof presentation which is also discussed in this section. In the final section of Chapter 2, *learning from texts and multimedia presentations*, I discuss literature that has considered learning from texts and learning from multimedia environments.

1.4.2 Chapter 3: E-Proof

Chapter 3 holds discussion on *e-Proofs* from two different view points. Firstly, by reporting a web-based survey and three focus-group interviews with undergraduates who were exposed to e-Proofs during their real analysis module, I explore undergraduates' perceptions on e-Proofs and what they think is the impact of reading e-Proofs on their learning. Secondly, I explore Lara Alcock's intentions and expectations from e-Proofs through a semi-structured interview. In short, this chapter highlights the issue of effectiveness of e-Proofs based on a self-reporting research strategies which is a very common research approach in effectiveness study in education (Dynarski et al., 2007; Yin, 1992, 1994).

1.4.3 Chapter 4: The experimental methodology

Chapter 4 is about the experimental methodology which I used for two empirical studies reported in Chapter 5 and Chapter 7. This chapter starts with a discussion of why an experimental research approach was a suitable methodology to address the research goals. Validity and reliability are essential criteria of research and I explain how these issues were addressed to obtain maximum validity and reliability in my studies. I also discuss the ethical considerations addressed while planning and conducting the research.

1.4.4 Chapter 5: Experiment 1: The effectiveness of the two ways of presenting proofs on undergraduates' proof comprehension

Chapter 5 reports the first experiment in which I compared the effectiveness of e-Proofs on undergraduates proof comprehension with traditional written textbook proofs. A between-subjects research design was used in which two groups of undergraduates read either an e-Proof version of an unseen proof or a written textbook version of that proof for comprehension. Their comprehension was measured by an open-ended comprehension test twice — immediately after reading the proof and after two weeks. I found that, the immediate impact of reading an e-Proof and a textbook proof were essentially the same, however the long term impact of reading an e-Proof was worse than reading a textbook proof. After reporting the outcomes of the experiment, a follow-up discussion highlights possible conjectures as to why the long term effect of e-Proofs are poor. To support the argument, I planned the second experiment which is reported in Chapter 7.

1.4.5 Chapter 6: The eye-tracking methodology

Chapter 6 describes the eye-tracking methodology that I used in the second experiment in which I investigated how undergraduates read and comprehend e-Proofs and written textbook proofs. I discuss the issues related to the *think aloud methodology* to highlight why the *eye-tracking methodology* was more suitable for exploring undergraduates' reading comprehension processes compared to think aloud research approaches. I describe the assumption of the eye-tracking methodology along with some terminologies related to this methodology that I use widely in Chapter 7. The Tobii T120 eye-tracker was used for my study and I describe briefly how I recorded and analysed the eye gaze data using the eye-tracker.

1.4.6 Chapter 7: Experiment 2: Reading comprehension of e-Proofs and textbook proofs

In Chapter 7, I report the second experimental study in which I investigated the reading comprehension processes of e-Proofs and written textbook proofs. A within-subjects research design was used in which participants (who were exposed to e-Proofs during their real analysis course) read both an e-Proof version and a textbook version of several proofs. Participants' eye-movements were recorded through the remote Tobii T120 eye-tracker to explore similarities and differences when undergraduates read these two different proof presentation methods. This experiment allowed me to explain why reading e-Proofs was less effective in terms of long term learning compared to reading written textbook proofs.

1.4.7 Chapter 8: Conclusion

In the last chapter, Chapter 8, I review my findings from Chapter 3, Chapter 5 and Chapter 7 to draw overall conclusions. I discuss the methodologies used for studying effectiveness of e-Proofs and how my findings contribute to a body of knowledge on the issue of how to study the effectiveness of new educational technologies and resources in learning mathematics.

Chapter 2

Literature Review

2.1 Introduction

This literature review has three main sections — proof and proving, proof presentation and learning from texts and multimedia presentations. In the proof and proving section, I review why proofs are important in learning of mathematics; what students, especially undergraduates, think of mathematical proofs; and how they engage with proving activities, such as proof construction, proof validation and proof comprehension. I then discuss proof presentation, by firstly considering how proofs are presented in undergraduate classrooms and what alternative presentation methods have been proposed by mathematics educators. The e-Proof is a specific alternative proof presentation method which I discuss in detail in Section 2.3. Comprehending proofs is a matter of both linguistic and conceptual understanding ; therefore in the final section, I discuss the literature that has considered learning from the texts and learning from multimedia environments[1].

2.2 Proof and proving

Mathematical proofs have been discussed in the mathematics education literature from several points of view, for example, why proofs are needed from a philosophical view point (Rav, 1999); what role proofs can play in classroom practice from social view point (Hanna, 1989; Hemmi, 2006). However mathematicians and

[1] Because the e-Proofs are presented in a multimedia environment.

mathematics educators are agreed about the importance of proof in mathematics, especially given the structure the modern mathematics. However, what proof is still a debatable topic and there is no agreed answer among philosophers.

2.2.1 What is proof?

Proofs are everywhere in modern mathematics, to the extent that some have discussed proof as "the heart of mathematics" (Rav, 1999, p.6). Probably proof makes mathematics different from the other discipline of sciences — because of proof, once a conjecture is established and accepted by mathematicians as correct, it will remain correct; whereas observing a new phenomena may have the power to change a well established truth in physics or biology. In the early days of mathematics, it was not very different from other empirical sciences without proofs (Krantz, 2007). But since the advent of proof formalised mathematics (definitions, axioms, theorems, proofs), mathematics is no longer an empirical science, rather "mathematics is (i) coming up with new ideas and (ii) validating those ideas by way of proof. The timelessness and intrinsic value of the subject comes from the methodology, and that methodology is proof" (Krantz, 2007, p.33).

Proofs have been described in many ways in the literature:

- A mathematical proof is just a sequence of logical steps, following some (implicit) formal rules, to be judged on the merits of legitimacy of these steps (Rav, 2007).
- A mathematical proof is an argument that proves a theorem (Selden and Selden, 2003).
- "A proof is an argument needed to validate a statement, as argument that may assume several different forms as long as it is convincing" (Hanna, 1989, p.20).

For this thesis, I am considering 'proofs' that undergraduates usually experience in different domains of mathematics at the university level. Only well established proofs (that have significant impact on that domain of mathematics) are part of the undergraduate syllabus; proofs where the validity is still a debatable question among mathematicians are usually not introduced at the undergraduate level. So, for this thesis, I am considering only 'proofs' whose validity, by virtue of their source, seems to be beyond any question of doubt.

2.2.2 The role of proof in mathematics classroom

Many mathematics educators have discussed the various roles of proof in the learning of mathematics. For mathematicians, proof is a communicating tool through which they convince their colleagues about their research (Hersh, 1993; Rav, 1999). Hersh (1993) argued that conviction is not a problem for students as they are convinced easily; therefore, in classroom, proofs are presented to explain why a statement is true. A good proof, as Hersh suggested, would convince a skeptical mathematician and also explain to a naive student. Proof has been characterised in terms of its role and importance in mathematics classroom practice mainly as follows:

- *Verification or Justification*: Verifying the truth of a proposition is a primary purpose of a proof (Bell, 1976; de Villiers, 1999). Once a proposition is verified and accepted by the mathematics community as truth in the form of proof, it remains unchanged timelessly.
- *Explanation (or promoting understanding)*: A proof is expected to convey an insight into why a proposition is correct (Bell, 1976; de Villiers, 1999). Moreover, Hanna (1995) claimed that a proof that fails to provide some ideas to the reader 'why' a theorem is true adds little to the reader's understanding and therefore may not even be very convincing. For classroom practice, proofs promote mathematical understanding by explaining how propositions can be established as proofs.
- *Systemisation*: The organisation of results into a deductive system of axioms, major concepts and theorems, and minor results derived from these is an important characteristic of mathematical proof (Bell, 1976; de Villiers, 1999).
- *Communication (or bearer of knowledge)*: Proof can promote new mathematical knowledge, for example, one can learn new proof techniques by reading proofs (Hanna and Barbeau, 2008; Knuth, 2002; Rav, 1999). Rav (1999) claimed that learning (or illustrating) new proof techniques is one of the key reasons why mathematicians read proofs (or produce proofs).

Although proof has several functions, it is not always possible to distinguish these functions from one another, because "in some cases certain functions may dominate others, while in some cases certain functions may not feature at all" (de Villiers, 1999, p.11). But certainly a good proof, as Manin (1981) said, makes its reader wiser.

2.2.3 The perception of proof at university level compared to school level

School mathematics primarily involves calculation performed upon specific mathematical objects, whereas university mathematics is more formal with plenty of proofs (Alcock and Simpson, 2002b; Harel and Sowder, 1998; Moore, 1994). In many countries mathematical proofs are not in the school mathematics syllabus and because of the little exposure of proof at school level, students' perceptions of proof can be quite poor — they often demonstrate vague ideas about what constitute a mathematical proof (Healy and Hoyles, 2000; Knuth, 2002; Martin and Harel, 1989; Segal, 1999). Research shows that the majority of school students accept inductive arguments (proof by examples) as valid proofs and many students hold two different perceptions of proof (conviction and validity) simultaneously. Students may see inductive arguments as convincing, but recognise that this kind of argument may not be accepted by teachers (or by the mathematics community) as a valid proof (Healy and Hoyles, 2000; Segal, 1999). On the contrary, some students believe that deductive arguments always constitute a proof and would be accepted by their teachers. Moreover, many students (even many elementary school teachers) get convinced by the 'formal look' of a deductive argument rather than its 'content' (Harel and Sowder, 1998; Healy and Hoyles, 2000; Martin and Harel, 1989; Segal, 1999).

It is also accepted by mathematics educators that a natural way of learning is by induction (Fischbein, 1982; Harel and Sowder, 1998), but it is also widely accepted that students need to develop the skill of proving in deductive way; especially to understand the structure of mathematics at university level (Moore, 1994; Vinner, 1976). It has been argued by many mathematicians and mathematics educators that the notion of proof should be introduced at the school level (Bell, 1976; Hanna and Jahnke, 1993; Tall, 1989).

To overcome the newly entered university students' naive perceptions of proof, many universities offer proof oriented courses (often called 'bridge' courses or 'transition to proof' courses) in which students are introduced to the notion of formal proof and various proof techniques (for example, proof by induction, proof by contradiction) in their first year of university. By introducing such courses, it is expected from students that they would appreciate and recognise the value of proof in mathematics and would be able to do proof related activities by themselves in more advanced courses. However, it has been found that even after attending such introductory proof-related courses many undergraduates still face difficulties with proving activities, such as, how to start proving a conjecture, how to use definitions in proofs etc (Moore, 1994).

2.2.4 Proof construction

Constructing proofs is a common and regular activity for mathematics undergraduates. Research shows that students have several problems with proof construction: how to begin a proof (Moore, 1994); how to derive a proof from a definition (Moore, 1994; Weber, 2001); applying relevant knowledge while proving (Weber and Alcock, 2004); applying an appropriate strategy and the right idea at the right time when proving (Raman, 2003; Weber, 2001). Tall (1991) argued that, in classrooms, teachers often present a ready-made proof to the students, and consequently the proving process gets little attention in classrooms. Students experience the end product of a proving process, but are not always told explicitly the process through which mathematicians construct proofs. According to Tall, that is one of the reasons behind students' difficulties in proving activities and he suggested that students need teaching in advanced mathematical thinking which may lead them towards a successful proof construction.

Proof construction is a complex mathematical activity with logical, conceptual and problem solving dimensions (Weber, 2005). However a constructed proof only becomes a proof when the mathematics community accepts it as a proof (Hanna, 1990; Manin, 1981). Mason described three steps through which one can establish a proof: by convincing yourself, by convincing a friend and finally by convincing an enemy (Mason et al., 1982). Similarly, for Harel and Sowder (1998), the proof construction process has two subprocesses: (i) *ascertaining* is the process through which an individual removes personal doubts about the truth of a proposition and

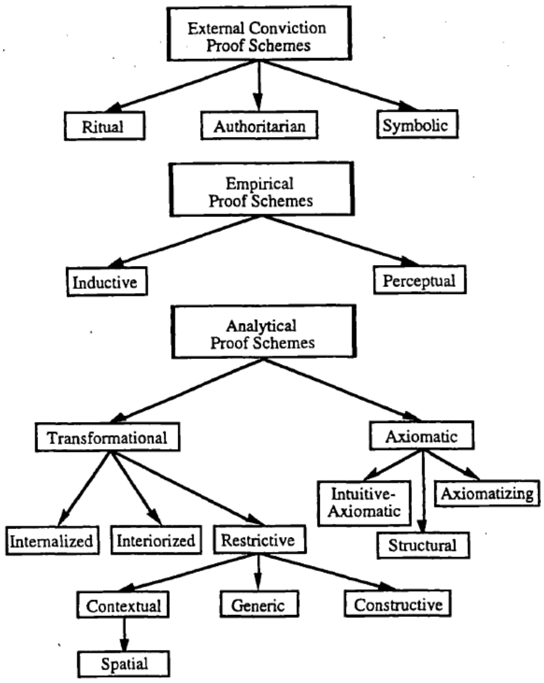


FIGURE 2.1: Categories of students' proof schemes (Harel and Sowder, 1998, p. 245)

(ii) *persuading* is the process through which an individual removes others' doubts. These authors proposed students' proof schemes in which they categorised undergraduates' proving processes (based on the above definitions). Their proposed proof schemes have three categories — (i) the external conviction proof scheme, (ii) the empirical proof scheme and (iii) the analytical proof scheme (See Figure 2.1). These proof schemes have identified several behaviours of students while constructing proofs, which have been also found across the mathematics education literature. In the next section I will discuss Harel and Sowder's (1998) students' proof-schemes in detail and also refer to other studies that reported similar proof schemes.

The proof schemes

The first caregory of Harel and Sowder's (1998) proof schemes is the external conviction proof scheme. This scheme represents behaviour where students' thinking approaches are influenced by issues that are not directly related to the subject knowledge while proving a conjecture. This proof scheme has three subcategories:

the ritual proof scheme, the authoritarian proof scheme and the symbolic proof scheme.

In the ritual proof scheme, students are convinced by the surface appearance of an argument without checking its correctness. For example, students may be convinced by a false deductive argument over a valid narrative argument. Similar tendencies of accepting a false-argument as a valid proof were found among school students (Healy and Hoyles, 2000; Segal, 1999), preservice school teachers (Martin and Harel, 1989) and higher secondary school teachers (Knuth, 2002); because they thought the argument 'looks like a proof'.

The authoritarian proof scheme is the second subcategory of the external conviction proof scheme. The authoritarian proof scheme covers students' dependency on the authority which could be books, class-notes or teachers. Harel and Sowder (1998) observed that some students relied on their teachers (or on study materials) as if those teachers (or those materials) were the sole source of knowledge acquisition. As a consequence, while proving, such students would always wait for their teachers' answers or go through their notes for solutions; but would rarely show confidence in their own abilities.

The third subcategory of the external conviction proof scheme is the symbolic proof scheme, which has identified that many students have a tendency to start a problem solving task by manipulating symbolic expressions involved in the problem. Those students spent almost no time in understanding the problem situation or the meaning of symbols and their functional importance in the problem. The authors argued that students have picked up this tendency during their school years and that it is extremely difficult for them to give up the habit.

The second category of Harel and Sowder's (1998) proof schemes is the empirical proof scheme which has two subcategories: the inductive proof scheme and the perceptual proof scheme. Under the inductive proof scheme, Harel and Sowder (1998) described students' well-known tendency to prove a conjecture by using one or more examples. Many research studies suggested that the majority of students at school level (Chazan, 1993; Healy and Hoyles, 2000; Segal, 1999) and also at the beginning of undergraduate level (Recio and Godino, 2001) prefer to use inductive proofs, although being aware of their limitations. Selden and Selden (2003) argued that students are aware of what constitutes a valid proof, but that they are not able to construct a proof using anything else but examples. Many mathematics educators accept the fact that thinking inductively comes naturally (Bell, 1976; Harel and Sowder, 1998; Segal, 1999), but their concern is "students proof schemes do not develop beyond the empirical proof scheme" (Harel and Sowder, 1998, p. 252). Some students seem to accept an inductive argument even after finding a counterexample, believing that "the counter example is just an exception" (Harel and Sowder, 1998, p. 253).

In the perceptual proof scheme, which is the second subcategory of the empirical proof scheme, students tried to establish the certainty of a conjecture through their perceptual observations — by using elementary and often inadequate mental images (Harel and Sowder, 1998). Mathematicians may also use their perceptual observations as "key ideas" for proving theorems (Raman, 2003). Research shows that for solving problems or for proving theorems many students use their 'concept image' rather than 'concept definition'[2] (Edwards and Ward, 2004; Tall and Vinner, 1981; Vinner, 1991) and having inadequate concept images does not help students to solve problems but misleads them (Edwards and Ward, 2004).

The third and final proof scheme is the axiomatic proof scheme in which Harel and Sowder (1998) described students' proving processes by using logical deductions. This proof scheme has two subcategories— the transformational proof scheme and the axiomatic proof scheme. The transformational proof scheme involves goal oriented and anticipatory mental operations which involve transformations of images as part of the deductive process. The axiomatic proof scheme covers students' awareness and understanding of the role of initially accepted results (or axioms) in verifying or proving the certainty of a conjecture.

Undergraduates' proving processes

Students taking an inductive approach while proving a theorem is a common scenario (Harel and Sowder, 1998; Healy and Hoyles, 2000; Recio and Godino, 2001; Segal, 1999). Many mathematics educators have investigated how undergraduates construct their own proofs, especially when they take a deductive approach (Raman, 2003; Weber, 2001; Weber and Alcock, 2005). Among several processes that undergraduates use for constructing deductive proofs, I will discuss three processes in particular: the procedural, the syntactic and the semantic.

[2]A concept image does not necessarily need to be a visual image but concept image is "the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes"(Tall and Vinner, 1981, p.152).

- *The procedural proving process*: In a procedural proving process, "undergraduates sometimes attempt to locate a proof of a statement that is similar in form and use this existing proof as a template for producing a new one"(Weber, 2005, p.353). Through this process students can produce a correct proof by mimicking an existing proof but this may result in little understanding.

- *The syntactic proving process*: In a syntactic proving process some undergraduates attempt to write a proof by manipulating correctly stated definitions and other relevant facts in a logically permissible way in response to a proving task (Weber and Alcock, 2004). Raman (2003) described the idea behind this proving process as 'procedural idea' which leads to a formal proof without a connection of informal understanding. However, Alcock and Inglis (2008) found evidence which suggested that a syntactic proving process can lead to a meaningful informal understanding.

- *The semantic proving process*: In a semantic proving process, some undergraduates use their informal or intuitive representations of relevant concepts to understand the statement to be proven and use that understanding for constructing formal proof (Weber and Alcock, 2004). According to Raman (2003), the idea involved behind this proving process is the 'heuristic idea' which gives a sense of understanding.

Weber (2001) argued that even though undergraduates may have sufficient understanding and knowledge in a domain of mathematics, they may nevertheless fail to a construct proof successfully. He suggested that because, "there are many inferences one can derive, but most of these inferences will not be relevant" for proving [p. 101]; therefore undergraduates need to use "strategic knowledge". He described "strategic knowledge" as "knowledge of how to choose which facts and theorems to apply" [p. 101] . Weber's study suggested that undergraduates' approaches were often inadequate for proving as they did not know powerful proof techniques and how to apply those techniques successfully to prove a proposition. Moreover, some undergraduates were also unable to recognise which theorems are important and when to use them and when not; Weber suggested that the lack of such strategic knowledge is one of the causes of their difficulties.

From a similar viewpoint, Raman (2003) argued that different ideas are involved in the proving process; but that the type of ideas undergraduates' use are mainly heuristic and procedural. In heuristic ideas, undergraduates' use informal intuitive thoughts that lead the proving process and give a sense of understanding; whereas procedural ideas are formal logical thoughts which lead towards a formal deduction and give a sense of conviction. But for mathematicians, Raman argued that proofs are essentially about "key ideas" which is a balance of heuristic ideas and procedural ideas. Moreover, key ideas give both a sense of understanding and a sense of conviction; however undergraduates are often not aware of key ideas involved in proofs.

2.2.5 Proof validation

Besides proof construction, another activity that has received attentions in several studies is proof validation. In the present section, I will discuss studies in which 'proof validation' is used as a process of checking whether a purported proof is correct or not. Selden and Selden (1995) described proof validation as a complex activity which involves many sub-activities including making affirming assertions, asking and answering numerous questions of oneself, and perhaps even constructing subproofs. These authors also argued that, it is a regular activity for mathematicians as part of their mathematics practice through which they learn new mathematics. But "there has never been a single set of universally accepted criteria for the validity of a mathematical proof" (Hanna, 1995, p. 44). And one of the consequences is although it is purely a mathematical activity, many non-mathematical issues can influence the process of proof validation. For example, who is the author of a proof can influence the outcome of a proof validation process. In Weber's (2009) study, a couple of mathematicians rejected an argument because of a gap in the argument and that proof was constructed by a student. Those mathematicians accepted that they would accept such argument as a valid proof if it came from a mathematician. Based on those mathematics' judgements that which proofs are valid or invalid, Weber pointed out that "processes involved in proof validation are highly dependent on contextual factors, including the mathematical domain in which the proof is situated, the community that is evaluating the proof, and the author of the proof" (Weber, 2008, p. 451). Thurston (2006) expressed a similar view that the validity of some proofs cannot be determined independently from the context and community in which they were produced.

Many mathematics educators suggested that proof validation is essential for learning and comprehending proofs (Alcock and Weber, 2005; Selden and Selden, 1995; Weber and Alcock, 2005). However it is not a common classroom activity. Several studies have explored that even though undergraduates experience plenty of proofs, they are poor in proof validation tasks. Selden and Selden (2003) examined undergraduates' proof validation capabilities by presenting four different proofs (student-constructed) of a theorem and asked them to determine which constituted a valid proof. Initially less than half of the participants classified the proofs correctly. After receiving participants' judgements on valid or invalid arguments, the authors met the participants in a reflective interview session in which participants were partially guided to reflect and reconsider their judgements. As a consequence, the authors reported that undergraduates' responses changed dramatically and 81% were able to judge the valid proof. The authors claimed that instructions could improve performance in validation tasks, but they did not mention what kind of instruction they used during interviews and what kind of instruction could be effective for proof validation tasks.

Selden and Selden (2003) did not emphasise the process by which undergraduates validate proofs and why they accept or reject an argument as a valid proof. These issues were investigated by Alcock and Weber (2005) who focused on the way undergraduates attempt to validate proof and the reasoning they used when validating proofs. The statement "$(\sqrt{n}) \to \infty$ as $n \to \infty$" and its proof (an incorrect proof) were presented to the 13 undergraduates who were asked to check the proof and make corrections to it where appropriate. Three undergraduates rejected the proof instantly as they identified that a false warrant[3] was used in the proof. They also produced counterexamples to support their arguments. Three other students rejected the proof by arguing that no definition was used in the proof. These students demonstrated a tendency to use definitions without recognising when they are needed. Students' tendency of start solving problems or proving theorems or judging the truthfulness of statement by writing down definitions first have been found in several studies (Alcock and Simpson, 2002a; Harel and Sowder, 1998; Pinto and Tall, 1999). Other participants accepted the proof as valid until they were asked to pay their attention to a particular point (where a false warrant was used). The authors suggested that inadequate knowledge in relevant subject area is a primary reason behind students' failure in validation task.

[3]The justification of why a data (a previous statement) supports a conclusion (a next statement). In the next paragraph 'warrant' will be discussed.

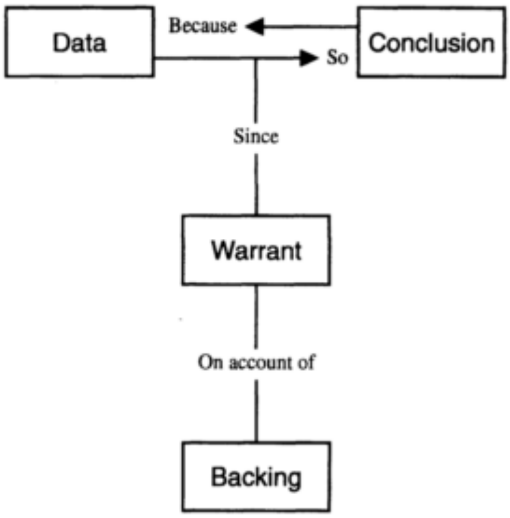


FIGURE 2.2: Toulmin's model of argumentation (Weber and Alcock, 2005, p.35)

Therefore the question arises what kinds of skill or knowledge is needed in proof validation tasks besides adequate knowledge in the domain of proofs. Selden and Selden (1995) suggested that understanding the logical structure of a statement is essential for validating proofs. Weber and Alcock (2005) argued checking warrants is key when judging whether each assertion is used in a proof is correct or not. These authors argued that an implication can be true logically but cannot be valid in order to prove a theorem. The example they used in the paper was: "if 7 then 1007" is logically true but not valid mathematically. Weber and Alcock used four main components of Toulmin's model of argumentation — data, conclusion, warrant and backing (see Figure 2.2) — to explain how the validation of a mathematical argument can be checked with clarity. One can draw a conclusion based on data and the explanations of why the data supports the conclusion is known as a warrant. If a warrant is not sufficient to draw a conclusion then 'backing' provides additional support to the warrant. In mathematics "if-then" structure is common, but while validating a proof, it is important to check whether that 'if-then' structure is warranted to establish a proof or not .

But to successfully complete a task of proof validation, it is important to comprehend a proof first without an understanding of how statements in a proof work, someone might not be able to justify whether a given argument is a valid proof or not.

2.2.6 Proof reading and comprehension

This section will shift focus from the activities such as proof construction and proof validation that have already got a great deal of attention from mathematics educators, to the activity of reading mathematical proofs which has received so far little attention in mathematics education research (Inglis and Mejia-Ramos, 2009; Selden and Selden, 1995).

Reading is the first step of learning, but teaching reading in mathematics is not something that teachers commonly do (Shepherd, 2005), even though "teaching reading in mathematics is different than teaching reading because teaching reading in mathematics means helping students make sense of — and learn from ... mathematics text" (Barton et al., 2002, p.1). The conceptual density of mathematics texts is one the major difficulties for reading and comprehending mathematics texts (Barton et al., 2002; Davis et al., 1995). According to Davis et al. (1995, p.313), "the absorption of a page of mathematics on the part of the professional is often a slow, tedious, and painstaking process".

What makes a mathematics text different from texts of other areas is the language of mathematics — a combination of words and symbols. The symbolic aspects of mathematics makes mathematics distinguished from other subjects. In mathematics, words and symbols often have a definite meaning, a meaning that sometimes is different than in other non-mathematical situations and and they cannot be read unless readers know the meaning (Pimm, 1989)[4]. Österholm (2006) conducted an experimental study in which the author compared undergraduates and high secondary school students' reading comprehension of three texts — a mathematical text with symbols, a mathematical text without symbols and a historical text. The results indicated a similarity in reading comprehension between the mathematical text without symbols and the historical text, whereas the comprehension between the two mathematical texts was different. Österholm (2006), therefore, concluded that reading a mathematical text is a special kind of activity that needs to be learned. Other mathematics educators also echoed that suggestion: that it is important to teach undergraduates to read and comprehend mathematical texts (Cowen, 1991; Selden and Selden, 1995; Shepherd, 2005), especially at university level, and that reading comprehension also needs to be examined (Cowen, 1991).

[4]For a simple example, take the sentence 'what is the difference between 24 and 9'. In mathematics, 'the difference' means the subtraction of 24 and 9.

The role of definition in comprehending proof

Compared to school level mathematics, university mathematics is more formal, definitions-axioms-theorems oriented and deductive in nature. Theorems and proofs are often derived from definitions. In mathematics, definitions are more formal, contain mathematical language and mathematics symbols and express the idea by using less words.

In mathematics classrooms, teachers often explain the condensed ideas expressed by the definition in a less formal way by using everyday language, by drawing diagrams or by providing examples (Weber, 2004). Those informal or semi-formal explanations lead students to create their own image associated with the definition. This formation of concept which could differ from the formal definition is referred as 'concept image' by Tall and Vinner (1981). The concept image could be in any form, the visual representation to the collections of impressions or experiences, "the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes". The notion of 'concept definition' and 'concept image' proposed by Tall and Vinner (1981) are influential concepts as it have been observed widely that some students use their concept image more often than the concept definition.

How mathematics majors used definitions when studying real analysis course and group theory was investigated by Edwards and Ward (2004). They found that undergraduates' understanding of the role of definition was different from mathematicians. The authors argued mathematicians see definitions as stipulated (objects with fixed properties), whereas for some students, they "have to make the definition" (Edwards and Ward, 2004, p.415). The authors also noticed that even though students were provided definitions to solve problems; they did not use them until they were deliberately asked to use definitions. Sometimes students were able to state and explain a definition correctly, but did not use them properly to solve problems. Other research studies also pointed out students' difficulties in applying definitions (Moore, 1994; Weber, 2001); even though those students knew the correct definitions. When using a definition, some students often relied on their concept image rather than concept definition —"when the concept definition conflicted .... with concept image, the concept image won" (Edwards and Ward, 2004, p.417).

Most of the mathematics modules at undergraduate level are theorem-proof oriented in which definitions have a crucial role to play. But students face great difficulties with theorems and proofs, and one of the factors behind their difficulties is not knowing definitions and not being able to use definitions properly (Moore, 1994).

Mathematical reasoning

Reasoning is known as a basic requirement in comprehending mathematics and the NCTM (2000) put emphasis on proof and reasoning in the school mathematics curriculum. Reasoning in general involves inferences that are drawn from principles and from evidence, whereby the individual either infers new conclusions or evaluates proposed conclusions from what is already known (Johnson-Laird, 1994).

A range of empirical studies have investigated students' mathematical reasoning ability at school level to university level. Rodd (2000) argued that the logical implication in the form of 'if p then q' is the most basic structure for establishing a mathematical truth. Hoyles and Küchemann (2002) investigated how school students determine whether a statement of logical implication is true or not. They collected data twice from 63 different schools in England with students of Year 8 and one year later when these students were in Year 9. The authors presented two statements — an implication and its converse (the first one was false and the second one was true) and asked students whether these statements were saying the same thing or not. The two statements used in this study were:

- If the sum of two whole numbers is even, their product is odd.
- If the product of two whole numbers is odd, their sum is even.

The results of the study suggested than only 13% students said these two statements are not the same, between two years, 15% students changed their answers to 'yes same statements' to 'no, not same' and 71% of students stated that both statements were saying the same thing. A year later, 60% students still maintained their statements that two statements were saying the same thing. Overall, students' reasoning skills were rather poor.

Selden and Selden (1995) investigated the third and fourth year undergraduates' abilities in unpacking the logic of mathematical statements. These authors believed that unpacking the logic is a vital skill for reading and comprehending proofs. They presented four informal statements, two were correct and two were incorrect, to their students and asked them to write down logically equivalent formal statements. They believed that this task could explore students' understandings about the logical structure of a statement. Less than 10% students was completed the task successfully. This result suggested that undergraduates are poor in comprehending the logical structure of statements. The authors also argued that lack understanding of the logical structure of a statement, would reflect less success in proving related activities.

Proof comprehension model

Reading and comprehending proofs can be extremely difficult because the gaps between the statements in proofs are sometimes big. A proof may become impossibly long if every detail is included in that proof (Davis et al., 1995; Lamport, 1995). Moreover, proof is a combination of many aspects of mathematics such as definitions, logical relations, which indicates that comprehending a proof demands multiple levels of understanding. Here, I will discuss two reading comprehension models — Lin and Yang's (2007) reading comprehension model for geometric proofs and Mejia-Ramos et al.'s (2010) proof comprehension model for undergraduate mathematics.

Lin and Yang (2007) interviewed mathematics teachers to identify what are the factors students need to comprehend when reading proofs to come up with the hypothetical model of reading comprehension. They proposed this model for geometric proofs, however, this model can fit into all levels of mathematics proof. Their model has four levels and five facets of comprehension (see Figure 2.3). The four levels are: surface, recognising elements, chaining elements and encapsulation (represented by the rectangular boxes in the Figure 2.3 from bottom to top). In the bottom level, i.e., in the surface level, understanding of the terms and symbols used in a proof is required, but without analysing how and why they are used. The next level is recognising elements. To reach this level, it is required to understand the premises, conclusion or the other properties used in the proof implicitly in this

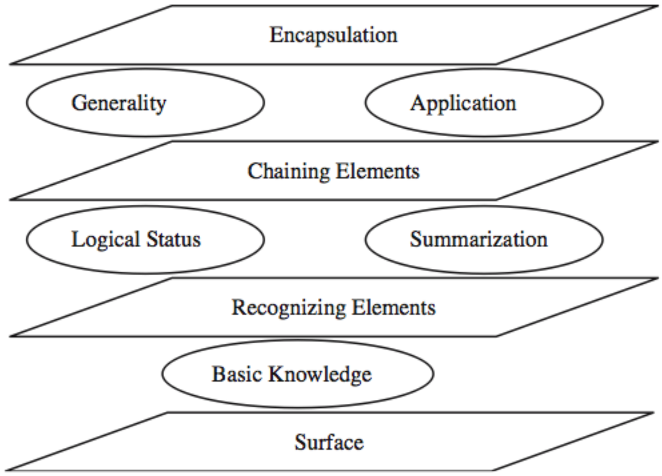


FIGURE 2.3: The hypothetical reading proof comprehension model (Yang and Lin, 2008, p.71). The rectangle boxes represent four level of comprehension and the five facets of comprehension lie between the different level of comprehensions.

level. The third level is chaining logically, in which, according to the model, identifying logical links between those premises, conclusion and properties are needed. Finally, the fourth level is encapsulation where interiorizing the proposition and the proof as a whole is needed in order to apply that in a relevant situation.

Within Yang and Lin's (2008) model, the five facets of proof comprehension lie between the four levels that discussed above. However, the authors did not explicitly mention the difference between a facet and a level. These five facets are (i) basic knowledge; (ii) logical status; (iii) summarisation; (iv) generality and (v) application.

- *The facet of basic knowledge*: This facet refers to recognising and understanding the meaning of terms, symbols are properties used in a proof and the understanding of the algebraic operations involved in a proof. They placed basic knowledge between level of surface comprehension and level 2, comprehension of recognising elements (see Figure 2.3).

- *The facet of logical status*: This facet refers to the understanding of logical links between statements, premises, conclusion. They put the facet of logical status between comprehension level of recognising elements and comprehension level of chaining elements. They claimed that by linking arguments into logical chains students are approaching the comprehension level of chaining elements.

- *The facet of summarisation*: This facet refers to the understanding of the critical ideas involved in the proof. Knowing the critical ideas in a proof was identified as approaching the level of chaining elements, therefore the authors placed this facet between comprehension level of recognising elements and comprehension level of chaining elements.

- *The facet of generality*: This facet refers to understanding a proof as a whole, understanding what is validated in a proof and justifying the correctness of the proof.

- *The facet of application*: This facet refers to knowing how to apply a theorem and its proof in other situations. Both generality and application were placed between comprehension level of chaining elements and level of encapsulation because both derived an orientation towards the creation of new knowledge.

Mejia-Ramos et al. (2010) proposed a proof reading comprehension model which is an extended model of Yang and Lin's (2008) for undergraduate-level proofs. This model has six dimensions of proof comprehension that identify important aspects of proof understanding. Mejia-Ramos et al. (2010) however put more emphasis on how, at the university level, proofs should be assessed by asking different types of questions which could reflect the different dimensions of undergraduates' proof comprehension.

- *Meaning of terms and statement*: According to Mejia-Ramos et al.'s (2010) model, understanding the meaning of individual terms and statements used in a proof is fundamental and this dimension is analogous to Yang and Lin's (2008) surface level comprehension. There are three features within this dimension that should be assessed — (i) understanding the meaning of a theorem statement, (ii) understanding the meaning of the individual statement in the proof and (iii) understanding the meaning of the terms in the proof.

- *Justification of claims*: This dimension of proof comprehension involves grasping how new assertions follow from previous ones. For assessing that a reader can be asked to make explicit justifications of statements used in the proof. The authors claimed that this dimension of comprehension is similar to the third level, chaining elements, of Yang and Lin's (2008) model.

- *Logical structure*: This dimension of proof comprehension identifies how statements are logically linked or the relationships between different components in a proof.
- *Higher-level ideas*: This dimension involves breaking down a proofs into modules and chaining those modules together to produce a proof summary. Mejia-Ramos et al. (2010) argued that understanding the central ideas of a proof and making and identifying a good summary of a proof can assess a reader's grasp on the higher-level ideas of that proof.
- *General method*: The dimension of proof comprehension recognises the understanding of what method is used in a proof, knowing when that method is applicable.
- *Application to examples*: This dimension refers to the understanding whether readers are able to apply specific examples by following the sequence of a proof.

Both Yang and Lin's (2008) and Mejia-Ramos et al.'s (2010) models can be used to assess proof comprehension success.

Assessing proof comprehension

Comprehending proofs is key for learning proofs, and how well students comprehend a proof is usually assessed in examinations by asking them to reconstruct that proof. This method often does not reflect the students' level of comprehension, because to reconstruct a proof many students rely on their memory with little or no understanding of a proof (Conradie and Frith, 2000; Houston, 1993). An alternative way through which students' proof comprehension can be assessed as suggested by many mathematics educators is comprehension tests (Conradie and Frith, 2000; Houston, 1993; Mejia-Ramos et al., 2010). Comprehension tests are hugely used in language courses. The main features of this method are that students are presented a theorem and its proof and then required to answer questions based on that proof. Conradie and Frith (2000) argued that comprehension tests can give more precise evaluation of students' understandings at all levels. The questions that these authors suggested for comprehension tests can be used (i) to test understanding of basic knowledge (the facet of basic knowledge, in the sense of Yang and Lin (2008)), (ii) to test understanding of specific steps in a proof (the facet of logical status), (iii) to test understanding of concepts used in a proof (the facet of summarisation), (iv) to test understanding of applying the conclusion of a proof (facet of application) or (v) to test understanding of some more subtle aspects of a proof (facet of generality) and so on. Therefore, it can be argued that in a comprehension test, questions can be set in such a way that could require different facets of proof comprehension and therefore this test could be an useful methodological tool of examining students' different facets of proof comprehension.

However, Hodds et al. (2014) conducted a study in which he assessed undergraduates understanding of a mathematical proof by using the proof comprehension model proposed by Mejia-Ramos et al. (2010). A factor analysis which was conducted to show the influence of different questions (dimensions of proof comprehension), suggested there was only one factor. That means their study suggests that proof comprehension is unidimensional.

Whereas Conradie and Frith (2000) proposed comprehension tests contain open ended questions; Mejia-Ramos et al. (2010) also suggested multiple choice comprehension tests to assess undergraduates' proof comprehension, which they suggested as an useful assessment tool especially for large scale studies. Moreover, Houston (1993, p.60) argued that "comprehension tests in mathematics are one way of encouraging students to "read mathematics with understanding"."

2.2.7 Summary of Section 2.2

- Proofs are central in mathematics and they are all over university mathematics. But many undergraduates have serious difficulties with proof related activities, in particular proof construction and proof validation.
- Though undergraduates spend a huge amount of time reading and comprehending proofs, how they engage with the reading and comprehension process has received comparatively little attention from mathematics education researchers.
- Definitions and logical reasoning have a crucial role in learning of mathematical proofs.

- Comprehending proofs demands several levels (or dimensions) of understanding, and comprehension tests might be an effective way to assess undergraduates level of proof comprehension, but it seems that a single one dimensional scale is sufficient to assess how well a reader has comprehended a proof.

2.3 Proof presentation

There are primarily two resources through which undergraduates are introduced to mathematical proofs — in live lectures and through textbooks. In this section, first I discuss Weber's (2004) case study in which he described how a lecturer presented mathematical proofs at the undergraduate level. Subsection 2.3.2 focuses on how proofs are presented at undergraduate textbooks; however to date very little research has been done in this area. Then the focus will be on what are the alternative ways proofs can be presented to students; here I discuss Leron's structural method of presenting proofs and Alcock's computer based e-Proofs.

2.3.1 Presenting proofs in mathematics classrooms

One of the main resources for the introduction of proofs for undergraduates is their mathematics classes, where lecturers present proofs either in a traditional manner with blackboard and chalk or by writing on slides and projecting those slides onto a screen, with or without handing out study materials to the students. Undergraduates spend a significant amount of time watching lecturers presenting theorems and proofs in classrooms (Weber, 2004). How proofs are presented to undergraduates, however, has received little attention in the mathematics education research (Inglis and Mejia-Ramos, 2009).

Weber (2004) investigated how proofs are presented in mathematics classroom practice at the undergraduate level. Weber observed three different teaching approaches offered by the lecturer Dr.T (as Weber called him) when presenting proofs. Dr.T took different teaching approaches depending on what kind of proofs he was presenting to the class. Different proofs have different purposes and therefore presenting them differently was equally important for him, as per Weber's observation. Dr.T intended to draw students' attention to different aspects for different proof presentations.

His first teaching approach was the *logico-structural*, in which Dr.T put an emphasis on logical thinking — he started writing a proof with the assumptions of the theorem and then used definitions logically to derive the conclusion. He demonstrated the process of the proof construction and wanted students to learn from definitions by using simple logical thinking, how the gap between the first line (the given assumptions) and the last line (the conclusion) of the proof can be filled.

The second teaching approach, Dr.T took, according to Weber, was a methodical *procedural* approach when proving a proposition such as "$\lim_{x\to\infty} \frac{n+1}{n} = 1$". His goal was to teach students a particular proof technique which is very important in real analysis. He wanted his students to be able to start writing such a proof without thinking too much and to be able to complete the proof construction by following a systematic procedure.

The third teaching approach was the *semantic* in which Dr.T emphasised 'understanding the key concepts' of the proof rather than concentrating on the proof technique or how to write a proof. This teaching approach appeared at the end of the real analysis course where the concepts become harder. He handed out the written proofs instead of constructing the whole proof on board. He spent more time on explaining key concepts involved in the proofs by presenting diagrams on board. Understanding the concepts was the focus in this teaching approach.

In another study, Weber (2010) interviewed university-mathematicians about how they would present proofs of "$\sqrt{2}$ is irrational" or "a monotone bounded sequence converged". Several strategies emerged: presenting proofs with examples, comparing between examples and proofs, drawing diagrams for explanations, focusing on new proof techniques or on the new ideas involved in the proofs, or reviewing related concepts involved in the proofs before presenting the proofs.

In classrooms undergraduates are usually offered a lot more explanations verbally alongside diagrams, examples etc compared to the study materials they usually get from their teachers. A lecturer can provide much richer and more specific explanations while presenting proofs in classrooms compared to what is available in textbooks. Moreover, students get several suggestions — what to learn from a particular proof, what are the new ideas or how and where to apply proof techniques and what to start a proof when meeting similar sort of statement for proving, how definitions are used in proving etc. Those sort of suggestions are not always explicitly provided in textbook proofs, though textbooks do often provide diagrams, examples etc.

2.3.2 Proof presented in textbooks

During lectures students often exhibit different approaches — they might try noting down as much of what they experience from the lectures or they might write the points that they think are important, or some may write nothing and rely instead on memory or the provided study materials (Abowd et al., 1997). When lectures are over, the presented explanations and information may be lost (Alcock, 2009) and students need to rely either on the study materials provided by their lecturers or on the mathematics textbooks or on the notes they took themselves. This means, that written (textbook) proofs have an important role to play in the learning of proof at the undergraduate level. Surprisingly, then, I have not found any research paper that has specifically discussed the way proofs are presented in mathematics textbooks.

