Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author’s right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author’s permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
Does pre-training reduce the cognitive load of learning complex science information in authentic settings: science classrooms?

Carolyn Haslam

Abstract

This study investigated the cognitive effects of employing the strategy of pre-training on students learning of complex ideas in science. A preliminary study confirmed that learning graphing skills in physics was a suitably complex and appropriate context for the main study. Students from three classes in each of eight different secondary schools were involved in learning science ideas using multimedia presentations in the topic of motion, which includes graphs, in their own classrooms. The students were all novices for this topic. This study contrasted the use of pre-training as a strategy to reduce the cognitive load and enhance the understanding and learning of complex information with no pre-training. The experimental treatment group received pre-training before being presented with the science ideas (teaching), the other two groups only received the teaching, Treatment group 2 once through and Treatment group 3 twice through. Cognitive load measures (during and after learning) were employed and performance measures included the difference between pre and post test scores, for high and low level learning, and for calculation and graphing ideas. In addition instructional efficiency was calculated. The effect of the pre-training strategy on males and females, different decile schools, and different ethnicities in New Zealand were also investigated. Results suggest that pre-training reduced the cognitive load and improved learning when complex information in science was presented and increased instructional efficiency for all students irrespective of gender, decile, and ethnicity when learning in an ecologically valid setting.
Dedication

This thesis is dedicated to Edwin Roy Davidson (Ed) ONZM
Acknowledgements

Firstly I would like to acknowledge the enormous contribution of my main supervisor Dr. Richard Hamilton. For his patience, expert advice, ongoing support and encouragement. The completion of this thesis is a testament to his guidance.

Thanks also to the schools who were involved in this study, to the teachers who willingly gave me access to classes so I could teach their students. Also to the students who were involved and who put in the effort to learn the information.

Thanks to Peter and Mum for their interest, support and encouragement.

In addition, sincere appreciation to Andrew Rogers who proof read my thesis.

Lastly, I would like to acknowledge Academic Consulting for editing and checking the Results, Discussion and Conclusion Chapters.
# Table of Contents

Chapter 1: Introduction ................................................................. 1

Study 1 .................................................................................................. 2

Study 2 .................................................................................................. 3

Organisation of the Thesis ................................................................. 5

Chapter 2: Study 1 .............................................................................. 8

- Literature review ........................................................................... 8

- Method ............................................................................................. 13

  - Materials and procedures ......................................................... 14

  - Analysis ....................................................................................... 17

- Results ............................................................................................. 18

  - Textbook analysis ........................................................................ 19

  - NCEA level 1 Science exam analysis ........................................... 20

  - Graphing questionnaire analysis ................................................ 21

Discussion and Conclusion ................................................................. 23

Chapter 3: Literature Review ............................................................. 25

- The Cognitive Processing System ..................................................... 25

  - Working memory ......................................................................... 25

  - Long-term memory ...................................................................... 27

- Cognitive Load Theory .................................................................... 29

  - Extraneous cognitive load ............................................................ 33

  - Intrinsic cognitive load ................................................................ 33

  - Germane cognitive load ............................................................... 34

  - Measuring cognitive load ............................................................... 36

  - Using cognitive load measures and performance measures to
calculate instructional efficiency .................................................... 40

  - Recent advances in cognitive load theory .................................... 43

  - Critique of cognitive load theory .................................................. 45

- The Cognitive Theory of Multimedia Learning (CTML) ..................... 56

  - Processing pictures and words in CTML ...................................... 57

  - Main assumptions ......................................................................... 57

  - Meaningful learning in CTML ...................................................... 60

  - Categories of cognitive processing and basic multimedia principles ........................................................................ 60

  - Advanced multimedia learning principles ................................... 63

  - Boundary conditions ..................................................................... 64
Analysis ........................................................................................................................................ 145
  Participant ability (mathematics, reading and spatial ability), behaviour (class level), and prior knowledge measures analyses .......... 146
  Research Question 1: Cognitive load analysis ......................................................... 147
  Research Question 2: Performance measures analysis ............................................. 149
  Research Question 3: Instructional efficiency measures analysis .................. 149
  Research Question 4: The influence of gender, decile, and ethnicity on dependent measures analyses .................................................. 151

Chapter 5: Results ...................................................................................................................... 155
  Participant ability and prior knowledge measures .................................................. 156
    Mathematics and reading ability ......................................................................... 157
    Spatial ability ....................................................................................................... 157
    Pre test measures ................................................................................................ 159
    Summary ................................................................................................................ 159
  Behaviour Rating scale measures ............................................................................. 160
  Cognitive load measures ............................................................................................. 160
    Summary ................................................................................................................ 163
  Performance measures ................................................................................................. 164
    Learning .................................................................................................................. 164
    High versus low order questions ........................................................................... 165
    Calculation versus graphing questions .................................................................. 166
    Summary ................................................................................................................ 166
    Instructional Efficiency ............................................................................................ 167
    Summary ................................................................................................................ 169
  The influence of gender, decile, and ethnicity on dependent measures .... 169
    Gender ................................................................................................................... 169
    Decile .................................................................................................................... 171
    Ethnicity ................................................................................................................ 178
    Summary ................................................................................................................ 180

Chapter 6: Discussion ............................................................................................................... 181
  Research Question 1 ................................................................................................. 181
  Research Question 2 ................................................................................................. 185
  Research Question 3 ................................................................................................. 188
  Research Question 4 ................................................................................................. 190
    Gender findings ....................................................................................................... 190
    Decile findings ........................................................................................................ 192
    Ethnicity findings .................................................................................................... 193
  Spatial ability ............................................................................................................... 193
  Theoretical and educational significance ................................................................. 194
    Theoretical Implications ......................................................................................... 194
    Educational implications ......................................................................................... 196
  Future Research ........................................................................................................ 198
    Target content areas ............................................................................................... 198
Specific methodological changes ................................................................. 199
Comparisons with other methods .............................................................. 200
Different target populations .................................................................... 201
Participant characteristics ..................................................................... 201
Improving overall performance .............................................................. 202
Conclusion .................................................................................................. 203

Chapter 7: The Challenges of Doing Research in Schools (A Personal Account) ................................................................. 204

The challenges of making contact and getting consent .................. 205
The challenges of getting Ethics approval for research in schools ....... 205
The challenges of making initial contact with schools and explaining the research ................................................................. 207
The challenge of gaining consent to work with students ............... 210

The challenges of conducting research ........................................... 211
Being organised ...................................................................................... 212
Being consistent .................................................................................. 213
Maximising the data collection ............................................................... 214
Recognising their contribution ............................................................... 215
The challenges of increased time commitments and timetabling visits to collect data ................................................................. 215
Using technology in schools ................................................................. 218
The challenges of managing the class ...................................................... 219
The challenges of teaching ESOL students ............................................. 221

Appendices .............................................................................................. 223

Appendix Table of contents ................................................................. 223

References ............................................................................................... 301
List of Tables