Textbooks are one of the most vital resources for learning at any level. The mathematics textbooks at undergraduate level mostly contain proofs which are "natural language arguments which deductively establish the correctness of theorems ..., although [they are] often shorter and more detailed than the proofs found in mathematics research journals" (Selden and Selden, 1999, p.1). In a similar manner, Rav (1999) noted that the textbook proofs are less rigorous than the proofs from mathematics research papers. But the same proof could vary vastly from textbook to textbook; in terms of what symbols and notations are used or whether the explanations are short or long narrative etc, so it may be difficult to draw general conclusions like this reliably.

Here I consider the two proofs of Rolle's theorem to show that even through two proofs are proving the same theorem their presentations may differ widely. My first example is from the book, 'A first course of mathematical analysis' by J. C. Burkill, in which Rolle's Theorem was presented without any diagram (see Figure 2.4). This book did not offer the geometrical interpretation of Rolle's theorem.

The next example is from the book called 'Mathematical analysis and techniques' (Vol 1) by A. Page (see Figure 2.5). Here six diagrams were used in support of

Theorem 4.52. (*Rolle*). *If*

(1) *f is continuous in the closed interval $a \leqslant x \leqslant b$,*

(2) *f' exists in the open interval $a < x < b$,*

(3) $f(a) = f(b)$,

then there is a value c, with $a < c < b$, for which

$$f'(c) = 0.$$

Proof. Let $M = \sup f(x)$ and $m = \inf f(x)$ for $a \leqslant x \leqslant b$. Let $f(a) = k$.

If $M = m = k$, then $f'(c) = 0$ for every c between a and b.

Suppose that either $M > k$ or $m < k$, say the former. By theorem 3.6, there is a value c, with $a < c < b$, for which $f(c) = M$.

By (2) $f'(c)$ exists. We shall prove that $f'(c) = 0$.

If $f'(c) > 0$, by theorem 4.51, f is strictly increasing at c and there are values of x to the right of c for which $f(x) > f(c)$. This contradicts the fact that $M = f(c)$ is the supremum of f.

Similarly, if $f'(c) < 0$, there would be a value of x to the left of c at which $f(x) > M$.

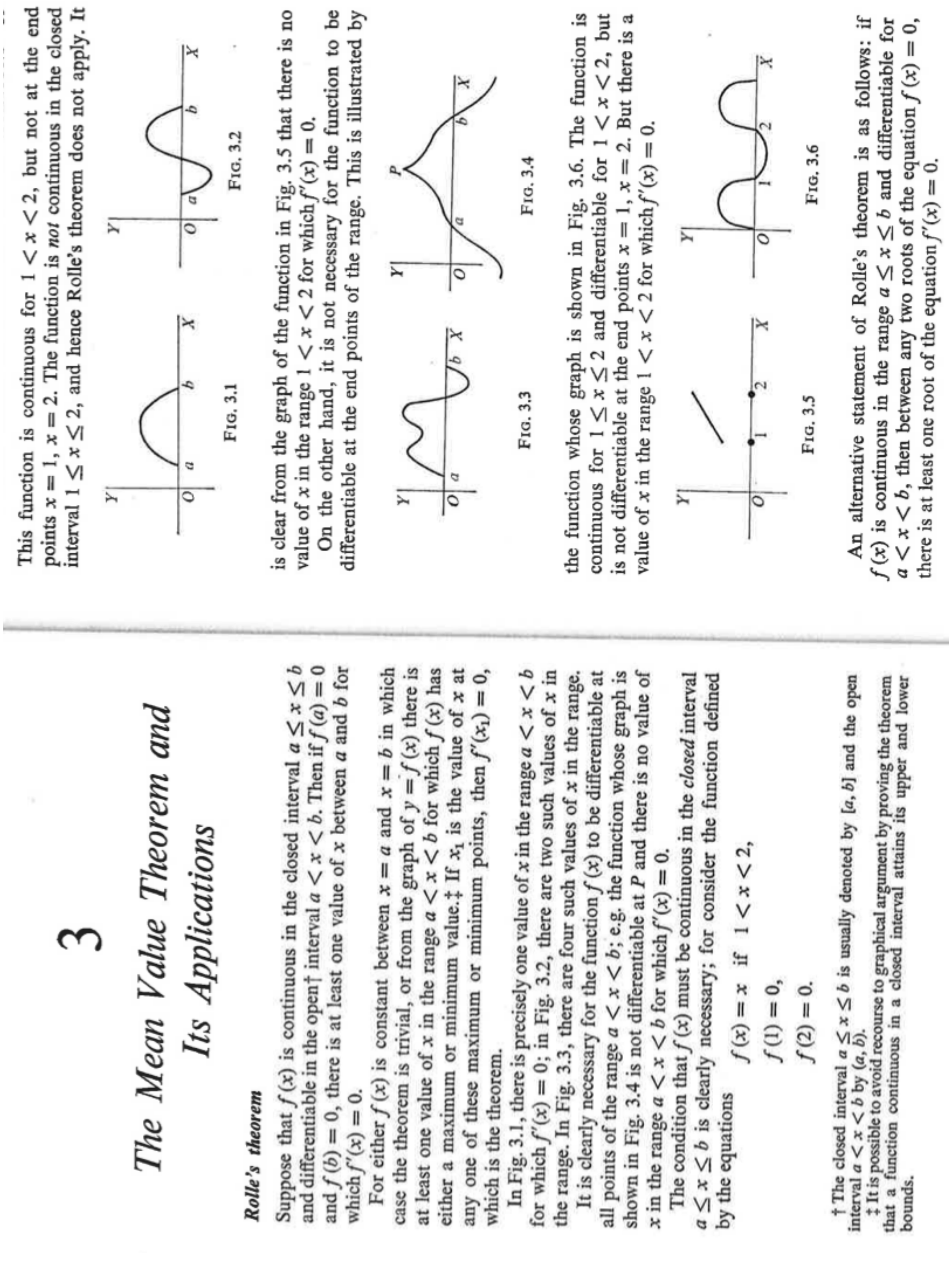
Therefore $f'(c)$ must be equal to 0. ∎

FIGURE 2.4: Burkill (1978, p.73-74) presented Rolle's theorem and its proof in the above manner.

Rolle's theorem and its proof. A detailed discussion of the geometrical interpretation of the theorem was offered in the book — the author explained how continuity and differentiability play a crucial role in the theorem and how it can be interpreted and verified by diagrams. Compared to the first textbook, the second provided a subtantially more detailed explanation.

Often the logical gaps between statements in a proof are left out[5], especially because of the demand of "maximum information in minimum space" (Davis et al., 1995, p.314). Weber and Alcock (2005, p.38) have expressed concern that textbook proofs at the undergraduate level do not always explicitly discuss warrants, "textbooks on logical reasoning introduce the notion of material implication in a highly structured manner (e.g., in terms of rules for inference and truth tables), but usually give no explicit discussion of the need to consider warrants when reading proofs". They also argued that undergraduates do not necessarily have the skill to check the warrants, which may cause lack of understanding (see page 20). This means leaving students with incomplete information is a potential drawback of textbook proofs.

[5]For example, in the second line of the first example of Rolle's theorem (see Figure 2.4), it says "if $M = m = k$, then $f'(c) = 0$ for every c between a and b", therefore it is a reader's job to understand that when $M = m = k$ then $f(x)$ is a constant function and therefore $f'(x) = 0$.



3

The Mean Value Theorem and Its Applications

Rolle's theorem

Suppose that $f(x)$ is continuous in the closed interval $a \leq x \leq b$ and differentiable in the open† interval $a < x < b$. Then if $f(a) = 0$ and $f(b) = 0$, there is at least one value of x between a and b for which $f'(x) = 0$.

For either $f(x)$ is constant between $x = a$ and $x = b$ in which case the theorem is trivial, or from the graph of $y = f(x)$ there is at least one value of x in the range $a < x < b$ for which $f(x)$ has either a maximum or minimum value.‡ If x_1 is the value of x at any one of these maximum or minimum points, then $f'(x_1) = 0$, which is the theorem.

In Fig. 3.1, there is precisely one value of x in the range $a < x < b$ for which $f'(x) = 0$; in Fig. 3.2, there are two such values of x in the range. In Fig. 3.3, there are four such values of x in the range.

It is clearly necessary for the function $f(x)$ to be differentiable at all points of the range $a < x < b$; e.g. the function whose graph is shown in Fig. 3.4 is not differentiable at P and there is no value of x in the range $a < x < b$ for which $f'(x) = 0$.

The condition that $f(x)$ must be continuous in the *closed* interval $a \leq x \leq b$ is clearly necessary; for consider the function defined by the equations

$$f(x) = x \quad \text{if} \quad 1 < x < 2,$$
$$f(1) = 0,$$
$$f(2) = 0.$$

† The closed interval $a \leq x \leq b$ is usually denoted by $[a, b]$ and the open interval $a < x < b$ by (a, b).

‡ It is possible to avoid recourse to graphical argument by proving the theorem that a function continuous in a closed interval attains its upper and lower bounds.

This function is continuous for $1 < x < 2$, but not at the end points $x = 1$, $x = 2$. The function is *not* continuous in the closed interval $1 \leq x \leq 2$, and hence Rolle's theorem does not apply. It

FIG. 3.1 FIG. 3.2

is clear from the graph of the function in Fig. 3.5 that there is no value of x in the range $1 < x < 2$ for which $f'(x) = 0$.

On the other hand, it is not necessary for the function to be differentiable at the end points of the range. This is illustrated by

FIG. 3.3 FIG. 3.4

the function whose graph is shown in Fig. 3.6. The function is continuous for $1 \leq x \leq 2$ and differentiable for $1 < x < 2$, but is not differentiable at the end points $x = 1$, $x = 2$. But there is a value of x in the range $1 < x < 2$ for which $f'(x) = 0$.

FIG. 3.5 FIG. 3.6

An alternative statement of Rolle's theorem is as follows: if $f(x)$ is continuous in the range $a \leq x \leq b$ and differentiable for $a < x < b$, then between any two roots of the equation $f(x) = 0$, there is at least one root of the equation $f'(x) = 0$.

FIGURE 2.5: Page (1947, p.58-59) presented Rolle's theorem and its proof along with the six diagrams.

2.3.3 Leron's structural proofs

Since is it well known that students often struggle to understand proofs, mathematics educators have given different suggestions about what might be beneficial for learning of proofs (for example, Weber and Alcock (2005) have suggested that checking warrants when reading a proof gives better understanding); but not many suggestions have been given on how to present proofs to undergraduates. Leron (1983) proposed a particular way of presenting proofs, which he called the *structural* method. He argued that presenting proofs by the structural method could help students to comprehend better than reading through textbook proofs. Usually in textbooks, proofs are presented in a step-by step manner — from hypothesis to conclusion, which Leron called the *linear* presentation. The idea underlying the structural method is to breakdown a proof into levels, proceeding from top to bottom (see Fig. 2.6), in a hierarchical style. The top level gives the overview of the proof without technical (i.e, notational, computational etc) details. The second level elaborates on the generalities of the top level, supplying proofs for substantiated statements, detail general descriptions, specific constructions for objects and so on. The lower level will provide further details if needed, for example, a subproof of a lemma that can fill a gap in the main proof. Each level contains one major idea of the proof which clearly and explicitly connects to its appropriate place in the total hierarchy. This structural presentation will give more clarity on how segments are linked globally and locally. Leron suggested that breaking down a proof into smaller segments would help students to grasp the idea of each segment with more clarity.

Mejia-Ramos (2011) conducted a study in which 12 undergraduates were asked to read and comprehend two proofs — one linear presentation and one structural presentation. The majority of undergraduates who took part in the study reported being comfortable and confident when they were reading a linear presentation of the proof rather than reading a structural presentation of the proof. Those undergraduates also performed better in open ended comprehension tests when they read the linear version of the proof compared to the structural version. However, it is not wise to generalise the outcomes of this study for several reasons. First, there were only 12 participants which was statistically not enough to generalise the results. Secondly, participants had seen the structural version of the proof first time during the study and under a restricted time, it was not possible for them to understand the purpose of presented proof that particular way. And this could

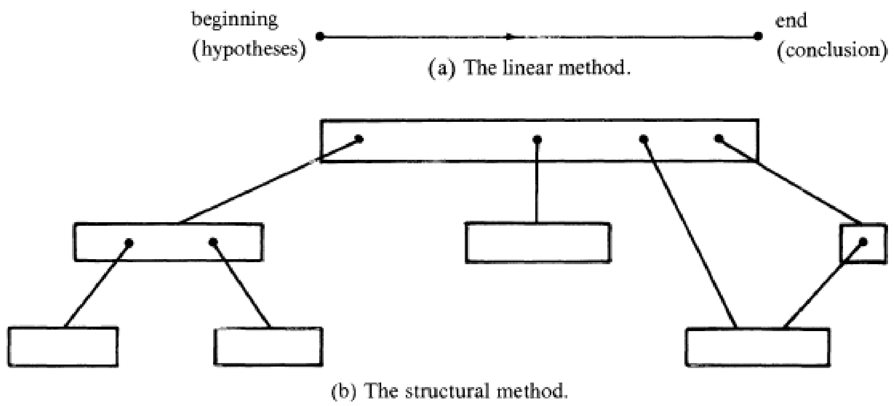


FIGURE 2.6: The linear method and structural method of presenting proofs (Leron, 1983; p.175).

be a reason of that they did not like this new presentation of proofs and did not get benefit from it. This means, so far there is no strong evidence exists that can justify or deny the effectiveness of the structural proof presentation.

Alcock (2009) has also used the the idea of 'breaking down a proof into different segments' when presenting proofs in the format of *e-Proofs* by using an electronic medium. In the next section, I discuss e-Proof in detail.

2.3.4 The e-Proof

Alcock (2009) has presented proofs using an electronic format which she called *e-Proofs*. In e-Proofs, proofs are presented in a multimedia environment using texts (visual medium) and the verbal explanations (audio medium). In the e-Proof, Alcock (Alcock, 2009; Alcock and Inglis, 2010; Alcock and Wilkinson, 2011) breaks down a proof into a sequence of screens. Each screen shows the theorem and the whole proof, but much of this is greyed out to focus attention on particular lines (see Figure 2.7, for an example). In each slide, there is an attached audio file which explains how that particular portion locally and globally works within the proof. This facility provides the extra information that students need for comprehending a proof. Alongside the audio explanations, as the audio proceeds, boxes and arrows appear for highlighting logical relationships and indicating links among the lines.

Figure 2.7 shows the fifth slide of the e-Proof of Rolle's theorem, in which Line 5 and part of Line 1 and Line 3 are highlighted to grab readers attention and to indicate the connection between these parts. The attached audio explains, "in

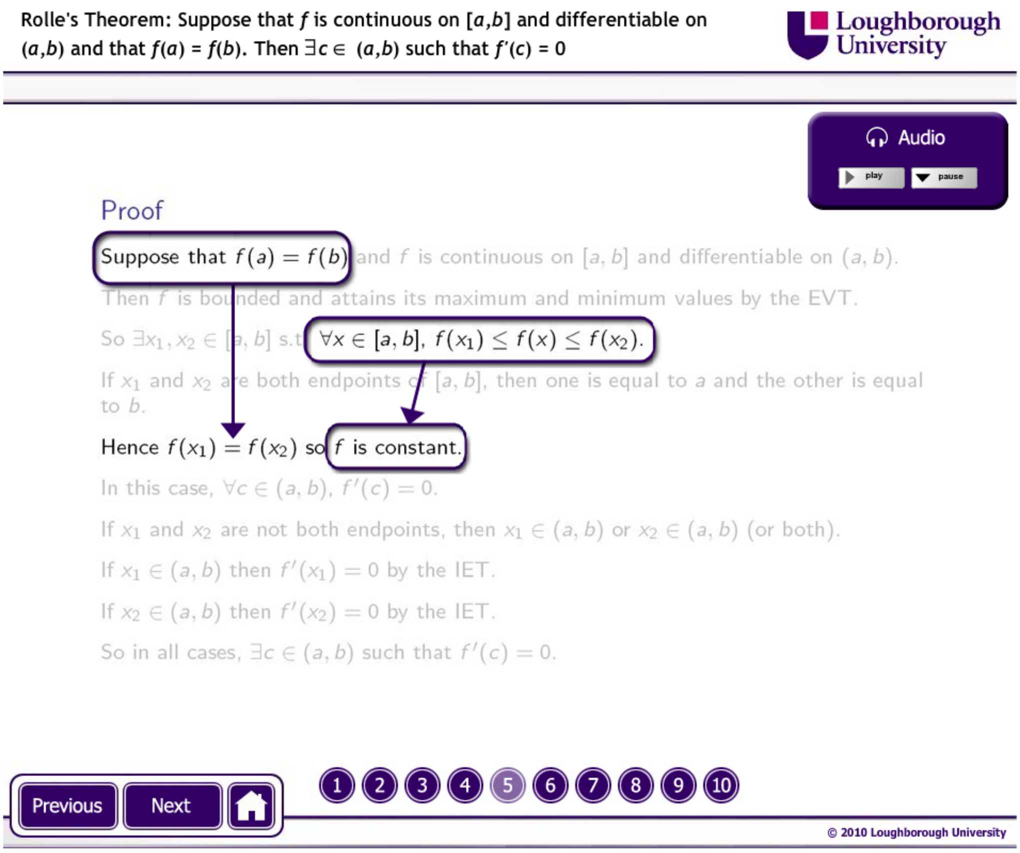


FIGURE 2.7: The screen sort of the fifth slide of the e-Proof of Rolle's Theorem.

the fifth line we deduce this means that $f(x_1)$ is equal to $f(x_2)$ because of earlier assumptions that $f(a)$ equals $f(b)$. We can thus conclude that f is constant because the earlier inequality must be equality in this case" and the arrows and boxes appear as the audio proceeds to indicate more specific links between lines.

We might ask why Alcock wanted to present proofs in such a way for undergraduates. Alcock (2009) argued that even though lecturers explain and illustrate proofs (by explaining warrants for each line, explaining the structure of the proof etc) in detail in live lectures; that might not be enough in terms of helping students to comprehend proofs given in live lectures. Three drawbacks of live lectures that she was concerned about were: (i) undergraduates do not always know where to put attention when reading a proof, (ii) their lack of understanding on logical relationships between lines and (iii) finally, only the written proofs are not helpful enough for undergraduates, they need extra explanations in comprehending proofs.

In the e-Proof presentation format, lines that are linked are highlighted to focus undergraduates' attention on particular parts at a time, which addressed Alcock's first concern. In e-Proofs undergraduates can listen to the extra audio explanations and can view (when playing the audio) arrows and boxes that appear to indicate the logical links between highlighted lines — these therefore address her last two concerns.

Alcock considers e-Proofs as an beneficial resource for proof comprehension because,

- "Attention is directed precisely;
- Logical relationships are explicitly highlighted by coordinating the appearance of annotations with audio commentary;
- Audio can be replayed as many times as the student wishes;
- Visuals are clear and without extraneous distractors;
- Small sections of audio can be scripted and recorded separately for clarity;
- Annotations appear one at a time and do not permanently add content, so the integrity of the proof is preserved without clutter or more to learn;
- Navigation to a specific point of difficulty is straightforward." (Alcock, 2009, p.5)

In her article, Alcock (2009) presented some of the feedback on e-Proofs that she gathered from the undergraduates, who have used e-Proofs both in live lectures and in private study, through questionnaires. The questionnaire survey indicated that undergraduates value having e-Proofs as a resource, which they said help them in several ways: to understand particular proofs, to understand why lines are valid in proofs etc. The survey also reported that undergraduates believed that e-Proofs are easier to comprehend than the written textbook proofs.

So, undergraduates like to use e-Proofs, but that does not necessarily mean that they learn better from e-Proofs. In fact, research indicates that students' interests and satisfaction with multimedia courses are negatively correlated with their achievements at the end of course (Bernard et al., 2004; Clark and Feldon, 2005). One of explanations offered for this finding is that students prefer such recourses based on the expectation that learning from such resources would require less effort. As a consequence, the learning outcomes are poor. In case of e-Proofs, survey data indicates that students believe that learning from an e-Proof is easier than learning from a written proof (Alcock, 2009). It is therefore important to examine the effectiveness of e-Proofs on proof comprehension, especially its effectiveness compared to the usual resources — textbook proofs (written proofs) or proofs presented in live lectures. The primary aim of this thesis is to investigate whether a resource like e-Proof which apparently seems an efficient way of presenting proofs, is beneficial or harmful for proof comprehension.

2.3.5 Summary of Section 2.3

- Proofs are presented in classrooms with elaborate explanations (verbal and written), along with diagrams, examples, depending on the proofs. Lecturers explain more specifically what a proof is about, what to learn from a proof, how definitions are used in proving, what is the technique used in a proof, how to use that technique elsewhere and so on.

- Textbook proofs vary vastly. Written textbook proofs often leave gaps in the argument (for example, often textbook proofs do not explicitly discuss warrants), which is therefore a reader's responsibility to infer the gaps to comprehend proofs.

- Alcock's (2009) computer based e-Proofs are an alternative way of presenting proofs in the multimedia environment. She argued that e-Proofs are better for comprehension than watching proofs presented in live lectures or than written textbook proofs.

- Evidence showed that undergraduates prefer to have e-Proofs as an additional resource. But the effectiveness of e-Proofs on undergraduates' proof comprehension is a different question which needs to be examined. This is the primary issue that this thesis is going to address.

2.4 Learning from texts and multimedia presentations

In Section 2.3, I have discussed different proof presentations — proofs presented at live lectures, written proofs in textbooks and the computer-based e-Proofs. Mathematical understanding is both a linguistic and a conceptual matter (Vergnaud, 1998), therefore it is important to consider the literature that has discussed learning from texts. Moreover, it is also important to consider multimedia learning, as an e-Proof is a multimedia representation in which not only written texts but also audio explanations are involved.

2.4.1 Text comprehension

There have several theories that explain text comprehension, but the theory I discuss here one of the most often cited: the *text comprehension model*. This theory was initially developed by van Dijk and Kinstch and later by Kintsch (Kintsch, 1993; Kintsch, 1998; van Dijk and Kintsch, 1983). This theory described the mental representations generated from a text, from recognising words to constructing a situational representation of that text.

The theory suggests that to comprehend a text, readers need to construct three levels of mental representations of the text — (i) the surface component, (ii) the textbase model and (iii) the situation model.

- *The surface component:* The text comprehension starts with understanding the meaning of the words and phases used in the text. The exact words or phases of the text are recorded in the reader's memory and this is known as the surface component of the mental representation. Kintsch argued that a surface component is always present when a reader is reading a text — "it is generally the case, that at least some of the exact words and phases are remembered at least for a time" (Kintsch, 1998, p.105).

- *The textbase model:* The textbase is the semantic and rhetorical structure of the text, that is, it "consists of those elements and relations that are directly derived from the text itself" (Kintsch, 1998, p.103), but without adding anything that is not explicitly mentioned in the text. A text can be recalled or summarised even if only a text-base is constructed, without understanding what the text is about.

- *The situation model:* The situation model is the mental representation of what a text is about. In order to understand a text, readers elaborate the text by using their prior knowledge and experience to achieve a personal interpretation of the text that is related to other information held in their long-term memory. A situation model is "a construction that integrates the text-base and relevant aspects of the comprehender's knowledge" (Kintsch, 1998, p.107). It contains causal chains of events that unfold the key points of the text; however, the situation model may differ widely among readers and also depending on the purpose of the reading.

To explain how this text comprehension model can fit for the comprehension of mathematical texts, next I will consider the following statement:

If x^2 and x^3 are continuous functions, then $x^2 + x^3$ is also continuous.

At the surface level, a reader needs to comprehend what the meaning of 'continuous function' is and what the meaning of the symbols (x^2 or x^3) in the statements are. For the level of text-base comprehension, a reader should have an understanding that x^2 and x^3 are both continuous function and because both of them are continuous their sum, that is $x^2 + x^3$, is also continuous. But having a text-base level comprehension, does not mean the reader will understand explicitly why the sum of x^2 and x^3 should be a continuous function. This suggests that a reader can easily recall the information contained in the statement with a limited understanding, but that they may not be able to employ the information in some other way. The comprehension of a text will be completed if the reader can construct a situation model when he/she can integrate the relevant prior knowledge to create an overall picture of the statement. Suppose, a reader knows what the diagrams of x^2, x^3 and $x^2 + x^3$ look like and can make a link between the diagrams and continuous functions. Alternatively, perhaps the reader is aware of the fact that all polynomial functions are continuous and makes the connection between the continuous functions and the polynomial functions, though the statement does not mention the term polynomial. These sorts of connections lead towards constructing the situation model and result in more sophisticated and global understanding.

Kintsch (1998) suggested that to learn from a text, the construction of a complete and elaborate situation model is essential. He argued that learning from a text and remembering a text is not equivalent. Remembering a text means readers can reproduce the text in some form and to do so a well constructed text-base level comprehension is needed. However, learning from a text requires understanding of the subject matter to that extent where readers would be able to use the information provided by the text productively in other ways other than in reproducing the text. Moravcsik and Kintsch (1993) reported that their participants remembered a text reasonably well but failed to understand it at the level of the situation model. Conversely, Kintsch (1998) cited the experimental study by Bransford et al (1972), where participants had no memory of the text, neither the surface component nor the text-base, but they "had understood the text very well and were able to form stable situation models on the basis of which they could answer questions correctly and make inferences about the text" (p. 105). Their study also demonstrated that long-term memory of a text is poor if the constructed situation model is incomplete. A situation model is retained in memory much longer than a text-base or a surface level comprehension.

Moreover, for problem solving tasks the situation model plays a significant role: "the basis for problem solving is not the text-base directly, but the model derived from it" (Van Dijk and Kintsch, 1983, p.341). To comprehend a mathematical proof, having both a text-base model and a situation model of that proof is essential, as that will allow a learner to reproduce the proof with an understanding of why and how that proof works. In written proofs (see Section 2.3.2) often some conceptual gaps are left out for readers and therefore it is readers' responsibility to fill the gaps and build up essential situation models to comprehend proofs.

2.4.2 Factors that influence text comprehension

Two factors that directly influence the text comprehension process are: how the text is organised and who the readers are.

Structure of the texts

In the research area of reading comprehension, many studies have been conducted to explore the effect of differently structured texts on readers' comprehensions. A text can be organised as a high-coherence text by adding more information that links most of the ideas explicitly or as a low-coherence text by leaving more coherence gaps in the text. McNamara and Kintsch (1996) noted that

> text coherence refers to the extent to which a reader is able to understand the relations between ideas in a text. This is generally dependent on whether these relations are explicit in the text. Thus, the general

> approach to increasing text coherence is to add surface-level indicators of relations between ideas in the text. Such modifications range from adding low-level information, such as [..], to supplying background information left unstated in the text. (McNamara and Kintsch, 1996, p.248)

To illustrate how mathematical statement can be a high or a low coherence text, I am considering here two statements:

(i) n and $n+1$ are relatively prime.
(ii) Since n and $n+1$ are consecutive integers, they have no factor in common except 1; therefore they are relatively prime.

Compared to the first statement, the second statement is an example of high-coherence text, because the reasons why n and $n+1$ are relatively prime is explicitly explained. In the first statement, however, readers are left with 'n and $n+1$ are relatively prime'. This means, it is the readers who need to make the links that 'n and $n+1$ are relatively prime' because 'n and $n+1$ are consecutive integers, which means they have no factor in common except 1'. Clearly, to comprehend the first statement, the low-coherence text, more inferences need to be made to fill the coherence gaps. On the other hand, the second statement, the high-coherence text, is easier to comprehend as most of the information is provided in the statement.

It seems that it is easier to construct text-base models as well as situation models from high coherence texts. However, research shows that the high-coherence texts are not always beneficial for learning (Goldman, 1997; Kalyuga and Sweller, 2004; McNamara et al., 1996; McNamara and Kintsch, 1996). Comprehending from a high-coherence text makes the comprehension process easier which does not necessarily facilitate learning. Research has also shown that readers performed better in reading comprehension related tasks, such as, making summary of a text, solving problems when they read from a comparatively low-coherence text than a high coherence text (McNamara et al., 1996).

Cognitive load

Before discussing how readers' prior knowledge can influence their comprehension of a text, I want to discuss briefly two types of cognitive load: *intrinsic* and *extraneous*. Different types of cognitive load have different types of influence on learning. Therefore, it is worth discussing them here, as I refer them in the next section many times. The cognitive load theory is developed on the basis of the assumption of human's limited working-memory capacity.

- *Intrinsic cognitive load* is the mental work imposed by the complexity of the content of a study material and it can be high or low depending on the amount of element interactivity in the material needed to be learned (Chandler and Sweller, 1991; Paas et al., 2003; Sweller, 1994; Sweller and Chandler, 1994). Element interactivity means that several knowledge elements need to be coordinated in memory to understand that material. For example, intrinsic load will be high for an undergraduate to learn the concept of limit in real analysis compared to the concept of least common multiple (l.c.m). The intrinsic load of a study material can not be altered directly, but the load can be managed by decomposing complex tasks into subtasks and by activating prior knowledge related to the material.
- *Extraneous cognitive load* can be high or low depending on how study materials are presented or designed. This load is also referred as inefficient cognitive load as high extraneous cognitive load interferes learning. If a text is presented in a way that helps readers to understand the study material, the extraneous cognitive load is low and that is helpful for learning. Research has shown that adding extra explanation does not necessarily reduce extraneous cognitive load that helps readers to learn better (Paas et al., 2004; Sweller, 1994), rather readers' level of expertise plays a important role which I discuss in the next section.

Therefore, while preparing study materials educators ultimately aim to balance the cognitive load so that total load cannot exceed the working memory resources available, if learning is to happen.

Reader's prior knowledge

Another key factor is readers' levels of expertise that can decide what kind of text structure would be beneficial for readers (Goldman, 1997; Kalyuga and Sweller, 2004; McNamara et al., 1996; McNamara and Kintsch, 1996). Readers with low prior knowledge may not get the benefit from low-coherence texts, but may instead do better with high coherence texts. When reading a low-coherence text, those readers might not be able to infer what is required to fill all the conceptual gaps and thus comprehend the text. The high-coherence texts provide extra support that those readers need while constructing the text-base and the situation models to learn from the texts. High prior-knowledge readers, however, may learn better from low-coherence texts: McNamara et al. (1996) argued that while reading a low coherence text, readers need to engage actively to infer the unstated information and that process can make a positive effect on text comprehension.

Some researchers explained why different text structures are effective for different readers in terms of cognitive load theory (Chandler and Sweller, 1991; Kalyuga et al., 2003; Kalyuga and Sweller, 2004; Sweller, 1994). The high prior knowledge readers possess well constructed schemas[6] which reduce the cognitive demands on the limited-capacity working memory. Moreover, well constructed schemas help readers to fill the gaps of a low-coherence text, which may not be an effortless process for those who have poorly structured schemas. The low prior knowledge readers, therefore, need additional support to complete the necessary cognitive processes that they might not able to do themselves. But when nonessential extra information is provided to the high prior knowledge readers, they will end up processing redundant information. This processing (which demands more space in the limited working memory) creates an extraneous cognitive load (Sweller, 1994), interrupting learning. Chandler and Sweller (1991) showed that when students are capable of learning from explanatory diagrams, then adding texts imposes an extraneous cognitive load on them. As a consequence processing those redundant texts not only interrupts learning but also had negative effects. The authors called this a redundancy effect.

Kalyuga et al. (2003) used the term *expertise reversal effect* to explain the ineffectiveness of high coherence texts for high prior-knowledge readers. They suggested that this effect occurs "when an instructional procedure that is relatively effective for novices becomes ineffective for more knowledgeable learners" (Kalyuga and Sweller, 2004, p.558).

[6]A schema is a cognitive structure that helps organise and interpret information (Sweller, 1994). There are two ways through which information are organised and interpreted: schema acquisition and automation. When something new is learned, the process of learning (acquired schemas) is slow and needs considerable thought and effort. A familiar material can be processed automatically without conscious effort, means the automatic processing requires less memory space in the limited working memory.

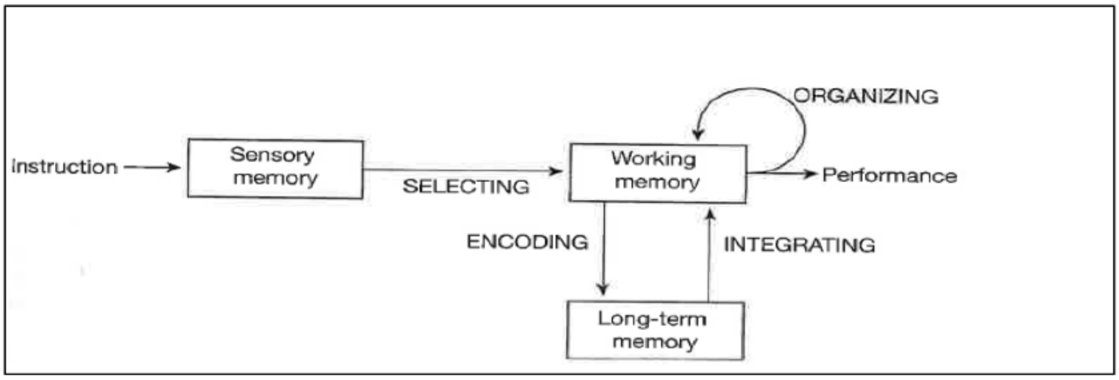


FIGURE 2.8: Three cognitive processes of multimedia learning (Mayer, 2009, p.183)

An expert reader, however, is capable of using different reading strategies to concentrate on the central ideas of a text. Research shows that those readers who are capable of using different strategies when needed understand better than those who cannot do so. Self-explanation is an effective reading strategy that can help learners to develop good understanding of the material study across many domains (Ainsworth and Th Loizou, 2003; Chi et al., 1994). When students spontaneously self-explain, they learn more than those who do not, because a self-explanation strategy encourages readers to infer left out gaps to make sense of a text. This also helps the readers to construct a situation model of that text, which is beneficial for learning. Moreover self-explanations are usually more effective than explanations provided by others.

2.4.3 Multimedia learning

Previously I have discussed different proof presentations and, in Section 2.3.4, I discussed what is involved in an e-Proof. The e-Proof is a multimedia resource. Therefore, while considering learning from e-Proofs, it is important to consider multimedia learning. In this section, I discuss Mayer's (2001) theory of multimedia learning in which the benefits of presenting study materials in the multimedia environment have been discussed. Mayer viewed learning as an active process, in which learners are actively engaged in cognitive processing to construct mental representations of the studied material. He argued that there are three essential cognitive processes that learners go through when learning from multimedia presentations.

The three cognitive processes are — *selecting, organising* and *integrating* (see Fig. 2.8). The first cognitive process is *selecting* in which information (auditory and visual) goes from the sensory memory to the working memory. The second cognitive process is *organising*, in which the information is mentally rearranged into a coherent cognitive structure. The final cognitive process is *integrating*, in which new information is integrated with the relevant knowledge retrieved from the long-term memory to create new knowledge. This new knowledge then enters into the long-term memory where it is stored for long term use.

The explanation Mayer (2001) offered for why multimedia presentations are more effective than presenting material in one medium for learning was based on the assumptions of Paivio's (1986) dual coding theory and Baddeley's (1992) theory on limited working memory, and his proposed active cognitive processing assumptions discussed above. The dual code theory says that visual information (texts, pictures) and verbal information (verbal explanations) are entered into and processed in our cognitive system through two separate channels. The limited working memory theory says that a limited amount of information can be processed in working memory at a time. In a multimedia environment, the visual information, that is, what we see through our eyes goes to the visual sensory memory and the information entering via ears goes to the auditory sensory memory (see Fig. 2.9). In working memory, newly entered information is rearranged into a coherent cognitive structure — an auditory-based model for the auditory information and a visual-based model for the visual information. The newly organised visually-based model and auditory-based model are then integrated with the existing knowledge in the working memory to create new knowledge. The capacity of working memory is limited. Therefore, in a multimedia environment, maximum use is made of both the visual and the auditory channels rather than putting load into one single channel. It is this that is hypothesised to make the multimedia environment more learner friendly.

Mayer suggested that five effects need to be considered while designing instructional materials in a multimedia environment:

- The *multimedia effect:* It is better in terms of learning if study material is presented in two mediums rather than one. In several empirical studies, Mayer and his colleague (Mayer and Anderson, 1991, 1992) showed that students generated significantly more solutions in problem solving tasks, when

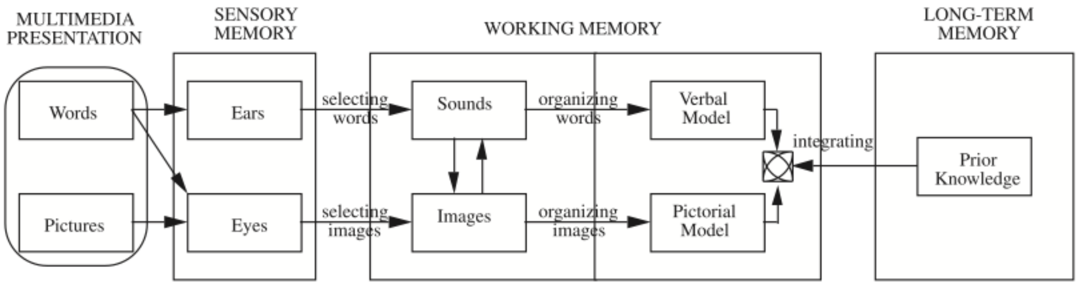


FIGURE 2.9: Cognitive theory of multimedia learning (Mayer,2001; p.44), here words refer towards the verbal objects and pictures as visual objects.

the study material was presented both in the visual medium and the verbal medium, rather than only visual medium. Mayer called it the *multimedia effect*. Clark and his colleagues (Clark, 1994; Clark and Feldon, 2005), however, argued that it is not the multimedia but the instruction methods that can influence learning. That is, if the same instruction is presented in two different mediums the impact of that instruction on learning will be the same. Clark also referred to Bernard et al.'s (2004) meta-analysis report which indicated that there are weak learning advantages or no differences when comparing traditional live and multimedia distance learning. Thus the multimedia effect is controversial.

- The *contiguity effect:* Students learn better when the visual and the verbal information are well synchronised. Mayer and Moreno (1998b) argued that well synchronised visual and verbal information must be in the working memory at the same time in order to facilitate the integration of referential links between them.

- The *split-attention effect:* Students learn better when texts and pictures are presented in two separate mediums. This means that texts (Mayer used 'word' instead of text) should be presented in auditory medium rather than a visual medium. Mayer and Moreno (1998a, p.318) explained that, "when pictures and words are both presented visually (i.e., a split-attention situation), learners are able to select fewer pieces of relevant information because visual working memory is overloaded. When words and pictures are presented in separate modalities, visual working memory can be used to hold representations of pictures and auditory working memory can be used to hold representations of words".

- The *coherence effect:* Students learn better when unnecessary information is not provided in a multimedia presentation. Mayer et al. (2001) reported a study in which those students who were presented with the same information in the auditory medium and in the visual medium performed more poorly in a retention test than those who received no repeated information in the multimedia presentation. A similar sort of effect has also found by text comprehension researchers (discussed in Subsection 2.4.2 which is known as the redundancy effect (Kalyuga et al., 2003; McNamara and Kintsch, 1996). According to cognitive load theory, the unnecessary information increases cognitive load on the working memory and consequence poor learning outcomes.

- *Individual differences:* Mayer (2001) argued that multimedia effects, contiguity effects and split-attention effects depend on individual differences in the learners. Students with limited prior knowledge tend to show stronger multimedia effects than students who have higher levels of prior knowledge. High prior knowledge students are able to generate their own mental images while listening to an audio explanation or reading a text, without having both.

Mayer and his colleagues did not come across of any disadvantage of multimedia presentations but suggested the positive impact of multimedia on learning if the instructions are designed carefully by following the above five principles. The claim Mayer and his colleagues made about the effectiveness of multimedia instructions on learning was based on the immediate post-test results of either a retention test or a transfer test completed by their participants. Therefore, they should claim about the *immediate* impact of multimedia instructions on learning, rather than claiming it in more general sense. Because they did not provide any evidence to claim the long term impact of multimedia instructions on learning compared to a instruction presented in a single medium. Learning brings understanding and knowledge which stay long in our memory, therefore it is important to consider long term effect of a new resources on learners, no matter how it is presented.

Schnotz and his colleagues, however, suggested that multimedia facility can have both a positive and a negative impact on learning (Schnotz, 2002; Schnotz and Rasch, 2005), which is discussed in the next section.

Enabling and facilitating effects in multimedia learning

Schnotz and Rasch's (2005) research outcomes indicated that the impact of the multimedia animation presentations on learning are not always beneficial, but can be harmful. These authors suggested two different effects that a multimedia animation can have on learning: the *enabling* effect and the *facilitating* effect. Multimedia environment with an animation facility provides more information than a static one-mode presentation. This extra information can have a positive enabling effect, if that helps learners to engage with the cognitive processes that those learners may not be able to do without extra help. The authors also noted that students with high prior-knowledge are able to use the extra information as an enabling function, but not students with low prior-knowledge. Conversely, the extra information and animation facility can make the cognitive processes easier. This is called the facilitating effect on learning. However the facilitating effect does not necessarily have a positive impact on learning and may lead to a shallow processing of the presented material, when learners engage with the cognitive processes less deeply than they otherwise would. Not only that, it can lead learners to what Betrancourt (2005) called the 'illusion of understanding'. Primarily the facilitating function seems beneficial for the low prior knowledge students, because they need that extra support offered by the multimedia presentation to perform the necessary cognitive processes; however, the impact of such learning does not last for the long time. For the high prior-knowledge students, this facilitating function has negative impact because that extra support stops those students to engage with the cognitive processes that they are capable of doing by themselves without any extra help.

It seems from the above literature that a multimedia resource needs to be well balanced based on several issues — who the learners are and what their level of knowledge is, how much information to add to facilitate learning, how to use different mediums to minimise unnecessary cognitive load etc.

2.4.4 Summary of section 2.4

- A mental representation of a text is essential for comprehending a text — especially constructing the textbase model and the situation model of the text.

- A high coherence text contains more implicit ideas in a text which may make the text comprehension process easier but this is not necessarily beneficial for learning. It also a reader's level of prior knowledge that plays a crucial role in learning. Reading strategy such as self-explanation has been shown to have a positive impact on learning.

- Study materials presented in a multimedia environment are better for learning than when they are presented in one single medium (either visual or verbal), because visual and verbal information are processed separately in working memory.

- In a multimedia environment, study materials can be presented more dynamic way (for example, animation) which can make the cognitive processes, that are essential for learning, easier. Therefore the mental efforts required to learn such materials are less and as a consequence the learning outcomes are poor.

- Multimedia presentations can be effective for learners, if they are designed appropriately.

- Alcock (2009) e-Proofs have been designed in such way that they appear to meet the principles of multimedia learning.

Chapter 3

The e-Proof

3.1 Introduction

E-proofs, a multimedia representation of proofs, have been used in the real analysis course of Loughborough University since 2009, aiming to improve proof comprehension at the undergraduates level. The aim of my thesis is to evaluate whether e-Proofs fulfil the purpose of improving undergraduates' proof comprehension or not.

To evaluate new interventions in education, methods commonly employed are observations, expert's reviews, students' surveys and other qualitative and descriptive approaches (Dynarski et al., 2007; Yin, 1994). Students' satisfaction is widely used to measure the impact of new interventions. I adopted the common practice of measuring students' satisfaction with e-Proofs to explore "what do undergraduates think is the impact of using e-Proofs on their proof comprehension?". To do so, I used self-reporting methods such as interviews and surveys.

In the Literature Review chapter, I discussed e-Proofs based on the published articles written by Alcock (Alcock, 2009, 2010; Alcock and Inglis, 2010; Alcock and Wilkinson, 2011). Since e-Proofs are a key interest of this thesis, I interviewed Alcock for better understanding of her intentions of creating e-Proofs and her expectations from it. This chapter discusses particularly the following aspects:

1. What are e-Proofs for?
   (a) What motivated Alcock to create e-Proofs?

   (b) What role should e-Proofs play in undergraduate education?

   (c) What are the factors that influenced the design chosen for e-Proofs?

   (d) How does Alcock use e-Proofs in her undergraduate lectures and what do undergraduates' think of it?

   (e) How, in Alcock's view, should undergraduates interact with e-Proofs to get most benefit from them?

   (f) What are Alcock's perceived limitations of e-Proofs?

2. What do undergraduates likes and dislikes about e-Proofs?

3. What do they think of the impact of e-Proofs on their proof comprehension?

To address the above questions, I used focus group interviews to explore what undergraduates think of the impact of using e-Proofs (related to Question 2 and 3), a web-based survey to explore what undergraduates like and dislike about e-Proofs (related to Question 2) and a semi-structured interview with Alcock to understand the purpose of e-Proofs (related to Question 1). I report the outcomes of these three studies separately and at the end of this chapter, I merge them to highlight the overall findings.

3.2 Focus group interviews

To get some insight about what undergraduates think of e-Proofs and why they think so (to explore Question 2 and Question 1.b, Question 1.d) three focus group interviews were planned and conducted. Another aim of this study was to explore the common views of undergraduates on e-Proofs which then would be used for a web-based survey addressing a bigger sample.

3.2.1 Method and methodology

Focus group interview is a very popular self-reporting research method and I employed this method for the following reasons:

- The focus group interview is a convenient way to collect data from several participants simultaneously for exploring their experiences of a particular topic (Kitzinger, 1995).
- It also allows researches to explore not only what participants think about a topic but how they think and why they think that way.
- Moreover, focus group interviews can generate a wider variety of different views on a particular topic than an individual interview can, because focus group discussion encourages individuals to interact and share their spontaneous opinions. Therefore, "the researcher may stand a chance of ending up with more realistic accounts of what people think" (Bryman, 2004, p.348).

3.2.2 Participants

Second or third year mathematics undergraduates (single or joint honours) of Loughborough University who were enrolled for the real analysis course (in the first semester) in the 2009/10 academic year were invited to take part in this study. Eleven undergraduates (8 females and 3 males) took part in the three focus group discussions. The first focus group had five participants (3 females, 2 males), second group had four participants (all females) and for the third focus group interview only two participants turned up on the day (one female and one male). All of them were familiar with e-Proofs as they had been using them in the real analysis module.

3.2.3 Procedure

The focus group interviews were lead by Tony Croft, a mathematician and a mathematics education researcher of Loughborough University. Having Tony as a moderator allowed me to take note and observe, which would have been impossible if I had moderated. Moreover, Tony had several years experience being a moderator in focus group discussions and undergraduates knew him as a Professor in the School of Mathematics, therefore it was believed they would be more likely to open up to share their views during the discussion. Three group interviews were held in the first week of December, 2009. The study took place in a quite room. The discussions lasted 45 minutes on average and were audio recorded.

Tony asked questions to address those issues that I mentioned earlier to him. The focus group interviews were started with a general discussion of participants' experiences about the real analysis module. Then the discussion moved towards more specific issues about e-Proofs — what they think of using e-Proofs during lectures, how they use e-Proofs in their private studies and what they think of the impact of using e-Proofs on learning, what they think of the purposes of creating e-Proofs.

3.2.4 Data analysis

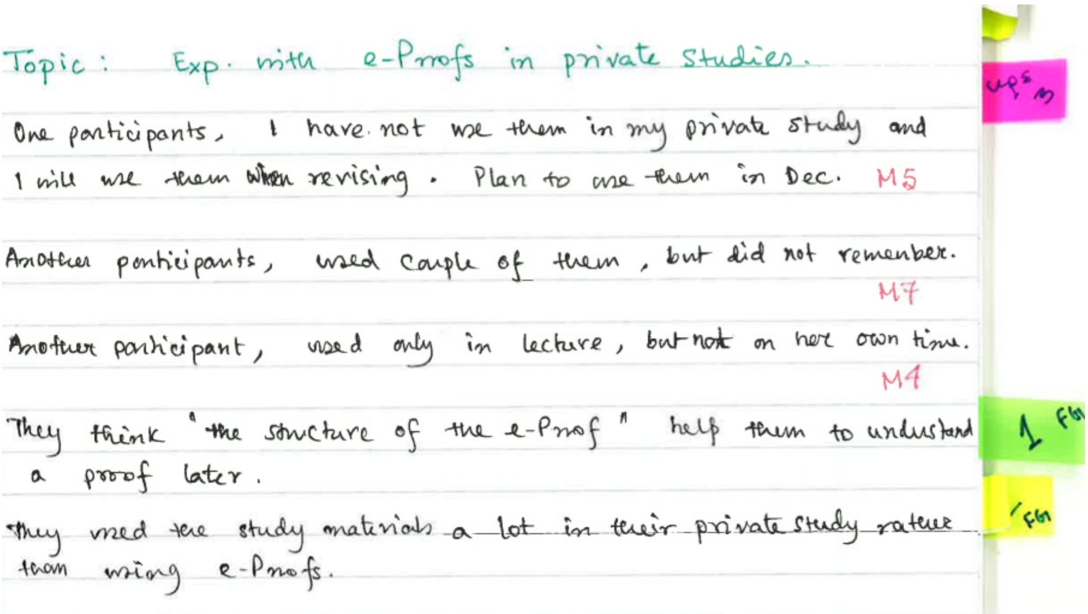
I first listened to the recording several times. I made summary of each topic for all three interviews. Krueger and Casey (2000) suggested that the purpose of a study should drive the data analysis and analysis begins by going back to the intention of the study. This conception helped me managing the data and getting rid of irrelevant information. I only transcribed those parts of each interview that I identified as relevant to fulfil the aim of this chapter. I mainly focused on those sections where participants talked about their experiences, their views, and their expectations from e-Proofs.

I used a thematic analysis approach (in the sense of Bryman, 2004) for analysing the focus group interview data, because I was aware of those issues that I wanted to explore about e-Proofs (as stated at the beginning of this chapter and also those that were not explicitly mention in Alcock's published articles). Cohen et al. (2007, p.468) suggested that using the research questions to lead the analysis "is a very useful way of organising data, as it draws together all the relevant data for the exact issue of concern to the researcher, and preserves the coherence of the material". I kept the themes of the structured interviews intact while analysing my data and presented my results under those themes. I used M1, M2 and M3 as pseudonyms for male participants and for female participants F1, F2, ..., F7.

Figure 3.1 shows a sample of how I summarised a theme of participants' views on using e-Proofs during private study, from the second focus group interview data.

Figure 3.2 shows another example, where participants shared their views about using e-Proofs during live lectures. I used blue and orange colours to highlight opposite opinions of participants.

FIGURE 3.1: A sample of how I summarised a topic (or a theme) from the interview of one of the focus groups.



3.2.5 Results

All participants showed a great deal of enthusiasm for, and interest about e-Proofs and expressed their preference for having e-Proofs in the real analysis module. It seemed that all participants believed that e-Proofs were helpful to them for understanding proofs.

Undergraduates' experiences with the real analysis module

Although the focus of this chapter is on exploring what undergraduates think of particularly about e-Proofs, it is important to briefly discuss what undergraduates think of the real analysis module, because this is the only module in which e-Proofs are used.

When asked about the experience of the real analysis module, all participants agreed that they were enjoying the module more than the other modules. They said that this module was 'very well-structured', 'clear', 'accessible'. To make this module enjoyable, these undergraduates gave a lot of credit to the lecturer and describer her as a 'very good lecturer'.

FIGURE 3.2: Another example of summarising a theme from the second focus group interview data.

Facilitator: Experience of using e-Proofs in lectures 10min

UG1 (F1): I think it just another way of presenting proof, sometimes we just get proofs are written down and we just copy them and then with the e-Proof you get explained step-by-step and it focuses on specific, the proofs highlight bits and bobs and yah Lara is always talking on the top of it as well as on the audio explaining it even more.

UG2 (M1): Ya, I agree with that, that breaking it down (proof) let you focus on each bit individually not 15 lines of this proof which is like daunting introductory then the proofs and the conclusion bit it much more easier to understand.

UG3 (F2): It is good to breaking the lecture size, you are not doing a same style of things in every a single lectures.

Question: What is the different about watching Lara in real life and watching her in e-Proof in the same room.

UG4 (F3): .....e-Proofs you are seeing written down and highlighted and also hearing her voice saying on the top of it as well, it is just kind of like presenting it in different way.

UG5 (M2): I am not sure about the recording. It seems more interactive if she is actually sparking herself rather than playing the recording.

UG4 (F3): I think I agree with that. I like the set out of the e-Proofs, but I prefer if Lara is actually talking herself and explaining it, each step herself. E—Proofs good for revision purposes.
I personally less concentration when it (audio recording) being played than if Lara is talking herself and doing her hand gestures.

> M1: She seems enthusiastic, some lecturers just stand and talk and you get bit bored, but Lara is enthusiastic and keeps you occupied and keeps you engaged with the whole lecture.

Furthermore, participants tended to agree that Alcock's lecturing style made the module easier, understandable and enjoyable.

> M2: Very accessible, means told you a lot about like why exactly, what everything means, exactly why it is there.

> F2: She puts a lot more into than some lecturers give us, when she like revises the topics, like goes back over and over how individual sections fit together.

Overall, all eleven participants confirmed that this was a 'good' module and they liked it more than other modules — they liked their lecturer and her lecturing style as well as the structure of the module.

Undergraduates' views on using e-Proofs in live lectures

When asked to share their experiences of using e-Proofs in live lectures, the participants expressed different opinions and they also had different reasons to like or not to like e-Proofs in live lectures. Out of 11, four participants expressed negative views regarding the use of e-Proofs in live lectures. They said they would rather prefer to listen to the explanations from lecturer in a live lecture than watching e-Proofs.

Those participants who liked having e-Proofs in lectures had different reasons for liking it. F2 liked the use of e-Proofs in a lecture because for her that brings variation in a lecture :

> It is good to break the lecture size, you are not doing the same style of things in an every single lecture.

In contrast, F1 preferred the presentation of the e-Proof compared to the usual way of presenting proof in which the lecturer writes proof either on the board or on a projector and students need to copy it down:

> I think it just another way of presenting the proof, sometimes we just get a proof is written down and we just copy them and then with the e-Proof you get it explained step-by-step and it focuses on specifics.

For the same reason M1 also preferred e-Proofs in live lectures:

> Breaking it [proof] down lets you focus on each bit individually, not 15 lines of this proof which is like a bit daunting.

All participants liked the idea of "breaking down a proof into sections", used in the e-Proofs and all agreed that particular set-up helped them to 'understand proof', and that made proofs 'clearer' and 'less scary'.

But participants shared different views on use of the audio explanation in lectures. Some participants expressed their concentration were poor when the audio was on.

> F3: I like the set out of the e-Proofs, but I prefer if Lara [Alcock] is actually talking herself and explaining it, each step herself.

And her concentration was much better when “Lara is talking herself and doing her hand gestures”.

Two participants preferred the interaction between the lecturer and students, for them e-Proofs did not have that interactive effect.

> F4: When the recorded audio is on, the interaction is lost.

> M2: I am not sure about the recording, it seems more interactive if she is actually speaking herself rather than playing the recording.

Three participants said it would be enough if the lecturer introduced the e-Proofs once, and then students can use them in their own time.

> M3: I do not think its a good use of lecture time, when you just listen to the audio because like I said, you can do that in your own at home.

Sometimes, as participants knew that e-Proofs were available to use any time, some of them “switched off when e-Proofs are on”.

Using e-Proofs in private studies

Though undergraduates showed enthusiasm and positive views about e-Proofs, only two participants (M3 and F8) were using the e-Proofs on a regular basis in their private studies. Three participants said they used one or two e-Proofs occasionally (but failed to recall which e-Proofs), whereas the rest of them did not use them in their own time at all. But they all said that they would be using them in the revision period before the final exam and in the Christmas break. The majority of participants (all except M3) confirmed that they preferred to use the study materials in private studies compared to e-Proofs.

M3 said a larger part of his private study consisted of learning from e-Proofs. He usually did not use notes when he was using e-Proofs. When asked how he used e-Proofs, he said:

> I have my pen and paper, I use the audio and I press play and then I listen through each one two or three times, then I write down a few bits said and then in my own words write down what is happening and I ended up writing down what is going on and keep writing ... eventually I am just basically writing up the proof, but I understand it.

Where M3 mentioned he liked to spend hours on e-Proofs, F2 expressed a different way to use e-Proofs. She said, if she has 5-10 minutes in hand, she would prefer to watch e-Proofs to kill the time.

F8 used most of the e-Proofs available but for a sort period of time (ten minutes). She said she used e-Proofs for a change because they are different from the rest of her study materials. She, however, expressed negative views about the audio explanations and preferred to watch an e-Proof without audio. She said she used the audio only the first time but not after that and mentioned that the audio explanation "does not work for me at all".

Referring to the audio explanations, F1 said that she usually went through the proof herself and then listened to the audio explanations selectively and made notes if necessary, because for her the e-Proofs involved explanations that are very basic which she avoided.

> generally speaking, like, I find with e-Proofs, there are always bits [audio explanations] which are quite trivial.

In summary, most of the participants did not use e-Proofs in their private studies on a regular basis but believed that e-Proofs would be a very useful resource for revision purposes.

Impact of e-Proofs on learning

When asked if there is any impact of e-Proofs on their learning, all participants claimed that e-Proofs definitely have a positive impact on their understanding of proofs. M1 said that especially for longer proofs, he got frustrated, but in case of an e-Proof by "breaking it down makes it simpler and more understandable" and gave "a new way of learning proofs and theorems". For F2 an e-Proof is "more accessible and [is] not such a big scary long thing any more."

All participants seemed more confident about the real analysis module than any other modules. They also claimed that they understand proofs in the real analysis module better than any other modules. Most of them claimed to be more confident about proofs than before and more confident about applying the 'breaking down' strategy to understand a proof.

There was also an emotional impact of e-Proofs on those participants. The majority of them mentioned that "having e-Proofs is good as an extra resource". They felt emotionally secured by thinking that they would get help from e-Proofs if they struggled to understand proofs during the revision period, when there is no other help available.

Purposes of e-Proofs

When asked what they think of the purpose of creating e-Proofs, two purposes came out — promoting understanding and resource for revision and all participants seemed to agree on these. "To encourage us to understand" proofs was the key purpose for these participants. They also believed that e-Proofs make understanding and memorising proofs easier:

> F4: To help us understand long proofs, so you can actually remember and understand meaning of what happening behind them.

"E-Proofs are for revision purposes" — this came out quite strongly from the very beginning of each focus group discussion. These participants believed that they might need help before the exam and e-Proofs will be the essential resource for that time. F6 said that "a lot of people will use them for revision" and the rest of the participants shared the same view.

> F3: You are home doing your own work and you are stuck on something, it is quite hard sometimes to figure out without someone else's input, if you got the e-Proof there, it may be taught to you.

In summary, all participants agreed that the purposes of creating e-Proofs were to aid the understanding of proofs and to provide an essential resource for revision.

3.2.6 Summary

Overall, all participants were extremely positive about e-Proofs, and they even expressed a lot of interest and enthusiasm about the real analysis module. When asked about the impact of having e-Proofs on their confidence level, the majority of them mentioned that they felt confident about the proofs in the real analysis module than any other module. Interestingly, all participants claimed to be confident about the module than any other modules they were doing. But the credit behind their confidence level, does not go to the e-Proofs only. During the discussions several factors appeared to have a strong impact on their confidence level [1]:

- The lecturer herself had a strong impact on those participants. Her enthusiasm and lecturing style came out repeatedly in the discussions and reflected as a key factor that makes this particular module interesting and enjoyable.
- For participants the design of the module was well-structured and clearer than any other module. This structure also helped them to keep track of the topics covered in the module.
- The regular class tests was another factor that participants mentioned frequently as a positive aspect of the module. Those tests demand regular studying and practicing which made those participants confident about the module.
- Tutorials were also very much appreciated by the participants and they mentioned that attending lecturers and tutorials made the difficult module enjoyable.

[1] Since detail discussion on these factors is beyond the interest of this chapter, I have not included any quotes here. However the quotes used in the subsection 3.2.5 demonstrated the influence of the first two factors on participants.

3.2.7 Conclusion

The undergraduates who took part in the focus group discussions agreed that e-Proofs were useful for them. All of them saw the great potential of e-Proofs as a resource for the revision period. Even though all participants mentioned e-Proofs as a very useful and helpful resource, in reality most of them did not use e-Proofs during their private studies. Besides that all of them claimed that e-Proofs had a strong impact on their understanding of proofs, as well as they claimed to be more confident about Proofs especially in the real analysis module.

Since all participants expressed their interest and enthusiasm about e-Proofs, that certainly reflects their feeling towards e-Proofs but this does not imply that e-Proofs were educationally effective. This question arises, especially because most participants admitted that they did not use e-Proofs regularly in their own time.

3.3 A web-based survey

Based on the students' feedback on e-Proofs from the focus group interviews, a web-based survey was conducted to explore more general views on e-Proofs. The aim of this web-survey was to get a more general representative understanding of how much undergraduates liked or disliked e-Proofs compared to the small scale focus groups.

3.3.1 Method and methodology

The web-based survey has the potential to reach greater numbers of participants with minimum organisational constrains, costs and time (Bryman, 2004; Cohen et al., 2007). In the questionnaire, undergraduates were asked to give their responses to various statements using Likert scales, which provides a range of responses to a given question from strongly disagree to strongly agree (Cohen et al., 2007). This questionnaire allows to explore respondents' views about a given topic.

3.3.2 Participants

All second or third year mathematics undergraduates (single or joint honours) of Loughborough University who were pursuing the real analysis module in 2009-2010, were asked through emails to complete the survey on e-Proofs. The survey was uploaded in the Bristol Online Survey website (http://www.survey.bris.ac.uk/). Thirty eight undergraduates took part in the survey during the month of May.

3.3.3 Materials

In the *web-based survey*, there were 28 statements in the questionnaire (see Figure 3.3). In the survey participants were asked to respond to the statements using a 5-point Likert scale. This scale had the following choices: strongly disagree, agree, neutral, disagree and strongly disagree. There were eight statements (1, 4, 5, 7, 8, 17, 22 and 23) on e-Proofs that were set to explore how much undergraduates like e-Proofs. Out of these eight statements, four (1, 4, 5 and 23) were reverse statements. For example, the statement "showing e-Proofs in lectures was a waste of time" is reversed in the meaning from the overall direction of the survey, that is, how much undergraduates like e-Proofs.

The statements on e-Proofs were designed based on the undergraduates' feedback — the questionnaire survey, which was published in Alcock's (2009) paper and the focus group interviews discussed earlier. However, the design of this web-based survey was more balanced than the previous questionnaire. In the previous questionnaire, only positive statements about e-Proofs were asked; whereas this web-based survey had both negative and positive statements on e-Proofs.

Two statements, 15 and 20, were addressed as neutral, as they did not reflect the issue of learning from e-Proofs and therefore, excluded from the data analysis. The other 18 statements (2, 3, 6, 9, 10, 11, 12, 13, 12, 15, 18, 19, 21, 24, 25, 26, 27,28) were taken from an assessment scale known as "the need for cognition" to measure "the tendency for an individual to engage in and enjoy thinking" (Cacioppo and Petty, 1982, p.116). These statements were set to investigate if there is a correlation between how much undergraduates like e-Proof and how likely they engage with cognitively demanding tasks. Out of the 18 statements on the need for cognition, nine were reverse scored statements (2, 3, 9, 10, 11, 13, 19, 21, 28).

Figure 3.3: The questions were set in the questionnaire for undergraduate to answer in the web-based survey.

1. I would not use e-Proofs if these were provided for other courses.
2. I like tasks that require little thought once I've learned them.
3. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.
4. e-Proofs made studying proofs less daunting.
5. Showing e-Proofs in lectures was a waste of time.
6. I would prefer complex to simple problems.
7. I understand more by studying a proof on paper for myself than I do when using an e-Proof."
8. e-Proofs helped me learn what kinds of structure to look for when studying new proofs.
9. I only think as hard as I have to.
10. Thinking is not my idea of fun.
11. I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.
12. I prefer my life to be filled with puzzles that I must solve.
13. I prefer to think about small, daily projects to long-term ones.
14. I usually end up deliberating about issues even when they do not affect me personally.
15. I used e-Proofs to make extra notes about proofs in Analysis.
16. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.
17. e-Proofs helped me understand where the different parts of a proof come from and how they fit together.
18. I really enjoy a task that involves coming up with new solutions to problems.
19. Learning new ways to think doesn't excite me very much.
20. Resources for studying proofs should force me to come up with my own explanations rather than providing them.
21. It's enough for me that something gets the job done; I don't care how or why it works.
22. e-Proofs made my learning more effective overall.
23. e-Proofs were not useful to me as a revision resource.
24. I like to have the responsibility of handling a situation that requires a lot of thinking.
25. I find satisfaction in deliberating hard and for long hours.
26. The idea of relying on thought to make my way to the top appeals to me.
27. The notion of thinking abstractly is appealing to me.
28. I feel relief rather than satisfaction after completing a task that required a lot of mental effort.

3.3.4 Data analysis

Participants' answers for each statement were scored in the Likert scale and the final score for each participants was a tally of the individual's point from]each statement. An individual scored four if he/she strongly agreed with a positive statement on e-Proofs or on the need for cognition scale. For example, if a participant strongly agreed with the statement "e-Proofs made studying proofs less daunting", he/she will be given 4 points since this statement is scored positively. But, if he/she strongly disagreed with the statement "showing e-Proofs in lectures was a waste of time", they would get 4 points because this statement is scored negatively. For the analysis, all eight statements on the e-Proofs scale were scored together as a group and all 18 statements on the need for cognition scale were considered together as a group. Using the 5-point scale (0 to 4), the highest possible score on e-Proofs scale was 32 (eight statements multiplied by 4 points each) and the hight score on the need for cognition scale was 72 (18 statements multiplied by four points each).

A higher score in e-Proof scale indicates that participants like to have e-Proofs. A high score in the need for cognition scale indicates participants like to engage in thinking about tasks, enjoy the thinking process and motivated to apply their thinking skills with little prompting (Cacioppo and Petty, 1982).

3.3.5 Results

The mean score on the e-Proof scale was 25.53 out of 32. This indicates that those undergraduates who took part in the survey had extremely favourable views on e-Proofs. A one-sample t-test showed that the mean score was significantly higher than 24 (75%), $t(37) = 2.871, p = 0.007$.

There was no significant correlation between the of e-Proof scores and the Need for Cognition scores, $r = 0.112, p = 0.503$. This indicates that how much undergraduates like to have (use) e-Proofs and how much they enjoy doing cognitively demanding tasks were not related, suggesting that e-Proofs are not viewed favourable by only those who enjoy difficult cognitive tasks.

There was a significant correlation between students' exam marks on the analysis module and the Need for Cognition score, $p = 0.05$, which indicates that if undergraduates prefer to engage with cognitively demanding tasks they are more likely to perform better in exam. However, no significant correlation was found between the exam marks of the analysis module and how much students liked the e-Proofs.

3.3.6 Conclusion

The web-based survey was conducted to investigate undergraduates' views about e-Proofs, and to determine if there is a correlation between how much they like e-Proofs and how likely they are to engage with cognitively demanding tasks. Thirty-eight undergraduates, who used the e-Proofs when pursued the real analysis module, took part in the survey. The outcomes of the survey indicate that undergraduates had very favourable views on e-Proofs (even significantly more than 75%). And that their views about e-Proofs was not correlated with the need for cognition scale.

3.4 An interview with Alcock

I discussed Alcock's views on e-Proofs based on her published articles (Alcock, 2009, 2010; Alcock and Inglis, 2010; Alcock and Wilkinson, 2011) in the Literature Review Chapter. Since e-Proofs are of key interest in this thesis, I wanted further investigation to understand Alcock's rationale for creating e-Proofs, what motivated her to chose the design of e-Proofs and what were her views on how undergraduates should be using e-Proofs.

3.4.1 Method and methodology

The intention of conducting the interview with Alcock was very specific, and for that reason I planned a semi-structured interviewing approach. This approach suits my purpose, as Bryman (2001, p.323) stated when "the research is beginning the investigation with a fairly close focus, rather than a very general notion of wanting to do research on a topic, it is likely that interviews will be semi-structured ones, so that the more specific issues can be addressed". The interview was held in an informal way involving a "two-person conversation initiated by the interviewers for the specific purpose of obtaining research-relevant information and focused by him [her] on content specified by research objectives of systematic description, prediction, or explanation" (Cannell and Kahn (1968) as cited in Cohen et al. (2007, p.351)).

The topics I covered in the interview were: how Alcock came up with the idea of e-Proofs, what roles e-Proofs can/should play in undergraduate mathematics, what factors influenced the design for the e-Proofs, and how undergraduates should use e-Proofs to get the most benefit from them. The interview lasted for 42 minutes and was audio recorded. Then the audio recording was transcribed for the data analysis.

3.4.2 Data analysis

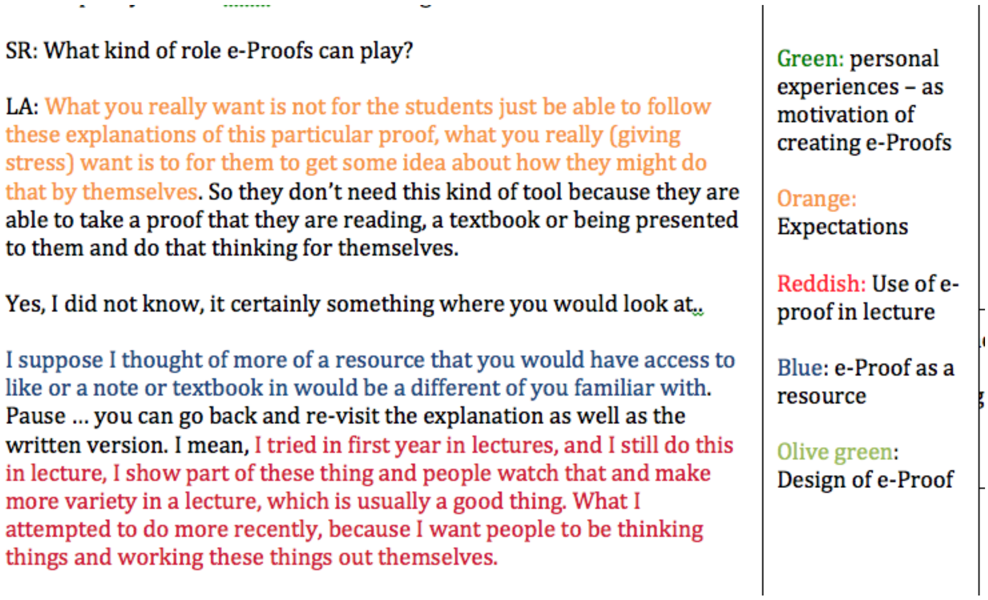
To analyse the interview data I took a similar approach to that which I took for analysing the focus group interviews: a thematic analysis approach. Due to the structured approach to the interviews, I had a sense what type of themes and sub-themes would be generated for each set of questions. While transcribing the interview as I became familiar with the data, those themes and sub-themes became more prominent.

The research questions (stated in Section 3.1) were used as main themes, such as the motivation for creating e-Proofs. It allows me to highlight those themes and sub-themes that reveal the relevant insight needed for this particular chapter.

Initially, I identified seven themes: the motivation for creating e-Proofs, the role of e-Proof at the undergraduates level, the use of e-Proofs in live lectures, the design of e-Proofs, how to read an e-Proof to get most benefit out of it, and limitations of e-Proofs. Initially I highlighted the transcribed data in various colours for identifying different themes. Once I identified those themes, I looked carefully for sub-themes within each theme, which I did manually by making notes. Because it was single interview, I found it easy to do the process of identifying similar or different patterns manually within a theme to create sub-themes.

Figure 3.4 shows an example of how I highlighted different themes by using different colours.

FIGURE 3.4: A sample of how in the transcribed interview data different themes were highlighted in different colours.



SR: What kind of role e-Proofs can play?

LA: What you really want is not for the students just be able to follow these explanations of this particular proof, what you really (giving stress) want is to for them to get some idea about how they might do that by themselves. So they don't need this kind of tool because they are able to take a proof that they are reading, a textbook or being presented to them and do that thinking for themselves.

Yes, I did not know, it certainly something where you would look at..

I suppose I thought of more of a resource that you would have access to like or a note or textbook in would be a different of you familiar with. Pause ... you can go back and re-visit the explanation as well as the written version. I mean, I tried in first year in lectures, and I still do this in lecture, I show part of these thing and people watch that and make more variety in a lecture, which is usually a good thing. What I attempted to do more recently, because I want people to be thinking things and working these things out themselves.

Green: personal experiences – as motivation of creating e-Proofs

Orange: Expectations

Reddish: Use of e-proof in lecture

Blue: e-Proof as a resource

Olive green: Design of e-Proof

3.4.3 Results

Motivation for creating e-Proofs

- *Undergraduates often struggle to understand proofs.* Alcock was very much aware that most undergraduates have serious difficulties in learning mathematical proofs saying "that's not controversial information at all", both from her teaching experience and her background as a mathematics education researcher. She "thought about those things before — what you would be doing if you try to understand a proof" and this motivated her to create the e-Proofs which she believed could guide undergraduates how to read proofs for understanding.

- At the undergraduate level, the activities students mostly engage with are reading and understanding proofs. She mentioned that, even though the ultimate goal of teaching proofs at undergraduate level is to teach students how to construct proofs to achieve that goal the first step to learn is knowing how to understand a proof through reading.

  > What I really want is that students learn to construct proofs, however, the activity they are engaging more, certainly with something like analysis, is trying to make sense of the proofs that they are given and that was what I was focusing on.

Also being a mathematic education researcher she was concerned that little research has been done on proof comprehension, whereas plenty of research has already been conducted in the area of proof construction. This also motivated Alcock to keep her focus on the area of proof comprehension.

- It is important for undergraduates to understand what they should be thinking while reading a proof. To make sense of a mathematical proof and understanding why and how a proof works, lots of thinking is required when reading. Alcock suspected that undergraduates are often not aware of that process and they have hardly any experience of "thinking about things", especially the way mathematicians' think while reading proofs. Therefore, by presenting the e-Proofs, Alcock's aim was to demonstrate "the process an expert might go through when reading a proof" to the undergraduates.

The role of e-Proofs at the undergraduate level

- *The e-Proof is an explanatory tool.* Alcock described e-Proofs as an explanatory tool where the attached audio provides extra explanations about how lines fit together in a logical level, reminds if definitions, lemma or other theorems are used in a proof and explains how and why etc. She believed that the explanations a lecturer gives in a live lecture by discussing each line of a proof is "normally lost" when the lecture is over and "so essentially they [undergraduates] end up with a textbook proof" which is not enough help for undergraduates to understand a proof.

- *The e-Proof is a tool that can guide undergraduates how to read a proof.* Alcock argued that e-Proofs can guide undergraduates how to develop the skill of reading mathematical proofs by breaking down a proof, finding logical links between lines. Her aim was "to give them sort of a model, the kind of thing you would look for [...] when you can't read a proof."[2]. She also expected that when undergraduates become aware of the thinking process involved in reading proof, they might not need any further guidance from e-Proofs.

- *Capture an explanation of the process an expert might go through when reading a proof.* Alcock said that demonstrating the way mathematicians read

[2]Obviously, by "can't read a proof" Alcock meant not knowing how to read proofs for comprehension

a proof to undergraduates could help students to understand what they should be thinking and where they should be looking at when reading a proof. Her assumption was that displaying mathematicians' ways of thinking is important in terms of learning effectively how to read proofs. Note that this suggestion relies upon the assumption that the author of e-Proof knows how mathematicians read proofs, and that there is no large between-mathematician differences in this respect.

- *The e-Proof demonstrates that a proof is a combination of a few ideas.* From her teaching experiences, Alcock experienced that undergraduates often have a tendency to consider a proof as difficult and scary based on how many lines are in a proof, not based on the content of the proof. Through e-Proofs, Alcock wanted to draw undergraduates' attention into the main ideas involved in a proof rather than considering each line as a separate idea:

  > This [initial part of a proof] is all set up, this is only introducing this idea, and then working with it in a obvious way, this is putting two things together to give you the conclusion.

- *An useful resource for the revision period before the exam.* Alcock said that undergraduates "like to have extra resources", and that they consider e-Proofs to be especially helpful in the revision period before the exam.

- *The e-Proof is a confidence boost-up resource and an emotional support for undergraduates.* In Alcock's view, one of the reasons undergraduates "cite them [e-Proofs] as a very positive thing" is because e-Proofs give them "the feeling of security" as they know:

  > there is an extra explanation available to me and I can go and get it when I am ready.

  This is especially the case during revision when no help is available from the lecturer.

Using e-Proofs in live lectures

- *Use of e-Proofs brings variation in a lecture.* Alcock has been using e-Proofs in her lectures of a course in real analysis for the last four years. In the real analysis course, there are plenty of proofs that she presented to the undergraduates. She believes that having variation in lecture is a good thing which help students to avoid boredom and engage more actively in the classroom. Using e-Proofs can help to bring that variation in her teaching.

- *Encouraging undergraduates to think about proofs.* Alcock tried to use e-Proofs in a way where undergraduates are encouraged to read though a proof before getting any help from e-Proofs or from the lecturer. She suggested that she had often handed out a paper copy of an e-Proof[3] to the class, allowing undergraduates a few minutes to read and discuss it, then run through the e-Proof or the parts of e-Proof which she thought of most relevant to discuss. She spent "a bit longer [time] than I was doing initially" when she used to play an e-Proof (or part of it) and used to add extra explanation on the top of that.

- *Undergraduates have different views about using e-Proofs in lectures.* As Alcock said, some students like having e-Proofs in a live lecture, some students said they tend to drift off if they are watching the whole e-Proof in the class. Being aware of those different views, Alcock mentioned that she tries to use e-Proofs in a way to satisfy undergraduates more.

Design of the e-Proof

- *An important feature of an e-Proof is breaking down a proof into slides.* Alcock wanted to record the process an individual goes through when reading each line of a proof:

  > rather than just leaving them [undergraduates] with the finish product, and I suppose thats what I wanted, [...], I wanted some sort of way of recording [each step].

  That is why in an e-Proof format, each slide highlights a particular portion of the proof and an accompanying audio explains how that portion links with other parts of the proof. Boxes and arrows appear in addition to draw reader's attention — showing the process of what a reader should be thinking when reading a particular line at a time, before considering all lines together.

[3]A paper copy of e-Proof allows undergraduates to read only the textual content of an e-Proof.

- *The audio explanation is another key feature of the e-Proof.* When asked how she decided how much explanations to provide in e-Proofs, Alcock admitted that she did not follow any systematic process, rather her intuition plays a vital role for designing any individual e-Proof. She thought about providing explanation in a way that she should do when a student is struggling to comprehend the proof.

  > [What] I tried to do was just, is pretty much what I would do very intuitively if I sat with a students who is struggling, who is just completely lost with some particular proof, what I would do, well I will point out some links, [...], explaining those links and lines in a way if I do it in lecture; so I did not have a system, perhaps I should have a system.

  Therefore, the audio explanations are intended to provide most of the essential information that a weaker student needs while reading and understanding a proof.

- *Alcock said that she used her own reading strategies as a model for the design of e-Proofs.* She said, "I just did it in a way that what kind of thing do I see that I am looking for" in a proof. And what she usually thinks when reading a proof is:

  > here is a big formula that appears in the beginning and you tend to think what did that come from? Or why would one introduce that? And I would be thinking why would they want that? And I would be tracking down and thinking ooh I see when you get to here this is what happens, so sort of links forward one and anticipatory like that, [...], others are more backward explanatory links where you know you just say well this came from the fact that we assume this is up here look.

  In e-Proofs, such links (anticipatory and backward) are highlighted by using boxes or arrows along with the audio explanations.

- *The design of e-Proofs is more flexible than written proofs.* From her teaching experiences, Alcock knew that undergraduates are often scared of lengthy proofs. The problem she identified for the written proof is:

> If you want to add more explanations you kind of have to do that with narrative, which potentially if you do that within the proof makes it very long. If you do it outside the proof means you sort of stole the information and then bringing it to mind again when you looking at the particular line of the proof.

The design of the e-Proofs gives the opportunity to provide information either in written words or in the audio format. This flexibility also allows the length of a proof to be kept shorter by providing some information in the audio format. Moreover, without compromising the length of a proof, an e-Proof offers other facilities too:

> the extra things [boxes and arrows] appearing and disappearing [in an e-Proof] would not interfere with the written thing, you could make it bit shorter, you could help people see what the links are without having to be extremely explicit.

- *The e-Proof is presented in a multimedia environment.* Alcock used two media — visual medium (written texts) and verbal medium (audio explanation) to present each proof, she gave little attention to what research has shown is effective in multimedia environments. Instead, she focused on presenting and explaining proofs in a way that undergraduates (even the weaker students) can comprehend as well as develop skills that is needed in comprehending proofs.

- *Alcock made various decisions about the layout of e-Proofs.* She wanted to fit proofs into a single slide which she believed would be helpful for undergraduates. Also she said that with the proofs in real analysis, she "could reasonably" fit a proof into a slide with some adjustments.

  > Some of the information can go in the audio, so if you do need a bit of space you can take out perhaps small explanatory things and then re-add that in your audio.

  She said she made many decisions about spacing and about layout based on her intuition, rather than following a systematic procedure and her intuition played an important role in the creation of e-Proofs.

How to read an e-Proof to get most benefit out of it

- *In Alcock's view, e-Proofs need to be used intelligently.* Alcock expressed her concern that e-Proofs would not be helpful for passive learners.

  > I do say, use it intelligently, there is no point just sitting there and watching through something, well what you want to be doing reading it and trying to understand it yourselves.

- *Undergraduates should be independent learners.* When asked what she would suggest undergraduates do when using e-Proofs to get most benefit from them, Alcock said:

  > if you have a good go with something [e-Proofs] first, you might be able to then use it more intelligently because you would be thinking, you would have some sense of . . . what going on here, do not know why this is done this way and then that might help you to do that to fill in the gap yourselves.

- *Undergraduates should use the additional explanations when needed.* Alcock believed that undergraduates should be familiar with the text of the proof as well as they should be aware of on which parts they need help and then that would be the right time for them to get extra help from the audio explanations. Therefore, listening to the audio explanation would be more effective when "you are familiar enough with the text [because] I suppose you can digest the explanation quite quickly".

Limitation of the e-Proof

Considering the limitations of e-Proofs, Alcock admitted that she did not pay explicit attention to the theorem statement. She was aware that undergraduates often "are not very alert to if-then statements" which is one common structure of theorem statements. Alcock said,

> very recently I understood how little students pay attention towards theorems, before they start looking at proof

and

> I guess it become very clear now and I will do [put explicit emphasis on theorem statements] that more now if I am doing [creating e-Proofs] again.

Some contradictory views

- Alcock said that in an e-Proof, explanation for each line is given in a way that if a student who is struggling to understand a proof would get benefit from its e-Proof. In her view, those undergraduates who struggle to understand proofs appreciate e-Proofs the most.

  > People [who] appreciate it most are the ones who are not really ready and who are using it as a crutch — 'something is there for me' without having to think it through for themselves so much.

  It seems from the above quote that these undergraduates might not be able to put effort in understanding proofs and could be passive learners. According to Alcock, however, passively watching an e-Proof would not be useful for learning proofs. This indicated that those (weaker) undergraduates who are appreciating and using e-Proofs most, may not be getting the benefit that Alcock thought of while creating the e-Proofs.

- Alcock believed that undergraduates who are already good and understand proofs well probably get most out of an e-Proof by seeing it. In her view, those undergraduates become more aware of that where they should be looking for when reading a proof. However the question is if they are already good enough in comprehending proofs, they might know what is needed and where to look while reading proofs. The additional explanations (which Alcock explained from a very basic level of and especially for those students who are not capable of understanding a proof without help), therefore, are redundant for them and might not have positive impact on their learning.

3.5 General Discussion

In this chapter, I presents the focus group interviews and the web-based survey to explore undergraduates' views on e-Proofs and a semi-structured interview with Alcock to get insight into her viewpoint.

Overall findings of the focus group interviews and the web-based survey suggested that undergraduates liked e-Proofs a lot and believed that they are helpful and useful in learning of mathematical proofs. Undergraduates also claimed that e-Proofs 'definitely' had impact on their learning and that experiencing e-Proofs made them more confident about proofs. However, there are some noticeable contradictions regarding what undergraduates claimed about e-Proofs and how they actually used e-Proofs.

- All the undergraduates who took part in the focus group interviews expressed a lot of enthusiasm and willingness to have a resource like e-Proofs (even in some other modules), however, most of them (9 out of 11) admitted that they did not regularly use e-Proofs in private study. Real Analysis was a module of 11-weeks and the focus group interviews took place during the tenth week. Those undergraduates had almost completed the module when they were invited for the focus group discussion and admitted that the real analysis module means 'lots of proof'. But for these ten weeks, most of them did not use e-Proofs at all in their own time. It seems that either they did not give enough importance to e-Proofs for comprehending proofs or they had sufficient knowledge of reading and comprehending proofs themselves without getting any help of e-Proofs.
- Participants in the survey claimed that e-Proofs 'definitely' had a positive impact on their learning and made them more confident about proofs. Only two out of 11 undergraduates used them regularly in their private studies and other two claimed to use e-Proofs once or twice in their own time but failed to recall which. The rest had only seen e-Proofs in lectures and did not use them in their own time at all. Moreover, three of those undergraduates who did not use e-Proofs at all, mentioned that they did not like the use of e-proofs in lectures and they "switched off when e-Proofs are on" for various reasons. Considering that, it seems that their claim about e-Proofs make them confident and had positive impact on their learning could be an 'illusion' rather than a realistic claim.

The web-survey on e-Proof suggested that undergraduates preferred to have e-Proofs. The extent to which undergraduates liked e-Proofs was not correlated with their final exam marks on that particular module. This result does not support the claim about the positive impact of e-Proofs on learning. It is important to mention that there are other examples of self-reporting surveys, which reported that students' preferences with multimedia courses are negatively correlated with their achievements at the end of courses (Bernard et al., 2004; Clark and Feldon, 2005). Though I did not find such negative correlation, the importance of a further investigation on the impact of e-Proofs on proof comprehension becomes clearer.