<table>
<thead>
<tr>
<th>Table number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Means and SD for time spent on graphing in science Years 9-11</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Studies investigating the Pre-training and the Isolated-elements strategies</td>
<td>74</td>
</tr>
<tr>
<td>3</td>
<td>TIMMS means of the four main ethnic groups in NZ from 2002/2003 to 2010/2011</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>Means in PISA 2006 for NZ students by ethnicity</td>
<td>101</td>
</tr>
<tr>
<td>5</td>
<td>Percentage of New Zealand students in the highest and lowest proficiency levels in PISA 2006 by ethnicity</td>
<td>102</td>
</tr>
<tr>
<td>6</td>
<td>Achievement Standard 90940 Results for Māori, Pasifika, NZ European and Asian</td>
<td>105</td>
</tr>
<tr>
<td>7</td>
<td>Achievement Standard 90944 Results for Māori, Pasifika, NZ European and Asian</td>
<td>106</td>
</tr>
<tr>
<td>8</td>
<td>Achievement Standard 90948 Results for Māori, Pasifika, NZ European and Asian</td>
<td>107</td>
</tr>
<tr>
<td>9</td>
<td>Summary of Secondary Schools involved in the experimental treatment</td>
<td>120</td>
</tr>
<tr>
<td>10</td>
<td>SOLO taxonomy classification of the pre test questions</td>
<td>123</td>
</tr>
<tr>
<td>11</td>
<td>Treatment groups used in Study 2</td>
<td>134</td>
</tr>
<tr>
<td>12</td>
<td>Online Cognitive load question and Times 1 to 4 relating to Treatment groups</td>
<td>141</td>
</tr>
<tr>
<td>13</td>
<td>Means and Standard deviations for, PAT Mathematics and Reading, Spatial ability and Pre test for all Treatment groups.</td>
<td>158</td>
</tr>
<tr>
<td>14</td>
<td>Means and Standard deviations for Spatial ability and Gender</td>
<td>159</td>
</tr>
<tr>
<td>15</td>
<td>Means and Standard deviations for the Online and Summative cognitive load measures for all treatment groups</td>
<td>163</td>
</tr>
<tr>
<td>16</td>
<td>Means and Standard deviations for the</td>
<td>165</td>
</tr>
</tbody>
</table>
Performance measures for all treatment groups

17 Means and Standard deviations for the Instructional Efficiency measures for all treatment groups 168

18 Means and Standard deviations for the Instructional Efficiency measures for Treatment 1 and 3 comparing the first and second times of teaching respectively 168

19 Means and Standard deviations for Treatment and Gender 171

20 Means and Standard deviations for Online and Summative cognitive load for Treatment and Decile 174

21 Means and Standard deviations for Difference and Efficiency for Treatment and Decile 175

22 Means and Standard deviations for Treatment and Ethnicity 179
# List of Figures

<table>
<thead>
<tr>
<th>Figure number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Percentage of graphs in a sample of 14 New Zealand science textbooks for use with Year 9-11.</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>Total number of graphs in New Zealand national exams at Level 1 from 2004 – 2011.</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>Teacher reasons given in response to the questionnaire for spending more time on graphing in science in Years 9-11.</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>Teacher reported issues concerning the teaching of graphing skills</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>Slide 9 A Sample slide from Initial Phase PowerPoint</td>
<td>121</td>
</tr>
<tr>
<td>6</td>
<td>Slide 3 and 9 from pre-training PowerPoint</td>
<td>125</td>
</tr>
<tr>
<td>7</td>
<td>Slide 5 and 14 from the teaching PowerPoint</td>
<td>126</td>
</tr>
<tr>
<td>8</td>
<td>Slide 7 and 11 from the Spatial ability PowerPoint</td>
<td>132</td>
</tr>
<tr>
<td>9</td>
<td>Paper Folding Test Question 1 on Spatial Ability Tests Answer Sheet</td>
<td>133</td>
</tr>
<tr>
<td>10</td>
<td>SPSS generated interaction plot for summative question 2 (effort to learn) for decile x treatment</td>
<td>176</td>
</tr>
<tr>
<td>11</td>
<td>SPSS generated interaction plot for summative question 5 (pace) for decile and treatment</td>
<td>176</td>
</tr>
<tr>
<td>12</td>
<td>SPSS generated interaction plot for difference (learning for treatment x decile)</td>
<td>177</td>
</tr>
<tr>
<td>13</td>
<td>SPSS generated interaction plot for efficiency for treatment x decile</td>
<td>177</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Research into problem-solving by Sweller in the 1970’s led to the development of the cognitive load theory (Sweller, 1988, 1989, 2010a; Sweller, van Merriënboer, & Paas, 1998). This theory has been actively researched internationally for over 40 years and as a result a comprehensive set of guidelines and principles for the presentation of instructional materials have been developed. Initially, the research emphasis was on simplifying instructional materials so that making sense of these materials would not overload the human mind. However, more recent research has focused on strategies within the cognitive load framework that enable learners to make sense of complex or difficult material without causing cognitive overload (Pollock, Chandler, & Sweller, 2002). Parallel research by Mayer into the learning of instructional materials that include pictures and words (multimedia) which takes into account cognitive load theory principles and guidelines has also resulted in strategies to reduce cognitive overload when learning difficult or complex information (Mayer, 2005b; Mayer, Mathias, & Wetzell, 2002; Mayer, Mautone, & Prothero, 2002; Mayer & Pilegard, 2014). This thesis will explore the use of one of these strategies called pre-training, in an ecological setting after initially establishing the suitability of graphing as a complex context.

The focus of this thesis is on the use of graphing skills in a science context by secondary school students in New Zealand. The thesis consists of two parts, Study 1 - an assessment of the nature and extent of graphing within science classrooms, relevant science textbooks, and science national tests employed within New Zealand secondary schools and Study 2 - an intervention aimed at facilitating the acquisition of graphing skills in an authentic educational setting in New Zealand secondary schools.

Graphing skills involve the skill of construction and interpretation of graphs. In science, graphs are used extensively by scientists to present scientific data in a concise manner, to show patterns and relationships in data collected, and to aid in the analysis of such data (Ates & Stevens, 2003; Hipkins, 2011). In addition,
graphs are used extensively in newspapers and magazines distributed to the general public (Galesic & Garcia-Retamero, 2011; Trumbo, 1999). Learning graphing skills and science concepts are important aspects of scientific literacy, however, understanding and creating graphic representations and learning science concepts are complex processes and intellectually demanding (Hipkins, 2011; Mautone & Mayer, 2007; Shah & Freedman, 2011). According to cognitive load theory this could overload students and interfere with further learning. One way suggested by research into cognitive load is to present the material essential for understanding the concepts in two stages. In stage one names and characteristics of the main parts or ideas are presented to provide the learners with some prior knowledge but no understanding of the concepts. Stage two is when all the material required for full understanding of the concepts is presented. This is the pre-training strategy (Mayer, 2005b; Mayer & Pilegard, 2014) or the isolated-interacting elements learning approach (isolated-elements strategy) (Pollock, Chandler, & Sweller, 2002).

The remainder of this chapter will provide an overview of the thesis, reasons for the choice of research topic, and justification outlined, the method and research questions will be explored and the unique contribution to the research environment detailed.

**Study 1**

Graphing is a skill used extensively to communicate in science (Ates & Stevens, 2003) and one of the Objectives of the Nature of Science strands in the New Zealand Curriculum is called “Communicating in science” (Ministry of Education, 2007). Research in science education places significant importance on the teaching and learning of graphing skills (Shah & Freedman, 2011; Shah & Hoeffner, 2002) but it is a difficult set of skills to teach and learn (Mautone & Mayer, 2007; Canham & Hegarty, 2010). Despite these difficulties, there has been no science education research in New Zealand on the nature and extent of graphing use in science classrooms, textbooks, and national exams. Study 1 involves document analysis in order to assess the nature and extent to which graphs are used in textbooks and NCEA level one examinations and the collection of survey data to provide information on the extent of graphing in science Years 9-
11 at secondary school within a wide range of schools in Auckland and Northland. This study will look at the content areas (disciplines) of physics (physical world), chemistry (material world), and biology (living world) in the science learning area within the New Zealand Curriculum (Ministry of Education, 2007). When referring to Year 11 this includes the physics, chemistry and biology content within science national exams. The research questions which will be explored in Study 1 are:

1. What is the nature and extent of graph use in secondary science textbooks?

2. How does graph use and difficulty based on national science exam question type (achieved, merit or excellence) vary across the science disciplines of physics, chemistry, and biology?

3. What is the nature and extent of graph use in secondary science classrooms and to what extent is graphing perceived as an important skill by secondary science teachers?

4. What common difficulties do teachers report students’ facing when working with graphs?

The answers to these research questions will help establish the perceived importance of graphing as a skill in science teaching and will be used to inform and describe the context for Study 2.

**Study 2**

This is an intervention study using the strategy of pre-training to help students in secondary school science learn complex information including the skills of graphing by artificially reducing the cognitive load of learning this material. Science is not an easy subject to teach and learn. Many of the concepts especially in the senior school are complex. In New Zealand one of the main achievement issues in science is the underachieving of Māori and Pasifika students and this is evident in the New Zealand Qualifications Authority (NZQA)
data regarding the national exams where lower numbers of Māori and Pasifika students earn merit and excellence (the two highest) grades in national exams especially in physics (NZQA, 2012, 2013, 2014). Material assessed for merit and excellence questions is complex. One of the aims of this research is to better prepare students for complex science content and potentially provide a way to raise the achievement of all students using the strategy of pre-training but in particular the achievement of Māori and Pasifika students. Another related issue is the potential contributions of gender difference in science achievement favouring males on the impact of pre-training.

All previous studies into the effects of pre-training have been conducted in a laboratory setting with undergraduate students learning content unrelated to their tertiary study. Mayer (2005a) suggested that this strategy needed to be trialled in a real setting i.e., in a classroom, with students learning realistic information in segments of greater than 10 minutes, and including cognitive load measures. Study 2 involves 606 Year 9 and 10 students in nine Auckland secondary schools, and assesses the effects of a pre-training strategy to help science students prepare to and learn complex information which includes graphing skills in the physics topic of motion. The selected students are novices, that is, they will not have studied the topic of motion at secondary school. However, it is a topic which they will be required to study either as part of their course later in the same year or for national exams early the following year so prior exposure to this material will be beneficial. Outcome measures of interest in this study are the cognitive load of the pre-training strategy intervention and related instruction, near and far transfer of target information, and the efficiency of learning. Appropriate participant measures will also be employed to attempt to control for potential moderating variables (prior knowledge and skills). There are three treatment groups, the experimental condition receives teaching including pre-training and a PowerPoint presentation of to-be-learned information, another group receives the PowerPoint presentation once (normal teaching practice), and the third group receives the PowerPoint presentation twice. Study 2 aims to explore the following research questions.
1. Does pre-training reduce the cognitive load associated with learning complex science information involving graphs as measured during and after learning?

2. Within the context of learning complex information involving graphs, what is the effect of pre-training on secondary science students’ overall learning, performance on different levels of question (high versus low order), and performance on different content questions (calculation versus graphing)?

3. Does pre-training increase instructional efficiency for secondary science students when learning complex information involving graphs?

4. Are there differential effects on cognitive load, learning (difference scores) and instructional efficiency with respect to gender, decile and ethnicity when using pre-training in a secondary science context involving graphs?

The two aims of this study are firstly, to investigate whether the strategy called pre-training (based on the principles of cognitive load theory) reduces cognitive load and helps secondary science students learn complex information within an authentic classroom setting, secondly, to assess the usefulness and value of on-line and post measures of cognitive load in authentic educational settings. Chapter One will conclude with an outline of the content in the remaining chapters of this thesis.

Organisation of the thesis

Chapter 2: Study 1.

This chapter will outline in detail Study 1. It will include the Literature review covering research into graphing, difficulties students experience when constructing and interpreting graphs, and ways to address these difficulties while teaching the skills of graphing, the factors that influence graphing ability and gender differences
with respect to graphing ability. The remaining sections will look at the Method, Results, Discussion and Conclusion for Study 1.

Chapter 3: Literature Review Study 2.

This chapter includes a review of the relevant literature associated with Study 2, the two main theories associated with this study are discussed in detail, namely cognitive load theory and the cognitive theory of multimedia learning (CTML). Also included are discussions of the main assumptions, recent advances, future directions, and a critique of both theories including a description of pre-training and the isolated-elements teaching approach. This study involves the presentation of pictures and words so a section discussing the Integrated Text and Picture Comprehension (ITPC) model of making sense of pictures and words is included as a comparison with the CTML. The strategy of pre-training is a preinstructional strategy which has similarities to advance organisers and graphic organisers which are discussed in the next section. As this study also involves the learning of complex information in physics the next section discusses a model for the teaching of complex skills, the four-component instructional design (4C/ID) model. This section is followed by a section on the impact of spatial ability on graphing and science achievement as well as the different patterns of performance and behaviour of male and female students within and related to science courses. The last section discusses Māori and Pasifika achievement in science in New Zealand. The evidence comes from international and national studies and data from national exams. Reasons for this underachievement are discussed as are New Zealand Government initiatives to lift achievement of Māori and Pasifika students.

Chapter 4: Method Study 2.

This chapter discusses the research design, the participants, the materials, and equipment used, the procedure, including justification of the cognitive load measures used in Study 2. In this chapter details of the planned analysis will be outlined.
Chapter 5: Results Study 2.

In this chapter results will be discussed in the following six sections, and will be organised according to the research questions posed. After an introductory section the results of analyses regarding participant ability and prior knowledge will be discussed, the next three sections will report on analyses regarding cognitive load measures, learning, and instructional efficiency measures. The final section will report on the influence of gender, decile, and ethnicity on the strategy of pre-training.

Chapter 6: Discussion and Conclusion Study 2.

This chapter will discuss the research findings with respect to cognitive load (online and summative), learning, and instructional efficiency and the influence of pre-training on the potential moderating factors of gender, decile, and ethnicity and how the findings compare to closely related studies. Then it will consider the contributions to the two theories that are associated with this study (cognitive load theory and CTML) as well as explore the potential benefits of implementing the strategy of pre-training in schools. After this, future directions for further study into the strategy of pre-training to expand and refine the findings of this study will be suggested. Lastly, this chapter will conclude with a summary of the overall findings of Study 2.

Chapter 7: The Challenges of doing Research in Schools- A personal account.

The final chapter is a personal account by the researcher of the challenges of doing ecologically valid research in educational institutions with whole classes of students. It looks in detail at the dilemmas of maintaining a positive relationship with the school for future research, fulfilling the research requirements of a University, and being mindful of the primary purpose and needs of teachers, pupils, and schools.
Chapter 2

Study 1

“On the plus side there’s more white space.”