The interview with Alcock explored the motivations behind the creation of e-Proofs — reading and comprehending proofs is the activity undergraduates mostly engage with at the university level but often their lack of knowledge in this task causes substantial problem in the learning of proofs. The e-Proof provides, in Alcock's view, a 'model' to guide undergraduates what experts do when reading each line in a proof. It seems that she believed that there is no large difference how mathematicians read proofs. However, recent studies have shown regarding reading proofs there is no uniform approach among mathematicians (Inglis and Alcock, 2012; Weber and Mejia-Ramos, 2011).

In Alcock's view the weaker undergraduates appreciate e-Proofs the 'most'. But that does not necessarily mean they learn 'most' from e-Proofs (which I have already discussed in Page 75). It can be argued that high achieving undergraduates might not need the additional help offered in e-Proofs; rather excess information could cause the 'redundancy effect' (in the sense of Kalyuga et al. (2003) and McNamara et al. (1996), discussed in Chapter 2) and poor learning outcomes.

The e-Proof is a multimedia representation of proofs, depending on the students' level of expertise multimedia representation could be both useful and harmful (discussed in Chapter 2). For example, 'trivial' explanations provided in an e-Proof, might have a negative effect on a high-achieving undergraduate who is capable of generating such explanations himself/herself. However, when the audio explanations refer to a definition or a lemma that is not explicitly written in the e-Proofs, it could be useful for those undergraduates who might not be able to recall the definition or the lemma themselves. Considering e-Proofs from the multimedia learning prospective was essential, though Alcock admitted to put little attention on these issues.

Finally, this chapter shows that that self-reported studies can open up views about e-Proofs, pupil like and dislike, but cannot provide evidence whether e-Proofs are educationally effective or not. It seems that undergraduates are not reliable to judge the impact of e-Proofs on learning. Therefore self-reporting methods, survey and interviews, were not appropriate and further research was required to ensure impact of e-Proof as a pedagogical tool.

3.6 Summary of the chapter

- It seems that undergraduates were enthusiastic about e-Proofs, however most undergraduates did not use them in their private studies at all. They also claimed that e-Proofs had strong impact on their learning of proofs, which is quite an unrealistic claim, especially when most undergraduates watched them only once during live lectures. Therefore, the impact of e-Proofs on undergraduates proof comprehension was not answered so far and this needs investigation.
- Alcock created e-Proofs to guide undergraduates in the task of reading and comprehending proofs. She added explanatory information for each line in an e-Proof in a way that even weaker undergraduates should get benefit from it. In Alcock's view, brighter undergraduates might 'get the most out of' an e-Proof, but it can be argued that if they already capable of understanding a proof themselves, they might not need a resource like the e-Proof. Moreover, they might find that the additional information provided in an e-Proof are redundant, which may harm their learning.
- On the other hand, according to Alcock, undergraduates who struggle to understand proofs, appreciate e-Proofs the most. But how much those weaker undergraduates learn from e-Proofs is a different question which needs further investigation.

In the next chapter, I will discuss the experimental methodology which was used to investigate the impact of reading an e-Proof compared to reading a written textbook proof on undergraduates' proof comprehension and to compare undergraduates' reading comprehension process of these two proof presentations.

Chapter 4

Experimental methodology

4.1 Introduction

In this chapter, I describe the experimental methodology used in the remainder of this thesis. At the beginning, in Section 4.2, I state the research questions that this thesis wanted to explore. Section 4.3 describes what the experimental research approach can explore, including various experimental designs used in research. Section 4.4 explains what experimental designs were used in this thesis and why those were suitable to address the research questions set for the thesis. Validity and reliability are essential criteria of research and Section 4.5 discusses how these issues are addressed to obtain maximum validity and reliability in my studies. In Section 4.6, I discuss the ethical considerations that I addressed while planning and conducting the research. At the end, I highlight the key points of this chapter in the form of a short summary.

4.2 Research questions

This thesis aimed to evaluate the effectiveness of e-Proofs on undergraduates' proof comprehension. In Chapter 3, I reported the purpose of creating e-Proofs and what undergraduates think of the impact of using e-Proofs on their proof comprehension, by using self-reported research methods. Findings showed that undergraduates believed that 'e-Proof' is a very useful resource that 'definitely' has the positive impact in learning proofs. However when it was about using the e-Proof in the regular basis, most undergraduates preferred to use the class notes than e-Proofs. Their claim seemed unrealistic and no evidence was found to indicate educational impact of e-Proofs. Therefore, the aim was to directly compare the impact of reading e-Proofs with reading written textbook proofs. For that, I considered experimental method for studying the research question *whether the impact (immediate and long term) of reading an e-Proof is better compared to reading a written textbook proof on undergraduates' proof comprehension.*

4.3 Experimental methodology

The main reason for considering an experimental research approach for my study was that it allows causal inference(Christensen, 2003). More precisely, it allows me to find out whether undergraduates perform better *because* they read an e-Proof or a textbook proof. In an experiment, a researcher measures the effects of manipulating one variable on another variable (Bryman, 2004; Campbell et al., 1963; Christensen, 2003). The variables that are manipulated by researchers are referred to as independent variables and those that are measured are dependent variables. That means, an independent variable is the variable hypothesised to be one of the causes of a presumed effect. In a study, there could be more than two independent variables as different conditions or treatments and researchers measure the impact of those different conditions on participants.

In my experiment, the aim was to investigate undergraduates' proof comprehension under two conditions — when they read an e-Proof and when they read a textbook proof. And under these two conditions, I measured their proof comprehension through a written comprehension test. Before discussing the particular research design that I used for my study, in the next section I describe between-subject and within-subject experimental design and their advantages and disadvantages.

4.3.1 Between-subjects and within-subjects research design

I will discuss two main experimental research designs — the *between-subjects* research design and the *within-subjects* research design.

A *between-subjects* research design is one in which independent groups of participants receive one of the conditions. This research design allows a researcher to compare the effect of different conditions on two or more groups. However, there is always a possibility that the participants in different groups are different enough to influence the effects of the conditions. There are mainly two techniques, *matching* and *randomisation*, that researchers use to guarantee as few differences as possible exist among participants in various conditions groups.

Matching is one technique in which participants' characteristics are matched in various conditions and those participants who had equal or very similar characteristics could be assigned as a group. One problem with this technique is that researchers cannot match every characteristic because it is not always possible to know all the characteristics that need be matched.

Randomisation is a more common technique used by researchers to ensure the formation of equivalent groups of participants. Randomisation ignores the characteristics of the participants and each participant has an equal and unbiased opportunity to be in any of the condition groups. Thus the groups are probabilistically equivalent.

In a *within-subjects* research design the same group of participants experiences all the conditions set for a study. There are two advantages of this research design. Firstly, it is comparatively easy to increase the number of participants in a within-subject design compared to a between-subjects design, because in between-subjects all participants need to divide into several groups. Since the numbers of participants increase in within-subjects design, that increases the statistical power of the experiment. Secondly, because participants are the same in the different conditions, individual difference factors do not effect the dependent variable.

However, the disadvantage of the within-subjects design is the *carryover effect*, which means one condition may effect participants' performances in other conditions. Two basic types of carryover effects are *practice* and *fatigue*. When participants are asked to do similar tasks in each condition, then if participants' performances are found to be better on a later condition, this is referred to as a practice effect, but not because of different conditions. If the performances in later conditions are worse, then that could be because of tiredness and that negative effects is referred to as a fatigue effect.

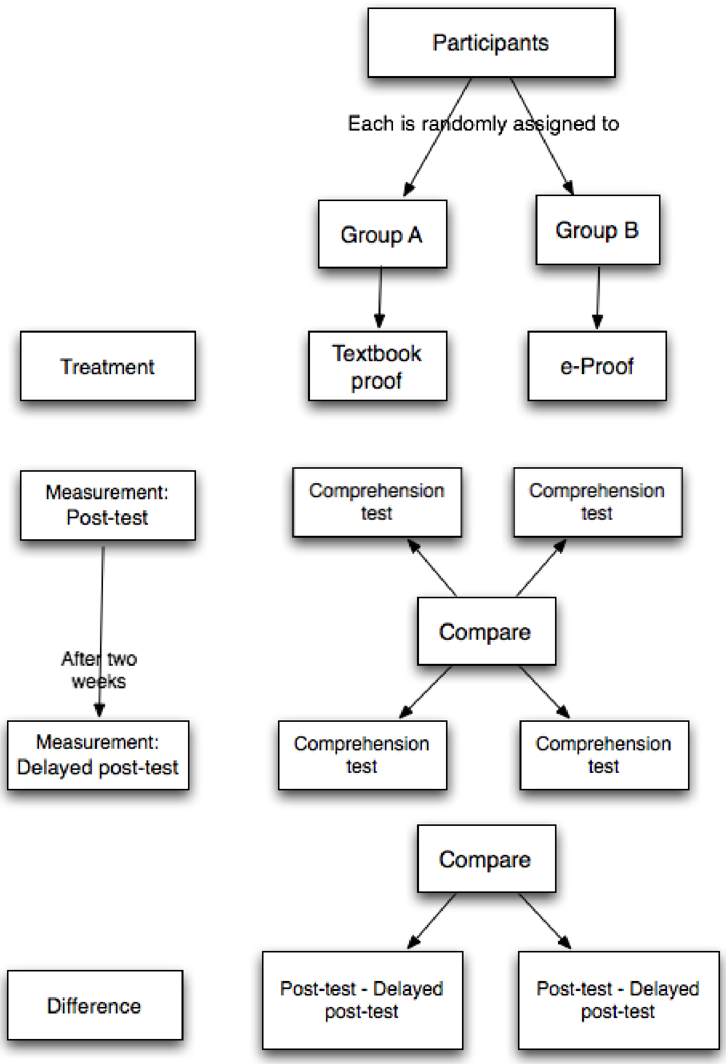
4.3.2 Some experimental research designs

Here I describe some of the experimental research designs commonly used in many research disciplines; psychology, medicine for example.

- In the *one group research design*, one group of participants receive a treatment condition and researchers conduct a pretest and a post-test before and after the group receives the treatment. The difference between the pretest and the post-test result is taken as an indication of the effectiveness of the given treatment condition.
- The *post-test only research design* contains the basic components of most of the experimental research plan in which the dependent variable is measured only once and after different condition groups receive different treatments.
- In the *pre-post test research design* participants are randomly assigned to groups and pretested. Then each group receives different treatments and after which they are tested again. And the difference between pre and post test for those groups are statistically analysed to measure the effect of different treatment conditions.
- In the *posttest-delayed posttest research design*, participants are randomly assigned to different treatment groups and after receiving different treatments they are tested twice — one immediately after receiving the treatment and one after a certain time interval, one or two weeks later for example. The difference between the post and the delayed post-test are used for assessing the long term impact of different treatments on those groups.
- The *factorial design* is one in which researchers observe the effect of two or more independent variables on the dependent variable.

Depending on the research questions each of these research designs has its own advantages (Christensen, 2003). For example, one group research design does not allow researchers to compare the effect of a treatment with a control group in which participants receive either no treatment or a standard treatment. In the pre and post-test design, by conducting a pretest researchers can then directly observe the changes in participants' behaviours as a result of treatment effect, which the post-only design might not be able to indicate clearly.

FIGURE 4.1: An overview of the experimental design of the study which was aimed to compare the effectiveness of two different proof presentations on undergraduates' proof comprehension.



4.4 Experimental design used in the thesis

In my first experiment, I investigated if undergraduates perform better in proof comprehension tasks when they read an e-Proof compared to when they read a written textbook proof. I was also interested to compare the long term reading impact of these two proof presentations. Based on my research interest, a between-subjects research design was adopted. Undergraduates were randomly assigned (to produce probabilistic group equivalence) to two groups who received either a textbook version or an e-Proof version of a proof for comprehension. Their proof comprehension was then tested twice — immediately after reading the proof and two weeks later. The post-test and in the delayed post-test were used to measure the effects (immediate and long term) of these proof presentations on undergraduates' proof. This experimental design allows me to say the performances (better, worse or same) are caused by the proof presented to the groups. Figure 4.1 gives an overview of the design used for the first study and more detailed discussion on the design is given in Chapter 5.

My second study aimed to investigate how undergraduates' read and comprehend proofs, that is, the processes involved in the activity of reading proofs. For this study a within-subjects research design was chosen in which each participant (from two groups selected randomly) read both versions of proofs. Both groups read the same proofs but different versions (e-Proof and textbook proof). For example, if the first group read the textbook version of Theorem 1, then the second group read the e-Proof version of Theorem 1. This research design allow me to observe the similarities and differences that exist when undergraduates read two different proof presentations. More detail description of the research method for the second study is given in Chapter 7.

4.5 Validity and Reliability

It has been advised that a good experimental research must obtain some criteria to make the research valid and reliable. First, I discuss what makes a research valid and then Subsection 4.5.2 addresses issues that makes a research study reliable.

4.5.1 Validity

The essential criteria for experimental research are having *internal validity*, *external validity* and *construct validity* (Christensen, 2003; Elmes et al., 2011; Robson, 2002).

Firstly, the *internal validity* is an essential criteria for the experimental research, in which researchers need to make sure that the manipulations actually *cause* the outcomes of a study. That is, if a study can plausibly demonstrate that the observed changes or differences in a dependent variable is because of different treatment conditions, it is referred to as having internal validity. In my study, there could be a possibility that the two groups of participants perform differently in both tests not because they are asked to read two different versions of proof, but they have very different characteristics as individual groups. To remove this individual characteristics effect, participants were randomly selected for both groups to produce probabilistic group equivalence. Thus any observed differences would be due either to the manipulation or random chance. And the probability of random chance causing the observed differences can be calculated with hypothesis tests.

The second criteria is *external validity*, also known as generalisability — the extend to which findings can be generalised from the specific sample in a study to to some target population. Assessment of external validity is essential because in research "the long-range goal is to understand the basic underlying laws of behaviour"(Christensen, 2003, p.399), beyond the immediate purpose of establishing relationships between the independent variables and the dependent variables. For example, participants of my studies were undergraduates of Loughborough University, pursuing either a single or joint honours including mathematics. The outcomes of this thesis might not be able to be extended to the students of primary schools. They should not be limited to the undergraduates of Loughborough University, but also to such sample groups who have similar educational background and educational exposure.

The third criteria is *construct validity* — the degree to which the independent and dependent variables accurately measure that they are intended to measure. In my first study, a comprehension test (containing nine questions) was used to measure undergraduates' proof comprehension. This test was designed by using the reading comprehension model of proofs proposed by Yang and Lin (2008), in which they argued that the understanding of five facets are essential for comprehending a proof as a whole. The comprehension test was designed in a way that can reflect the understanding of the different facets of proof comprehension and presumably how well undergraduates comprehend a proof based on their written answers.

In the comprehension test, what question appears first can influence performances on later questions and to avoid any such possibility that can reduce construct validity, a randomisation technique was used. That means, all the nine questions used in the test was randomised between participants, so the probability of appearing any of the questions at the beginning of the test was equal and unbiased. Moreover, to obtain construct validity, the time for the reading a proof and answering the test was fixed.

4.5.2 Reliability

Reliability is another essential criteria that researchers must consider before presenting the outcomes of a study to the world. A research should be conducted by maintaining a consistency thoughout while conducting a study. For example, if participants in a study receive different instructions before doing the same task, that could influence their performances. Researchers should clearly state the method and procedure of a study, so that others can reproduce the study to confirm whether the outcomes are similar to the original study. If repeating an experiment by following the similar procedure can produce similar outcomes to a previous study, then that would refer as that the study has high reliability (Bryman, 2004; Christensen, 2003; Robson, 2002). In this thesis, the method and the procedure of each experiment was explained elaborately and as clearly as possible.

Since in my first study I wanted to measure undergraduates' comprehension of proof by using a written comprehension test, therefore it was crucial to maintain a consistency while marking their answers. To maintain that consistency, marks for each question were set before starting the marking process. What kind of answers would get what kind of marks was also decided beforehand. Detailed discussion about the marking process of the comprehension test is given in Chapter 5.

4.6 Ethics

Mathematics education research involves human participants and therefore considering ethical issues regarding human values are essential. Research institutes such as universities have their own committees who provide guidelines about ethical issues that researchers need to consider while planning research. When planning a research, Bryman (2004) has suggested four ethical issues to consider — whether there is harm to participants, whether there is lack of informed consent, whether there is an invasion privacy and whether deception is involved. Since in my study, the issue of invasion privacy was not relevant to address, here I discuss the other three issues.

- *Avoiding harm to participants*: Research should be conducted in a way so that participants do not suffer from any negative effect, such as, stress, depression etc (Bryman, 2004). In my study, I was investigating whether reading an e-Proof has a better effect on undergraduates' proof comprehension in term of answering the comprehension test compared to reading a written textbook proof. However, my study was not aimed to observe an individual's understanding of a given proof and did not comment on performances of an individual. Moreover, to avoid the issue of harm by comparing undergraduates' performances, identities of individuals were kept confidential.

- *Informed consent*: Informed consent means participants should be given all the relevant information about the study that they would take part. Bryman (2004) quoted Social Research Association (SRA), ethical guide line which suggested that researchers must not deliberately hide any information that are likely to affect participant's willingness to take part in the study. Researchers have the responsibility to make the participants aware about their 'right to withdraw at any time'. That is, participants have the right to withdraw from a study at any stage, even without giving any reason and to withdraw data even after finishing the study. Undergraduates of my study were informed about the study how the study would be conducted by their lecturer of the real analysis module in advance. They were also informed that it was not compulsory for them to take part in the study. But they were encouraged to take part as that would help me as a researcher to explore learning of mathematical proofs at the university level.

- *Deception*: Deception occurs when participants are informed something that is different from the actual study. For example, if participants are informed about the study and what it is intended to find before conducting a study, but the actual study intended to find something different then that would not be ethically accepted. There is sometimes a tendency to hide information to limit participants' understanding of what the research is intended to find so that participants can respond more naturally to the experimental treatment. However, from an ethical perspective, researchers must explain a study before asking participants to take part, but they can keep it minimal to avoid any kind of influence on participants. For example, the hypothesis was set for my first experiment was — the group of participants read an e-Proof version of a proof would perform better in the comprehension test compared to the group who read a written textbook proof. But exposing that hypothesis to the participants might cause a negative stereotype threat for the participants of the textbook group. Many studies (Aronson et al., 1999; Steele et al., 2002) have shown participants' performances vary significantly under a stereotype thread condition, such as women are poor in mathematics compared to men. By informing participants that the aim of my study to investigate undergraduates' comprehension of mathematical proof when a proof is presented differently, I provided them sufficient information without increasing any stereotype threat. After the study, participants were given all materials that were used in the study.

4.7 Summary of the chapter

- To investigate the effectiveness of an e-Proof and a textbook proof on undergraduates' proof comprehension and to explore the reading comprehension process of these two proof presentations, I adopted the experimental research approach, because this research approach allows a researcher to measure the effects of manipulating one variable on another variable.

- To find out effectiveness of reading an e-Proof and reading a textbook proof, a between-subjects design was used in which two groups of undergraduates read one of the versions of the proof. Participants for each group were selected randomly to established probabilistic equivalence. To measure undergraduates' proof comprehension, a written comprehension test was used twice — immediately after reading the proof (post-test) and two weeks later (delayed post-test).

  To explore how undergraduates comprehend e-Proofs and textbook proofs, a within-subjects research design was used in which each participants read both e-Proofs and textbook proofs.

- To maximise the validity of my research, I considered issues concerned with internal validity, external validity and construct validity. For example, for internal validity, it was essential to assure that the outcomes of my first study caused because of two different treatment conditions — reading an e-Proof and reading a textbook proof, but not because the different characteristics of two groups of participants. To avoid such effect a randomisation technique was used to produce groups that have probabilistic equivalent characteristics.

- A study would be called reliable if repeating that study produces similar outcomes as the original study. To make a study repeatable, it is essential for researchers to state clearly the research method and procedure and maintain a consistency throughout while conducting the study. In the first experiment, since a comprehension test was used to measure undergraduates' proof comprehension, I followed consistent procedure (discussed in Chapter 5) for marking the test to increase the reliability of the marking process.

- Ethical considerations are essential while planning and conducting research, especially when humans are involved in the research. Participants of my research were adult undergraduates. They gave informed consent before the experiment took place. They were also told that it was not compulsory to take part in the study but their participation allowed me as a researcher to get insight into the learning of mathematical proofs at undergraduate level. They were also aware of their right to withdraw their participation at any time without given a reason.

Chapter 5

Experiment 1: The effectiveness of e-Proofs and written textbook proofs on undergraduates' proof comprehension

5.1 Introduction

In this chapter, I report on the first experimental study in which the effect of reading e-Proofs on undergraduates' proof comprehension was compared with written textbook proofs. After briefly discussing the aim of the study in Section 5.2, Section 5.3 discusses the design of the study, and describes the materials and tasks used for the study in Section 5.4 and participants in Section 5.6. Before describing the procedure of the study in Section 5.7; section 5.5 reports a pilot study which was conducted before the main study, especially to check the comprehension test that was designed to examine undergraduates' comprehension of an unknown proof. After presenting the outcomes of the study in Section 5.9, a follow-up discussion presents in Section 5.10 and the chapter ends with a short summary which highlights the key issues and main conclusions.

5.2 Aim of the study

Chapter 3 has shown that undergraduates who experienced e-Proofs during a module of real analysis; they tended to claim that e-Proofs had strong impact on their learning of mathematical proofs. However their claim seemed unrealistic (discussed in Chapter 3) and no evidence has been found so far to support it. Consequently, it was also important to compare the impact of reading e-Proofs, a new way of presenting proofs, with written textbook proofs, a traditional way of presenting proofs, while considering the issue of effectiveness. This study aimed to investigate the following research questions:

- Is the impact of reading an e-Proof on undergraduates' proof comprehension better than the impact of reading a written textbook proof?
- What was the long term impact of these two proof presentations on undergraduates' comprehension of proofs?

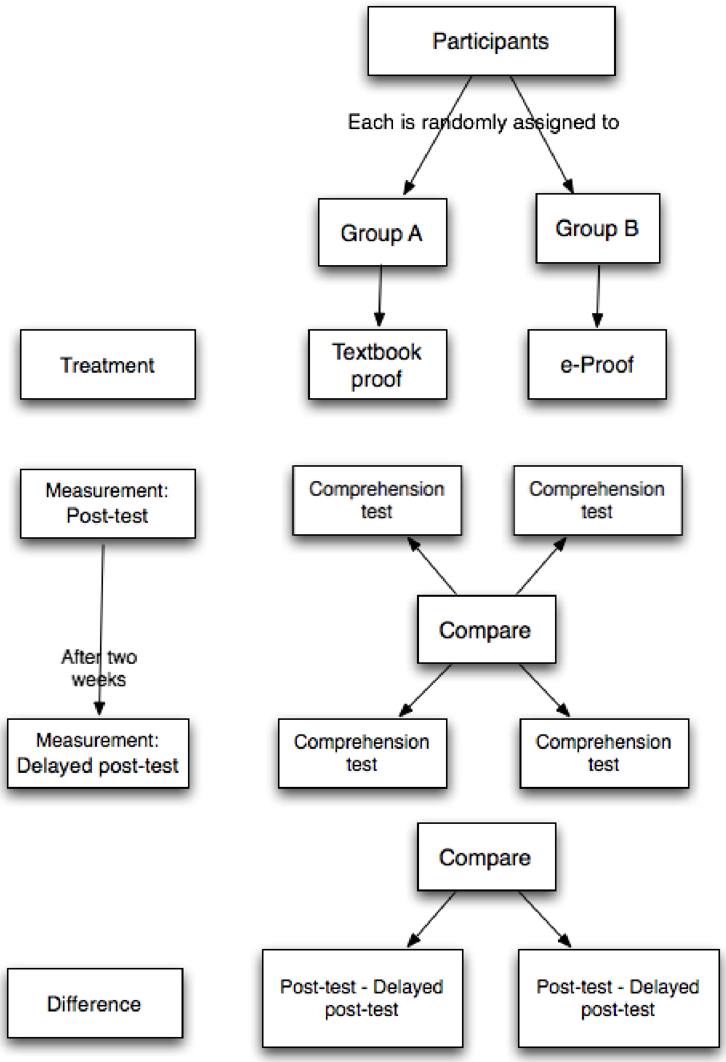
5.3 Design of the study

The aim of this study was to directly compare the impact of reading e-Proofs with reading written textbook proofs to explore whether e-Proofs are more effective on undergraduates' proof comprehension than written textbook proofs. Therefore, an experimental research approach[1] was adopted to confirm or to refute the hypothesis that undergraduates would perform better in proof-related tasks when they read an e-Proof compared to a written textbook proof.

A between-subjects research design was adopted in which participants are randomly assigned to two groups where each group received either an e-Proof presentation or a written textbook presentation of a proof of an unknown theorem for comprehension. This randomisation technique is used to produce probabilistic equivalence between groups (Christensen, 2003). The impact of reading the two different proof presentations was evaluated based on undergraduates' performances on a comprehension test twice – the post-test (immediately after reading

[1] A detail discussion of why an experimental research approach was chosen for the study was presented in the previous chapter.

FIGURE 5.1: An overview of the experimental design of the study.



the proof) and the delayed post-test (two weeks later). The difference between the post-test score and the delayed post-test score for both groups were used to examine the long term impact of an e-Proof and a written textbook proof. Figure 5.1 gives an overview of the design of the experimental study.

5.4 Materials

The proof chosen for the experimental study was the proof of Cauchy's Generalised Mean Value Theorem (GMVT). There were several reasons to chosen the theorem.

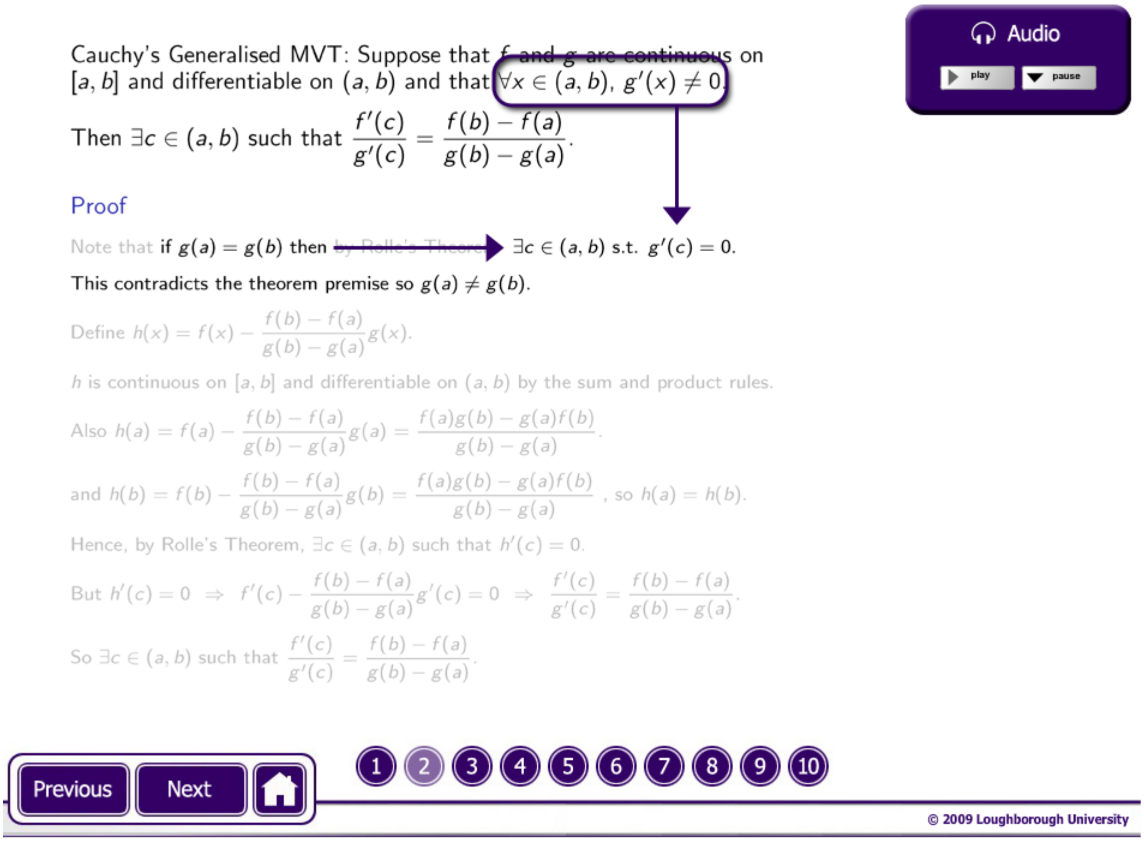
- The proof of Cauchy's GMVT is a well-balanced proof which is not too easy nor too hard considering the proof comprehension task for undergraduates.
- Most importantly, the e-Proof version of Cauchy's GMVT has all the typical characteristics that previous e-Proofs have had. E-Proofs "focus on supporting comprehension by explicating the relationships among the theorem premises and conclusions, the individual lines of the proof, and external information such as established definitions and theorems" (Alcock and Wilkinson, 2011, p.4). In case of Cauchy's GMVT also the attached audio with each slide gives explanation to provide additional information beyond the written texts of the proof. For example, the audio explains how Rolle's Theorem and the Mean Value Theorem are linked to the proof and used in order to prove the theorem. The boxes and arrows appears to link between lines (e.g. between the theorem statement and the proof) to draw reader's attention. Therefore, the e-Proof of Cauchy's GMVT can be seen as representative of the wider class of e-Proofs.
- The plan was to run the experiment during a lecture of the real analysis module. Rolle's Theorem and the Mean Value Theorem are required in order to understand Cauchy's GMVT, so the experiment was planned in the sixth week of the module; by the time undergraduates (who were pursuing the module) has seen Rolle's Theorem and the Mean Value Theorem in lectures, but not Cauchy's Generalised Mean Value Theorem (GMVT).

5.4.1 Cauchy's GMVT

The written texts of the e-Proof of Cauchy's GMVT were identical to the written textbook version of the proof. The e-Proof version, however, had ten slides and each slide highlighted a particular line and those parts of the proof that are directly associated with that line. The attached audio with each slide gives explanation to provide additional information beyond the written texts of the proof. The boxes and arrows appears to link between lines or to draw reader's attention. Figure 5.2 represents the snap shot of the second slide of the e-Proof of Cauchy's GMVT and Figure B.1 shows a snap sort of the written textbook proof of Cauchy's GMVT.

Also a short version of Cauchy's GMVT was prepared for the study which was handed out before the comprehension test. Undergraduates who took part in the

FIGURE 5.2: A snap sort of the second slide of the e-Proof of Cauchy's GMVT.



experiment had seen the proof for the first time during the experiment and they were required to answer a comprehension test based on the proof. Since the aim of the experiment was intended to examine participants' understanding but not to test their memory, a short version of the proof was presented to them during the test. This short version of the proof had minimum explanation compared to the written textbook proof. Moreover, the comprehension test refers to some expressions that were included in the short version of the proof, $h(x) = f(x) - \frac{f(b)-f(a)}{g(b)-g(a)} g(x)$ for example. Figure 5.4 presents the short version of the proof.

An additional information sheet of definitions and theorem statements were also prepared for those participants who read the written textbook version of Cauchy's GMVT. This information sheet contained definitions of continuity, differentiability and theorem statements of Rolle's Theorem, Mean Value Theorem and properties of continuity and differentiability which were used to prove the theorem. To comprehend Cauchy's GMVT, one might need to recall these theorems and lemmas. And the intention of providing the additional sheet was to give all participants equal chance to recall those relevant information, rather than depending or their

FIGURE 5.3: A snap sort of the written textbook proof of Cauchy's GMVT.

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

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

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

Proof

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

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

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

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

Also $h(a) = f(a) - \dfrac{f(b) - f(a)}{g(b) - g(a)} g(a) = \dfrac{f(a)g(b) - g(a)f(b)}{g(b) - g(a)}$.

and $h(b) = f(b) - \dfrac{f(b) - f(a)}{g(b) - g(a)} g(b) = \dfrac{f(a)g(b) - g(a)f(b)}{g(b) - g(a)}$, so $h(a) = h(b)$.

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

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

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

own memories. Appendix B.3 presents the materials used in the additional information sheet.

Finally a comprehension test was prepared on Cauchy's GMVT, which is discussed next.

5.4.2 The comprehension test

A comprehension test was designed to examine undergraduates' proof comprehension of a proof of Cauchy's GMVT. The test contained nine questions. To

FIGURE 5.4: A snap sort of the short version of the proof of Cauchy's GMVT.

Suppose that f and g are continuous on $[a, b]$ and differentiable on (a, b). Suppose also that $\forall x \in (a, b)$, $g'(x) \neq 0$.
Then $\exists c \in (a, b)$ such that $\dfrac{f'(c)}{g'(c)} = \dfrac{f(b) - f(a)}{g(b) - g(a)}$.

Proof

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

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

3) Define $h(x) = f(x) - \dfrac{f(b) - f(a)}{g(b) - g(a)} g(x)$.

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

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

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

7) But $h'(c) = 0 \Rightarrow \dfrac{f'(c)}{g'(c)} = \dfrac{f(b) - f(a)}{g(b) - g(a)}$.

8) So $\exists c \in (a, b)$ s.t. $\dfrac{f'(c)}{g'(c)} = \dfrac{f(b) - f(a)}{g(b) - g(a)}$.

comprehend a proof, according to Lin and Yang (2007), five different facets of proof comprehension are involved — basic knowledge, logical status, summarisation, generality and application (discussed in Chapter 2). The eight questions were designed in such a way as to reflect five facets of proof comprehension. Table 5.1 gives an overview of the questions and the associated facets of proof comprehension.

There were two questions, Question 4 and Question 8, on the facet of basic knowledge for reflecting basic knowledge such as algebraic manipulation. Questions 2 and Question 5 were designed for reflecting the undergraduates' understanding of logical links in the proof, which refers to the facet of logical status. Understanding the critical ideas used in the proof is essential for answering Question 1 and Question 3 and that kind of understanding was labeled as the facet of summarisation in Lin and Yang's (2007) proof comprehension model. Question 6 and Question 7 were intended to measure respectively the facet of generality — facet of justifying

Table 5.1: The table shows which questions were set to examine which facet of proof comprehension.

Facet of proof comprehension	Question
Basic knowledge	Question 4, Question 8
Logical status	Question 2, Question 5
Summarisation	Question 1, Question 3
Generality	Question 6
Application	Question 7

correctness and identifying what is validated by the proof and the facet of application — and the facet of identifying the different premises and knowing to apply for the same premises in other cases.

The nine questions set for the comprehension test are given below:

1. (a) How does line (1)[2] contradict the theorem premise?

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

2. Which of the following properties are used in the proof?
   If functions p and q are continuous on $[a, b]$ and differentiable on (a, b) then

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

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

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

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

3. Write a short paragraph to explain why we set

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

4. With h defined as in the proof, what is $h'(x)$?

5. Where[3] does the proof show that h satisfies the conditions for Rolle's Theorem?

[2]Here line (1) indicates the first line of the short version of Cauchy's GMVT (see subsection B.2) that undergraduates received at the beginning of the comprehension test.

[3]All participants received a short version of the Cauchy's GMVT (see subsection B.2), while answering the comprehension test.

6. (a) Imagine that the premises of Cauchy's GMVT read:
   Suppose that f and g are continuous on $[a, b]$ and differentiable on (a, b) and that $\forall x \in (a, b), f'(x) \neq 0$.
   What would the conclusion say?
   (b) If you were proving the new version (i.e. (a)) directly, how would you define $h(x)$?
7. Find an interval where $f(x) = 1 + x^2$ and $g(x) = x^2 - 1$ satisfy the premises of Cauchy's Generalised MVT.
8. Show that

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

9. Write down in your own words a short summary of the proof.

However one question, ("write down in your own words a short summary of the proof".) was excluded from the quantitative data analysis. It proved hard to develop a reliable mark scheme for this question, so the analysis focussed only on Question 1 to Question 8.

5.5 Pilot study

Before conducting the experiment a small pilot study was conducted, which helped me to reflect and to reconsider several issues — especially the design of comprehension test of Cauchy's GMVT.

5.5.1 Aim of the pilot study

The aim of this pilot study was to get an overview of the following issues:

- To check whether the comprehension test was well balanced or not (i.e. not too easy or not too hard). If the questions were too easy then that might not be able to reflect whether the different proof presentations have different impact on undergraduates' proof comprehension. A similar argument holds if the questions were too hard. Therefore a balanced comprehension test was essential for fulfilling the purpose of the experiment.

- To check the clarity of the questions set for the comprehension test, that is, to check whether the phrases used in the questions were understandable and to eliminate ambiguity and difficulties in wording if necessary.
- To check how much time participants would take to read the proof and to answer the comprehension test.
- To check what kind of written answers to expect in the main study and how to mark the answers.

5.5.2 Participants of the pilot study

Four third year undergraduates (2 females and 2 males) from the mathematics department of Loughborough University took part in the pilot study. Two of them were pursuing a MMath degree and the other two were studying B.Sc. Mathematics. All participants were familiar with Cauchy's GMVT as they successfully completed the real analysis module in their second year of study in the academic year of 2008-09.

5.5.3 Method and procedure of the pilot study

Participants were invited individually for a task-based interview. Firstly, participants were asked to read and comprehend the written textbook version proof of Cauchy's GMVT and to return it to the interviewer when they finished. There was no time restriction for that reading task. When participants finished reading the proof, the comprehension test paper was given to them along with a short version of Cauchy's GMVT and additional information sheet which contained related theorem statements, axioms and definitions. While answering the questions, participants were asked to think aloud. Thinking aloud is a widely accepted technique of gathering information about human's thinking process while engaging in an activity (Ericsson and Simon, 1998). According to Ericsson and Simon (1998), a think aloud protocol provides evidence of the sequence of thoughts in a wide range of tasks. The purpose behind the individual interview was to reveal those issues mentioned above. On average the reading task took around 10 minutes to finish, and answering the comprehension test took 29 minutes. However, the average answering time included the time when undergraduates were explaining their thoughts and clarifying those questions asked during the task, besides producing written answers to the test. All four interviews were audio-recoded and video-recorded.

5.5.4 Outcomes of the pilot study

In the first two interviews, I noticed that both participants had difficulties with understanding the phrasing of Question 6(a). Since Question 6(b) was directly related to the Question 6(a), it seemed meaningless to them. Therefore, Question 6 was re-phased after the second interview. The rephrasing (as written in subsection 5.4.2) made the question understandable, as in the next two interviews undergraduates did not find any difficulties in understanding the question.

After observing the four undergraduates in the pilot study, it was decided that in the main study undergraduates would receive 15 minutes for the reading task and 30 minutes to answer the comprehension test.

5.6 Participants of the main study

The second and third year mathematics undergraduates (single or joint honours) of Loughborough University, who were enrolled for the real analysis module in the academic year of 2009-2010, were invited to take part in the study. Those undergraduates had been introduced to e-Proofs by their lecturer during the analysis module. There was 32 and 25 undergraduates who took part initially as participants in the textbook group and in the e-Proof group respectively. For the delayed post-test, there was 28 participants in the textbook group and 21 participants in the e-Proof group. Only the 49 participants who took part both in the post-test and the delayed post-test were considered during the data analysis.

5.7 Procedure of the main study

The main study took place during a lecture in the sixth week of the real analysis module in the academic year of 2009-2010. Undergraduates were informed by their lecturer about the experiment during a previous lecture and they were also told taking part in the study was not compulsory to them and the study was entirely for research purpose in the area of proof comprehension at undergraduate level. They were divided randomly into two groups — the textbook group and the e-Proof group[4]. Undergraduates received an email from their lecturer before the day of the study in which they were informed about which group they were in and to which room they should be going for that particular day.

There were 32 and 25 undergraduates who turned up on the day of experiment for the textbook group and for the e-Proof group respectively. Both groups were told they would be given an unfamiliar proof which they have to comprehend and then they need to answer a written comprehension test based on that proof. The textbook group received the written textbook version of Cauchy's GMVT (see Appendix B.1). This group also received an additional information sheet which contained definitions of continuity and differentiability and statements of relevant theorems (see Appendix B.3). The additional information was provided to the group, so that they could integrate the information while comprehending the proof, because the focus of the study was to explore to what extent undergraduates comprehend an unfamiliar proof, but not to test their existing subject knowledge. On the other hand, the e-Proof group sat in a computer lab where the e-Proof of Cauchy's GMVT was on the computer screen. Both the textbook group and the e-Proof group were asked to comprehend the proof and they were told that they had 15 minutes to comprehend the proof. After 15 minutes the reading group were asked to return the written textbook proof and for the e-Proof group their computers were logged off automatically.

[4]There was a third group called lecture group in the experiment. The participants of this group attended a live lecture on Cauchy's GMVT during the experiment. This group performed significantly better in both the post-test and in the delayed post-test compared to the other two groups. But I did not include the lecture group in this chapter because of a vital methodological error by the lecturer and me in this condition. The lecture group had a 21 minutes lecture on Cauchy's GMVT whereas other two group had 15 minutes to comprehend the proof. Moreover, this thesis aimed to examine the effectiveness of the e-Proofs as a new resource compared to a traditional resource textbooks rather than comparing it with live lectures. The characteristics of a live lecturer are so different from an e-Proof, that they are not easily comparable.

After comprehending the reading task, both groups received the short version of Cauchy's GMVT and the comprehension test question paper which contained nine questions. There was blank spaces in the question paper to write the answer for each question. Both groups had 30 minutes to answer the comprehension test.

To find out the longer term effects of the written textbook proof and the e-Proof on undergraduates' proof comprehension, a delayed post test was conducted after two weeks during a lecture. All undergraduates who came to the lecture received the same comprehension test question paper (which contained 9 questions) along with the same short-version of the Cauchy's GMVT and the same additional information sheet of definitions and theorem statements. This time they had 20 minutes to answer the comprehension test. There were 80 undergraduates in the lecture, among them 28 took part as participants of the textbook group and 21 were in the e-Proof group.

5.8 Marking the comprehension test

All the answer sheets for the post-test and the delayed post-test were checked and marked (except Question 9 — "write down in your own words a short summary of the proof"), yielding a score out of 18. The distribution of marks between each question was decided while marking of answer sheets from the pilot study. This helped to maintain a consistency in marking process.

Question 1 had two subquestions of one mark each. In case of Question 2 (which of the following properties are used in the proof?), undergraduates were required to choose the properties used in Cauchy's GMVT among four options and there was one mark for each option. Undergraduates would get four marks if they choose properties (a) and (c) that has been used in the proof and properties (b) and (d) that were not used in the proof. Here are two examples of undergraduates' answers to Question 2 presented in Figure 5.5 and Figure 5.6 which scored two and four respectively.

Question 3 (write a short paragraph to explain why we set $h(x) = f(x) - \frac{f(b)-f(a)}{g(b)-g(a)} g(x)$.) had three marks. Three examples which show what kind of answers received what kind of scores in the test are presented below. Figure 5.7 shows an answer which got one mark out of three, Figure 5.8 and Figure 5.9 are the examples of answers scored two and three respectively.

FIGURE 5.5: The answer given below scored two out of four. Two options were identified correctly in the answer, that is, property (a) that was used in the proof and property (d) was not used, whereas property (c) was used instead of property (b). Therefore, this answer scored two out of four.

Which of the following properties are used in the proof?
If functions p and q are continuous on $[a, b]$ and differentiable on (a, b) then

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

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

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

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

(a) and (b) are used. The sum rule and multiple rule.

FIGURE 5.6: For choosing the correct two options (a) and (c) and left out other two that were not used in the proof, this answer scored four.