Permission given by Mark Anderson to use this cartoon

Literature Review

Graphs are tools that scientists, mathematicians, geographers, other academics, and professionals use extensively to communicate information to their peers and the general public (Trumbo, 1999). However, people differ in their ability to read graphs, often referred to as graphic literacy (Okan, Garcia-Retamero, Cokely, & Malando, 2012; Okan, Garcia-Retamero, Galesic, & Cokely, 2012). Galesic and Garcia-Retamero (2011) found that about one third of the public in the USA and Germany that they surveyed had low graphic literacy (the ability to understand information presented in graphs, i.e. graph comprehension or interpretation) and Brasell and Rowe (1993) found that at least one fifth of high school physics students lacked adequate graphing skills. The context for this study is science and so further reference to the use of graphing will be related to this discipline only. Graphing is an important skill (that includes the skills of interpretation and
construction of graphs) that students studying science need to know (Bowen & Roth, 2003; Brasell & Rowe, 1993; Hipkins, 2011; Mokros & Tinker, 1987; Shah & Hoeffner, 2002) but one that they find difficult (Bowen & Roth, 1998; Culbertson & Powers, 1959; Hipkins, 2011; Mautone & Mayer, 2007; Shah & Hoeffner, 2002; Woolnough, 2000).

Graphs are spatial aids (Moore, 1993; Okan, Garcia-Retamero, & Galesic, 2012) or graphic representations (Schnotz, 1993) that are used extensively by scientists to present scientific data in a concise manner (Brasell & Rowe, 1993; Onwu, 1993; Roth, Pozzer-Ardenghi, & Young Han, 2005; Shah, 2002), to show patterns and relationships in data collected (Roth & Bowen, 2003), and to aid in the analysis of scientific data (Roth, Bowen, & McGinn, 1999; Shah & Hoeffner, 2002). Graphs are used extensively in scientific articles, particularly in biology and popular media (Roth, Bowen, & McGinn, 1999; Shah & Freedman, 2011; Shah, Freedman, & Vekiri, 2005; Tairab & Al-Naqbi, 2004), but are found less frequently in science textbooks (Roth, Bowen, & McGinn, 1999). Learning graphing skills is an important aspect of scientific literacy (Ates & Stevens, 2003; Bowen & Roth, 2003; Hipkins, 2011; Shah & Hoeffner, 2002) and educational researchers have begun to focus on learning environments to help students become literate in reading, producing, and using graphs in science (Glazer, 2011; Roth, Pozzer-Ardenghi, & Young Han, 2005). Understanding graphic representations is a complex process (Friel, Curcio, & Bright, 2001; Gerber, Boulton-Lewis, & Bruce, 1995; Hipkins, 2011). Their complexity is related to the many different processes learners must undertake to understand the graph (Shah & Carpenter, 1995; Shah, Mayer, & Hegarty, 1999). Graphs are complex representations which often include many interacting elements which need to be looked at simultaneously to understand the information presented in the graph; these include axes, units, points and lines of different shapes size and colour, keys, and labels (Glazer, 2011; Roth, Pozzer-Ardenghi, & Young Han, 2005). Novices in graphing therefore often experience cognitive overload due to this complexity (Brasell, 1987). A number of studies have investigated the use of computers to reduce the complexity and hence cognitive load on learners (Adams & Shrum, 1990; Berg & Phillips, 1994; Brasell, 1987; Mokros & Tinker, 1987).
When looking at the graphing sub-skills of construction and interpretation, students find interpretation the most difficult (Leinhardt, Zaslavsky, & Stein, 1990; Mokros & Tinker, 1987), but when constructing graphs students are often confused by the scales on the axes (Shah & Carpenter, 1995). Research in classrooms has highlighted a number of difficulties that students experience when both constructing and interpreting graphs in science (Berg & Phillips, 1994; Biechner, 1994; Mokros & Tinker, 1987; Roth, Bowen, & McGinn, 1999; Shah & Carpenter, 1995; Shah & Hoeffner, 2002). The main difficulties highlighted in the literature include confusion between the slope and height of a line graph, reading the graph as a picture (reading the slope of the graph as upward motion), and confusion when interpreting a graph between focusing on a point or an interval. In addition, difficulties also arise if the context of the information presented in the graph is unfamiliar, the students are not familiar with graphing conventions (Friel, Curcio, & Bright, 2001; Glazer, 2011; Hipkins, 2011; Roth & McGinn, 1998; Shah & Freedman, 2011), and if the graph does not have its origin at 0 (Beichner, 1994). These difficulties can affect achievement in science as students perform lower than expected on questions which include graphs (Forster, 2004; Onwu, 1993; Roth & McGinn, 1997; Woolnough, 2000). Interpretation and construction of graphs is an important skill in physics which has been found to significantly influence performance in physics (Forster, 2004).

A suggested reason for students’ lower levels of achievement in graphing related tasks and disciplines is the lack of time spent on teaching and practise of graphing skills in science classrooms (Brasell & Rowe, 1993; Curcio, 1987; Friel, Curcio, & Bright, 2001; Fry, 1981; Phillips, 1997; Shah & Hoeffner, 2002). Effective teaching and practise of graphing skills can have a positive effect on improving science student’s graphing skills (Ates & Stevens 2003; Brasell & Rowe, 1993; Friel, Curcio, & Bright, 2001; Hipkins, 2011; Phillips, 1997; Mautone & Mayer, 2007; Shah & Hoeffner, 2002; Tairab & Al-Naqbi, 2004). Examples of ways to encourage competency of graphing skills are:

• using professional development sessions to make teachers aware of the difficulties students experience and encourage explicit teaching of those skills not just using graphs as part of the context being taught
• teaching graph design
• teaching why different graphs are used for different purposes
• critiquing and comparing different graphs used in textbooks and the media
• using questions to guide students to relate the content of the graph to the context it is used in to assist them in identifying trends and generalising
• using real data that the students collect to practise their graphing skills
• using cognitive aids (scaffolds) to guide cognitive processing to help students make sense of graphically presented information.

Some researchers view graphing ability as being related to practise alone and suggest that it is akin to learning a language, where if you don’t know how to speak a language it is not related to your ability but a lack of exposure and practice (Roth & McGinn, 1997). However, Shah and Hoeffner (2002) disagree and suggest that a well developed graphing schema (that comes from practice) and also the characteristics of the graph (e.g., type of graph and content knowledge related to the information presented in the graph) determine graphing success. This is very important in a science context which uses graphs such as the physics topic of motion (mechanics).

Another factor which can influence success with interpreting and constructing graphs is spatial ability (Gerber, Boulton-Lewis, & Bruce, 1995; Moore, 1993; Trickett & Trafton, 2006; Uttal et al., 2013). Spatial visualization has been found to be an important skill in solving physics problems especially those relating to the interpretation of kinematics (motion) graphs as these involve building and manipulating mental images of the motion that is represented by lines on a graph (both speed-time and distance-time graphs) (Kozhevnikov, Motes, & Hegarty, 2007). Shah (2002) and Culbertson & Powers (1959) also found that graph format (type) influenced the difficulty students experienced in graph comprehension and construction. Of the two types of graphs commonly used to display information students find line graphs more difficult to draw and interpret than bar graphs (Crooks & Flockton, 1995, 1996; Onwu, 1993; Shah & Carpenter, 1995; Wavering 1989; Zacks & Tversky, 1999). In addition, Stewart, Cipolla, and Best (2009) and Canham and Hegarty (2010) found that extraneous information and graph design (2-dimensional or 3-dimensional) can influence the difficulty students experience in graph comprehension. That is, students find the addition of extraneous
information, especially detailed graphics and 3-D graphs more difficult when interpreting complex information.

Studies have also found that gender differences exist in mathematics and science with respect to graphing tasks especially in physics topics (Forster, 2004; Hein, 1997). The study of physics involves applying mathematical ideas and constructing and interpreting graphs. Beeken (2014) and Deacon (1999) have described graphs as the visualization of mathematical ideas related to the physical world. Mathematical ability can also influence science achievement (Snow, 2010). Graphing skills are taught in both mathematics and science but it has been found that students do not often transfer graphing skills learned in mathematics to graphing in science (Roth, Bowen, & McGinn, 1999; Woolnough, 2000). It has been found that males perform consistently higher on data analysis and graphing tasks than females (Casey, Nuttal, & Pezaris, 2001; Gamer & Engelhard, 1999; Lummis & Stevenson, 1990). Hyde, Fennema, and Lamon (1990) found that diagrammatic information was more easily processed by males and this includes graphs.

In a New Zealand context the National Educational Monitoring Project (NEMP) has been monitoring children aged 8-9 and 12-13 years old (Year 4 and 8) in science in New Zealand schools since 1995. NEMP reports from assessment in 1995, 1999, 2003 and 2007 found that students make good progress on graphing skills between Year 4 and 8, but overall students find interpretation more difficult than construction (Crooks & Flockton, 1995,1996; Crooks, Smith, & Flockton, 2008; Hipkins & Kenneally, 2003; Praat, 1999).

Despite extensive research into graphing world-wide for an extended period of time there has been little research on graphing in science education in New Zealand and none looking into the prevalence of graphing use in different science disciplines in junior and Year 11 science teaching, textbooks, and national exams at Level one. Based on the reported difficulties students have and the importance of graphing in science, particularly in physics, a preliminary study of graphing skills was conducted to determine firstly, the amount of space given to graphs in textbooks, secondly, the number of graphs in the National Certificate of
Educational Achievement (NCEA) Level one national science exams, and thirdly, the amount of time teachers spend on graphing.

The results of Study 1 will be used to help frame an Intervention study that includes teaching the skill of graph interpretation and construction when introducing the topic of motion, which includes speed-time and distance-time graphs from the science Achievement Standard 90940 at Year 11. It is hoped that explicit teaching on graphing, the use of visuals, and reducing cognitive load will enhance the learning of complex ideas in the physics topic of motion (mechanics).

Study 1 aims to investigate the following Research Questions.

1. What is the nature and extent of graph use in secondary science textbooks?

2. How does graph use and difficulty based on national science exam question type (achieved, merit or excellence) vary across the science disciplines of physics, chemistry, and biology?

3. What is the nature and extent of graph use in secondary science classrooms and to what extent is graphing perceived as an important skill by secondary science teachers?

4. What common difficulties do teachers report students’ facing when working with graphs?

**Method**

This study involved both document analysis and a survey to determine the importance placed on the development of graphing skills in learning science, in national exams, and in science textbooks. The document analysis was carried out due to a lack of available data in New Zealand on graph use and the development of graphing skills in science. In addition, graphing is an important skill as identified by the New Zealand Council for Educational Research (NZCER), which is valued
in a New Zealand science context and importance is placed on the teaching and mastering of these skills (Hipkins, 2011). Given that graphing is viewed as an important skill and is valued within the New Zealand context (Hipkins, 2011), it would be expected that one would see graphing featured more widely within the textbooks used in secondary school science classrooms and that graphing would be an important focus of national exams (NCEA).

The quality of data obtained will depend on the number and range of textbooks involved and the data collected. For this reason, the most commonly used textbooks were selected and all of these were included in the analysis. Regarding the NCEA exams, all the external NCEA exams from 2004 to 2011 were included, 2004 was the earliest year that a full set of three external exams were available at the time of this study. Secondly, a questionnaire was distributed as this was a quick way to collect large amounts of information without having to make appointments to visits the teachers involved in their schools and this meant that they could reply when they had time to complete the questionnaire. The questionnaire was anonymous to encourage honest responses. There is the disadvantage that questions cannot be explained and misunderstandings could occur when answers are written by the respondent and also when recorded by the researcher, however, in this study respondents were given the option of clarifying their answers and adding comments at the end of the questionnaire. The questions were also simplified and many of them required answers to be circled as a way of reducing any confusion. The purpose of surveying the teachers was: a) to ascertain if teachers attributed the same value and importance to graphing skills as New Zealand educational researchers, and b) to compare teacher value to the implicit value or importance evidenced by the appearance of graphing within relevant textbooks and national exams.

**Materials and Procedures.**

**Textbooks.** Fourteen science textbooks that students in Years 9-11 taking a typical science course commonly used were selected for analysis. The textbooks were published by the five largest science textbook publishing companies that are represented in New Zealand. Of these 14 science textbooks, nine were produced for use with Year 9 and 10, three for use with Year 11, and
two were designed to assist with the teaching of science skills over the range from Years 9-13 but aimed predominately for use with Year 11. Approximately 15 teachers in Auckland and Northland schools were verbally surveyed to determine the textbooks used with Year 9-11 science classes and indications were that the 14 textbooks selected for inclusion in this study were all widely used as science texts in secondary schools in Auckland and Northland and therefore appropriate to use in an analysis looking into the extent to which graphs are used in science textbooks.

**NCEA Level 1 science exam.** In New Zealand NCEA is the official secondary school qualification. NCEA is a standards-based assessment system which has three levels, Level 1 is the first level. Students must achieve a set number of credits by passing achievement standards to gain each level in NCEA (science is only one of the subjects in which students can gain credits, others include e.g., mathematics and english). Some achievement standards are internally assessed and others are externally assessed. The three external achievement standards in science are physics (mechanics), chemistry (acids and bases) and biology (genetics). It is these three science achievement standards that are the focus of this analysis as they represent all the external science exams (three per year) for eight years from 2004 to 2011. So a total of 24 Level 1 NCEA science exams were used for this analysis (NCEA will be discussed in more detail in Chapter 3). The reference numbers for these achievement standards are Science (90940 - mechanics), Science (90944 – acids and bases) and Science (90948 - genetics). Verbally asking approximately 15 Auckland and Northland science teachers which achievement standards were taught in a typical science course, indicated that these three external achievement standards were taught in most schools as part of a science course at Year 11 in Auckland and Northland, hence making them appropriate to include in an analysis looking at the extent of graph use in national science exams at Year 11.