1. Which of the following properties are used in the proof?
If functions p and q are continuous on $[a, b]$ and differentiable on (a, b) then

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

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

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

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

Properties c and a

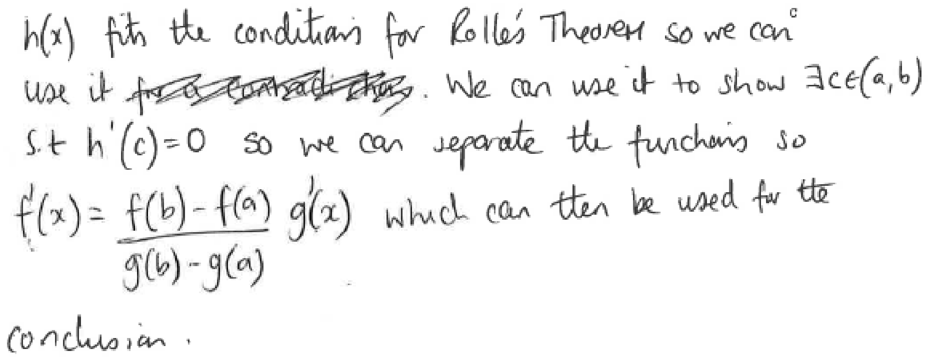
FIGURE 5.7: The example given below shows one of the answers to Question 3, that an undergraduate wrote in the post test and the mark given for such answer was one out of three.

So we can get h(a)=h(b) and prove the rest of the proof.

FIGURE 5.8: The marks given for answer presented below was two out of three.

We set h(x) as the above so that we have the conditions for Rolles theorem, including h(a)=h(b)

FIGURE 5.9: This is an example of answer of Question 3 which scored three out of three.


h(x) fits the conditions for Rolle's Theorem so we can use it ~~for a contradiction~~. We can use it to show $\exists c \in (a,b)$ s.t $h'(c)=0$ so we can separate the functions so $f'(x) = \frac{f(b)-f(a)}{g(b)-g(a)} g'(x)$ which can then be used for the conclusion.

Question 4 had one mark in which the expression for $h'(x)$ was required as a correct answer. Question 5, (where does the proof show that h satisfies the conditions for Rolle's Theorem?), had 2 marks because the answer would be completed if one mentioned (i) h is continuous on $[a, b]$ and differentiable on (a, b) and (ii) $h(a) = h(b)$. If undergraduates wrote either of them, they would get one out of two.

Question 6 had two subquestions of one mark each. Question 7, (find an interval where $f(x) = 1+x^2$ and $g(x) = x^2-1$ satisfy the premises of Cauchy's Generalised MVT), had two marks. Undergraduates would get two marks if they gave an example of a *closed* interval which does not include zero, whereas a correct example of an *open* interval would get only one mark. There were two marks in Question 8 for performing the steps of algebraic manipulation to establish the relationship: $f(a) - \frac{f(b)-f(a)}{g(b)-g(a)} g(a) = \frac{f(a)g(b)-g(a)f(b)}{g(b)-g(a).}$

5.9 Results

Since the aim of the experiment was to compare the performances of the e-Proof group and the textbook group in both tests, an analysis of variance (ANOVA) with one *within-subjects factor, time* (post-test and delayed post-test), and one *between-subjects factor, group* (with two levels, the textbook group and the e-Proof group) was conducted to analysis the performances of both groups. The analysis revealed that the main effect of time was highly significant, $F(1, 47) = 28.213, p < 0.001$, which means both groups performed significantly better in the post-test (with an average score of 10.7) than the delayed post-test (an average score of 8.9).

TABLE 5.2: Mean scores (out of 18) of each group in the post-test and the delayed post-test and their the average score across both tests.

Group	Post-test	Delayed Post-test	Average
Textbook group	10.4	9.4	9.9
e-Proof group	11.2	8.4	9.8

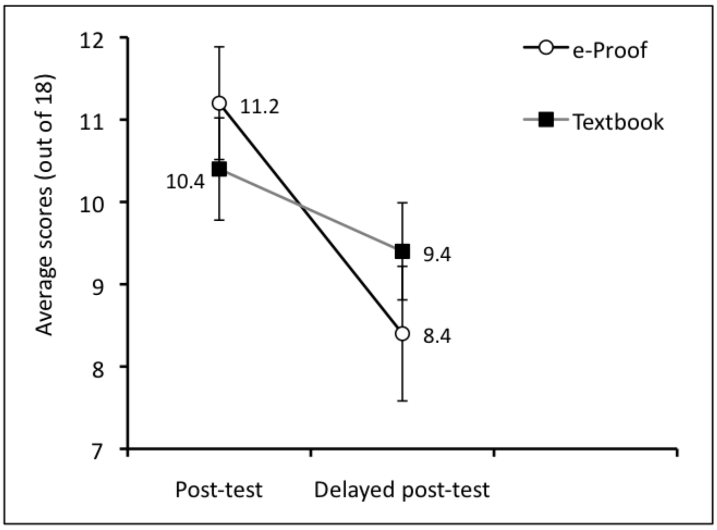


FIGURE 5.10: This graphs presents the average scores (out of 18) of the e-Proof group and the textbook group in the post-test and the delayed post-test. The graph also shows the significant interaction effect of time×group. Error bars represent ±1 SE of the mean.

The interaction effect of time × group was significant, $F(1, 47) = 5.659, p = 0.021$. This result indicates that the drop off in the score for the e-Proof group (16%) in the delayed post-test was significantly higher than the drop off in the score for the textbook group (6%). This means that the e-Proof group had significantly lower retention than the textbook group in the delayed post-test.

However, the main effect of group was not significant, $F(1, 47) = 0.006, p = 0.938$. This indicates that there was no significant difference in the performances between the e-Proof group and the textbook group across both tests. In the post test, the mean score of the e-Proof group was 11.2 and 10.4 for the textbook group; whereas in the delayed post-test, the mean score of the e-Proof group and the textbook group were 9.3 and 8.4 respectively. The mean scores in the both tests for each group are also shown in Table 5.2.

In sum, both groups performed significantly better in the post-test than the delayed post-test, however there was no difference in the performance of the comprehension test when undergraduates read the e-Proof version of Cauchy's GMVT or the textbook version. The score drop off in the delayed post-test compared to the post-test for the e-Proof group was significantly higher than the drop off in the score of the textbook group (with a significant interaction effect). This suggests that although the immediate impact of reading an e-Proof was essentially same as reading a written textbook proof, the long term impact of reading e-Proof on undergraduates' proof comprehension was poorer than reading a textbook proof.

5.9.1 High and low achieving undergraduates

One important question was regarding achievement[5] — was the drop off in scores for the high achieving undergraduates in the delayed post-test just as much as for the lower achieving undergraduates? Those 49 undergraduates who took part in both tests were split into two categories — high achieving and low achieving, based on their achievements in the final exam of the real analysis module. There were 23 undergraduates in the high achieving category (those above the median score) and 26 in the low achieving category (those at or below the median score).

To compare the performances of the undergraduates, a 2×2×2 ANOVA was conducted with one within-subject factor time (post-test and delayed post-test) and two between-subject factor group (e-Proof and textbook) and achievement (high and low). The interaction effect of time × group × achievement was not significant, $F(1, 45) = 0.043, p = 0.836$, which indicates that the between-conditions drop off scores from the post-test to the delayed post-test for the high achieving undergraduates and for the low achieving undergraduates were essentially same. Therefore, it was not only the low achieving undergraduates who performed poorly in the delayed post-test but also the high achieving undergraduates, especially those who were in the e-Proof group.

There are two graphs presented showing the performances of the high achieving (see Figure 5.11) and low achieving undergraduates (see Figure 5.12) in both tests. Figure 5.11 shows that the performances of those high achieving undergraduates who read the e-Proof of Cauchy's GMVT, were dropped off significantly in the delayed post-test compared to those high achieving undergraduates who read the textbook version of that proof.

[5]In the sense of achievement in exam, here high achieving undergraduates refer to those undergraduates who achieved high scores in the final exam of the real analysis module

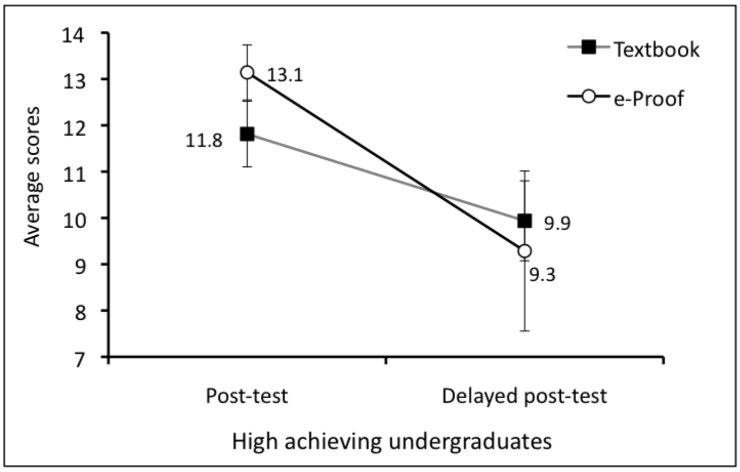


FIGURE 5.11: This figure presents the average scores of the high achieving undergraduates in the post-test and in the delayed post-test both from the e-Proof group and the textbook group. Error bars represent ±1 SE of the mean.

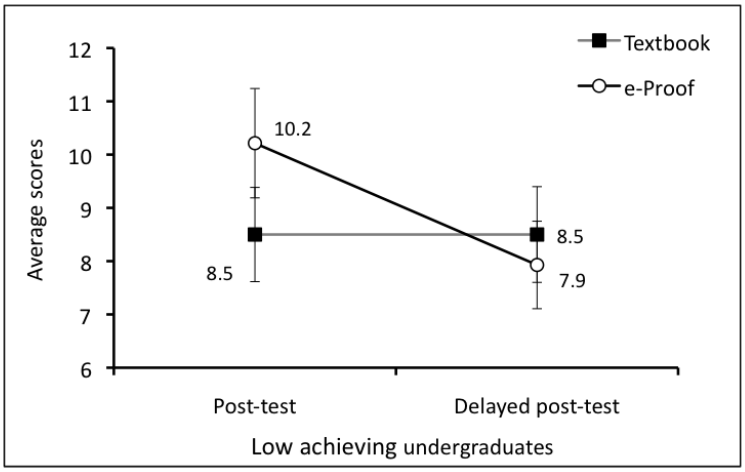


FIGURE 5.12: The figure presents the average scores of the low achieving undergraduates in both tests for the e-Proof group and the textbook group. Error bars represent ±1 SE of the mean.

The detention in the performances in the delayed post-test was significantly higher for the low achieving undergraduates who read the e-Proof than those who read the textbook version of the proof (Figure 5.12). Therefore, in sum, the long term impact of reading an e-Proof on both the high and the low achieving undergraduates was poorer than reading a written textbook proof.

5.10 Discussion

The aim of this study was to investigate whether an e-Proof has better impact on undergraduates' proof comprehension than a written textbook proof. Before the experiment, it was expected that the undergraduates, who would read the e-Proof version of Cauchy's GMVT, would perform better in the comprehension test than those who read the textbook version of the proof. This expectation was logical in two ways:

- The e-Proofs were created to help undergraduates to comprehend proofs in a better way as e-Proofs provide additional comprehensive explanations and more. Therefore the group who received the e-Proof of Cauchy's GMVT had the chance to comprehend the proof better than the textbook group, who, in contrast, received a written proof without any additional explanation.
- Secondly, undergraduates gave very positive reviews about e-Proofs and made strong claim about the positive impact of e-Proofs in surveys (both in the web-based survey that I reported in Chapter 3 as well as in the survey reported by Alcock (2009)) and in focus group discussions.

However, the outcomes of the study exhibited a different scenario — the immediate impact of reading an e-Proof and a textbook proof on undergraduates' proof comprehension was essentially the same, but the long term impact of reading an e-Proof was significantly poorer than reading a textbook proof. There are two key issues that I will discuss in order to explain why the impact of e-Proof was poorer than the impact of a textbook — firstly, the structural differences in these two proof presentations and secondly, the activities required to comprehend them.

Firstly, the structure of these proof presentations are very different. Here by the term 'structure of the proof presentation', I mean how much information is given in the proof presentation and how that information is presented. In the e-Proof of Cauchy's GMVT, the proof is presented through ten slides in which each slide highlights a particular line and the attached audio provides additional explanation about that line. For example, take the first line of the proof — "note that if $g(a) = g(b)$ then by Rolle's Theorem $\exists c \in (a, b)$ s.t. $g'(c) = 0$". The attached audio of the first line explains — "in the first line we observe that $g(a)$ equals to $g(b)$ the Rolle's theorem can be applied to g. This is because we know from the theorem premise that g is continuous in the open interval (a, b) and differentiable in the closed interval $[a, b]$. Applying Rolle's theorem tells us that there exists c in the open interval (a, b) such that $g'(c) = 0$". Moreover, arrows appear to synchronise between the written texts and the audio explanations. On the other side, the first line of the written textbook proof of Cauchy's GMVT only provides — "note that if $g(a) = g(b)$ then by Rolle's Theorem $\exists c \in (a, b)$ s.t. $g'(c) = 0$".

The e-Proof is an example of high coherence text[6] in which most of the information required to comprehend the proof is explicitly provided through the audio explanations (in the sense of McNamara and Kintsch (1996)). In contrast, the written textbook proof (with no additional explanation) is a low-coherence text, where the coherence gaps are bigger. Therefore, to comprehend the written textbook proof, undergraduates would need to generate more inferences to fill coherence gaps (Weber and Alcock, 2005), than those who read the e-Proof version of the proof. Research has suggested that depending on the level of students prior-knowledge, low or high coherence text can have different impact on students (Chi et al., 1994; McNamara et al., 1996; McNamara, 2001). Low prior knowledge students should be benefited from high coherence text (such as e-Proof) as they need additional support to fill the gaps in a text. One may conjecture that high coherence text allows low prior knowledge students to complete the cognitive processes that (they are not capable of doing without help) required for comprehending the text. In case of e-Proof however, the additional help did not facilitate comprehension for low achieving undergraduates (see Figure 5.12). Conversely, High prior knowledge students get a benefit from low coherence text as they are capable of filling the coherence gaps that are not explicitly stated in the text. But proving too much information can cause a reverse effect for them as found in my study. Excessive help in e-Proof did not improve proof comprehension for high achieving undergraduates. These findings question Alcock's presumption that e-Proofs would be beneficial for both high and low achieving undergraduates. Next I discuss what kind of activities are needed to comprehend an e-Proof and a written textbook, which I consider as a vital issue.

An e-Proof not only has more detailed information for each line of the proof than its textbook version, but also that information is presented using two mediums — the visual medium (written texts) and the audio medium (audio explanations).

[6]A mathematical proof can be considered as text — combinational of mathematical ideas expressed by using words and mathematical symbols.

That means, an e-Proof is a multimedia representation of a proof. According to Mayer (2001), study materials presented using both audio and visual mediums are better for learning than materials presented in a single medium, a written textbook proof, for example. Mayer argued that when information is presented both in audio medium and in visual medium, it is processed separately into our cognitive system through two separate channels, which reduces the cognitive load and facilitates learning. Therefore, according to the theory of multimedia learning (which I have discussed in detail in Chapter 2), the presentation of an e-Proof should facilitate learning. Conversely, only the visual medium is used for presenting a written textbook proof and the cognitive load would be comparatively high on our cognitive system while comprehending a textbook proof and that load is not good in terms of learning. However, the experiment showed a different outcome that e-Proofs had worse long term impact on learning than textbook proofs.

It is the research of Schnotz and his colleges (Schnotz, 2002; Schnotz and Rasch, 2005) who have shown that presenting material through multimedia with animation facility to make the process of learning easy, could have positive (enabling) effect only if learners need it to engage with the cognitive processes that they may not be able to do without that extra help. But because they are capable, extra help allows them to devote less in the cognitive processes than they should, and that shallow processing leads to a poor learning outcomes. Schnotz referred this negative impact on learning as the *facilitating effect*. It seems that undergraduates of the e-Proof group might be capable of comprehending a proof by their own, but additional facilities make the reading comprehension process much easier for them and allow them to put less cognitive effort than they should, which leads to the poorer proof comprehension compared to the textbook group.

Clearly, to comprehend a written textbook proof undergraduates need to generate their own explanations (Weber and Alcock, 2005). The activity of generating inferences might encourage them to engage more with the task of proof comprehension than those who had explicit explanations in an e-Proof. Research (Ainsworth and Burcham, 2007; Chi et al., 1994; Hodds et al., 2014) has shown that students who spontaneously self-explain when they study learn more than those who do not. Moreover, self-explanations are usually more effective than explanations provided by others. One of features of the e-Proof is "explanations provided by others", whereas comprehending a textbook proof demands more self generated explanations and probably the self-explanation factor had a positive long term effect on the textbook group's proof comprehension. The conjecture I set was: comprehending e-Proofs needed less cognitive effort when compared with comprehending textbook proofs.

Therefore, the second study was designed and conducted, to investigate *why* e-Proofs had poorer long term impact on undergraduates' proof comprehension compared to written textbook proofs and I believed that investigating reading proof comprehension process of these two proof presentation would give an answer.

5.11 Summary of Chapter 5

- A new resource, 'e-Proofs' were designed to guide undergraduates to comprehend a proof. Undergraduates claimed that e-Proofs had positive impact on their comprehension of proofs. However, the self-reporting data did not provide the evidence that can justify the claim. This chapter reports the study that was designed to compare directly the impact of reading an e-Proof with a traditional written textbook proof.

- Forty-nine Undergraduates took part in the study in which they were randomly divided into two groups — one group read an e-Proof version of Cauchy's GMVT and the other group read the textbook version of that proof. Their comprehension of the proof was examined twice — immediately after reading the proof (the post-test) and two weeks after (the delayer post-test after).

- There was no significant difference found in the performances of the comprehension test when undergraduates read either an e-Proof or a textbook proof. However, the drop off in the score for the e-Proof group in the delayed post-test was significantly higher than the drop off in the score for the textbook group (see Figure 5.10). This suggests the long term impact of reading e-Proof on undergraduates' proof comprehension was poorer than reading a textbook proof.

- The structure of an e-Proof is highly coherent in which most of the information are provided using two mediums — audio medium and visual medium. The facilities available in an e-Proof in terms of amount of information and the way it presented might make the process of proof comprehension easier than comprehending a textbook proof.

- To comprehend a textbook proof, a reader needs to fill the coherence gaps in the proof. To bridge those gaps readers needs to generate explanations and research (Chi at el, 1994; Ainsworth and Burcham, 2007) shows that the activity of self-explanation has long term benefit in term of proof comprehension. In contrast, an e-Proof itself provides those explanation rather than allowing its readers to put the effort, which might cause the poorer long term impact on undergraduates' proof comprehension.

Chapter 6

Eye-tracking methodology

6.1 Introduction

This chapter discusses the eye-tracking methodology that I adopted for the second experiment reported in the next chapter. This experiment was focused on how undergraduates read and comprehend e-Proofs and textbook proofs.

In this chapter, in Section 6.2, I report the general research interests of the second experiment. In Section 6.3, I discuss the issues related to the *think aloud methodology* to highlight why an *eye-tracking methodology* was more suitable to address the research questions than a think aloud research approach. Section 6.4 describes the assumption which is the basis of the eye-tracking methodology. I also introduce some terminologies related to this methodology which I will be using widely in Chapter 7. In section 6.5, I describe the Tobii T120 eye-tracker that was used for the study. I also describe briefly how I recorded and analysed the eye gaze data using the eye-tracker. Finally, this chapter ends by highlighting the key points of the eye-tracking methodology.

6.2 Research questions

The first experiment showed that the long term impact of reading an e-Proof was poorer on undergraduates' proof comprehension than reading a textbook proof, even though e-Proofs were designed particularly for helping undergraduates to comprehend proofs. The outcomes of the first study raised the question: 'why was the impact of an e-Proof was poorer than a textbook proof?', which leads to the second experiment. A possible reason as, I have argued in the Chapter 5, might be that: the additional information and the structure of the e-Proof makes the process of proof comprehension easier for undergraduates. As a consequence undergraduates might not need to devote as much cognitive effort as they need to for comprehending a textbook proof. Therefore, the hypothesis set for the second study was: *for undergraduates, comprehending an e-Proof is comparatively easier task than comprehending a written textbook proof.*

One of the main reasons of creating e-Proofs was to guide undergraduates how to read proofs for comprehension (Alcock, 2010). Direct instructions about how to read a proof (i.e, where to look at when reading each line of a proof) should have an impact on the way undergraduates allocate their attention across an e-Proof. For that reason, there must be some differences in terms of undergraduates' attention allocation when they read an e-Proof compared to a written textbook proof. Therefore, the second study aimed to clarify: *whether providing direct instructions through an e-Proof causes any differences in undergraduates' attentive behaviours compared to the way their attention is distributed across a textbook proof.*

Moreover, the existing literature of mathematics education provides little knowledge about how undergraduates read and comprehend proofs. Therefore, my second study had the potential to explore more about the process involved when undergraduates comprehend proofs. This study also allow me to explore what differences and similarities exist in the process of an e-Proof comprehension and a textbook proof comprehension.

6.3 Thinking aloud vs eye-tracking methodologies

To investigate how undergraduates read and comprehend two different proof presentations, the first consideration was how to gather information about the cognitive processes that undergraduates use while engaging in the activity of reading and comprehending proofs. One of the most useful ways of studying cognitive processes is the verbal protocol analysis, which is one of the most frequently used methodologies in the area of educational research. In verbal protocol analysis, information is gathered by asking a person to do a task and to verbally report all the thoughts that occur while doing the task (Ericsson and Simon, 1980).

The thinking aloud method has been used considerably (i) to develop theories of cognitive processes for various tasks, for example, problem solving (Schoenfeld, 1982) and (ii) to discover the patterns of behaviour while interacting with applications and documents, for example, how learners interact with online resources (Van Den Haak et al., 2003). Although the think aloud method is a popular research method of exploring human's cognitive processes, there are some drawbacks which were important considerations. First, I discuss those drawbacks to make an argument as to why the thinking aloud protocol was not a suitable research method for this particular study.

- The think aloud method is not necessarily useful for those activities that have become an automatic process, for example, reading. Reading is a automatic process for an adult reader and verbalising all thoughts while reading is not always possible. Moreover, people may think faster than they speak. In a study, Branch (2000) was told by her participants that they think faster than they speak and their thought process is much more complex than they verbalised.

- Proof comprehension is an activity which demands high cognitive processing (Selden and Selden, 2003). Research has suggested that thinking aloud is problematic for such an activity that has high cognitive demands (Branch, 2000). Constantly verbalising thoughts while doing the activity substantially increases the cognitive load. This high cognitive load takes participants' focus away from the activity that they are supposed to do and may impact upon performance.

- Research has shown that people cannot report accurately on their own cognitive processes (Nisbett and Wilson, 1977). Even experts are poor in reporting the process through which they proceed to solve problems, as they are expert in solving problems, but not in reporting what they are thinking while doing the task. Participants might not be able to report all the strategies they use while doing a task as some of them occur in a non-conscious level. In addition, sometimes people might report those mental events that did not actually occur while doing a task, but they would like to proceed to complete the given task (Nisbett and Wilson, 1977).

- Talking while doing a task can influence performance. This influence is known as the reactivity of the think aloud process. For example, thinking aloud might alter the natural reading process and it can influence positively the performance of a reading related task (Chi et al., 1994; Yoshida, 2008). Therefore, thinking aloud while reading and comprehending proofs might increase participants' efficiency in performing the task and therefore the data generated might not reflect the way actually those participants read and comprehend proofs when not thinking aloud.

- Thinking aloud is not a natural activity that people are used to doing. Because of that if participants keep silent for long period of time, the verbalisation might become unusable as a significant part of the cognitive process may be lost for that period of time (Ericsson and Simon, 1980). If researchers interrupt to encourage participants to continue the thinking aloud process that may cause error in the data collected through that process. For example, if participants are asked to explain their thought by researchers, then participants are forced to justify their thoughts. This interruption can break the natural thought process, according to Ericsson and Simon (1980).

Considering these drawbacks, it was clear that for a cognitively demanding task like reading and comprehending proofs, a think aloud protocol might not be efficient at exploring how undergraduates pursue the task of proof comprehension. Moreover, this study also wished to explore what similarities and differences exist in undergraduates' attentive behaviour when they read two different proof presentations. Some of the processes in reading and comprehending proofs are automatic and occur in a non-conscious level, so verbal reporting would not be able to reflect particularly those processes. An alternative way of studying undergraduates' reading comprehension processes is to capture their eye-movements while they engage with the activity of proof comprehension.

6.4 Assumptions of the eye-tracking methodology

The eye-tracking methodology has been used to study human behaviour while engaging in various activities such as reading (Rayner, 1998), searching and extracting information from web pages (Cowen et al., 2002; Goldberg et al., 2002; Pan et al., 2004), human-computer interaction (Jacob and Karn, 2003; Poole, 2005), comprehending and solving arithmetic word problems (Hegarty et al., 1992), validating mathematical proofs (Inglis and Alcock, 2012), reasoning (Ball et al., 2003), learning in a multimedia environment (Hyönä, 2010; Mayer, 2010).

According to Duchowski (2007), by tracking someone's eye-movements, we can follow their locus of attention. Since my second study wished to explore how reading an e-Proof is similar and different from reading a textbook proof, eye tracking data can provide some insight into what draws undergraduates' attention in e-Proofs and in textbook proofs and what they found hard or easy to comprehend, how much time they spend on reading each line, how they make links between lines etc.

The eye-tracking methodology is based on the assumption of the *eye-mind hypothesis* which says that there is a close link between the direction of human gaze and the focus of attention (Just and Carpenter, 1980). Especially for reading-related activities, there is a very close connection between eye-movement and cognitive process and eye-movements are tightly linked with moment to moment goals and subtasks (Land et al., 1999).

There are mainly two types of eye-movements — *saccades* and *fixations*.

- *Saccade:* Saccades are the fast movements in which the eye moves rapidly from one location to another. These are continuous and rapid movement of eye gazes which last a few tens of milliseconds. No information can be processed during a saccade (Rayner, 1998).
- *Fixation:* Between saccades, the eye is stationary, this period of time is refereed to as a fixation. Fixations involve relatively motionless gaze which lasts about 200-500 millisecond (ms), in which visual attention is aimed at a specific area of a visual stimulus. Visual information is processed entirely during fixations (Rayner, 1998).

FIGURE 6.1: This shows an eye tracking protocol for the proof comprehension task in which a participant read a given proof.

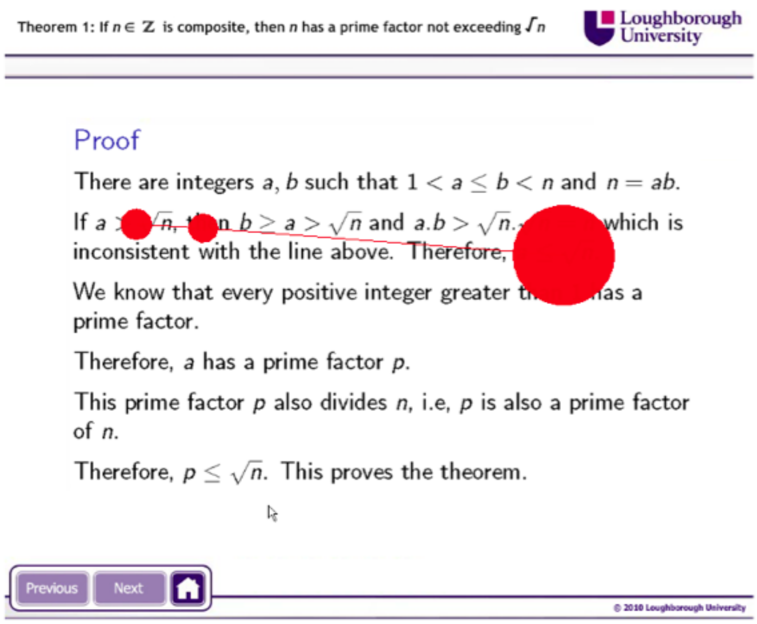


Figure 6.1 represents a snapshot of the recorded screen of the eye movement protocol for the proof comprehension task in which a participant reads the given proof. Three red spots represent fixations and the lines that connected these spots refer to saccades which occurs between two fixations. The size of the red spots indicates the duration of each gaze. The bigger spot indicates longer fixation duration.

6.4.1 Some terminologies of the eye-tracking methodology

Before discussing how eye-movement data can indicate what draws undergraduates' attention while reading two types of proof or how much effort undergraduates put to comprehend e-Proofs and textbook proofs, first I will be introducing some terminology that is used widely in rest of the thesis.

- *Area of interest (AOI):* An area of interest is an analysis method (a region based method) used in eye tracking. Researchers define areas of interest over certain parts of a displayed stimulus, and analyse only the eye movements that fall within such areas (Poole, 2005).

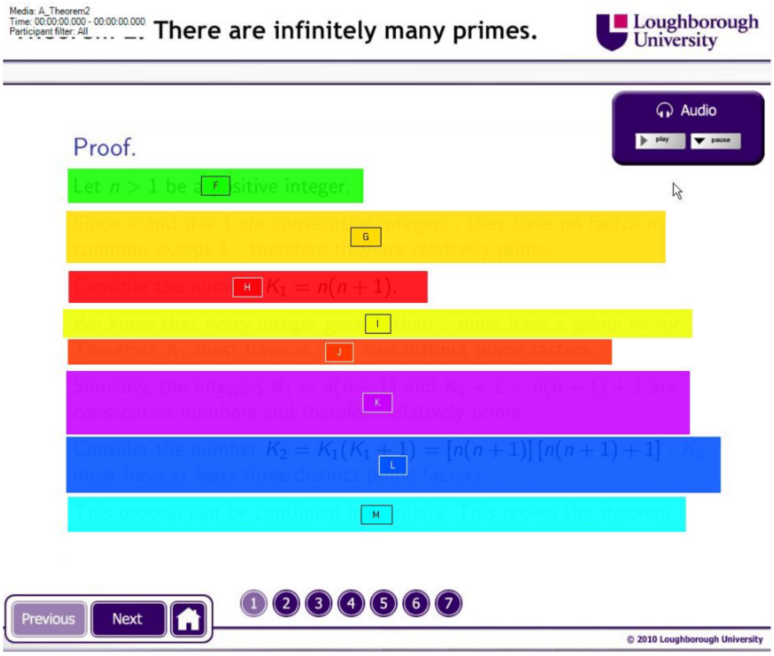


FIGURE 6.2: The pre-defined area of interests for Theorem2.

I adopted this method to analyse the eye movement data that I collected while undergraduates were engaged in the task of reading e-Proofs and textbook proofs for comprehension.

Figure 6.2 shows a screen which displays the e-Proof version of Theorem 2 used for the second study. The rectangles represent AOIs that I defined before analysing the data. Each AOI was identified with different colours and different names.

- *Fixation duration:* The amount of time one single fixation lasts is known as the fixation duration. This is linked to the processing time applied to the object being fixated (Just and Carpenter, 1980). When a text is conceptually difficult, fixation duration increases, that is, longer fixation duration indicates difficulties in processing the visual information (Rayner, 1998).

  Fixation duration is a commonly used dependent variable in those studies which aim to explore underlying cognitive processing involved while encoding visual information (Jacob and Karn, 2003; Poole, 2005; Rayner, 1977, 1998). Just and Carpenter (1980, p.330) argued that the eye remains fixated on a word as long as the word is bring processed. So the time it takes to process a newly fixated word is directly indicated by the gaze duration.

  For example, Rayner's (1977) study has shown that since the verb is the key for comprehending a sentence, the mean fixation durations were longer on verbs than the other parts of sentences. This means that the cognitive efforts readers put to comprehend verbs were higher than the other words. Similarly, Hyönä and Niemi's (1990) research has found that the average fixation durations decrease with repeated reading because less cognitive effort is needed to comprehend a text for the second time.

In my study, the fixation duration has been used to measure whether comprehending an e-Proof is easier than comprehending a textbook proof. Since the longer fixation duration indicates that readers have difficulties in processing the information, I was expecting to find that the mean fixation duration while reading a textbook proof would be significantly longer than when reading an e-Proof.

- *Dwell time:* Dwell time is the sum of all fixation durations that an observer spends on a pre-defined area (known as area of interest (AOI) on a displayed stimulus. This has been an useful measurement for many studies (Inglis and Alcock, 2012). It indicates how observers allocate their visual attention on the displayed stimuli. That is, what parts of the visual stimuli they find interesting or important so that the time spent on fixating on those areas are greater than the rest.

  In the case of the second study, my aim was to find if the additional information and different presentation format can make any difference in undergraduates' attentive behaviours while reading e-Proofs compared with reading textbook proofs. Moreover, in e-Proofs undergraduates are given direct instruction about how they should be reading a proof, for example, where to look while reading a line to understand how that line is linked with the rest of the proof. Conversely, in a textbook, information is presented in a linear manner. It is therefore readers who need to infer the links between lines which are not explicitly explained in the written proof. Therefore, measuring the dwell times for e-Proofs and textbook proofs allows me to compare directly whether undergraduates have the idea of how to read a proof or they need the additional help for comprehending a proof.

- *Fixation sequence analysis:* Fixation sequence analysis focuses on the order of eye-movements over the pre-defined AOIs. By observing transition between lines, researchers can reveal the cognitive strategies that a user adopts while processing information from a displayed screen (Inglis and Alcock, 2012).

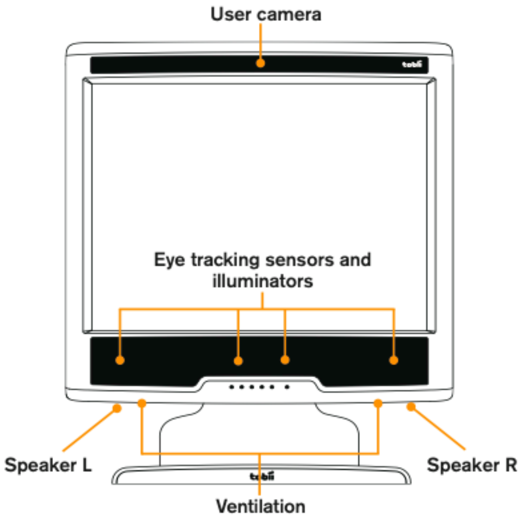


FIGURE 6.3: Tobii T120 the remote eye-tracker, taken from Tobii T120 Eye-tracker User Manual, (p.15).

For my study, my assumption was because of the presentation format of an e-Proof (that is, the audio explanations and highlighted parts for each line of the proof), readers are forced to look on particular parts as either they are highlighted or suggested to look by the audio explanations. This means the eye moves more frequently in both forward and backward directs while reading an e-Proof compared to reading a textbook proof.

6.5 The eye-tracker

The eye-tracker *Tobii 120* was used for the second study. It was integrated in a 17-inch TFT monitor and was designed for all types of eye-tracking studies, where the stimuli can be presented on the monitor. This is a remote eye-tracker, which does not require physical contact with the user's eye or eye-socket. A remote eye-tracker uses a reflector tracker — a beam of light is projected onto the eye, after which sophisticated cameras pick up the difference between the pupil reflection and known reference points to determine what the user is looking at (Duchowski, 2007). Figure 6.3 shows the Tobii T120 eye-tracker. The figure also shows where cameras and speakers are integrated on the TFT monitor.

Tobii Studio software was used for recording and analysing eye gaze data.

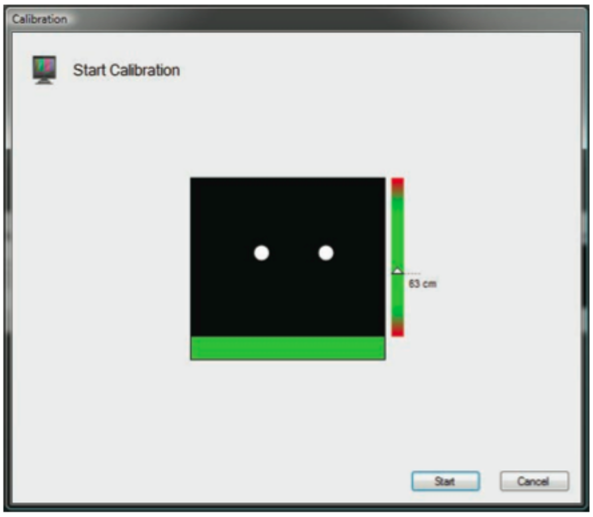


FIGURE 6.4: The figure shows a screen when calibration process starts.

- The study materials needed to be uploaded as a new project in Tobii software first. For my experiment, I included the initial instructions (that is what participants would be seen as the experiment proceeds), proofs (that they need to comprehend) and multiple choice comprehension tests (for each proof). Different types of stimuli can be used, such as image files, movie files, webpages, pdf files etc.

- When the study materials are uploaded, then that are ready for the experiment. Before conducting the experiment, each participant went through an automatic calibration process which ensured that the eye-tracker was able to capture both eyes. Figure 6.4 shows the screen when calibration process starts. Two white spots refer to the eyes and the side scale shows the distance between eyes and the computer screen. The ideal distance between the eyes and the screen should be within 60-70 cm.

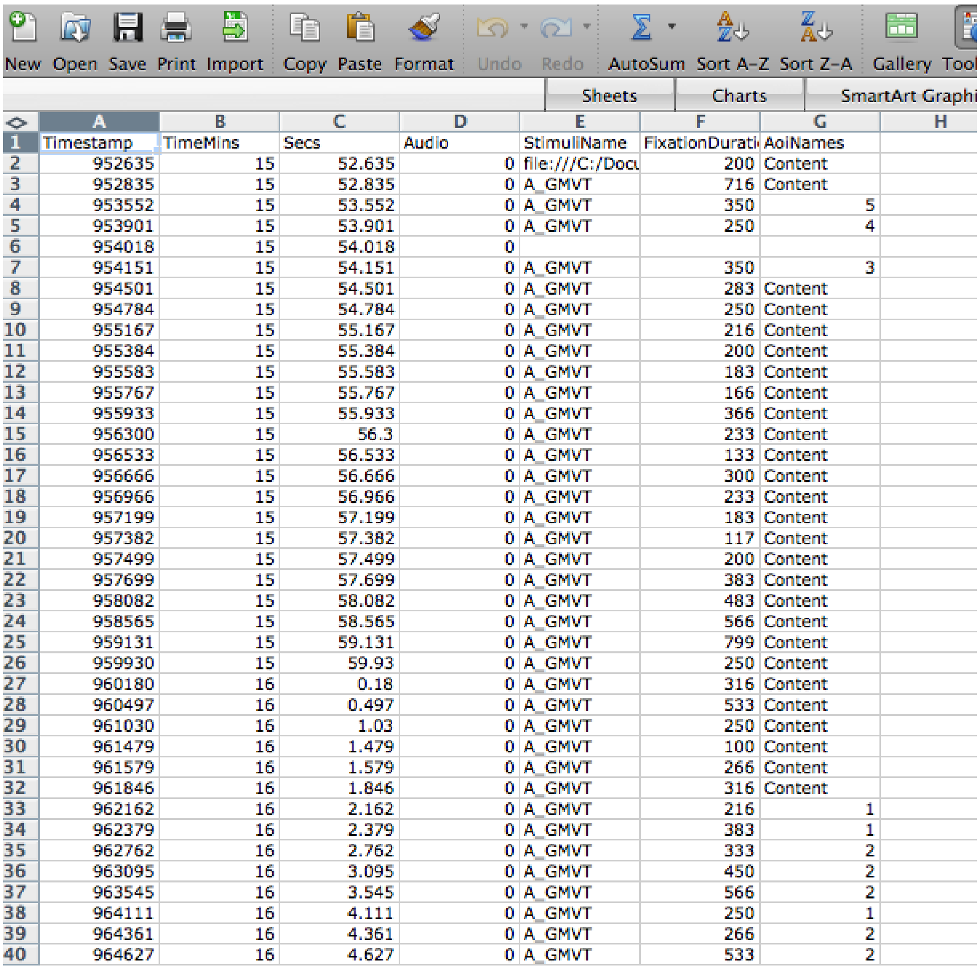
- Once a participant completes the experiment, how he/she proceeded through the given tasks can be replayed. By using the visualisation facility available in Tobii software, one can see a recorded experiment as many times as he/she wants to. The software also allows a researcher to observe directly how a participant's attention is distributed on a displayed stimulus through a *heat map*. A heat map uses different colours to show the number of fixations participants made in certain areas of a displayed stimulus or for how long they fixated within that area. Red usually indicates the highest number of fixations or the longest time, and green the least, with varying levels in

FIGURE 6.5: The figure shows the heat map of the second instruction page used for the experiment. The colours green to red indicate less to high attention that participant made while reading the instruction.

between. Figure 6.5 shows the heat map of the second instruction page used in my experiment.

- To compare how participants' eye-movements vary when they read an e-Proof or a textbook proof, I used statistical analysis. To conduct that analysis, first I extracted the eye gaze data into text files. Before extracting data into text file format, I defined area of interests for each proof (Figure 6.2 shows the defined AOIs for Theorem 2). I also created segments of each proof and for each participant.

  A segment represents a section of recording. I created the starting and ending point of each segment manually. Those segments were created because I was interested in those sections that contain the relevant gaze data.



	A	B	C	D	E	F	G	H
1	Timestamp	TimeMins	Secs	Audio	StimuliName	FixationDurati	AoiNames	
2	952635	15	52.635	0	file:///C:/Docu	200	Content	
3	952835	15	52.835	0	A_GMVT	716	Content	
4	953552	15	53.552	0	A_GMVT	350	5	
5	953901	15	53.901	0	A_GMVT	250	4	
6	954018	15	54.018	0				
7	954151	15	54.151	0	A_GMVT	350	3	
8	954501	15	54.501	0	A_GMVT	283	Content	
9	954784	15	54.784	0	A_GMVT	250	Content	
10	955167	15	55.167	0	A_GMVT	216	Content	
11	955384	15	55.384	0	A_GMVT	200	Content	
12	955583	15	55.583	0	A_GMVT	183	Content	
13	955767	15	55.767	0	A_GMVT	166	Content	
14	955933	15	55.933	0	A_GMVT	366	Content	
15	956300	15	56.3	0	A_GMVT	233	Content	
16	956533	15	56.533	0	A_GMVT	133	Content	
17	956666	15	56.666	0	A_GMVT	300	Content	
18	956966	15	56.966	0	A_GMVT	233	Content	
19	957199	15	57.199	0	A_GMVT	183	Content	
20	957382	15	57.382	0	A_GMVT	117	Content	
21	957499	15	57.499	0	A_GMVT	200	Content	
22	957699	15	57.699	0	A_GMVT	383	Content	
23	958082	15	58.082	0	A_GMVT	483	Content	
24	958565	15	58.565	0	A_GMVT	566	Content	
25	959131	15	59.131	0	A_GMVT	799	Content	
26	959930	15	59.93	0	A_GMVT	250	Content	
27	960180	16	0.18	0	A_GMVT	316	Content	
28	960497	16	0.497	0	A_GMVT	533	Content	
29	961030	16	1.03	0	A_GMVT	250	Content	
30	961479	16	1.479	0	A_GMVT	100	Content	
31	961579	16	1.579	0	A_GMVT	266	Content	
32	961846	16	1.846	0	A_GMVT	316	Content	
33	962162	16	2.162	0	A_GMVT	216	1	
34	962379	16	2.379	0	A_GMVT	383	1	
35	962762	16	2.762	0	A_GMVT	333	2	
36	963095	16	3.095	0	A_GMVT	450	2	
37	963545	16	3.545	0	A_GMVT	566	2	
38	964111	16	4.111	0	A_GMVT	250	1	
39	964361	16	4.361	0	A_GMVT	266	2	
40	964627	16	4.627	0	A_GMVT	533	2	

FIGURE 6.6: The figure shows the exported eye gaze data into an excel file.