**Graphing questionnaire.** A questionnaire was developed to collect the responses of teachers to assess the importance they place on developing students graphing skills in science and the difficulties that they encounter in this endeavour. The skill of graphing was divided into teaching and practice, and construction and interpretation. The questionnaire contained seven questions.
Questions 1-3 included completing 3 charts, which asked about the topics in which graphing skills were taught, the time spent on teaching and using graphing skills, and lastly, the total time allocated to science teaching per week in Year 9, 10 and 11. Question 4 was divided into 5 parts labelled a. – e: part a. asked if graphing was an important scientific skill and required a Yes/No answer; part b. asked teachers to rank their students difficulty regarding graph construction on a 7-point scale from very very easy to very very difficult; part c. was identical to part b. except it referred to graph interpretation. Parts d. and e. were identical except d. referred to graph construction and e. to graph interpretation. These two parts asked if students would benefit from spending more time on these two graphing skills, both required a Yes/No answer and then space was given to explain their answer. Question 5 involved teachers identifying which achievement standards were taught in a typical science class at their school. This was used to assess if schools did indeed offer the three external achievement standards, one of which was the focus of the analysis in part two of this study (Study 2). Question 6 asked teachers to identify the teaching methods used to develop graphing skills, options provided included white board, Smart board, PowerPoint presentation, computers/iPads, textbooks, worksheets, and lastly a space for other methods not specified. This question was included to assess if textbooks were used for teaching graphing skills. Question 7 was an open response question which was included to give teachers the opportunity to add clarifying comments regarding their responses or, if they felt strongly about any question to add additional comments. A copy of the questionnaire can be found in Appendix B.

The process of disseminating the graphing questionnaire began by sending Ethics sheets to the Principals of 51 secondary schools in Northland and Auckland. The targeted schools included a range of decile ratings from 1 to 10 and also included Catholic integrated schools, single sexed schools, private schools, and co-educational schools. When these were returned and consent given to approach the Science department the graphing questionnaire was sent to the Head of the Science Department (HOD Science) at the school along with the Ethics sheets and a return self addressed envelope. The Ethics sheets for Study 1 can be found in Appendix A. The HOD Science in a secondary school is also a science teacher so these two terms have and will be used interchangeably. This questionnaire
was only sent to the HOD Science at each school to be completed and returned, 35 questionnaires were completed and returned (69% return rate).

**Analysis.**

The analysis of data is divided into three parts, the first two involved document analysis of science textbooks and national exams, respectively, and the third part involved the analysis of teachers’ responses to a questionnaire.

**Textbook analysis.** Fourteen textbooks were reviewed individually in order to identify: a) the proportion of pages that contain graphs, b) the number of graphs of different sizes, e.g. half or quarter of a page, c) the average amount of space allocated to graphs, d) the proportion of pages of those that contain graphs that are used in teaching graphing as a skill (as distinct from having a graph to illustrate content or as practice and hence part of a question), and e) the proportion of graphs used in the three different science disciplines of physics, chemistry and biology, per book. In addition, the average number of pages between one graph and the next was also reported per book. The information was collated and then analysed. Frequency tables were drawn up and data entered for each book and then collated for all the texts. These were Year 9-11 textbooks from the main publishers and most were written for the 2007 New Zealand Curriculum (Ministry of Education, 2007).

**NCEA Level 1 science exam analysis.** Twenty-four NCEA Level 1 science exams were reviewed and the following aspects of the exams were noted and summarised: a) the proportion of papers that contain a graph or graphs, b) percentage of interpretation versus construction graphs, and c) the sub topic within each paper that most frequently contains graphs. The cognitive challenge of the question was also noted i.e., whether questions which include graphs predominantly involve higher or lower level thinking skills (Anderson & Krathwohl, 2001). In NCEA, during the years 2004 – 2011, the three levels of questioning used were low order questions (achieved questions) mid order questions (merit questions) and high order questions (excellence questions). It was expected that most graphs would be found in physics papers and that the questions will be higher level questions (merit or excellence), involving application eg. using multiple
formula to solve real-life problems. In addition, within the physics paper it was
expected that most questions on the topic of motion would contain a graph. Only
the 3 main science external exams i.e., physics, chemistry, and biology, were
included. For the years 2004 -2010, these were science achievement standards
(AS) 90191, 90188, and 90190, for 2011, 90940, 90944, and 90948 respectively.

Graphing questionnaire analysis. The statistical program used for this
analysis was SPSS, in which frequencies, descriptive data, and ANOVAs were
performed in order to explore differences between the different years, content
areas, and types of graphing skills focused on within the teaching cycle. The
comparisons of interest were: a) the percentage of time teachers spend on
graphing in any one year from Year 9 to Year 11, b) in which science discipline
(physics, chemistry, or biology) do teachers spend the most time on graphing, c)
the amount of time spent on graphing interpretation and graphing construction in
each topic and at each level (to determine if there are significant differences), d)
teacher’s rating of how important graphing is as a skill in science, e) the difficulty
teachers perceive students have with interpretation and construction of graphs,
and f) the methods teachers use to teach graphing. It was expected that the most
graphing would be done in Year 11 as preparation for NCEA exams and more in
physics that other disciplines at Year 11 as the topic of motion (mechanics) relies
heavily on the use of graphs.

Results

This study was concerned with determining the nature and extent of graph use in
secondary school science classrooms, textbooks used in teaching science, and
national exams, which assess learning in science. This section will be organised
according to the research questions and the findings will be presented to address
these questions as determined by analysis of textbooks and national exams and
surveys of Heads of Department: Science.
Textbook analysis.

Research Question 1. What is the nature and extent of graph use in secondary science textbooks?

The textbook analysis was aimed at answering Research Question 1. The analysis included 14 commonly used textbooks for teaching Year 9-11 science. The analysis focused on the size, topic used in, and type of graph in these textbooks. Only a small percentage of pages within the textbooks contained graphs ($M = 8\%$) and none of the graphs took up a whole page. Consequently, when the average amount of space over the 14 textbooks was calculated the total percentage space allocated to graphs per textbook was 1.23\%. When considering size, 90\% of the graphs were less than 1/8th page. When considering purpose, 59\% of the graphs were devoted to the use of graphing skills and formed part of questions that required students to use their knowledge of graphing in answering the questions. Consequently, they were not found in text which focused on the teaching of graphing skills. It was also found that most of the graphs were in physics and biology topics (26\% and 27\% respectively, See Figure 1). There were graphs in other topics e.g. earth science, introductory chapters, and astronomy, but this study only focused on the three main science disciplines of physics, chemistry, and biology. When searching through the textbooks for graphs it was found that on average there were 17.36 pages between one page containing a graph and the next page containing a graph. The total range was from 5.78 to 45.0 pages between graphs. Only two of the textbooks had dedicated sections concerning the teaching of graphing skills.

![Figure 1. Percentage of graphs in a sample of 14 New Zealand science textbooks for use with Year 9-11.](chart.png)
Research Question 2: How does graph use and difficulty based on national science exam question type (achieved, merit, or excellence) vary across the science disciplines of physics, chemistry, and biology?

The analysis of NCEA Level 1 science exams was aimed at answering Research Question 2. The analysis included 24 national science exam papers from 2004 - 2011 in the disciplines of biology, chemistry, and physics. The analysis focused on the size and type of graph in each discipline’s exams. There were 10 graphs in the 24 papers analysed, 9 of these were found in physics content questions and 1 in chemistry (See Figure 2). When considering the type of question associated with the graph, 86% of the graphs were linked to interpretation questions. All mechanics (motion) questions in physics contained a graph and most graphs were part of higher level questions (merit and excellence).

Figure 2. Total number of graphs in New Zealand national exams at Level 1 from 2004 – 2011.
Graphing questionnaire analysis.

**Research Question 3.** What is the nature and extent of graph use in secondary science classrooms and to what extent is graphing perceived as an important skill by secondary science teachers?