- Tobii T120 eye-tracker recorded a large range of data, but I selected only those data that were relevant for my research. For example, number of fixation on each AOI, or fixation duration on each AOI were the data that I was interested in, so I exported those information into text file format. These text files also open in Excel and SPSS. Figure 6.6 shows the exported text format data file. This figure shows the data extracted were — Timestamp (shows the time in milliseconds when the gaze data was collected), StimuliName (pre-defined displayed stimuli name), Fixation Duration (duration of each fixation, in miliseconds), AoiNames (pre-defined name of AOIs). I added three columns named TimeMins (gaze data recording time in min), Secs (gaze data recording time in sec) and Audio (0 and 1 refer as audio-off and audio-on respectively) later.

  This eye gaze data was then analysed using the statistical software SPSS, which is described later in Chapter 7.

6.6 Summary of the chapter

- Thinking aloud is a widely-used method to explore humans' cognitive processes while they are engaged in some activities. However, especially when the activity is cognitively highly demanding, such as comprehending mathematical proofs, thinking aloud is not an useful approach for gathering information. Moreover, reporting cognitive processes (when some of them are automatic and occur at the non-conscious level) is not always possible as the thinking processes are faster and more complex than can be verbalised.

- The eye tracking methodology is an alternative way to explore what differences and similarities exist when undergraduates read proofs presented in two ways — e-Proof and textbook proof. In addition, tracking undergraduates eye movements would allow me to compare directly undergraduates attentive behaviours while they read these different proof presentations.

- The eye-tracking methodology is based on the assumption of the *eye-mind hypothesis* which says that there is a close link between the direction of human gaze and the focus of attention.

- There are mainly two types of eye-movements — *saccades* and *fixations*. Saccades are quick eye movements to direct a viewer's eye to a visual target. Between saccades, the eye is stationary, this period of time is known as fixation and during fixation the visual information is processed.

- To analyse the eye movement data, mean fixation duration and dwell time can be used as dependent measurements. Longer fixation duration indicates that readers have difficulties in processing the visual information. Dwell time is used to refer to how readers distribute their attention over displayed stimuli.

- Tobii T120 remote eye-tracker was used for the experiment to record the eye gaze data.

Chapter 7

Reading comprehension process of e-Proofs and textbook proofs

7.1 Introduction

In this chapter, I report the second experimental study in which I investigated the reading comprehension processes of e-Proofs and written textbook proofs. I asked, what similarities and differences exist when undergraduates read these two proof presentations for comprehension. In Section 7.2, I explain the aim of conducting the experiment by restating the research questions. In Section 7.3, I describe the research method along with participants, materials used and the procedure of the study. Section 7.4 reports the outcomes of the study, followed by a discussion about those outcomes in Section 7.5. The chapter ends with a short summary which highlights the key points of this chapter.

7.2 Aim of the study

The second study is a follow up study of the first experiment. In Chapter 6, "The experimental methodology", I stated the research questions and explained why they are vital to explore further in order to resolve the issue of why e-Proofs appear not to be effective, especially on the long term basis, for undergraduates proof comprehension compared to traditional textbook proofs. Here, I restate those issues that I investigated in Experiment 2 before describing the research method adopted for the experiment.

- To recheck whether the immediate impact of reading e-Proofs and textbook proofs on undergraduates' proof comprehension are the same as I found in Experiment 1.
- To investigate Alcock's (2009) claim that undergraduates often do not know where to look when reading and comprehending a proof and they need additional help to grasp the skill of reading proofs which is displayed in e-Proofs.
- To investigate whether the design of e-Proof forces its readers to read a proof differently than the way they usually read proofs.
- To investigate whether the chance of getting distracted while reading an e-Proof is much higher compared to reading a textbook proof.
- To investigate how much effort undergraduates put to comprehend an e-Proof and a textbook proof.
- To investigate undergraduates' reading behaviours under two conditions — when reading e-Proofs, they listen to the given audio explanation and when they do not listen to the audio explanation.

Before reporting each result, I briefly discuss why that result is relevant for my study and how it addresses one of the research questions.

7.3 Research method

This study involved a within-subjects research design. A within-subjects design increased the power a lot and that, with eye tracking study, it is impractical to have a very large sample size. It had one within-subjects factor (e-Proofs and textbook proofs) and one between-subjects factor (Group A and Group B). Because of the within-subjects research design all participants experienced both the proof presentations — e-Proof and textbook proof. Four proofs were selected for the study. Two textbook proofs were set at the beginning of the task, because Alcock (2009) argued that the e-Proof was created to guide students how to read a

TABLE 7.1: The table shows which group read what proofs and in what order.

Group	Textbook proof	e-Proof
A	Th1, Th3	Th2, Th4
B	Th2, Th4	Th1, Th3

proof more effectively for better comprehension, two e-Proofs were set at the end of task to avoid the carryover effect of reading e-Proofs on the textbook proofs.

Participants were divided randomly into two groups — Group A and Group B. Group A saw the textbook versions of Theorem 1 (Th1) and Theorem 3 (Th3) first and then the e-Proof version of Theorem 2 (Th2) and Theorem 4 (Th4). Group B read proofs in the opposite order — the textbook version of Th2 and Th4 and then the e-Proof of Th1 and Th3 (see Table 7.1).

The eye-tracking methodology was used for the study. I discussed in detail the eye-tracking methodology in Chapter 6 as well as how I analysed the gaze-movement data, therefore I am not repeating the discussion about the eye-tracking methodology in this chapter.

7.3.1 Participants

The participants in this study were the second and third year mathematics undergraduates of Loughborough University, who had taken the module of real analysis. When I conducted the experiment, some of them had completed the module last year (academic year 2009-10) and others were pursuing the module in the current year (academic year 2010-11).

Thirty-four undergraduates (Male=16, Female=18) were recruited and were paid £8 for taking part in the study. The data was collected between the last week of October and the first week of December, 2010.

For analysing Theorem 1, one participant's eye-gaze data was removed because of the inconsistent recording (recorded less than 40%) due to some technical problems during the experiment (for example, excessive head movements or poor calibration). For similar reason, two participants for Th3 and one participant for Th4 were excluded from the analysis.

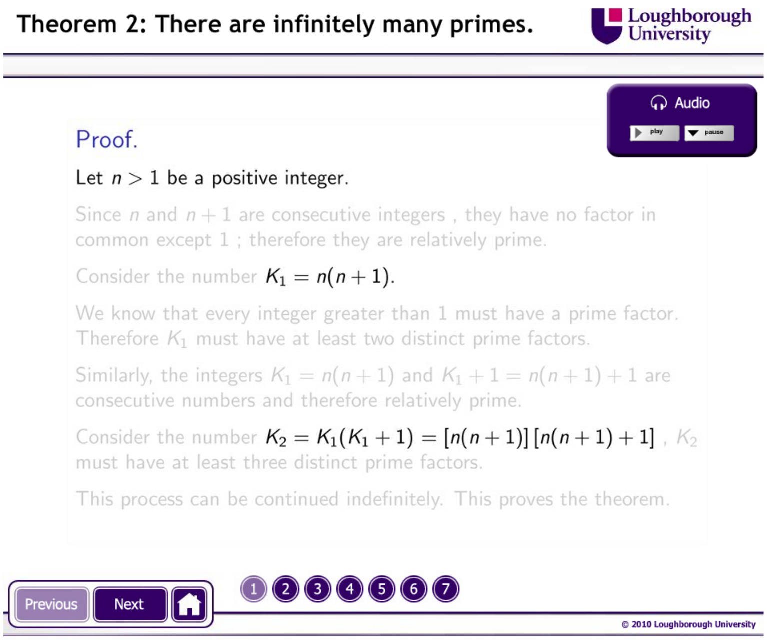


FIGURE 7.1: The snap shot of the first slide of the e-Proof version of Theorem 2.

7.3.2 Study materials

There were four mathematical proofs used for the experiment. Two of the four proofs were taken from the real analysis course (10MAB141) of Loughborough University and other two were elementary number theory proofs selected from Rose's (1984) book on number theory and a number theory website (http://primes.utm.edu/notes/proofs/infinite/Saidak.html). For the experiment, these two number theory proofs were slightly modified from how they were presented in the website. Both the e-Proof version and the textbook version of the four proofs were used in the study. Both the real analysis proofs already had an e-Proof version that had been used in the real analysis course of Loughborough University for last four years. The e-Proof version for the number theory proofs were created and reviewed by Lara Alcock, to ensure that they were faithful implementations of the e-Proof concept.

Figure 7.1 and Figure 7.2 show how the e-Proof version and the textbook version of Th2 were presented in the experiment.

For each proof, a multiple choice comprehension test was set based on Mejia-Ramos et al.'s (2012) framework. Those tests were set to ensure that participants should try to comprehend the proofs. The comprehension test for Th3 consisted

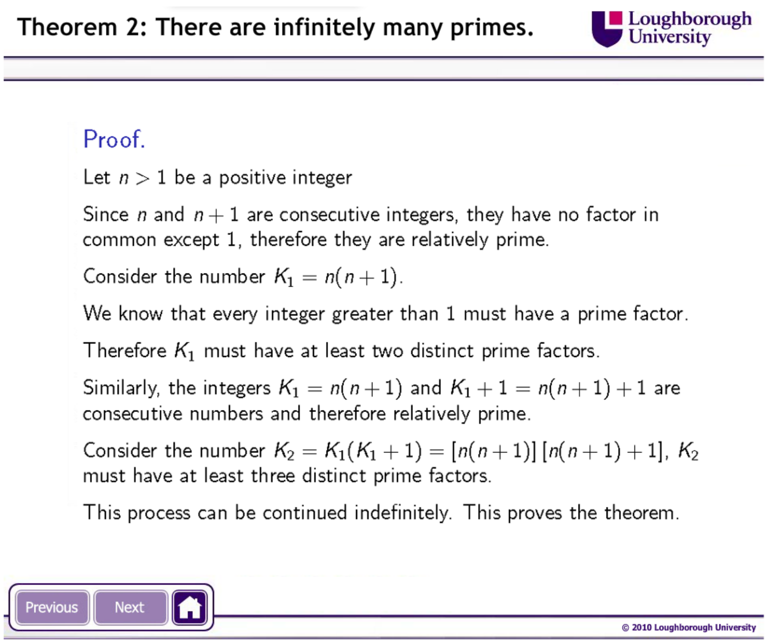


FIGURE 7.2: The snap shot of the textbook version of Theorem 2.

of five multiple choice questions and the three theorems each had three multiple choice questions.

7.3.3 Procedure

Each participant was invited individually and assigned to one of the groups (Group A and Group B) randomly. At the beginning of the experiment, all participants received oral instructions of what they would see during the experiment. Before each proof, there was an instruction page of what theorem and what kind of proof they were going to read (e-Proof or textbook proof) and how to proceed with the reading (for example, what key to use to move on to the next page). They were encouraged to read the proofs for as long as they wished, in order to maximise their comprehension. They were also told that when they had finished reading, they would need to rank their confidence level in a five-point likert scale from very poor to very good and to answer a multiple choice comprehension test. Each participant passed an automatic eye-tracking calibration before starting the reading task when they were seated in the chair facing the eye-tracker. They were also told to minimise their head movements as much as possible during the experiment.

7.4 Results

In this section, I report the results of Experiment 2. Since there are more than one research question that I wanted to address through this experiment, I briefly re-discuss the purpose of investigating these issues before reporting each of the findings.

7.4.1 Comprehension test

Although comprehension test scores were not the primary focus of the study, between-groups differences in scores on the comprehension test for each theorem were investigated using Kolmogrov-Smirnov Z tests (because tests were scored out of 3 or 5, parametric tests were not appropriate). No significant differences between the groups were observed for any proof: Theorem 1, $Z = 1.172, p = .128$; Theorem 2, $Z = 1.150, p = .142$; Rolle's Theorem, $Z = .190, p = 1.00$; GMVT, $Z = .380, p = .999$. In other words, like in Experiment 1, there was no evidence of the groups learning different amounts at the immediate post-test.

7.4.2 Visual attention allocation on e-Proofs and textbook proofs

Alcock (2009, p.4) argued that, while comprehending a proof presented by a lecturer in a live lecture, undergraduates' "attention may not be directed precisely enough, especially if the explanation involves looking at two lines in different places in the proof". This statement indicates that undergraduates might not have the skill of reading a proof that can bring understanding to them. To overcome such difficulties e-Proofs were created that can guide undergraduates (with its highlighted parts, arrows and boxes and also with the audio explanations) where to concentrate when reading each line of a proof. Even though the first experiment revealed the poor long term impact of reading e-Proofs on undergraduates' proof comprehension, e-Proofs might help undergraduates to learn how to read a proof. Therefore, in the second experiment, my first attempt was *to explore how undergraduates allocate their visual attention when reading a textbook proof comprehension compared to an e-Proof.*

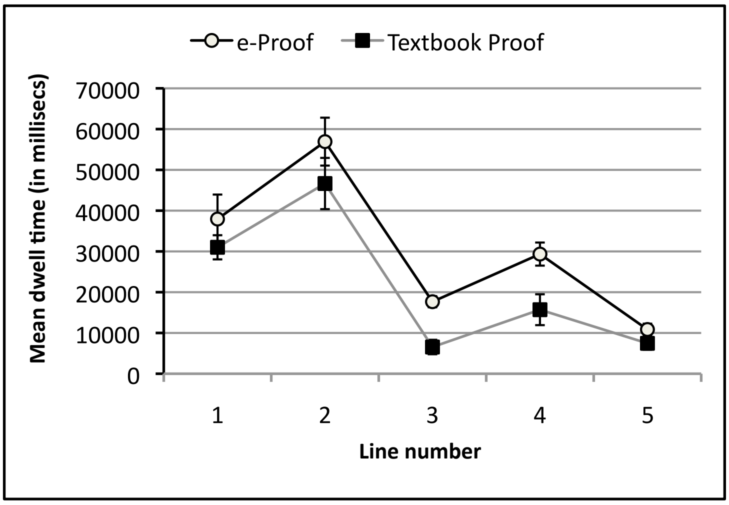


FIGURE 7.3: Dwell times on the each line of Theorem 1. Error bars represent ±1 SE of the mean.

For that, I analysed the dwell time (total fixation time) on pre-specified areas of interest (AOIs) for each proof[1]. This measurement indicates undergraduates' visual attention allocated during the proof comprehension and whether undergraduates' attention distributions while reading a textbook proof and an e-Proof were different or not.

For Theorem 1 (Th1), considering one within-subjects factor (line number: five lines of the proof, defined as five AOIs — 1,2,3,4,5) and one between-subjects factor (condition: e-Proof and textbook proof), a 5x2 Analysis of Variance (ANOVA) was conducted. The analysis showed that the main effect of line number was significant, $F(2.107, 65.311) = 49.999$, $p < 0.001$ (with Greenhouse-Geisser correction), which means the undergraduates spent significantly different times to comprehend each line. For example, the second line is a crucial line of the proof and the undergraduates took longer to comprehend Line 2 (mean 51632 ms) compared to the other four lines (average dwell time 19432 ms/line). The main effect of the type of proof was also significant $F(1, 31) = 7.91$, $p = 0.008$, which indicates that the reading time of the e-Proof of Th1 (with grand mean of dwell time 30.54 s/line) was significantly longer than the textbook proof (with grand mean 21.48 s/line). There was no significant interaction of line number × condition, $F(2.107, 62.311) = 0.646$, $p = 0.535$. Although reading the e-Proof version took longer than reading its textbook version, the attention distributions throughout the proof in both versions were similar (see Figure 7.3).

[1]Each line of proof was identified as one of the AOIs. For example, Theorem 2 has nine lines proof and therefore there were nine AOIs

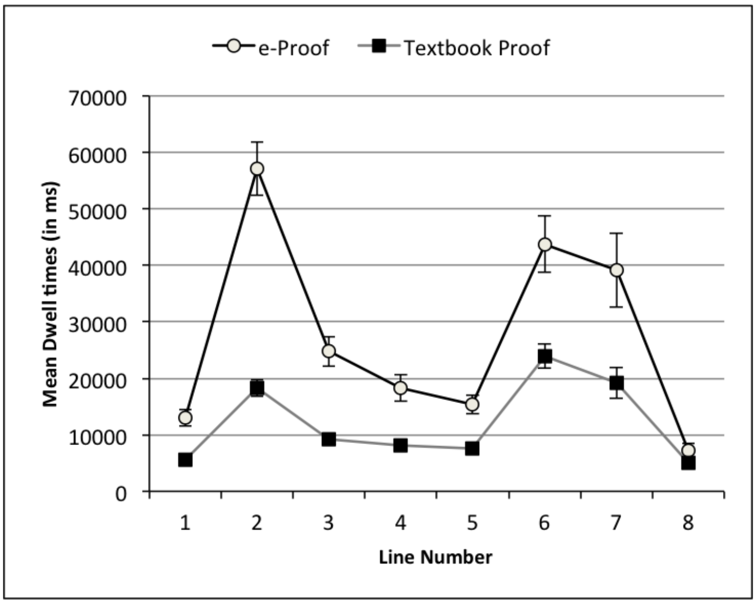


FIGURE 7.4: Dwell times on the each line of Theorem 2. Error bars represent ± 1 SE of the mean.

To analyse undergraduates' attention distribution for Theorem 2 (Th2), with one within-subjects factor (line number: eight lines of the proof) and one between-subjects factor (condition: e-Proof and textbook proof), a 8x2 ANOVA was conducted which showed that the main effect of line number was significant, $F(3.190, 102.086 = 55.422$, $p < 0.001$ (with Greenhouse-Geisser correction). That means that the attention distribution for each line was significantly different. Undergraduates took significantly longer time comprehending the e-Proof of Th2 (the grand mean of dwell time 27.338 s/line) than the textbook proof (the grand mean 12.1 s/line), $F(1, 32) = 34.377$, $p < 0.001$. The interaction effect of line numbers $\times$ types of proof was significant, $F(3.190, 102.086) = 12.34, p < 0.001$. However, a visual inspection of the attention distributions on each line of the proof in two different versions suggested that the distributions of attention between the two versions was approximately equal, with the interaction caused by substantially higher dwell times on line 2 for the e-Proof (see Figure 7.4).

Similarly, a 10x2 ANOVA was conducted to analyse the dwell time for Th3 with one within-subject factor (line number: ten lines of the proof) and one between-subject factor (condition: e-Proof and textbook proof). The main effect of line number was also significant for Th3, $F(4.392, 127.38) = 19.024$, $p < 0.001$ (with Greenhouse-Geisser correction); i.e, time spent on every line was different. The time undergraduates spent to comprehend the e-Proof with grand mean of dwell time 19.12 s/line was significantly longer than the textbook proof with grand

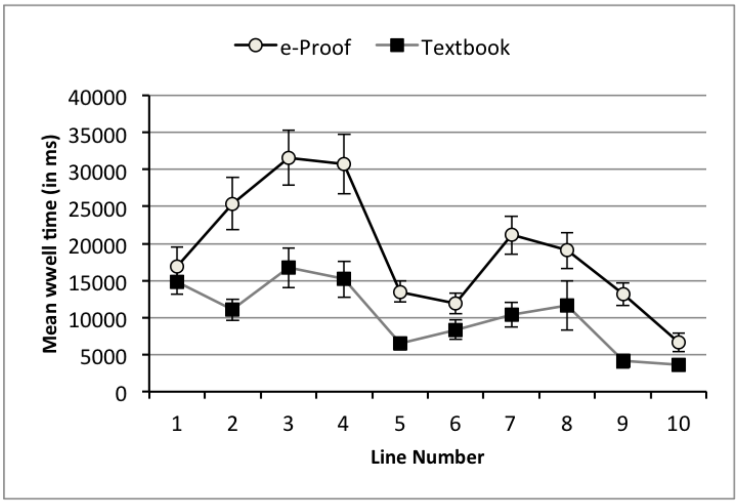


FIGURE 7.5: Dwell times on the each line of Theorem 3. Error bars represent ±1 SE of the mean.

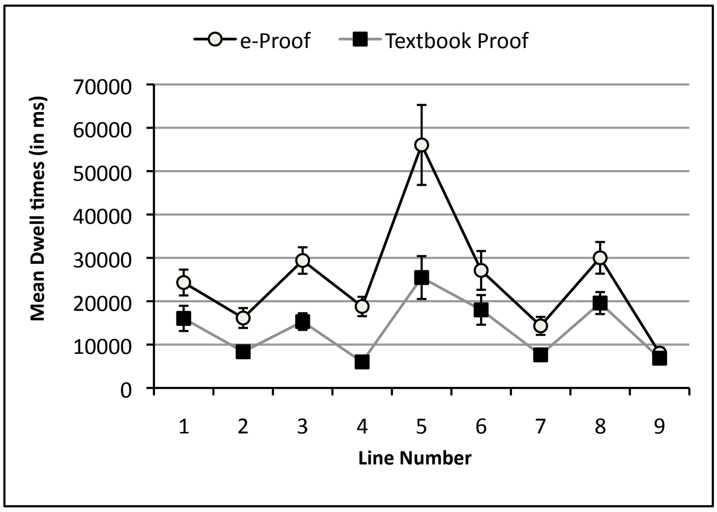


FIGURE 7.6: Dwell times on the each line of Theorem 4. Error bars represent ±1 SE of the mean.

mean of 10.23 s/line, $F(1, 29) = 21.902$, $p < 0.001$. There was a significant interaction effect of line number × condition $F(4.392, 127.38) = 19.024$, $p = 0.013$, though again a visual inspection of Figure 7.5 is showing that there is a reasonable similarities in terms of how attention distributed for the e-Proof and the textbook version of the proof.

A 9x2 ANOVA was conducted for Th4, with one within-subject factor (line number: nine lines of the proof) and one between-subject factor (condition: e-Proof and textbook proof). The result shows that the main effect of line was significant, $F(1.928, 59.78) = 21.515, p < 0.001$ (with Greenhouse-Geisser correction) and also the time undergraduates took for reading the e-Proof was longer (with grand mean of dwell time 24.91 s/line) compared to the textbook proof of Th4 (with grand mean of 13.69 s/line). The interaction effect of line number × condition was significant $F(1.928, 59.78) = 21.515, p = 0.038$, but if line 5 is excluded the interaction effect was no longer significant (with $p < 0.05$) which indicates that the attention distribution throughout the proof for the e-Proof and the textbook proof are quite similar except at Line 5 (see Figure 7.6). In the audio explanation of Line 5 of e-Proof, it was advised to readers to check the calculation of "$h(a) = f(a) - \frac{f(b) - f(a)}{g(b) - g(a)} g(a) = \frac{f(a)g(b) - g(a)f(b)}{g(b) - g(a)}$" and that advice was followed by readers as their spending time on in Line 5 jumped hugely in e-Proof. Despite the significant interaction, a visual inspection of Figure 7.6 indicates that the attention allocation across the both versions of the proof was very similar.

In summary, by analysing the dwell time of four proofs, I found that undergraduates spend significantly longer time to read the e-Proofs than the textbook proofs. They took longer time to comprehend those lines in the proofs (in both cases — e-Proof and textbook) which are crucial for the proofs and difficult to understand. However, the way undergraduates' visual attention distribute on both versions of a proof were essentially the same. This suggests that the skills of reading a proof for comprehension which Alcock wanted to provide through e-Proofs was some thing that undergraduates already largely had.

7.4.3 Reading patterns of e-Proofs and textbook proofs

There is limited research about how readers read proofs. However, many researchers suggested that reading mathematics is a special type of skill that one needs to learn (Davis et al., 1995; Österholm, 2006; Shepherd, 2005). And one of the reasons is reading mathematics is not linear, rather it needs more forward-backward, left-right directional reading than reading any other ordinary texts, such as a newspaper (Barton et al., 2002; Selden and Selden, 2003; Weber, 2009; Weber and Mejia-Ramos, 2011). Weber and Alcock (2005) argued that to understand a proof readers need to check warrant and to do so readers cannot read a proof in a linear manner. Through this experiment I would be able to check how different reading proofs are compared to reading normal texts such as instructions that participants read before start reading proofs. Reading an e-Proof might be different than reading a textbook proof, because of the design of the e-Proof (with highlights, arrows, boxes and the audio explanations). I am expecting to explore undergraduates' reading patters of e-Proofs and textbook proofs in terms of how their eye-gazes shift from line to line, in either a forward or backward direction. That is, how undergraduates' attention switches between lines, in what order and how frequently for two different proof presentations.

To analyse undergraduates' reading patterns of e-Proofs and textbook proofs, I did a fixation sequence analysis. The fixation sequence represents the series of eye-gazes on a displayed stimulus that can reveal the underlying cognitive strategies by observing the order in which eye-gazes jump from one AOI to another (West et al., 2006). Inglis and Alcock's (2012) study investigated experts and novices' approaches to reading mathematical proofs, in which the fixation sequence analysis was used to explore their reading patterns by following fixation transitions between lines. Their study found that the number of fixations mathematicians made while validating a proof was significantly higher than naive undergraduates.

The aim to use fixation sequence analysis in my study was to reveal if there are any similarities and differences in the way undergraduates' gaze transits from one AOI to another when they read two different proof presentations. That is, the focus was not only to highlight the number of fixations, but particularly how undergraduates' attention switches between lines — either forward and backward direction while reading e-Proofs and textbook proofs. The chances that a reader's eye-gaze might jump more frequently in both directions when reading an e-Proof compared to a textbook proof was higher, because of the design of e-Proofs.

For the fixation sequence analysis, I considered all four proofs together rather than each proof individually. To arrange the fixation sequence for all proofs, I used the eyePattern software (West et al., 2006) to get the sequences that only highlight the order in which a reader looked at AOIs. For example, a sequence would look like '12XX223X445555553538X8X597788', where 1,2, 3, 4, 5, 7, 8, and 9 represent eye-gazes that fall into those pre-defined AOIs for Theorem 4 and those eye-gazes which did not fall into any defined AOIs are represented by X. A sequence 12X indicates that eye gaze jumps from AOI 1 (i.e, Line 1) to AOI 2 ((i,e, Line 2) to an undefined area. When eye-gaze shifts from Line 1 to Line 2 and onwards that is referred as forward directional reading in this thesis and backward directional reading when eye-gaze jumps Line 2 to Line 1.

First, a between-line transition matrix was made from the fixation sequence for each participant for individual proof. Figure 7.7 represents a transition matrix

	A	B	C	D	E	X	Total
A	0	1	0	0	0	5	6
B	4	0	5	9	1	17	36
C	0	8	0	4	0	7	19
D	1	7	5	0	6	5	24
E	0	4	0	6	0	8	18
X	1	17	9	5	11	0	43
Total	6	37	19	24	18	42	146

FIGURE 7.7: An example of between-line transition matrix for Theorem 1 for an individual participant. A, B, C, D and E represent the AOIs and X represents undefined area, the numbers represent the number of fixations occurred while reading the proof.

based on a participant's fixation sequence for Theorem 1. Then I calculated the forward directional and backward directional between-line transitions for all participants for all four proofs. Since individual participant took different times to comprehend the given proofs, I considered the number of fixations per second for the analysis.

To explore the effect of forward and backward directional gaze transitions between-lines on different proof presentations, a 2x2 ANOVA was conducted with one within-subjects factor type of proof presentations (with two levels: e-Proofs and textbook proofs) and another within-subjects factor direction of gaze transition (with two levels: forward and backward directional gaze transitions). The main effect of type was significant, $F(1,31) = 17.09, p < 0.001$, which indicates that the number of between-line transitions for when undergraduates read e-Proofs (with average transition 0.45 per sec) was significantly higher than when they read textbook proofs (average transition .37 per sec). The main effect of direction was also significant, $F(1,31) = 49.697, p < 0.001$, which indicate that the number of forward directional transitions was significantly higher than backward.

However, there was a significant type×direction interaction effect, $F(1,31) = 17.344, p < 0.001$. I conducted post-hoc paired t-tests to compare the number of fixations in forward and backward directional between-line transitions for e-Proofs and textbook proofs. The result shows that there was no significant difference in forward (average .23 per sec) and backward (average .22 per sec) directional reading when undergraduates read e-Proofs, $t(31) = 1.799, p = 0.082$; however for textbook proofs the number of forward directional transitions (average .16 per sec) was significantly higher than backward (average .21 per sec). These outcomes explain significant the interaction effect of type and direction (see Figure 7.8).

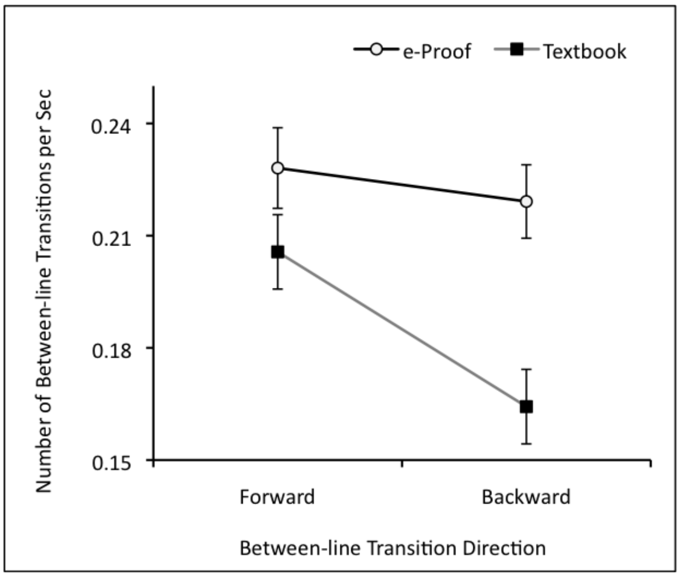


FIGURE 7.8: Between-line fixation transitions for e-Proofs and textbook proofs. Error bars represent ±1 SE of the mean.

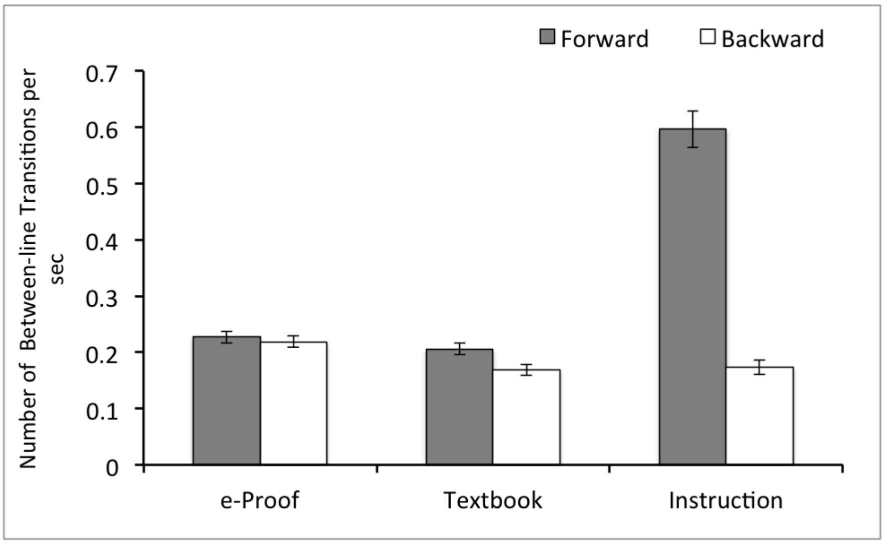


FIGURE 7.9: Between-line fixation transitions for e-Proofs and textbook proofs. Error bars represent ±1 SE of the mean.

Comparing reading patterns of instructions with proofs

I conducted a sequence analysis to explore how undergraduates read a page of instructions given during the experiment. Their instruction reading was compared with how they read e-Proofs and textbook proofs in terms of the between-line fixation transitions. A 3X2 ANOVA was conducted with two within-subjects factors type (with three levels: e-Proofs, textbook proofs and instructions) and direction (with two levels: forward and backward directional fixation transitions). The effect of type was significant, $F(1, 28) = 62.515, p < 0.001$ which indicates that the number of fixations while reading instruction was different from reading proofs. The effect of direction was also significant, $F(1, 28) = 292.866, p < 0.001$ which suggests that the tendency of forward directional reading was significantly higher than backward. The type×direction interaction effect was also significant $F(1, 56) = 182.236, p < 0.001$. To explain this interaction a paired t-test was conducted, which showed that when undergraduates read instructions the forward directional gaze transitions (mean 0.597 per sec) was significantly higher than backward (mean 0.173 per sec) compared to the proofs. Figure 7.9 shows a comparison of forward and backward directional reading between the e-Proof, textbook proof and instruction condition.

Mathematicians reading of mathematical proofs

This section has shown the difference between reading e-Proofs and reading textbook proofs. During the interview with Alcock (which I reported in Chapter 3), Alcock mentioned that e-Proofs display how expert mathematicians read and comprehend proofs, which allows undergraduates to grasp the skill of reading proofs. Therefore, the way mathematicians read proofs should be similar to the way undergraduates read e-Proofs, but not the way they read textbook proofs. To justify the above argument, I re-analysed Inglis and Alcock's (2012) data to get an insight of mathematicians reading patterns of proofs in terms of how their eye-gazes move between lines while reading a proof.

I considered two proofs — a long proof and a short proof to analyse if there is any difference in the forward and backward directional between-line transitions when mathematicians read proofs. Conducting a paired t-test for the long proof revealed that the forward directional between-line transitions are significantly higher than backward, $t(11) = 3.171, p = .009$. Similarly, for the short proof, a paired t-test showed that the forward directional transitions were significantly higher than backward, $t(11) = 3.232, p = .008$. These results suggest that while reading and comprehending proofs, mathematicians read in both directions forward and backward, but the tendency to read forward was significantly higher than reading backward, which was not the case for reading e-Proofs.

I plotted the ratio of forward and backward directional between-line transitions in four conditions — e-proof, textbook proof, instruction and mathematicians' reading of proof, to show the linearity of the reading processes of these conditions

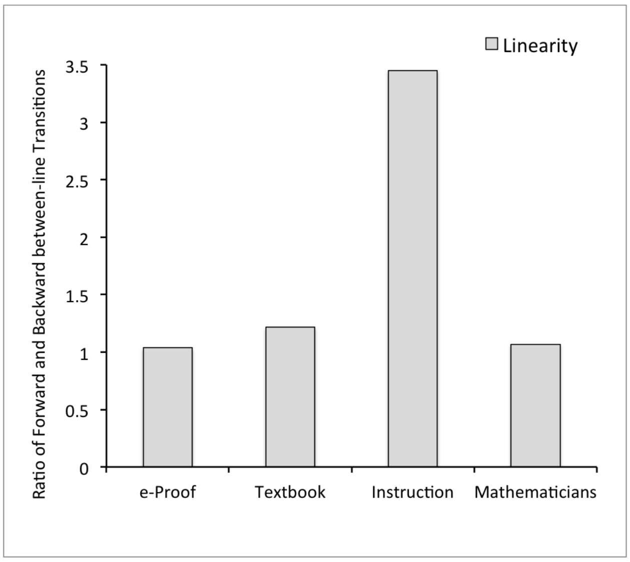


FIGURE 7.10: Figure shows the ratio of forward and backward directional between-line transitions, linearity of the reading processes under four conditions.

(see Figure 7.10). The results again show that reading mathematics (whether this is experts reading proofs, novices reading proofs, or novices reading e-Proofs) is substantially less linear than reading standard written English.

In sum, Section 7.4.3 showed that to comprehend proofs, readers need both forward and backward reading which is significantly different from reading a text, a page of instruction for example, which mainly demands forward directional reading (see Figure 7.10). Reading e-Proofs however were significantly different from reading textbook proofs. When undergraduates read e-Proofs, the number of between-line transitions in forward and backward direction were essentially same, however the forward directional transitions were significantly higher than backward in both conditions — when undergraduates read textbook proofs and when mathematicians read proofs.

7.4.4 First fixations on each line of e-Proofs and textbook proofs

Analysing the fixation sequence of e-Proofs and textbook proofs showed that the number of between-line transitions was significantly higher in the e-Proof condition than the textbook condition. This also suggests that the distractions while undergraduates read e-Proofs may be higher compared to when they read textbook

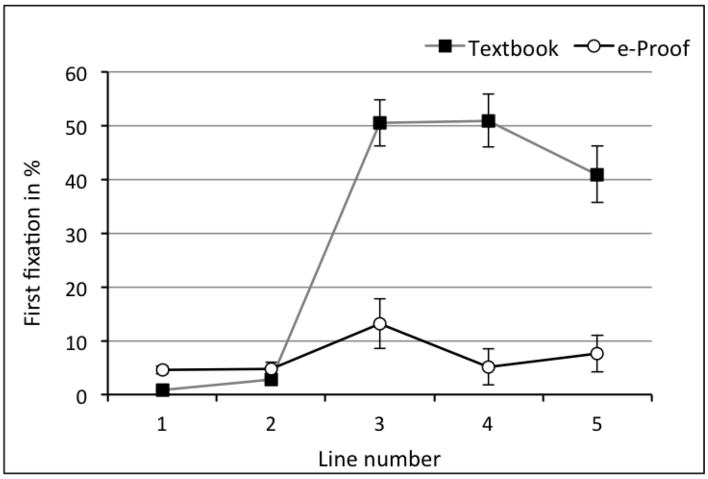


FIGURE 7.11: Figure shows the first fixation time (in % of the total time spent reading the proof) of each line of Theorem 1 in e-Proof and textbook conditions. Error bars represent ±1 SE of the mean.

proofs. The design of e-Proof may force undergraduates to fixate all over the proof more frequently than usually the way they read a textbook proof. In the case of e-Proofs, this distraction may cause weaker long term impact on undergraduates' proof comprehension, which I found in Experiment 1.

To investigate whether the design of e-Proof forces undergraduates to fixate on each line of a proof more rapidly compared to when they read a textbook proof, I calculated the time at which they first fixated on each line of each proof. Since there was a difference in how much time spent on each proof, I considered the percentage of their overall time spent comprehending each proof to check whether reading e-Proofs is a rapid reading process, whereas reading textbook proofs is a slow process.

For Theorem 1, the mean percentage of time spend to the first fixation of the last line of Th1 was 7.5% for e-Proofs and 40% for textbook proofs (see Figure 7.11).

For Theorem 2, the mean percentage of time spend to the first fixation of the last line of Th2 was 38% for e-Proofs and 45% for textbook proofs (see Figure 7.12).

For Theorem 3, the mean percentage of time spend to the first fixation of the last line of Th3 was 42% for e-Proofs and 60% for textbook proofs (see Figure 7.13).

For Theorem 4, the mean percentage of time spend to the first fixation of the last line of Th4 was 41% for e-Proofs and 60% for textbook proofs (see Figure 7.14).

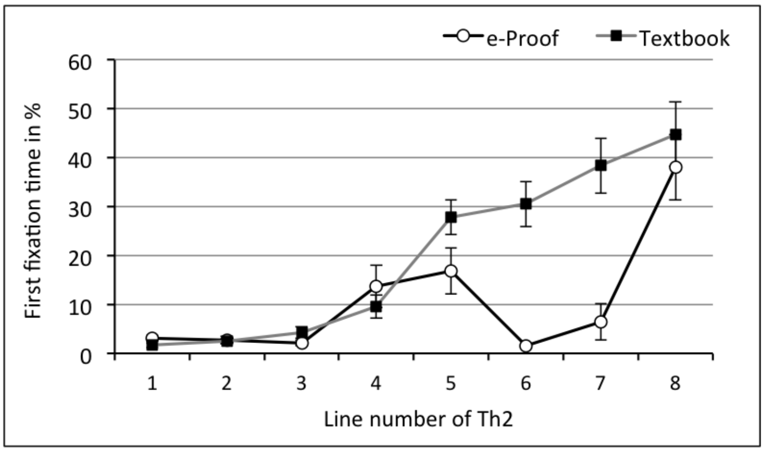


FIGURE 7.12: Figure shows the first fixation time (in % of the total time) of each line of Theorem 2 in e-Proof and textbook conditions. Error bars represent ±1 SE of the mean.

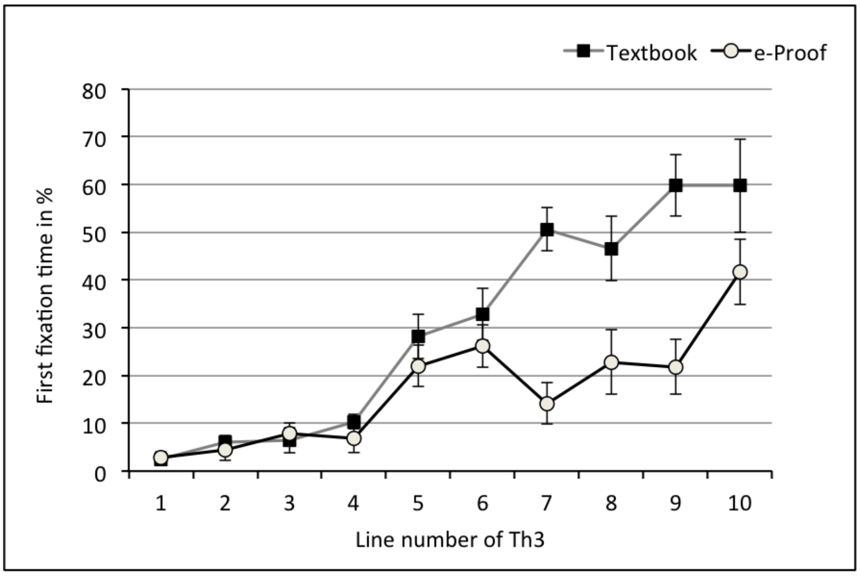


FIGURE 7.13: Figure shows the first fixation time (in % of the total time) of each line of Theorem 3 in e-Proof and textbook conditions. Error bars represent ±1 SE of the mean.

Overall considering all four proofs, the first fixation of last line of the e-Proof version was significantly different from the textbook condition, $t(33) = -2.291, p = .028$. This suggests that while reading e-Proofs eyes moved more rapidly to each line of the proofs, whereas reading textbook proofs was a comparatively slow process. The key reason behind this rapid reading process of e-Proof is its design — highlighted lines, audio explanations etc force readers to read a proof in a way that they might not typically do when reading textbook proofs.

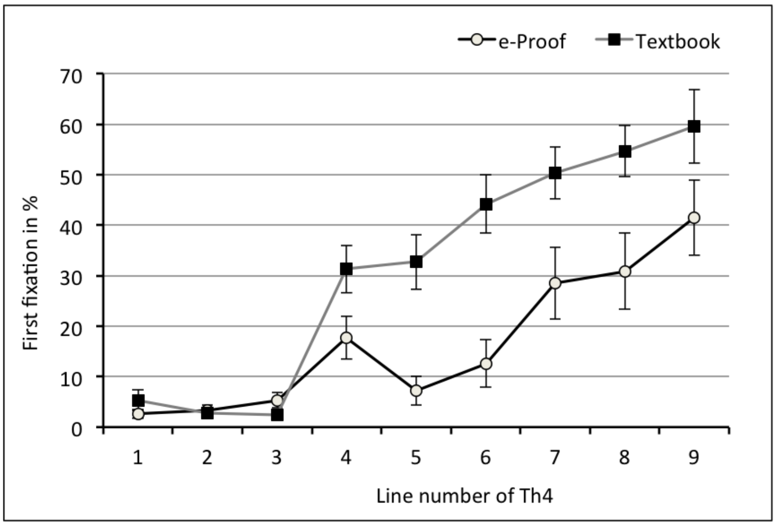


FIGURE 7.14: Figure shows the first fixation time (in % of the total time) of each line of Theorem 4 in e-Proof and textbook conditions. Error bars represent ±1 SE of the mean.

7.4.5 Processing effort for comprehending e-Proofs and textbook proofs

Based on the outcomes of Experiment 1, I made an argument earlier in Chapter 5 that e-Proofs are easier to comprehend for undergraduates compared to written textbook proofs. This means, to comprehend an e-Proof the processing effort needed is less than what is needed to comprehend a textbook version of the proof. To establish the above conjecture, I compared the efforts that undergraduates put to comprehend an e-Proof and a textbook proof by analysing the *mean fixation duration*. The longer fixation duration indicates that readers have difficulties in processing the information (Just and Carpenter, 1980; Poole, 2005). Therefore, I was expecting to find that *the mean fixation duration while reading an e-Proof is significantly lower than the mean fixation duration when undergraduates read a textbook proof*.

A 5x2 ANOVA was conducted to analyse the mean fixation duration for Theorem 1 with one between-subject factor (condition: e-Proof and textbook proof) and one within-subject factor (line number: five lines for Th1). The results showed that the main effect of line number was significant, $F(3.16, 97.958) = 40.464, p < 0.001$, which suggests that the efforts needed to comprehend different lines of the proof were significantly different based on their level of difficulties. But no significant main effect of the condition was found, $F(1, 31) = 0.32, p = 0.576$. This indicates

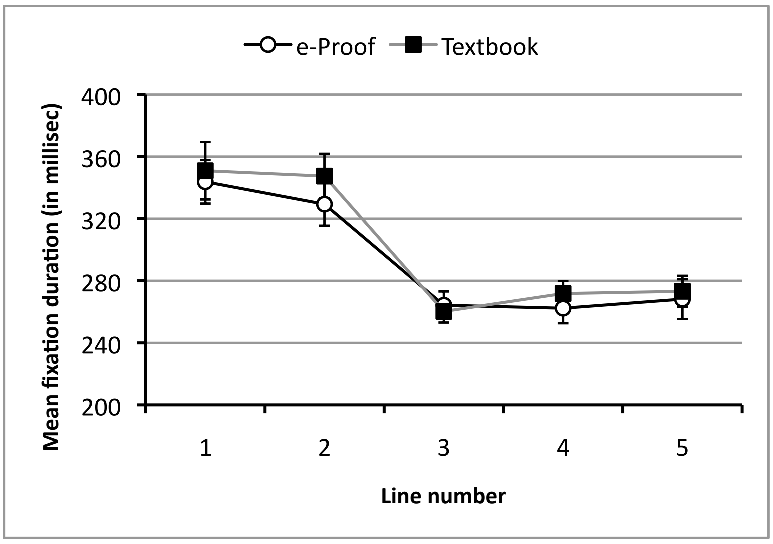


FIGURE 7.15: Mean fixation duration (in milliseconds) on the each line of Theorem 1. Error bars represent ±1 SE of the mean.

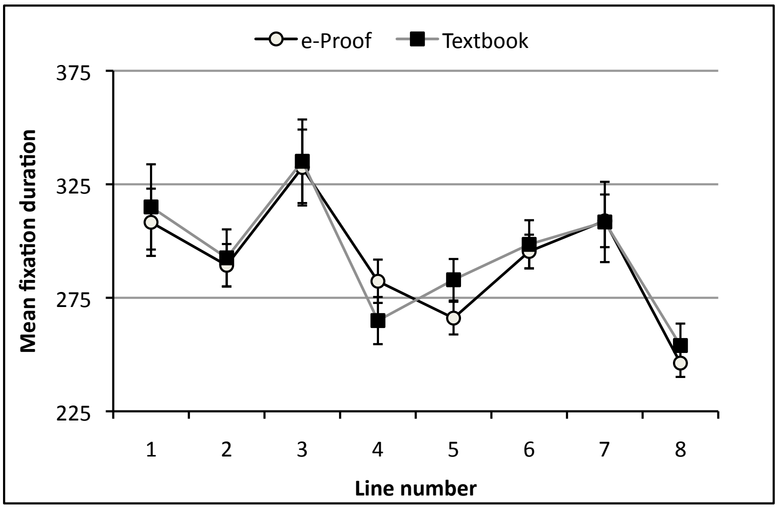


FIGURE 7.16: Mean fixation duration (in milliseconds) on the each line of Theorem 2. Error bars represent ±1 SE of the mean.

that processing effort required for comprehending e-Proof of Th1 (with grand mean of fixation duration 294 ms) and the textbook version (with grand mean of 301 ms) were essentially the same (see Figure 7.15). The interaction of line number × condition was not significant for Th1, $F(3.16, 97.958) = 0.364, p = 0.789$.

For Theorem 2, a 8x2 ANOVA was conducted with one within subject factor (line number: eight lines (AOIs) of the proof) and one between-subject factor (condition: e-Proof and textbook proof). A significant main effect of line number was found $F(4.35, 139.193) = 14.001, p < 0.001$. I also found the processing effort for e-Proof of Th2 (with grand mean of fixation duration 291 ms) was not

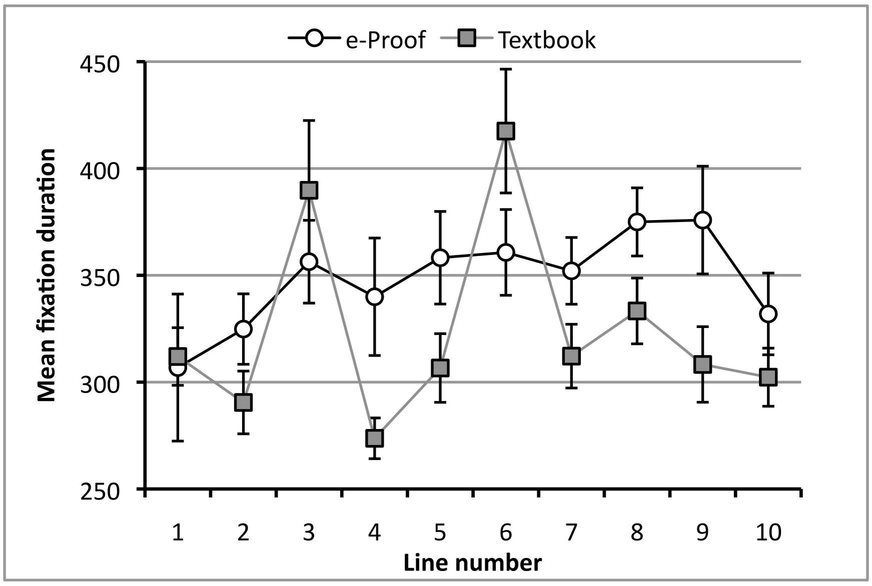


FIGURE 7.17: Mean fixation duration (in milliseconds) on the each line of Theorem 3. Error bars represent ±1 SE of the mean.

significantly different from the textbook proof (with grand mean of 294 ms) as I found the main effect of condition was not significant ($F(1, 32) = 0.057, p = 0.813$ (see Figure 7.16). The interaction of effect of line number × condition was not significant for Th2, $F(4.35, 139.193) = 0.475, p = 0.77$.

For Theorem 3, a 10×2 ANOVA was conducted with one within-subject factor (line number: ten lines of the proof) and one between-subject factor (condition: e-Proof and textbook proof). The main effect of line number was significant for Th3, $F(5.299, 153.658) = 4.356$, $p = 0.001$ (with Greenhouse-Geisser correction); which means the effort needed to process different lines are significantly different. However, no significant main effect of condition was found, therefore the processing effort across the e-Proof (grand mean 349 ms) and the textbook version (grand mean 325 ms) was not different $F(1, 29) = 1.459$, $p = 0.237$ (see Figure 7.17).

A similar result was found for Theorem 4. A 9×2 ANOVA was conducted with one within-subject factor (line number: nine lines of the proof) and one between-subject factor (condition: e-Proof and textbook proof), with a significant main effect of line number $F(4.616, 143.086) = 4.163$, $p = 0.002$ (with Greenhouse-Geisser correction) but no significant main effect of condition $F(1, 31) = 0.545$, $p = 0.466$. These results suggested that undergraduates' processing efforts for comprehending different lines were significantly different in both versions of the proof; however the effort needed to comprehend the e-Proof version (grand mean

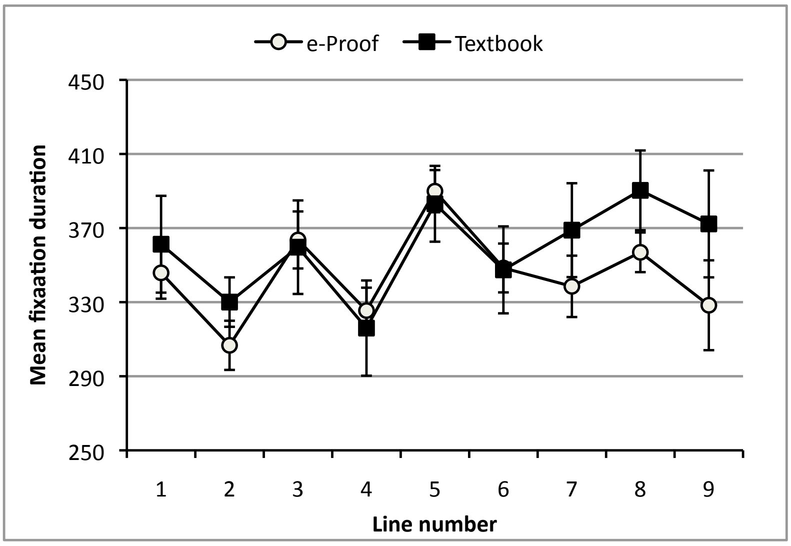


Figure 7.18: Mean fixation duration (in milliseconds) on the each line of Theorem 4. Error bars represent ±1 SE of the mean.

345 ms) and the textbook version (grand mean 359 ms) of Th4 was the same (see Figure 7.18).

In sum, the efforts undergraduates put to comprehend an e-Proof version of a proof and a textbook version were essentially the same. This finding does not support the conjecture set after Experiment 1.

Comparing the processing effort when audio explanation is on and off in e-Proofs with the processing effort in textbook proofs

Audio explanation is a key feature of an e-Proof. The audio explanations provide information to guide readers about how a proof works, that is, this audio feature directly helps readers to comprehend a proof which undergraduates often might not be able to do themselves (Alcock, 2009). Therefore, it was vital to observe how much effort undergraduates put in to comprehend a proof themselves without getting help from audio explanations provided in an e-Proof. For that reason, I investigated how much effort undergraduates put to comprehend the written texts of an e-Proof when they were not listening to the audio explanations and when they were listening to the audio explanation. For the rest of the thesis, I refer to “audio-off” as *when undergraduates read the texts of the e-Proofs without listening to the audio* and “audio-on” as *when undergraduates read the texts of the e-Proofs while listening to the audio explanations.* My conjecture is that the effort needed to comprehend e-Proofs when audio-off (in terms of the mean fixation

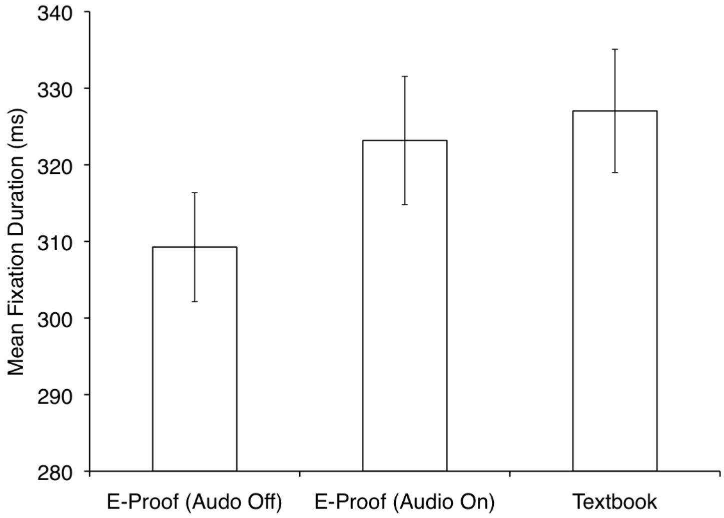


FIGURE 7.19: Mean fixation duration when undergraduates read e-Proofs with audio-on, audio-off and textbook proof. Error bars represent ±1 SE of the mean.

duration) should be significantly higher than when audio-on if undergraduates tried to comprehend a proof themselves.

To investigate how much effort undergraduates put to comprehend e-Proofs, the total time that undergraduates spend reading all e-Proofs was divided into two parts — when undergraduates read the texts of the e-Proofs while listening to the audio explanations (audio-on) and when they read the written textual parts of e-Proofs without listening to the audio (audio-off). For this analysis, I consider all the four theorems together rather than analysing each theorem separately, because this audio-on and audio-off are common features of every e-Proof.

An ANOVA was conducted with one within-subjects factor with three levels — audio-on and audio-off for e-Proofs and textbook proofs. This showed that there was a significant difference in the mean fixation duration for these three different conditions $f(2, 50) = 3.267, p = 0.046$. To get more insight, Bonferroni-corrected paired t-tests were conducted taking any two out of three conditions. The t-test showed that the mean fixation duration was significantly higher when undergraduates read textbook proofs compared to e-Proofs with audio-off, $t(33) = -2.992, p = .005$. This result indicates that the effort undergraduates put to comprehend written textbook proofs was significantly higher than when they read the texts of the e-Proofs without listening to the audio explanations. There was no significant difference that how much effort undergraduates put to comprehend textbook proofs and e-Proofs with audio-on condition, $t(25) = -.44, p = .664$. However the effort undergraduates put for reading e-Proofs when audio-on was significantly higher than when audio was off, $t(25) = -2.603, p = .015$. This suggests that undergraduates devote significantly less effort when they read the written texts of e-Proofs without listening to the audio explanations compared to other two conditions. This also indicates while comprehending an e-Proof, they relied on the given audio explanations rather than trying to comprehend the proof themselves. Figure 7.19 shows the mean fixation duration for all four proofs under the three conditions.

7.5 Discussion

Experiment 2 revealed three crucial points:

1. Undergraduates were apparently capable of allocating their attention in a manner similar to that Alcock desired, as their visual attention allocation when reading a textbook proof was similar to that when reading an e-Proof.

2. When an e-Proof version of a proof is available to undergraduates, they took a sit-back approach — they tried to comprehend the proof by following the given audio explanation rather than trying to comprehend it themselves. Listening to the given audio explanations might give undergraduates a sense of understanding but that understanding might not last for long period of time.

3. The design of e-Proofs forces its readers to adopt a different reading strategy — frequent and high number of gaze fixations than usually needed to comprehend a proof. This is a sign of distraction that happens during the learning process of e-Proofs.

The first point, therefore, raises the question whether undergraduates need e-Proofs at all. In this experiment, I did not find any evidence that can support Alcock's (2009) claim that undergraduates often do not have the skills of reading and comprehending proofs which was one of the key reasons of presenting e-Proofs to them (that is, directly help undergraduates in grasping those skills required to comprehend proofs) seems unreasonable.

The experiment showed that when undergraduates read e-Proofs they mostly relied on given audio explanations to comprehend a proof rather than putting effort to comprehend the proof themselves, even though they were capable of comprehending a proof themselves. Whereas, when they read textbook proof, they were forced to construct explanations themselves as no additional explanation is given. This indicates that even though undergraduates' level of understanding of a proof was essentially the same whether they read e-Proofs or textbook proofs, the nature of the activities involved to comprehend a proof are different. For e-Proofs undergraduates comprehend the given audio explanations, whereas to comprehend a textbook proof they have no choice but to self-explain. Research has shown that when students self-explain to comprehend a given text, they understand better than those who do not (Chi et al., 1994). And the activity of self-explanation has better impact on readers in terms of learning rather than when explanations are provided to comprehend a text (discussed in detail in Chapter 5). Schnotz and Rasch (2005) found that providing additional help through multimedia resources does not always have a positive impact on learning. Perhaps providing additional help makes the process of learning easier and thus students put less effort than they would otherwise, which leads to poor learning outcomes. The authors called it the facilitating effect which gives students a sense of understanding which is shallow in nature and does not create a knowledge that can last for a long period of time.

Giving additional explanations, highlighted links, boxes and arrows in an e-Proof facilitates the proof comprehension process and gives undergraduates a sense of understanding of the proof they are reading. But by not putting their own effort to infer how a proof works, they might not grasp the level of understanding which can be achieved by generating their own explanations. During the interview, Alcock also mentioned that undergraduates get benefit out of an e-Proof, if they try to comprehend a proof themselves first rather than listening to the given explanations (reported in Chapter 3).

Not only did undergraduates rely heavily on audio explanations for comprehending e-Proofs, their behaviour changed significantly when they read an e-Proof compared to when they read a textbook proof.

To explain why e-Proofs are less effective than traditional written textbooks especially if we consider its long term learning impact, I refer back to Mayer's (2001) theory of learning from multimedia resources. According to Mayer, there are

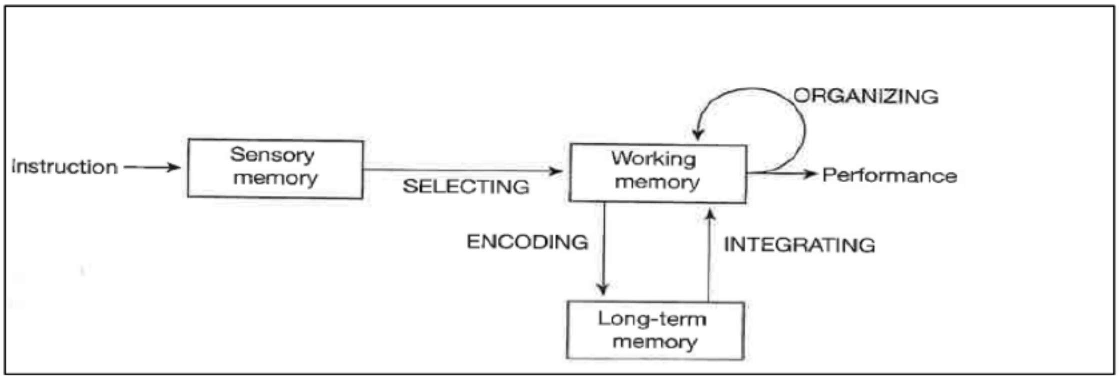


FIGURE 7.20: Mayer's (2001) model on cognitive theory of multimedia learning.

three stages of learning — selecting new information in audio and visual form, organising information to construct a coherence representation and integrating new information with the existing knowledge to create new knowledge. When undergraduates read an e-Proof, they engaged with multiple tasks — reading the written texts, listening to the audio, following the highlighted arrows and boxes, clicking the mouse for moving one slide to another and for listening to the audio. Because of these multiple tasks not only does reading e-Proofs need a longer time than reading textbook proofs (the dwell time results in Section 7.4.2), but they presumably also interrupt the organising process during the learning. This means that the new information needs to be structured in a coherent manner, but the frequent interruptions could distract that organising process. Moreover, the process of organising new information occurs in our working memory which has a limited capability to work with various tasks. While reading an e-Proof the continuous interruptions and load of performing multiple tasks in a limited memory space could plausibly prevent the new information from being formed in a structured way and thus this ill-structured information could not integrate with existing knowledge to create new knowledge. When new knowledge is constructed poorly, readers often struggle to reuse that knowledge, especially after a time gap. This account is therefore consistent with the poor long term impact of reading e-Proofs as I found in Experiment 1.

7.6 Summary of Chapter 7

- This chapter reports an experiment which was conducted to explore the similarities and differences that exist when undergraduates read e-Proofs and textbook proofs. A within-subjects research design was adopted for the study. Undergraduates eye-movements were observed by using an eye-tracker when they were engaged in the task of reading and comprehending proofs. They read two textbook proofs and two e-Proofs during the experiment, and they were asked to answer a multiple choice comprehension test after each reading attempt.

- The outcomes of the comprehension tests of Experiment 2 were similar to what I found in Experiment 1. That is, the immediate impact of reading e-Proofs or textbook proofs on undergraduates' proof comprehension was essentially the same.

- Alcock (2009) argued that undergraduates often do not have the skills of reading and comprehending proofs and that e-Proofs can directly help in grasping those skills. However, analysing the dwell times for each proof, I have found that the attention allocations across both versions of proofs are similar, which suggests that undergraduates know which aspects of proofs to attend to without any additional guidance such as that provided in e-Proofs. Therefore, it raises the question whether undergraduates need e-Proofs at all.

- Moreover, the design of the e-Proof demands significantly different reading styles in terms of more frequent and rapid saccades between lines compared to reading textbook proofs. This could potentially cause distractions. It may be these distractions while reading e-Proofs have a negative consequence in the form of poor long term impact on undergraduates' proof comprehension, which was found in Experiment 1.

- Analysing the mean fixation durations, I found that the amount of effort undergraduates put to comprehend an e-Proof and a textbook proof were essentially the same. However, when undergraduates read e-Proofs they put significantly more effort to comprehending the given audio explanations and less effort to read the textual parts of the e-Proofs without listening to the audio. This finding suggests that even though undergraduates were capable of comprehending a proof themselves, when they read e-Proofs they relied on given audio explanations rather than trying to comprehend themselves. In sum,*for e-Proofs undergraduates comprehend the given audio explanations* which explains how the proof works; whereas *to comprehend a textbook proof they have no choice but to construct explanations for themselves.*

- When undergraduates read textbook proofs no additional help is available and they have to generate their own explanations to comprehend a proof and this effort has a positive impact on proof comprehension in the long term.

Chapter 8

Conclusion

8.1 Introduction

This thesis is a single case study of the introduction and evaluation of new resources in higher education aiming to improve learning, where the computer-base *e-Proof* was chosen as the single case. More precisely this thesis aimed to explore the impact of reading e-Proofs on undergraduates' proof comprehension compared to conventional textbook proofs and what makes e-Proofs effective (or less effective).

The aim of this concluding chapter is to summarise the findings reported in Chapter 3, Chapter 5 and Chapter 7, and to discuss the research implementation of this thesis and reflect upon different methods of evaluating a new resource.

8.2 Review of findings

Chapter 3 primarily addressed two issues related to e-Proofs — (i) what e-Proofs are for and (ii) what undergraduates like and dislike about e-Proofs, and why. To investigate the above research questions, I used self-reporting research strategies — interviews and surveys. A semi-structured interview with Alcock was used to investigate what e-Proofs were designed to achieve. Three focus group interviews and a web-based survey were used to explore undergraduates' likes and dislikes about e-Proofs.

In Alcock's view, the e-Proof provides a 'model' to guide undergraduates in what they should be doing and thinking when reading each line of a proof. She believed that even the weaker students (who have a lack of knowledge of how to read a proof) should get benefit from the additional explanatory information attached with each line of e-Proofs. However she mentioned that e-Proofs need to be used intelligently, otherwise "there is no point just sitting there and watching through something [e-Proof]".

The web-based survey suggested that undergraduates had extremely favourable views on e-Proofs (significantly more than 75% satisfaction levels). Their views on e-Proofs were not correlated with the need for cognition scale, which suggests that e-Proofs are not viewed favourably by only those who enjoy difficult cognitive tasks.

In the focus group interviews, undergraduates shared positive views about e-Proofs and they liked to have e-Proofs as an extra resource. They seemed to believe that e-Proofs 'definitely' had an impact on their learning and that experiencing e-Proofs made them more confident about proofs. However, they admitted not using e-Proofs regularly during their private studies. Some of them had only seen e-Proofs during lectures and did not use e-Proof in their own time at all. Having e-Proofs seemed to give them a sense of security — that e-Proofs are always there for help whenever needed. It was clear that the self-reporting research methods certainly reflected undergraduates' positive feeling towards e-Proofs, but that did not constitute evidence of whether e-Proofs are educationally beneficial or not.

Chapter 5 reported the first experiment which compared the (immediate and long-term) impact of reading an e-Proof version and a textbook version of the same proof on undergraduates' proof comprehension. Participants in the experiment were randomly divided into two groups — e-Proof group and textbook group. The e-Proof group read an e-Proof version of Cauchy's Generalised Mean Value theorem (GMVT) and the textbook group read the textbook version of that same proof. Both groups then answered a written comprehension test twice — first immediately after reading the proof (post-test) and after two weeks (delayed post-test).

The outcomes of the experiment showed that there was no difference in the performance of the e-Proof group and the textbook group, both groups, however, performed better in the post-test than the delayed post-test. The e-Proof group had a significantly worse drop off in their performances in the delayed post-test compared to the post-test than the textbook group. In other words, those who read the e-Proof were worse at retaining their learning than those who read the textbook version of the proof. This indicates weaker long term impact of e-Proof on learning when compared with the textbook proof reading.

Moreover, high-achieving undergraduates[1] and low-achieving undergraduates both had significantly worse retention in the delayed post-test scores compared to the post-test when they read the e-Proof than those who read the textbook version of the proof. Alcock mentioned during the interview that low achieving students might get more help from e-Proof directly, whereas high-achieving undergraduates might grasp the skill of reading proofs the way expert mathematicians read proofs (reported in Chapter 3). Clearly, the findings did not support what Alcock claimed regarding the help that e-Proofs can offer to both high and low achieving undergraduates.

In sum, the long term impact of reading an e-Proof on both the high-achieving and the low-achieving undergraduates was poorer compared to reading a written textbook proof. This raised the question of 'why' the long term impact of an e-Proof was worse than the textbook proof, even though e-Proofs are specially designed to make the task of proof comprehension easier for undergraduates.

In the next experiment reported in **Chapter 7**, I investigated how undergraduates read e-Proofs and textbook proofs in order to find an explanation for why e-Proofs have poorer long term impact on undergraduates' proof comprehension compared with textbook proofs. A within-subjects research design was used for this experiment in which each participant read both e-Proofs and textbook proofs. Participants' eye-movements were recorded while reading to explore what similarities and differences exist in the reading comprehension processes associated with these two different proof presentations.

Firstly, dwell time analyses revealed that undergraduates took significantly longer to comprehend an e-Proof version of a proof compared to its textbook version, however the visual attention allocation across both versions of proofs were very similar; which suggested that undergraduates had a sense of where to focus their attention when reading a proof even if no additional help was available. The main

[1]In the sense of achievement in exam, here high (low) achieving undergraduates refers towards those undergraduates who achieved high (low) scores in the final exam of the real analysis module

motivation for Alcock to create e-Proofs was her presumption that undergraduates need help for reading and comprehending proofs (Alcock, 2009). However, the eye-tracking study did not find any evidence that supports Alcock's presumption regarding proof reading, rather it raises the question whether undergraduates need e-Proofs at all.

Dwell time analyses showed undergraduates were capable of reading and comprehending proofs themselves, however when they read e-Proofs they behaved as passive readers. Fixation duration analyses of e-Proofs (audio-on[2] and audio-off) showed undergraduates devoted significantly more effort while listening to the given audio explanations than when the audio was not played. Clearly, they put effort for understanding the given audio explanations of e-Proofs, but put significantly less cognitive effort for generating their own explanations to understand how a proof works. It is important to note that the activities involved in understanding given audio explanations and generating own explanations in order to understand how proofs work are different. The additional help provided in e-Proofs allowed undergraduates to take a back seat and it clearly did not promote active learning which is always appreciated and encouraged by teachers and educators.

The eye-tracking experiment provides evidence that e-Proof not only let undergraduates took an easier way to comprehend proofs by focusing on audio explanations, it was successful in altering undergraduates' reading behaviours in a way which can harm learning. The sequence analysis showed that, while reading e-Proofs, eyes jumped between lines significantly more than when they read textbook proofs. This was driven by a significant increase in backward transitions. The number of forward between-line transitions was significantly more than backward in case of reading textbook proofs, whereas the number of forward and backward directional transitions were the same for reading e-Proofs. This outcome was also supported by the analysis of the first fixation of each line, which showed while reading e-Proofs eyes move rapidly to each line of the proofs, whereas reading textbook proofs was a comparatively slow linear process. The key reason behind this rapid reading process of e-Proof is its design — highlighted lines, audio explanations etc force readers to read a proof in a way that they might not typically do when reading textbook proofs.

[2]Audio-on indicates those time period when undergraduates played the audio while reading e-Proof and audio off indicates those time period when they read e-Proofs without listening to the audio.

These factors suggest that the knowledge undergraduates gained from e-Proofs may have been constructed poorly and this was reflected in the delayed post-test in the first experiment. Mayer (2001) hypothesised that learning involves three cognitive processes — *selecting* new information in audio and visual form in our sensory memory, *organising* received information to construct a coherence representation in our working memory and *integrating* new information with the existing knowledge to create knowledge which then stores in our long term memory. But while reading e-Proofs the organising process might get over crowded with multiple tasks — reading the written texts, listening to the audio, following the highlighted arrows and boxes, clicking the mouse for moving one slide to another and for listening to the audio (evidence of high number of transition fixations) and as a result new information might not be structured coherently. Integrating such ill-structured information with existing knowledge can only create poor knowledge which has weak retention potential.

One of Alcock's (2009) aims was to show undergraduates how experts read proofs for comprehension. By creating e-Proofs, Alcock wanted to draw undergraduates' attention into particular parts of a proof while reading each line, which she believed is what mathematicians do for comprehending proofs. However, when mathematicians read proofs the between-line transitions in forward directions were significant more than backward transitions. Moreover, regarding between-line transitions, reading textbook proofs are more similar to the way mathematicians read proofs. It seems that mathematicians might not read the way Alcock assumed they do and perhaps there is no need to go back to previous lines so frequently. Such frequent transitions interrupt the reading process and that can affect learning. These findings question whether the design of e-Proofs demonstrates the skill an expert reader possesses efficiently or not.

Regarding text reading and reading mathematics, many researchers suggested that reading mathematics is different from reading regular text (Davis et al., 1995; Österholm, 2006). This experiment shows evidence that indeed reading proofs is significantly different from reading a text, in terms of how frequently eyes move between lines while reading texts and proofs. Reading a page of instruction for example mainly demands forward directional linear reading; whereas reading proofs requires both forward and backward non-linear reading.

In sum, the second experiment showed that undergraduates were capable of comprehending a proof themselves, but when they were given an e-Proof they relied on given audio explanations rather than trying to generate their own explanations. Moreover, the design of e-Proofs demands a significantly high number of between-line transitions which cause interruptions in the process of proof comprehension. These frequent interruptions and dependancy on audio explanations might be blamed for causing poor understanding of the proof over the long term compared with when undergraduates read a written textbook version of a proof.

8.3 Implications of this thesis

This thesis aimed to investigate the effectiveness of e-Proof in terms of undergraduates' proof comprehension. I chose to evaluate e-Proof as a single case for my thesis; however the findings have general implications in other area of educational studies. This thesis contributes to knowledge in two ways — adds *insight into the literature of undergraduates' proof comprehension* and presents *a methodology for evaluating educational interventions*.

This thesis clearly highlights that *students are not reliable at determining whether an educational intervention aids their learning*. This should be an important issue of concern because many educational interventions are entirely evaluated using student-satisfaction measurement. For example, the technology called *Lecture capture* is used widely in many counties believing its huge potential as a pedagogical tool. There are a few notable evaluation studies about its impact on student learning (Zhu and Bergom, 2007). Mostly those studies rely on students' self-reports about lecture capture (Brittain et al., 2006; Pinder-Grover et al., 2008; Zhu and Bergom, 2007). Though a significant amount of money has already been invested in the technology by many educational institutes in several counties, there is little evidence on whether lecture capture is an effective pedagogical tool in terms of improving learning or not.

In this thesis, I establish that self-reporting methods cannot be trusted entirely to ensure the effectiveness of a new technology. The self-reporting methods I used were focus-group interviews and web-based surveys and they clearly suggested that introducing e-Proof at undergraduates mathematics course would definitely be beneficial for undergraduates' proof comprehension. However, when I investigated further by comparing the effectiveness of e-Proofs with written textbook proofs, a different story emerged. If I had relied only on these self-report studies, I would have concluded my thesis by suggesting that e-Proofs should be introduced to all proof-based mathematics courses at university level. Clearly, what undergraduates reported about the effectiveness of e-Proofs showed their beliefs and views; not the reality. Self-reporting can bring out perhaps an incomplete story that may be driven by emotional accounts of participants. Studies that are entirely based on such methods could be an example of poor research.

This thesis also suggests that educational designers are not always reliable at determining whether their innovations will lead to effective learning. Researchers and educators often came up with new ideas of how to present study materials that can enhance learning. New pedagogical tools are mostly developed based on considering and presuming learners' strengths and weakness, about what can help them in learning etc. Intuition plays a key role for shaping new ideas into a new tool. In case of e-Proofs, Alcock (2009) had the belief that undergraduates have little knowledge of how to read a proof and they need help to acquire the skill of proof reading. Her intuition played a key role in creating e-Proofs and deciding how much help (audio explanations) to offer to them. E-Proofs have all the elements and potential that she wanted to provide to guide undergraduates overcoming their weakness in reading proof. Because she created e-Proofs considering undergraduates' needs, it was assumed that e-Proofs would deliver what Alcock intended to deliver. However, that was not the case as reported in the first experimental study. This clearly suggests that it is important to carry out an evaluation study to ensure the impact of a new tool rather than assuming its potential.

Experimental research is essential for any evaluation study. Moreover an experimental study designed with delayed post-test can play a vital role to ensure the impact of a new intervention on learning. Learning cannot be acknowledged as learning unless students continue to use the knowledge successfully in an appropriate situation in the future. In experimental studies, researchers use a post-test to examine the impact of a variable, but mostly ignore the delayed post-test. A delayed post-test with a time gap would help to understand better the long term impact of an variable on learning. For example, to understand the impact of multimedia presentation on learning Mayer and his colleagues ran series of experiments and the results of those experiments were based on the immediate post-test (Mayer, 2001) with no delayed post-test. In the case of e-Proof, it was the delayed post-test that showed the weaker long term impact of reading an e-Proof compared to reading a textbook proof, whereas the post-test study did not show any difference between them. My thesis highlights the great importance of including delayed post-tests, which can make the outcome of an experiment even stronger and clearer. Therefore, it is worthy to consider including delayed post-tests in future evaluations of educational innovations.

My thesis shows that many undergraduates have a sense of understanding of how to read a proof for comprehension. They spent more time on those crucial parts of proofs when reading for comprehension with or without additional help and in either case their levels of comprehension were similar. This clearly suggests that providing help does not necessarily promote learning, rather providing help when it is not needed could have a negative impact on learning. Learning is an active process, in which learners need to be actively engaged with the cognitive processes to construct new knowledge (Mayer, 2001). But in case of e-Proof, it seems that e-Proofs did not encourage active learning; rather with extra help undergraduates used less cognitive effort which is harmful.

The outcomes of this thesis clearly suggests the importance of considering how much help and what kind of help to offer to enhance learning. The aim should be not to offer direct help that can make the learning process easier, rather to provide help in a way that let learners actively engage with the cognitive process of learning. For proof comprehension, interventions to aid proof comprehension should focus on deepening students' engagement with the text of the proof, so that they can generate their own explanations in order to understand the proofs. Generating self-explanations has significant impact in improving proof comprehension as reported in Hodds et al. (2014). Their research has shown that undergraduates who were exposed to self-explanation training improved significantly in comprehending proofs compared to those who did not. This indicates such little help can encourage undergraduates to engage actively with the proof comprehension task that fosters learning.

8.4 Future works

The outcomes of the thesis are not only restricted to e-Proofs and proof comprehension, rather these can be extended in other areas of educational studies. New pedagogical innovations are emerging everyday aiming to enhance teaching and learning. There are many existing pedagogical tools that are in use, but there is little evidence of their positive influence in learning. On the other hand, there are many innovations that are believed to be effective based on students' satisfaction measurement, but this thesis clearly shows the importance of rechecking whether such measurement are ensuring educational outcomes or not. It is important work to evaluate such new or existing interventions to ensure their impact on teaching and learning. The methodology developed in this thesis could be adopted to carry on the evaluation study.

In the field of mathematical proofs, there are also such interventions, for example, generic proofs (Rowland, 2001) and structural-proofs (Leron, 1983). They were developed to present proofs in new ways aiming to improve proof comprehension. How these changes in proof presentation influence learning is mostly unknown. In future, I would be interested in looking into the impact of these different proof presentations and comparing between them by adopting the methodology developed in this thesis.

This thesis demonstrates that undergraduates have a sense of understanding of how to read a proof for comprehension and offering them direct help, such as providing explanations for each line of a proof, did not promote learning. To enhance leaning, it is not always necessary to make huge changes in conventional instructions or in study materials, some times small changes can be effective. For example, Diemand-Yauman et al.'s (2010) research has shown that superficial changes to learning materials, such as presenting reading material in a format that is slightly harder to read, could yield significant improvements in educational outcomes. They argued that such disfluency intervention acts as desirable difficulties which lead to deeper cognitive processing and improve learning. I believe it would be interesting to study the effect of fluency intervention on proof comprehension. If simple changes, such as font, can improve proof comprehension, that would be an important finding because such interventions are cost-effective and easy to adopt without changing the existing practice. It would also open up a new practice of studying what those desirable difficulties are that can boost students' cognitive activities with proof related tasks and that is what holds my interest as future work.

8.5 Concluding Remarks

This thesis contributes knowledge into the area of reading and comprehending proofs at undergraduate level and presents a methodology for evaluation studies on new pedagogical tools. The highlights of the thesis are:

- Undergraduates have a sense of understanding about how to read proofs themselves without any help. Too much help allows them to devote less cognitive effort than that they are capable of.
- Providing extra help does not always enhance learning, rather it can have a negative impact on learning.
- This thesis strongly indicates how important it is to carry out an evaluation study to ensure the effectiveness of a new intervention rather than assuming its potential based on student self-reports.
- This thesis presents a methodology that can be adopted in evaluating new pedagogical tools.

Appendix A

Interview with Lara Alcock

A.1 Interview with Lara Alcock

A.1.1 SR: How did you come up with the e-Proofs? How the whole thing started.

LA: I have been aware for long time that students struggle to understand proofs and that is not controversial information at all. Although, as you know, there not been work done on proof comprehension as such as students cannot construct proof very well.

And I suppose, I thought about those things before, what you would do if you try to understand a proof, and I have with previous classes given out some kind of guide, saying, try doing this, try doing that, try looking for the link between lines, try looking for overall structure, that sort of things. I think what kicked it off though (pause). One time I was doing revision lecture, and I just found myself breaking it [long proof] down, drawing lines around things, and saying look in this part, we just doing this, and then extra notes on relationships between lines, and really, going through what I might think while reading through a proof like this, almost modelling it sort off. This is the kind of things you would be thinking. And that just felt like something that I never quite done it that way before. Normally, when you are doing a first-run explanation, you know, you not particularly done it that way, I never really done it that way before, and the students seemed to respond well, and I came out of that thinking, probably I should do that more (laugh).

And then I went to a lunch time thing about teaching, I cant ever remember what it was now, but this guy from chemistry was presenting one of those things — it makes some ordinary board into a white board, you stick it downside and it records what you writing, and he said how he is using it to record himself constructing chemistry diagrams, so that he could put them together; because it was the construction process that was important, not the diagram that you ended up with. He wanted to put those on the web. And for some reason these things sort of came together in my head, you know, would not it be good to have something where you got this proof and you got the record of some of that kind of explanation in the same way, that would be different from how you might do if you constructing in a first place. And that just of came together as an idea for something would be useful, and I applied for the academic practice award, and then I got in touch with the people who support e-learning, and thats how I ended up making them [e-Proofs] in the first place. I wanted to provide something that would record the explanatory information, which is lost normally when you given a lecture, and then the students got the [written] proof, but they not got the access to the information again.

SR: Is your first idea of e-Proof different from the latest existing e-Proofs?

LA: Not really, one exists now the better one, what I had in mind. It is obviously when you start doing something like this, especially as sort of e-learning novice like me, you do not know what is possible and what is not. It feels like this is odd to be possible, and then of course as you try doing, you run into complication, and try to make something available using LEARN. You know, the first ones which are still on there [LEARN], are kind of clunky, you know, was not ideal by any means, in terms of how you interact with them. I wanted to make the flash version, that Lee had made of course much better. And the technology is improving all the time for all kind of presentation.

SR: But the basic idea of the e-Proof was same. Right?

LA: Ya pretty much, that has not change.

A.1.2 SR: What kind of role e-Proofs can play?

LA: What you really want is not for students just be able to follow these explanations of this particular proof, what you really [giving stress] want is to for them to get some idea about how they might do that by themselves. So they dont need this kind of tool, because they would be able to take a proof that they are reading, a textbook or being presented to them, and do that thinking for themselves.

I suppose I thought of more of a resource that you would have access to, like a note or a textbook, it would be different one that you are familiar with. You can go back [in case of e-Proofs], and re-visit the explanation as well as the written version. I mean, I tried in first year in lectures, and I still do this in lecture, I show part of an e-Proof and people watch that, and make more variety in a lecture, which is usually a good thing. What I attempted to do more recently, because I want people to be thinking things and working these things out themselves, I ask students to look at the paper copy of a theorem and proof, and try to make sense of themselves, and then use which of the part of the e-Proof I think of most relevant to help, to run through some of those things. And what I often see that, students are making notes on the work. I mean, again I dont know whether they are doing it thoughtfully or they are just copying from the screen, and again its like anything you know, I can give world's most brilliant lecture but people drifting of, they are not going to get benefit from it. So, ya, could go either way, I think. I guess, there would be a sort of bank of these things from people, form anywhere but the internet to get access to, if they find them using it, it might be useful for math centre website.

This year, when I used the main class, sometime I did use them [e-Proofs] sometime I didnt; sometime I just show the path of it, I though particularly difficult; sometime I run through the whole thing; so I dont thing I know enough about what is the best thing to do, whether I should use such things in lectures or not in lectures.

SR: Did you change your lecturing style?

LA: Ya, but I dont think, that is any more than my general lecturing style anyway. You know, from one year to the next, you always change things, and get people do more this type of activities. I mean, this is quite an intuitive process the way it goes. I mean, I make all sort of decisions about that [lecturing], and most (laugh) of them I have no evidence that can be any better, but you are aware of some sort of things, that its not quite working, that how you want it, and what is better when I did something a bit different, and so you try doing that more. But sometime when you do it more, its stopped being effective, because people get used to it and they just sit back. So its a fluid sort of process of giving lectures, and I am using resources like that. I did become aware of, from feedback of students as well, some of them said, we really like having e-Proofs in lectures, others said well it is good, I can see that it is good but I tend to drift off, you know, if you are watching the whole thing; so I tried to use them in a way that a bit more students' involvement and having a go for themselves first.

A.1.3 SR: Have you consider how e-Proofs can be beneficial over textbook proofs.