The analysis of the data gathered by the teacher questionnaire was aimed at answering Research Question 3. The analysis of this questionnaire focused on the time spent on the development of graphing skills, the difficulties students encountered when constructing and interpreting graphs, the teaching methods used when developing graphing skills and if teachers considered the time spent on graphing as adequate. The mean amount of time spent on graphing per year was about 7% for year 9 and 10 and increased to 9.25% for Year 11 within the context of science instruction. However, there is a large overall range reported by different schools of time spent on graphing (See Table 1). Across the three content areas, graphing occurs most often in physics for all year levels (Year 9 - 33.3%, Year 10 - 55.5%, Year 11 - 76.9%). Two separate ANOVAs found that: a) at Year 10, graphing construction occurred significantly more often in physics than in any other area \[F (4,18) = 3.147, p = .04\], and, b) at Year 11, graph construction \[F (2,25) = 4.543, p = .022\] and interpretation \[F (2,26) = 6.371, p = .006\] occurred significantly more often in physics. All of the teachers (100%) report that graphing is an important skill for science students and 77% say that they would like to spend more time on construction and 85% on interpretation. There were many reasons given for spending more time on graphing skills and teachers reported a number of issues concerned with teaching graphing skills (See Figures 3 & 4). Overall teachers reported that their students found interpretation more difficult than construction but the difference was not significant. \(M = 4.38\) for construction and \(M = 4.78\) for interpretation) (1 = extremely easy to 7 = extremely difficult). A summary of teacher responses to question 4 and 7 can be found in Appendix C.

There was no significant difference between methods used to teach graphing skills but teachers indicated that they do not use textbooks extensively for teaching graphing skills. Only 20% of teachers said that they use textbooks but all of these indicated that this was alongside other methods. The three most common
methods for developing graphing skills were whiteboards, PowerPoint presentations, and worksheets at 85%, 81% and 65% respectively.

Table 1  
*Means and SD for time spent on graphing in science Years 9-11*

<table>
<thead>
<tr>
<th>Total time per year</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 9</td>
<td>.01</td>
<td>.16</td>
<td>.08</td>
<td>.04</td>
</tr>
<tr>
<td>Year 10</td>
<td>.02</td>
<td>.20</td>
<td>.08</td>
<td>.04</td>
</tr>
<tr>
<td>Year 11</td>
<td>.02</td>
<td>.22</td>
<td>.09</td>
<td>.06</td>
</tr>
</tbody>
</table>

*Note:* Total time is the sum of construction and interpretation and are represented in fractions that can be converted to percentages i.e., .08 = 8%

*Figure 3.* Teacher reasons given in response to the questionnaire for spending more time on graphing in science in Years 9-11.


**Discussion and Conclusion**

In terms of the textbook analysis there is considerable variation between the use of graphs in different books. The lack of content devoted to graphs may be a reason that teachers reported that they do not often use textbooks for teaching graphing skills. The high incidence of graphs in biology topics in textbooks does not match the frequency of graphing questions in the external NCEA exams for biology, although it does however, reflect the use of graphs in biological articles as found by Roth, Bowen, and McGinn (1999).

The lack of graph construction or interpretation questions in NCEA level one biology and chemistry does not match the frequency of graphs used by scientists when communicating their work. As part of the Nature of Science strands “Understanding science” and “Communicating science” in the New Zealand Curriculum (Ministry of Education, 2007) graphing is an important skill that scientists use and so the lack of graphs in external exams does not align with the
New Zealand Curriculum (Ministry of Education, 2007). It is interesting to note that the word graph does not appear in any of the Achievement Objectives or Aims in the science learning area of the New Zealand Curriculum (Ministry of Education, 2007).

Overall there seems to be a discrepancy between the importance scientists and researchers place on graphing and the space allocated to graphs in textbooks, in external exams and the reported time science teachers spend on the teaching and practise of graphing skills in Years 9-11 science within the New Zealand context. As graphing is an important part of scientific literacy, which is the goal of science teaching especially for those students that do not go on to study science further than at high school as a core subject, it seems that graphing and the complexity of teaching graphing has been somewhat overlooked. It may be that the lack of graphs in the Year 11 NCEA external exams has reduced the perceived importance of teaching graphing skills. Research in science education certainly places importance on the teaching and learning of graphing skills and gives guidance on the most effective teaching and learning of these skills (Ates & Stevens, 2003; Friel, Curcio, & Bright, 2001; Hipkins, 2011; Shah & Hoeffner, 2002).

Study 1, which included the analysis of a questionnaire on teacher coverage as well as the analysis of science textbooks, suggests that the development of instructional graphing interventions is worthwhile within the current New Zealand science classroom. That is, given that graphing is difficult (a complex skill), and that time spent on graphing skills is limited (both in terms of textbook and teaching coverage), and that it is an important skill for learning science especially physics, it is critical that this gap be addressed via targeted instructional interventions. It is suggested that the application of cognitive load theory would be appropriate in this context given its success in facilitating the teaching of complex information efficiently and effectively so it would be useful in developing a targeted graphing intervention (Glazer, 2011). Within Study 2, current cognitive instructional theory and research (i.e., cognitive load theory) which has focused on the teaching of complex information will be employed as a vehicle for the development of a targeted intervention in a context which involves the skill of graphing, i.e. both interpretation and construction.
Chapter 3

Literature Review

Before introducing the cognitive load theory it is important to know about the system that it refers to, i.e., cognitive processing system, and in addition the constructs of working memory and long-term memory. Within the cognitive load section, the construct of cognitive load, how this is measured, and how this contributes to the calculation of instructional efficiency will be explained in detail. This section will conclude by outlining changes to the theory that have resulted from extensive research and critique, including a discussion of this critique.

The Cognitive Processing System

The cognitive architecture which people use to make sense of incoming information according to cognitive load theory is divided into two main parts, a limited working memory which was initially called short-term memory and a large possibly limitless long-term memory (Baddeley, 1992; Miller, 1956; Simon, 1974).

Working memory.

Working memory is where current information that we are conscious of is held and processed. It is therefore, working memory, which has the ability to limit the extent to which learners make sense of incoming new information. Processing information in working memory requires mental energy, which is limited. Limits also occur when the information is too complex or there is too much information to be processed at once (Baddeley, 1992). This is called cognitive overload. When cognitive overload occurs, processing stops and this interferes with understanding and learning. Miller (1956) and Baddeley (1992) suggest that working memory can only process seven items or elements at one time. Simon (1974) goes further and indicates that it could be as low as four or five. An element is defined as a piece of information that can be processed as one unit in working memory, this is different for novices and experts. If working memory is required to further process this information, for example, compare or organise, then the number of units that
can be held at any one time decreases, and it may be as low as two or three (Sweller, van Merriënboer, & Paas, 1998). In addition, there is also a limit on the time that these elements can be held in working memory without being rehearsed, this time limit has not been unanimously agreed upon but suggestions are between 20 and 30 seconds (Cowan, 2014; Peterson & Peterson, 1959).

Performing additional processing of material in working memory increases the interaction between the different units held in working memory and is referred to as “element interactivity.” Elements that interact need to be held simultaneously in working memory (Sweller, 2010a; Sweller & Chandler, 1994) and therefore an increase in element interactivity will increase the likelihood of cognitive overload (i.e., failure of the cognitive processing system due to task demands which exceed its capacity) (Halford, Mayberry, & Bain, 1986).