LA: Well, yes (stress). It was not so different, I guess, part of what I was thinking was that, I am imagining the student who goes to a lecture but write down what it says, but not perhaps always thinking about everything very carefully. So essentially, what they end up with a textbook proof, which then they have to interact with the same way, they would do they were reading a textbook at some point. So its not so terribly different, I mean, the problem with any kind of static written thing, or a problem, may be this is not a problem, may be this is exactly how everything should be; but certainly a textbook is very fixed, and so as the e-Proof off course, but one of the issues is that, if you want to add more explanations you kind of have to do that with narrative in a book, which potentially if you do that within the proof makes it very long. If you do it outside the proof, means you sort of stole the information, and then bringing it to mind again when you are looking at the particular line of the proof. I do not know, may be people are perfectly capable of that, but I suppose, I thought of at that time, that having the proof stay as it is, so that it is sort of something that you are looking at, but with the extra things appearing and disappearing wouldnt interfere with the written thing, you could make it bit shorter, you could help people see what the links were, without having to be extremely explicit and long window about explaining what they were and that the kind of things. I dont think, e-Proofs, the text in the e-Proofs looked pretty similar that what you would get in a textbook. I dont think that are substantially different, it depends on individual textbook, off course they have very different style, some of them are very tract, some of them are much more narrative and explanatory about things, you will see different things in textbooks.

The lecturer (chemistry) wanted to record the process you are going through when you constructed this thing, rather than just giving them the finish product, and I suppose thats what I wanted, I wanted go through the process, I wanted to have some sort of way of recording, do you know what I mean, capturing an explanation of the process one might go through when reading a proof.

SR: So e-Proofs are for helping in reading a proof.

LA: I dont know, whether I should done it in a different way, but what I really want, that students to learn to construct proof, however, the activity they are engaging more, certainly with something like analysis, is trying to make sense of the proofs that they are given, and that was what I was focusing. So I was sort of treating them, although they are variable, I was treating them something that a student would have and then it is sort of their job with their notes and do something with it. I wanted to give them sort of model, the kind of thing you would look for and add, when you cant read a proof.

A.1.4 SR: How did you decide how much information you give in the audio?

LA: (Laugh) Its interesting you know, I just did it. And I dont, hmm, I separating out by doing line by line check from understanding what global structure is, understanding what achieve by the module, or by component, or whatever you want to call it of the proof, (pause), so I did have those as separate things in my hear, but I did it. I did not have the system for deciding what I would explain, and what I would not explain, I suppose, what I tried to do was just (pause), is pretty much what I would do very intuitively if I sat with a students, who is struggling, who is just completely lost with some particular proof. What I would do? Well, I will point out some links, I think going through and it was done very much in an ad hock way proof by proof, explaining those links and lines in a way if I do it in live. So I did not have a system, perhaps I should have a system.

And now when I think about it, I did not have it, (pause) suppose I am explaining a particular proof, I suppose, I did not focus of general skill of what you should be looking for, may be something that could have been incorporated more effectively. But again, that would mean lot more flexibility, perhaps in held how want to interact with such things, means, you could imagine extremely sophisticated version, that asks you questions, then you choose between some options, or you did something or interacting within someway as a student and then, you know, it tells you actually you should be looking over, because of this is to do with that part of the theorem, it would not anywhere that stage. It is (e-Proof) an explanatory tool at this moment, presentation tool, not a sort of proper interacting thing.

I did not have the system, I just did it in a way that what kind of thing do I see that I am looking for, here is a big formula, that appears in the beginning, and you tent to think woo what did that come from, (laugh) and you know, or why would one introduce that, and then what I would be doing, I would be thinking why would they want that, and I would be tracking down, and thinking ooh I see when you get to here, this is what happens, so sort of links forward, and anticipatory like that, others are more backward explanatory links, where, you know, you just say, well this came from the fact that we assume this is up here, look.

SR: So the explanations in e-Proofs are very close to how you explain things in live lecture.

LA: Ya, obviously for every individual one, I thought lot more that I do in live lecture, because when you are recoding yourself, you want to have clear sentences and things like that, but that is more thinking about making the explanations good, rather than thinking about whether that was the best explanation, I dont know.

SR: Any other factors that influence the design of the e-Proofs.

LA: Various ones, which I wrote in other paper with Matthew.

Obviously, you certainly got these constrains, because you wanted to be quite large print on the screen, you wanted to fit it on one screen. One could make it work without that, but that kind of what I decided to work with, because it just seems easier, and because the proofs I was working with that time were off the length where you could reasonably do that. And it just seems like that would help with this having every thing lay down. I had to make lots of decision in and on an individual basis about spacing and about layout but that reason. But those are not really very different that what I would make in a hand written version anyway except that you want to compress it slightly sometimes, so I would make slide decision that I perhaps do slightly different if I was writing. Some of the information can go in the audio, so if you do need a bit of space you can take out perhaps small explanatory things and then re-add that in your audio. But again that no different if you would be deciding if you would just writing a proof in any other way really. It just focused your attention on those kind of aspects bit more.

I wanted the fronts clear and all the rest of it, I wanted the boxed and arrows and things not to interfere with anything else too much.

SR: What do you mean by not to interfere ?

LA: You just try to make sure that you design them so it is not coming over something.

I know thing, its unrelated to this question but I tell you anyway, I know realise that what I did, I focused not enough on the theorem at the beginning, you know. So I fairly recently, I understood that how little students pay attention towards theorems, before they start looking at proofs. So, I would make sure that each one has, this is the premises of the theorem, these are the assumptions and then very clear thing about this is a conclusion, and really would do that too, even it is just the first and the last line of the proof. That helps people realise that again, so I suppose I will do that more now.

That sort of things so obvious to you as an competent mathsy-person, that what we are doing here, and you know I come to realise that many of the students dont sort of aware of that. But you know, the theorems are even broken down like that, they are not very alert to if-then statements, off course I knew that anyway, so I dont know why I did not do that at that stage. But I guess it become very clear now, and I will do that more now if I am doing again.

Actual design hmmm constrains of the screen, and then the front size..

A.1.5 SR: While designing e-Proofs, did you consider the e-learning literature, multimedia learning?

LA: At that time, it was just an idea; it was a good opportunity to pursue something like that. I mean, of course, there is, maths is problematic anyway, because of the formatting and layout and things, and I have been to various e-leaning events where there has been discussion of that and sort of, it tends to be more in term of online testing rather than other kinds of resources of maths, you tends to get straight lecture like a math-centre website, that somebodys hand writing, so you dont have problem with the symbols or using CAA testing of various kinds and off course my colleagues who have done much more sophisticated work with that I have. No, I did not know great deal about the general literature about e-learning, I still dont know as much as I probably ought to about the sort of kind of prompts and interaction and things like that. No.

SR: Presenting a proof through both audio and visual media is a very effective way in terms o multimedia learning research.

LA: The intention was that, but it seems from our works that they dont necessarily learn more by interacting with these things. So I am certainly have become much more (pause), its seems like such a good idea, as these things always do, when one is doing; then you learn a lot about what you are not doing, in any particular type of presentation and the subsequence is that off course, I have read more, and become more aware of the general principle that people have used. Sometimes I dont know my impression is that, it is going to be a lot of it will be very dependent of the material of the students know rest of this. So this is complicated trying to work out what would be sort of best approach, although off course I can still hope for the something would be.

SR: From students feedbacks, did you get any picture?

LA: I dont have a systematic picture, but you know, people do cite them as a very positive things on feedback forms. Students like them, but they like when you do things, when do things that obviously intended to help them. So they like to have things to help things, they like to have extra resources. I dont know how much it is just they appreciate the effort you gone through to produce those things, they consider being helpful to them. But they also like the feeling of security, I think, and I think, the things like this contribute to this feeling of security, that there is an extra explanation available to me and I can go and get it when I am ready. So some of the positive feeling is probably that kind of emotional, some of the crutch, you know, something there to help me. I mean, they appreciate it and they feel like they learn from things, but then you know its not necessarily the same, as it being effective as a way of learning or the way of learning is a different stage. I should have said earlier actually; what I did have in my mind originally was like this would make one for everything, but then might be very particularly effective at this stage where students are first encountering lots of these kind of proofs and you want to get them to a point where they can learn how to break something down and understand it for themselves, after that you wanna give them anything, you want to doing that activity. So might be these students, some of them that are already at that stage and they can do it anyway, it might be that people appreciate it most are the one who is not really ready and who are using as a crutch, something is there for me without having to think it through for themselves so much.

SR: Do you think it might be good for 1st year students who have almost no experience with proofs?

LA: Yaa, may be, I dont know. I suppose that what I did think. Now I wonder whether people would just sort of rely on it too much. This is the thing, the people probably need the help, the one who is less likely to take useful from it, I would think as if the case of so many things, I can imagine that you might out quite a lot of people who are really quite good, and would be thinking of right lines anyway, and who needs little bit of support to sort of get to that point where they are confident that they can do this and they are not intimidated and (giving stress) if they working hard and if they are pretty on top of things would probably get lot out of seeing such things, because they be ooh I can see how things fit together and thats good and become more aware of that where they should be looking for. But I can also imagine students who use to be very passive in their learning and being told what to do, sitting in front of these things passively learning, all wash over without really engaging with the sense of now how could I do that myself next time. So may be, what one needs to like anything sort of much more systematic way of using things like that and then practicing with something else and then you know sort of developing this kind of skill for yourselves and testing that. This is making me again thinking, change all my teaching (laugh). You know, how little control I have on the process.

A.1.6 SR: Do you suggest in your class that how they should be reading e-proofs?

LA: Yes, but I would not surprise that goes in one ear and out the other, at least in one ear and out the other to the people who are not going to be doing the right thing anyway. So certainly, you can tell some people are really engaging and trying to think about detail of that. I mean, that is completely different from the transfer issue, all are then more able to do these to themselves with some kind of other proofs.

Ya, I do say, use it intelligently, there is no point just sitting there and watching through something, well, what you want to be doing reading it and trying to understand it yourselves, then use it as a resource, that is a backup where if you do not understand it, particularly part of this, you can get extra explanation, take ownership of yourselves. But that is part of the wider message of getting involved with mathematics and that kind of thing. And again you know in lecture there are 70% of people there, so no matter how much you say, there are lots of people who, what ever message you trying to send, are not getting that message because they are not there to hear it.

SR: What is your expectation from students that how they should read an e-proof to learn a new theorem?

LA:I First of all, I want them to try and if suppose they have a written version, I want them to try and understand it for themselves.

SR: Without getting the help from the audio explanation!

LA: I think, one of the other thing, you want people being independent. You want them to try things themselves, you want them to have a good go stuff. To put solution to the problem sheets, some people dont but I do, but I dont put them up until after the time which people should have a go. Whether that makes any difference or not I dont know, but there is definitely a sense of concern, what I think of some students probably do try but not very hard, and then get the solution, and read the solution, and convince themselves that they learn a lot by doing that, and they might have learn a lot by doing that, but they probably would not learn as much as they would done if they persevere and either succeed for themselves, or at least got through a point where there is very clear exactly what cause in the problem, and then look at the solution. E-Proofs are another things like that, if you have a good go with something first, you might be able to then use it more intelligently, because you would be thinking, and you would have some sense of — what is this bit, what going on here, I dont know why this is done this way, and then that might help you to do that to fill in the gap yourselves. But it would be better if you approached it already knowing, there is where my gaps are and this is why I dont understand. So you could look at that in order to specifically find a part that you want to. Or say this is a really hard one, you are really struggling, and the whole things make no sense to you, but to have a having a good go first, so when you watching that you are familiar enough with the text, I suppose, you can digest the explanation quite quickly. I want students to be working hard and thinking, and really that is what I want.

SR: It seems to me that you want students should be independent not e-Proofs dependent.

LA: Yes absolutely. When people first, some of them, not all of them, first meet this kind of thing, they have no idea what they be doing with it. If you asked them about things, you can tell that they are not really thought about it; like why this will be like this, they not got that routine or habit, why I am doing this, why I am doing that, how it fit together. I suppose, I am thinking of people at the beginning of their sort of that part of their mathematical career having access to this kind of thing and getting some sense of, ooh right that what I should be doing if I am going to understand this. So I imagine, people at that sort of stage and did not really imagine people further on.

SR: So this could be more effective for weaker students, especially when they might not get help directly from the teachers.

LR: what they say sometimes, and I had this very smart people who I know are working extremely hard, some of the people who did the best in the course, have remarked, how much work and how useful they have been and when especially when you are trying to revise on your own on the holiday period, when the lecturer is not there and you got to something extra to help you understand that. That is not different from going and looking at textbook, although there is always problem, I found very difficult when I was undergraduates; because all textbooks are bit different in the notations they use, the type of explanation they give and I found it very difficult if I was struggling with a particular thing, looking in textbook in that are often really help me anywhere near as much as you expect, because it much so much work to get into the way the expression thing that particular textbook had, that way more work than would have been appropriate that particular difficulty I was having. I suppose this particular thing (e-Proof) more tight to the particular cause, may be that is bad or is good, I dont know.

Certainly students do appreciate, some of them really are working, do appreciate that when there is not a person and there is at least some level of extra explanation about this.

SR: In e-proofs, in each slide more than one line or different portions are highlighted, and you want that students must look at those portions at the same time.

LA: I am fine with that. I thing that how it should be. For the reason I was talking earlier, couple of number of places, some of the typical analysis proofs where something just appears out nowhere near the beginning, you just define something in a certain way, and it seems very arbitrary that it come out nowhere. And certainly one of the first thing I will be doing, like I said before, I will look that and think, ok why, you know, may be I will track it down, and probably what I would be doing is looking at where just this thing reappears, and how is it use then. So I can see what going to happen, I dont know what the intermediate bit is, but what going to happen is, going to use it in this way at the end; so now, I can see why this thing is introduced at the beginning, which gives me more patient for the intervening steps. So I dont thing its a problem not to read it in order.

SR: But this might be a very expertise way to read a proof perhaps.

LA: Well I am hoping to students will learn. I suppose, I am hoping that seeing this will give you more sense of — it is possible, it is quite reasonable to ask that question. This conversation making me think, may be my audio things ought to be much more colloquial, now here is the thing, it seems come out from no where, lets have a look down towards the bottom and see if we can see how its used that will help us to understand why a person want to do that and highlight the next bit of the bottom and explain that and then go back, may be that would model the thinking process more and may be that more I should have focused on, I mean again these are questions one could ask.

SR: And that would be less formal and more close to a live lecture

LA: Ya, I suppose, what is happening when you start recording carefully, you do move towards being bit more formal, and sort of current about everything, and may be that good some way, but may be worse overall. I dont know. Certainly seems worth considering.

Appendix B

Study materials for Experiment 1

B.1 Cauchy's GMVT — the written textbook proof

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

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

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

Proof

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

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

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

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

Also $h(a) = f(a) - \dfrac{f(b) - f(a)}{g(b) - g(a)} g(a) = \dfrac{f(a)g(b) - g(a)f(b)}{g(b) - g(a)}$.

and $h(b) = f(b) - \dfrac{f(b) - f(a)}{g(b) - g(a)} g(b) = \dfrac{f(a)g(b) - g(a)f(b)}{g(b) - g(a)}$, so $h(a) = h(b)$.

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

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

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

B.2 Cauchy's GMVT — the short version

Suppose that f and g are continuous on $[a, b]$ and differentiable on (a, b). Suppose also that $\forall x \in (a, b)$, $g'(x) \neq 0$.
Then $\exists c \in (a, b)$ such that $\dfrac{f'(c)}{g'(c)} = \dfrac{f(b) - f(a)}{g(b) - g(a)}$.

Proof

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

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

3) Define $h(x) = f(x) - \dfrac{f(b) - f(a)}{g(b) - g(a)} g(x)$.

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

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

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

7) But $h'(c) = 0 \Rightarrow \dfrac{f'(c)}{g'(c)} = \dfrac{f(b) - f(a)}{g(b) - g(a)}$.

8) So $\exists c \in (a, b)$ s.t. $\dfrac{f'(c)}{g'(c)} = \dfrac{f(b) - f(a)}{g(b) - g(a)}$.

B.3 Additional information sheet: definitions and theorems

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

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

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

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

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

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

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

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

Appendix C

Theorems, proofs and comprehension tests used in Experiment 2

C.1 Theorem 1 and the comprehension test

Theorem 1 If $n \in \mathbb{Z}$ is composite, then n has a prime factor not exceeding $\sqrt{n}$.

Proof. There are integers a, b such that $1 < a \leq b < n$ and $n = ab$

If $a > \sqrt{n}$, then $b \geq a > \sqrt{n}$ and $ab > \sqrt{n}.\sqrt{n} = n$ which is absurd.

Therefore, $a \leq \sqrt{n}$.

We know that every positive integer greater than 1 has a prime factor.

Therefore, a has a prime factor p which also divides n.

Therefore, $p \leq \sqrt{n}$. This proves the theorem.

....

Q1. What method is used to show $a \leq \sqrt{n}$?

1. A backward method.

2. A contradiction method.

3. An inductive method.

4. An indirect method.

Q2. Why showing $a \leq \sqrt{n}$ is not enough to prove the theorem? Click the correct answer.

1. Because it was important to show that n has more than one factor, so we considered p.

2. Showing $a \leq \sqrt{n}$ is enough, but we show further $p \leq \sqrt{n}$ to make the argument more strong.

3. Because we do not know whether a is prime or not enough, but know p is a prime.

4. If we declare a as a prime we do not need to show $p \leq \sqrt{n}$.

Q3. Write n in terms of p.

1. Cannot possible to write n in terms of p.

2. Only possible if $a = p$, then $n = pb$.

3. We can write n in terms of p is many different ways.

4. None of these is true.

C.2 Theorem 2 and the comprehension test

Theorem 2 There are infinitely many primes.

Proof. Let $n > 1$ be a positive integer.

Since n and $n + 1$ consecutive integers, they have no common factor. Therefore, they are relatively prime.

Consider the number $N_1 = n(n + 1)$.

We know that every integer greater than 1 must have a prime factor.

Therefore N_1 must have at least two different prime factors.

Similarly, the integers $n(n + 1)$ and $n(n + 1) + 1$ are consecutive and therefore relatively prime.

Consider the number $N_2 = [n(n+1)][n(n+1)+1]$, N_2 must have at least 3 different prime factors.

This process can be continued indefinitely. This proves the theorem.

....

Q1. Which of the justification is <u>best</u> explain why "two consecutive numbers are relatively prime"

1. One of them is always odd and the other is even. Therefore they have no factor in common.
2. Because this is a rule of mathematics.
3. Because the difference between them is 1, therefore the only possible common factor is 1.
4. None of these is true.

Q1. Justify the following statements and click if true.

1. If N_3 has at least 4 different prime factors, N_4 will have exactly one more different prime factor.

2. If N_3 has at least 4 different prime factors, N_4 will have exactly one more but not different prime factor.

3. If N_3 has at least 4 different prime factors, N_4 might have more than one different prime factor.

Q3. Which of the following discussions do you think justify the proving method of the theorem.

1. This is an incomplete proof because it only showed two steps and then stated that "this process can be continued indefinitely" which is not enough to complete theorem.

2. This is a valid proof even though only two steps are shown in the proof. Because an induction method was used in the proof, we can claim "this process can be continued indefinitely. This proves the theorem".

3. This is an invalid proof because it only showed two steps and then stated that "this process can be continued indefinitely", but anything can happen while continuing the process which we do not know.

C.3 Theorem 3 and the comprehension test

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

Proof. Suppose that $f(a) = f(b)$ and f is continuous on $[a, b]$ and differentiable on (a, b).

Then f is bounded and attains its maximum and minimum values by the EVT.

So $\exists x_1, x_2 \in [a, b]$ s. t. $\forall x \in [a, b]$, $f(x_1) \leq f(x) \leq f(x_2)$.

If x_1 and x_2 are both endpoints of $[a, b]$, then one is equal to a and the other is equal to b.

Hence $f(x_1) = f(x_2)$ so f is constant.

In this case, $\forall c \in (a, b), f'(c) = 0$.

If x_1 and x_2 are not both endpoints, then $x_1 \in (a, b)$ or $x_2 \in (a, b)$ (or both).

If $x_1 \in (a, b)$ then $f'(x_1) = 0$ by the IET.

If $x_2 \in (a, b)$ then $f'(x_2) = 0$ by the IET.

So in all cases, $\exists c \in (a, b)$, such that $f'(c) = 0$.

....

Comprehension test

Q1. Does the assumptions of the Rolle's theorem hold on the interval

(i) $f(x) = x^4$ on the interval $[-1, 1]$. Yes ☐ No ☐

(ii) $f(x) = |x|$ on the interval $[-5, 5]$.

(iii)

$$f(x) = \begin{cases} x + 8 & \textbf{if } x \textbf{ is rational} \\ 2 & \textbf{if } x \textbf{ is irrational} \end{cases}$$

Q2. Which of the following gives the answer of "is the assumption that $f(a) = f(b)$ is necessary to a $c \in (a, b)$ s.t $f'(c) = 0$ for $f(x)$ on $[a, b]$".

1. Yes it is necessary, because to find $c \in (a, b)$ s.t $f'(c) = 0$, $f(x)$ must satisfy all the three assumptions.

2. Not necessary, because all the three assumptions are sufficient conditions but not necessary.

3. Depends on the function.

4. Depends on the interval.

Q3. Read the following statements and choose whether they are true or false?

1. Let $f(x)$ be a function continuous on $[a, b]$ and differentiable on (a, b). If there exists $c \in (a, b)$ such that $f'(c) = 0$, then $f(a) = f(b)$.

2. It follow from Rolle's theorem that there exists only one point c such that $f'(c) = 0$.

3. It follow from Rolle's theorem that there always exists more than one point c such that $f'(c) = 0$.

4. A constant function satisfies all assumptions of Rolle's theorem and its conclusion.

C.4 Theorem 4 and the comprehension test

Cauchy's GMVT Suppose that f and g are continuous on $[a, b]$ and differentiable on (a, b). Suppose also that $\forall x \in (a, b)$, $g'(x) \neq 0$. Then $\exists c \in (a, b)$ such that $\frac{f'(c)}{g'(c)} = \frac{f(b)-f(a)}{g(b)-g(a)}$.

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

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

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

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

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

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

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

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

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

Comprehension Test of Cauchy's GMVT

Q1. Find an interval where $f(x) = 1 + x^2$ and $g(x) = x^2 - 1$ satisfy the premises of Cauchy's GMVT? Choose the appropriate interval with right justification.

1. On the interval $[-1, 1]$
2. On the interval $[1, \infty)$ because interval must be continuous and differentiable.
3. On the interval $[1, 12]$ because it does not contain 0.
4. On the interval $(0, 2)$ because $f'(x) = 2x$ and $g'(x) = 2x$.

Q2. Which are the following properties used in the proof. If function p and q are continuous on $[a, b]$ and differentiable on (a, b) then

1. $p + q$ and kq are continuous on $[a, b]$ and differentiable on(a, b); where k is a constant.
2. $p + q$ and pq are continuous on $[a, b]$ and differentiable on(a, b).
3. $p+q$ and p/q are continuous on $[a, b]$ and differentiable on(a, b); where $q \neq 0$.
4. $p+q$ and kp/q are continuous on $[a, b]$ and differentiable on(a, b); where $q \neq 0$ and where k is a constant.

Q3. Rank the following short summaries of the proof from 1(best) to 4(worst).

1. Quotient of two derivatives of continuous function f and g is equal to a constant belongs to (a, b).
2. We establish that $g(a) \neq g(b)$ so that we can divide by $g(b) - g(a)$. We define h is terms of f and g and prove that h satisfies the premises of Rolle's theorem on the interval $[a, b]$. We apply Rolle's theorem to conclude that $\exists c \in (a, b)$ s.t, $h'(c) = 0$. We rewrite $h'(c) = 0$ in terms of f and g and rearrange to obtain the desired conclusion.
3. Let g be a straight line gradient $\neq 0$ and f is a function. Both f and g is differentiable, $(g(a) \neq g(b)$. Define $h(x)$ and $h(a) = h(b)$. By Rolle's theorem it is imples that there exists c in (a, b) s.t. $h'(c) = 0$.

4. On $[a, b]$ with continuous functions f and g, there is a point c, where the ratio between the gradients at this point is equal to the ratio between the straight lines $f(b) - f(a)$ and $g(b) - g(a)$.

Appendix D

Proofs used to re-analyse mathematicians' reading processes

D.1 The proofs used in Inglis and Alcock's (2012) study

FIGURE D.1: The short proof in Inglis and Alcock's (2012) study

Theorem. For any positive integer n, if n^2 is divisible by 3, then n is divisible by 3.

Proof. Assume that n^2 is an odd positive integer that is divisible by 3.

That is, $n^2=(3n+1)^2=9n^2+6n+1=3n(n+2)+1$.

Therefore, n^2 is divisible by 3.

Assume that n^2 is even and a multiple of 3.

That is $n^2=(3n)^2=9n^2=3n(3n)$.

Therefore, n^2 is a multiple of 3.

If we factor $n^2=9n^2$, we get $3n(3n)$; which means that n is a multiple of 3.

FIGURE D.2: The long proof in Inglis and Alcock's (2012) study

Theorem. There are infinitely many primes that can be written as $4k+1$ (where $k \in \mathbf{Z}$)

Proof. Suppose there are finitely many primes of the form $4k+1$.

Then these primes can be listed $p_1, p_2, p_3, \ldots, p_n$.

Define a number a as follows. Let $a=p_1p_2p_3\ldots p_n+4$.

Note that dividing a by 4 leaves remainder 1.

Every number that leaves remainder 1 when divided by 4 is divisible by a prime that also leaves remainder 1 when divided by 4.

However, for all i such that $1 \leq i \leq n$, p_i divides $p_1p_2p_3\ldots p_n$ and p_i does not divide 4.

Thus p_i does not divide a.

So dividing a by 4 leaves remainder 1 and a is not divisible by any prime that leaves remainder 1 when divided by 4.

This is a contradiction.

Bibliography

Abowd, G., Atkeson, C., Feinstein, A., Hmelo, C., Kooper, R., Long, S., Sawhney, N., and Tani, M. (1997). Teaching and learning as multimedia authoring: the classroom 2000 project. In *Proceedings of the Fourth ACM International Conference on Multimedia*, pages 187–198. ACM.

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

Ainsworth, S. and Th Loizou, A. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science*, 27(4):669–681.

Alcock, L. (2009). e-Proofs: Student Experience of Online Resources to Aid Understanding of Mathematical Proofs. *In Proceedings of the 12th Conference on Research in Undergraduate Mathematics Education*, Raleigh, NC, USA.

Alcock, L. (2010). Interactions between teaching and research: Developing pedagogical content knowledge for real analysis. *Learning Through Teaching Mathematics*, pages 226–244.

Alcock, L. and Inglis, M. (2008). Doctoral students use of examples in evaluating and proving conjectures. *Educational Studies in Mathematics*, 69(2):111–129.

Alcock, L. and Inglis, M. (2010). Visual considerations in the presentation of mathematical proofs. *Seminar.Net*, 6(1):Available: http://seminar.net/index.php/component/content/article/75/139–visual–considerations–in–the–presentation–of–mathematical–proofs.

Alcock, L. and Simpson, A. (2002a). Definitions: Dealing with categories mathematically. *For the Learning of Mathematics*, 22(2):28–34.

Alcock, L. and Simpson, A. (2002b). Two components in learning to reason using definitions. In *Proceedings of the 2nd International Conference on the Teaching of Mathematics (at the Undergraduate Level)*.

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

Alcock, L. and Wilkinson, N. (2011). e-proofs: Design of a resource to support proof comprehension in mathematics. *Journal of the International Society for Design and Development in Education*, 1(4):http://www.educationaldesigner.org/ed/volume1/issue4/article14/.

Aronson, J., Lustina, M., Good, C., KeoughClaude M, K., and Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35(1):29–46.

Ball, L., Lucas, E., Miles, J., and Gale, A. (2003). Inspection times and the selection task: What do eye-movements reveal about relevance effects? *The Quarterly Journal of Experimental Psychology*, 56(6):1053–1077.

Barton, M., Heidema, C., and Jordan, D. (2002). Teaching reading in mathematics and science. *Reading and Writing*, 60(3):24–28.

Bell, A. (1976). A study of pupils' proof-explanations in mathematical situations. *Educational Studies in Mathematics*, 7(1):23–40.

Bernard, R., Abrami, P., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P., Fiset, M., and Huang, B. (2004). How does distance education compare with classroom instruction? a meta-analysis of the empirical literature. *Review of Educational Research*, 74(3):379.

Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. *The Cambridge Handbook of Multimedia Learning*, pages 287–296.

Branch, J. (2000). Investigating the information-seeking processes of adolescents:: The value of using think alouds and think afters. *Library & Information Science Research*, 22(4):371–392.

Brittain, S., Glowacki, P., Van Ittersum, J., and Johnson, L. (2006). Podcasting lectures. *Educause Quarterly*, 3:24–31.

Bryman, A. (2004). *Social Research Methods*. Oxford University Press, Oxford.

Burkill, J. (1978). *A First Course in Mathematical Analysis*. Cambridge University Press, Cambridge, UK.

Cacioppo, J. and Petty, R. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1):116.

Campbell, D. T., Stanley, J. C., and Gage, N. L. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin Boston.

Chandler, P. and Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4):293–332.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, 24(4):359–387.

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

Christensen, L. (2003). *Experimental Methodolgy*. Allyn and Bacon, Boston, 9th edition.

Clark, R. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42(2):21–29.

Clark, R. and Feldon, D. (2005). *Five common but questionable principles of multimedia learning*. Cambridge University Press, Cambridge, UK.

Cohen, L., Manion, L., Morrison, K., and Morrison, K. (2007). *Research methods in education*. Routledge, Abingdon, Oxford.

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

Cowen, C. (1991). Teaching and testing mathematics reading. *American Mathematical Monthly*, 98(2):50–53.

Cowen, L., Ball, L., and Delin, J. (2002). An eye movement analysis of web page usability. *Proc. HCI*, Springer-Verlag, London:317–336.

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.

Davis, P., Hersh, R., and Marchisotto, E. (1995). *The mathematical experience*. Boston: Houghton- Mifflin.

de Villiers, M. (1999). The role and function of proof with sketchpad. *Rethinking Proof with The Geometer's Sketchpad*, page 310.

Diemand-Yauman, C., Oppenheimer, D., and Vaughan, E. (2010). Fortune favors the Bold (and the Italicize): Effects of disfluency on educational outcomes. *Cognition*.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. Springer-Verlag New York Inc.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., Murphy, R., Penuel, W., Javitz, H., et al. (2007). Effectiveness of reading and mathematics software products: Findings from the first student cohort. (NCEE Rep. No. 2007-4005). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Edwards, B. and Ward, M. (2004). Surprises from mathematics education research: Student (mis) use of mathematical definitions. *American Mathematical Monthly*, 111(5):411–424.

Elmes, D., Kantowitz, B., and Roediger, I. (2011). *Research methods in psychology (7th ed.)*. Wadsworth Pub Co, St. Paul, MN: West.

Ericsson, K. and Simon, H. (1980). Verbal reports as data. *Psychological Review*, 87(3):215–251.

Ericsson, K. A. and Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5(3):178–186.

Fischbein, E. (1982). Intuition and proof. *For the Learning of mathematics*, 3(2):9–24.

Goldberg, J., Stimson, M., Lewenstein, M., Scott, N., and Wichansky, A. (2002). Eye tracking in web search tasks: design implications. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, pages 51–58. ACM.

Goldman, S. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes*, 23(3):357–398.

Hanna, G. (1989). More than formal proof. *For the Learning of Mathematics*, 9(1):20–23.

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

Hanna, G. (1995). Challenges to the Importance of Proof. *For the Learning of Mathematics*, 15(3):42–49.

Hanna, G. and Barbeau, E. (2008). Proofs as bearers of mathematical knowledge. *ZDM*, 40(3):345–353.

Hanna, G. and Jahnke, H. N. (1993). Proof and application. *Educational Studies in Mathematics*, 24(4):421–438.

Harel, G. and Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. *Research in Collegiate Mathematics Education*, 7:234–283.

Healy, L. and Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 31(4):396–428.

Hegarty, M., Mayer, R., and Green, C. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology*, 84(1):76.

Hemmi, K. (2006). Approaching proof in a community of mathematical practice. *Unpublished PhD Thesis, University of Stockholm, Sweden.*

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

Highfield, K. and Goodwin, K. (2008). A review of recent research in early mathematics learning and technology. In *Proceedings of the 31st Annual Conference of the Mathematics Education Research Group of Australasia. Navigating Currents and Charting Directions*, volume 1, pages 259–264.

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

Houston, S. (1993). Comprehension tests in mathematics. *Teaching Mathematics and its Applications*, 12(2):60–73.

Hoyles, C. and Küchemann, D. (2002). Students' understandings of logical implication. *Educational Studies in Mathematics*, 51(3):193–223.

Hoyles, C. and Noss, R. (2003). What can digital technologies take from and bring to research in mathematics education? In Bishop, A., Clements, M., Keitel, C., Kilpatrick, J., and Leung, F., editors, *Second International Handbook of Mathematics Education*, pages 323–349. Dordrecht: Kluwer Academic Publishers.

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction*, 20(2):172–176.

Hyönä, J. and Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica*, 73(3):259–280.

Inglis, M. and Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education,*, 43(4):358–390.

Inglis, M. and Mejia-Ramos, J. (2009). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction*, 27(1):25–50.

Jacob, R. and Karn, K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In Hoyona, J., Radach, R., and Deubel, H., editors, *The Mind's Eye: Cognitive and Applied aspects of Eye Movement Research*, pages 573–605. Elsevier, North-Holland.

Johnson-Laird, P. (1994). *Mental models, deductive reasoning, and the brain.* In The Cognitive Neurosciences (M. S. Gazzaniga, Eds.), MIT Press, Cambridge, MA.

Just, M. and Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87:329–354.

Kalyuga, S., Ayres, P., Chandler, P., and Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1):23–31.

Kalyuga, S. and Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, 96(3):558.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge University Press, Cambridge, UK.

Kitzinger, J. (1995). Qualitative research: introducing focus groups. *British Medical Journal*, 311(7000):299.

Knuth, E. J. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education*, 33(5):379–405.

Krantz, S. G. (2007). The history and concept of mathematical proof. Avaiable on www.math.wustl.edu/ sk/eolss.pdf dated November,2012.

Krueger, R. and Casey, M. (2000). *A practical guide for applied research*. Sage Publications, London.

Kulik, J. (1994). Meta-analytic studies of findings on computer-based instruction. *Technology Assessment in Education and Training*, pages 9–33.

Kulik, J. (2003). Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say. Arlington, VA: SRI International.

Lamport, L. (1995). How to write a proof. *American Mathematical Monthly*, 102(7):600–608.

Land, M., Mennie, N., Rusted, J., et al. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11):1311–1328.

Leron, U. (1983). Structuring mathematical proofs. *American Mathematical Monthly*, 90(3):174–185.

Lin, F. L. and Yang, K. L. (2007). The reading comprehension of geometric proofs: The contribution of knowledge and reasoning. *International Journal of Science and Mathematics Education*, 5(4):729–754.

Manin, Y. (1981). A digression on proof. *Two-Year College Mathematics Journal*, 12(2):104–107.

Martin, W. G. and Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education*, 20(1):41–51.

Mason, J., Burton, L., and Stacey, K. (1982). *Thinking Mathematically*. Addison-Wesley Pub. Co., London.

Mayer, R. (2001). *Multimedia Learning*. New York: Cambridge University Press.

Mayer, R. (2009). Constructivism as a theory of learning versus constructivism as a prescription for instruction. In Tobias, S. and Duffy, T. M., editors, *Constructivist instruction: Success or failure?*, pages 184–200. New York: Routledge.

Mayer, R. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20(2):167–171.

Mayer, R. and Anderson, R. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83(4):484.

Mayer, R. and Anderson, R. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84(4):444.

Mayer, R., Heiser, J., and Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93(1):187.

Mayer, R. and Moreno, R. (1998a). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2):312.

Mayer, R. E. and Moreno, R. (1998b). A cognitive theory of multimedia learning: Implications for design principles. In *annual meeting of the ACM SIGCHI Conference on Human Factors in Computing Systems, Los Angeles, CA*.

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

McNamara, D. and Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22(3):247–288.

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

Mejia-Ramos, J. (2011). Do Leron's structured proofs improve proof comprehension? In *In Proceedings of the 14th Conference on Research in Undergraduate Mathematics Education*.

Mejia-Ramos, J., Fuller, E., Weber, K., Rhoads, K., and Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics*, 79(1):3–18.

Mejia-Ramos, J., Weber, K., Fuller, E., Samkoff, A., Search, R., and Rhoads, K. (2010). Modeling the comprehension of proof in undergraduate mathematics. *In Proceedings of the 13th Conference on Research in Undergraduate Mathematics Education*, Raleigh, NC, USA.

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, 27(3):249–266.

Moravcsik, J. and Kintsch, W. (1993). Writing quality, reading skills, and domain knowledge as factors in text comprehension. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 47(2):360.

Murphy, R., Penuel, W., Means, B., Korbak, C., Whaley, A., and Allen, J. (2001). E-desk: A review of recent evidence on the effectiveness of discrete educational software. *Menlo Park, CA: SRI International*.

NCTM (2000). *Principles and standards for school mathematics*, volume 1. National Council of Teachers of Mathematics.

Nisbett, R. and Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3):231–259.

Österholm, M. (2006). Characterizing reading comprehension of mathematical texts. *Educational Studies in Mathematics*, 63(3):325–346.

Paas, F., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, 38(1):1–4.

Paas, F., Renkl, A., and Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, 32(1):1–8.

Page, A. (1947). *Mathematical Analysis and Techniques*, volume 1. Oxford University Press, Oxford.

Pan, B., Hembrooke, H., Gay, G., Granka, L., Feusner, M., and Newman, J. (2004). The determinants of web page viewing behavior: an eye-tracking study. In *Proceedings of the 2004 Symposium on Eye tracking Research & Applications*, pages 147–154. ACM.

Pimm, D. (1989). *Speaking Mathematically: Communication in Mathematics Classrooms*. London, Routledge.

Pinder-Grover, T., Millunchick, J. M., and Bierwert, C. (2008). Work in progress: Using screencasts to enhance student learning in a large lecture material science and engineering course. In *proceesings of 38th IEEE Frontiers in Education Conference, Saeatoga Springs, NY.*

Pinto, M. and Tall, D. (1999). Student constructions of formal theory: Giving and extracting meaning. *Proceedings of the 23rd Conference of PME, Haifa, Israel*, 4:4–65.

Poole, A. (2005). Eye tracking in human-computer interaction and usability research: Current status and future prospects. *Chapter in C. Chaoui (Ed.): Encyclopedia of HCI. Pennsylvania: Idea Group, Inc.*

Raman, M. (2003). Key ideas: what are they and how can they help us understand how people view proof? *Educational Studies in Mathematics*, 52(3):319–325.

Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, 7(1):5–41.

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

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition*, 5(4):443–448.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372.

Recio, A. M. and Godino, J. D. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics*, 48(1):83–99.

Robson, C. (2002). *Real World Research: A Resource for Social Scientists and Practitioner-Researchers,*. Wiley-Blackwell, Oxford, 2nd edition.

Rodd, M. (2000). On mathematical warrants: Proof does not always warrant, and a warrant may be other than a proof. *Mathematical Thinking and Learning*, 2(3):221–244.

Rose, K. (1984). *Elementary Number Theory and Its Application [M].* Addison-Wesley.

Ross, S., Morrison, G., and Lowther, D. (2010). Educational technology research past and present: Balancing rigor and relevance to impact school learning. *Contemporary Educational Technology*, 1(1):17–35.

Rowland, T. (2001). Generic proofs in number theory. In In Cambell, S. and Zazkis, R., editors, *Learning and teaching number theory: Research in cognition and instruction*, pages 157–184. Albex Publishing,westport, CT.

Schacter, J. (1999). The impact of education technology on student achievement. *What the Most current Research Has to Say. San Francisco: Milken Exchange.*

Schnotz, W. (2002). Enabling, facilitating, and inhibiting effects in learning from animated pictures. *Paper presented at the Dynamic Visualizations and Learning, Tubingen, Germany*, Available: http://www.iwm-kmrc.de/workshops/visualization/schnotz.pdf.

Schnotz, W. and Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology Research and Development*, 53(3):47–58.

Schoenfeld, A. (1982). Measures of problem-solving performance and of problem-solving instruction. *Journal for Research in Mathematics Education*, 13(1):31–49.

Segal, J. (1999). Learning about mathematical proof: Conviction and validity. *Journal of Mathematical Behavior*, 18(2):191–210.

Selden, A. and Selden, J. (1999). The role of logic in the validation of mathematical proofs. *Tennessee Technological University Mathematics Department Technical Report*, 1:1–13.

Selden, A. and Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, 34(1):4–36.

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

Shepherd, M. (2005). Encouraging students to read mathematics. *PRIMUS*, 15(2):124–144.

Steele, C., Spencer, S., and Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, 34:379–440.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4):295–312.

Sweller, J. and Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12(3):185–233.

Tall, D. (1989). The nature of mathematical proof. *Mathematics Teaching*, 127:28–32.

Tall, D. (1991). *Advanced mathematical thinking.* Dordrecht, The Netherlands: Kluwer.

Tall, D. and Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12(2):151–169.

Thurston, W. (2006). On proof and progress in mathematics. *18 Unconventional Essays on the Nature of Mathematics*, pages 37–55.

Van Den Haak, M., De Jong, M., and Jan Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology*, 22(5):339–351.

Van Dijk, T. and Kintsch, W. (1983). *Strategies of Discourse Comprehension.* Academic Press, New York.

Vinner, S. (1976). The naive concept of definition in mathematics. *Educational Studies in Mathematics*, 7(4):413–429.

Vinner, S. (1991). The role of definitions in the teaching and learning of mathematics. *In D. Tall (ed.) Advanced mathematical thinking*, Kluwer Academic Publishers, Netherlands:65–81.

Wang, F. and Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development*, 53(4):5–23.

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, 48(1):101–119.

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

Weber, K. (2005). Problem-solving, proving, and learning: The relationship between problem-solving processes and learning opportunities in the activity of proof construction. *Journal of Mathematical Behavior*, 24(3-4):351–360.

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education*, 39(4):431–459.

Weber, K. (2009). Mathematics majors' evaluation of mathematical arguments and their conceptions of proof. In *Proceedings of the 12th Conference for Research in Undergraduate Mathematics Education*. Available: http://mathed.asu.edu/crume2009/proceedings.html.

Weber, K. (2010). The pedagogical practice of mathematicians with respect to proof. In *Proceedings of the 13th Annual Conference on Research in Undergraduate Mathematics Education*. Available: http://sigmaa.maa.org/rume/crume2010/Abstracts2010.htm.

Weber, K. and Alcock, L. (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics*, 56(2):209–234.

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

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

West, J., Haake, A., Rozanski, E., and Karn, K. (2006). eyepatterns: software for identifying patterns and similarities across fixation sequences. In *Proceedings of the 2006 Symposium on Eye tracking Research & Applications*, pages 149–154. ACM.

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

Yin, R. (1992). The case study method as a tool for doing evaluation. *Current Sociology*, 40(1):121–137.

Yin, R. (1994). *Case study research: Design and methods*, volume 2. Thousand Oaks,CA: Sage Publications.

Yoshida, M. (2008). Think-aloud protocols and type of reading task: The issue of reactivity in l2 reading research. In *Selected proceedings of the second language research Forum. Ed. Melissa Bowles et. al*, pages 199–209. Available: http://www.lingref.com/cpp/slrf/2007/paper1745.pdf.

Zhu, E. and Bergom, I. (2007). Lecture capture: A guide for effective use. *CRLT Occasional Paper*, (27).