The original model of working memory suggested that it is not a single entity but appears to be divided into three parts: the visuo-spatial sketchpad and the phonological loop, which are in turn co-ordinated by an attention controller, which is referred to as the central executive (Baddeley, 1992). The visuo-spatial sketchpad processes incoming visual and spatial information (and possibly kinaesthetic information) while the phonological loop processes incoming verbal information (Baddeley, 1992; Clark & Paivio, 1991; Mayer & Moreno, 2002; Paivio, 1986). It has since been suggested that working memory consists of a fourth component called the episodic buffer. This suggestion came after problems that arose with the current model when both explaining the combination of verbal and visual information in working memory and the integration of information from long-term memory with material being processed by the phonological loop or visuo-spatial sketchpad. This is problematic as the central executive is thought to have no storage capacity for this combination or integration. Therefore the episodic buffer which is under central executive control acts as a limited storage area where conscious processing occurs to integrate information from a number of different sources (Baddeley, 2000). Mayer and Anderson (1992) have found that presenting information visually and orally at the same time is superior to presenting it successively. This reduces the working memory capacity needed to process the information as each part of working memory processes different information, simultaneously reducing the total load on any one part when both
components are involved. So in summary, working memory is where information is held “ready for action” but sometimes this action is restricted as there are limits on the amount and how long this information can be acted on, this has the effect of restricting learning in some circumstances (Cowan, 2014).

If there were no other mechanisms in place then it would be nearly impossible to learn new complex information, as it would always overload working memory limits, however, the second main part of the cognitive processing system (long-term memory) is instrumental in overriding these limits in certain circumstances.

**Long-term memory.**

In contrast to working memory, long-term memory is believed to have limitless capacity (Baddeley, 1992). It is a storage place for all the information previously received and made sense of by working memory (all material that has been learned) and to a large extent long-term memory determines further learning (Kalyuga, 2010). Most of the time the information stored here remains inactive (Shiffrin & Schneider, 1977). The information stored in long-term memory is highly organised and stored in hierarchical “chunks” (Kalyuga, 2010; Simon & Gilmartin, 1973), referred to as schema (Simon & Simon, 1978). A schema is a cognitive structure which is hierarchically organised and contains all the interlinked information (elements) that have been learned about a particular concept and its use (Chi, Glaser, & Rees, 1982; Kalyuga, 2010; Sweller, 1988). Modifying existing schemas and constructing new schemas occurs within working memory. If working memory is overloaded then this will affect the process of schema construction and hence storage, which are crucial processes for learning. Depending on prior learning, schemas that are stored in long-term memory can be complex (many related schemas connected together as for experts) or simple (elements of low level schemas connected together as for new learners) (Shiffrin & Schneider, 1977). The difference between a novice and an expert in a particular field is the richness of the information held in schemas relating to a particular field (Chase & Simon, 1973; Kalyuga, 2010; Simon & Gilmartin 1973, Sweller, 2012), the degree of automation of these schemas (Kalyuga, 2010; Sweller, 1994) i.e., how quickly and automatically these schemas can be accessed from long-term
memory, and the extent to which metacognitive skills have developed in order to use this information in appropriate ways (Kalyuga, 2010).

Information can be transferred between working memory and long-term memory bidirectionally. Movement of information from long-term memory to working memory is called retrieval. Existing related schemas (prior knowledge) are retrieved, as new information is being processed and added to previously learned information in order to facilitate processing and understanding of this new information. Instruction that is preceded by helping learners to access relevant or related schemas already held in long-term memory eliminates the need to search for prior knowledge while learning, and therefore reduces the processing demands on working memory (Kalyuga, 2010). The use and development of schemas are instrumental in reducing cognitive load while processing new information. Schemas can be processed as single units (elements) in working memory thereby reducing mental energy and the number of elements held at any one time so reducing the possibility of cognitive overload (Sweller, 2010a). After practice, schemas can also be accessed automatically from long-term memory when required. When schemas are accessed automatically only minimal input is necessary from working memory, which reduces the mental energy required and consequently reduces cognitive load associated with processing the incoming information and integrating it with existing knowledge in long-term memory. Retrieval of existing knowledge (schemas) from long-term memory is essential to understanding new information and has important implications for understanding complex information (Kalyuga, 2010; Marcus, Cooper, & Sweller, 1996; Pollock, Chandler, & Sweller, 2002).

Understanding complex information involves making sense of information with high element interactivity (Marcus, Cooper, & Sweller, 1996; Sweller, 2010a). Information which is easy to understand (low element interactivity) can be easily held in working memory without the risk of cognitive overload. Since well established schemas can act as a single unit in working memory the effect of high element interactivity on understanding can be influenced by expertise (Sweller, 2010a; Sweller, van Merriënboer, & Paas, 1998). This indicates that high element interactivity is only of concern with respect to cognitive load, when working with novices. If working memory is unable to process information, that is, if there are
too many elements or if the interactivity between them is too high, then cognitive overload occurs and understanding will be affected (Marcus, Cooper, & Sweller, 1996; Sweller, 2010a).

Mayer (1984) has developed a model of meaningful learning, which involves selecting information, organising it to make sense of it, and integrating it with information already stored in long-term memory. A crucial step in understanding incoming information is organisation and integration, which occurs in working memory. If working memory is overloaded then this cannot occur and understanding will be affected. So schemas held in long-term memory assist us in making sense of new information, storing learned information, recalling and retrieving learned information, overriding the limits of working memory, and engaging in higher level processing.

In summary, our cognitive system consists of a limited working memory and a limitless long-term memory. Learning mechanisms like schema acquisition and automation reduce the load imposed on working memory when processing new information while at the same time integrating it with information already stored in long-term memory. These mechanisms may have developed through evolution to help our cognitive processing system compensate for the limited processing capacity of working memory (Sweller, 2005, 2010a). If working memory is overloaded by high element interactivity (too many elements requiring simultaneous processing) then no further processing will occur, i.e., understanding and learning will not occur.

**Cognitive Load Theory**

Cognitive load theory is an internationally recognized and researched, empirically based educational psychology theory which builds on other theories of cognitive processing. It describes psychological processes that occur in the human cognitive processing system as a result of instruction of which the intention is to promote learning (Plass, Moreno, & Brünken, 2010). It has had an impact on other learning theories, eg., the cognitive theory of multimedia learning (Mayer, 2005a, 2014a) and the cognitive affective theory of learning with media (Moreno, 2006) and the integrated model of text and picture comprehension (Schnotz, 2005,
There is prolific writing, citing and research activity with respect to cognitive load theory (Jones et al. 2010). It has stood the test of time, been challenged and survived, and over time it has been modified in response to critique and new research findings (See later sections for a discussion of this) (Paas, van Gog, & Sweller, 2010).

Cognitive load theory (Sweller, 1988, 1989, 2010a; Sweller, van Merriënboer, & Paas, 1998) is a theory generated from research into instructional design and its influences on the cognitive processing system. Cognitive load theory assumes that there is limited capacity in working memory and limitless capacity in long-term memory, that skilled performance is based on building organised cognitive structures (schemas) in long-term memory and being able to use these structures without involving working memory (schema automation). The implications of this theory are important in the design of instructional materials aimed at facilitating understanding and learning. This theory suggests that often the design and development of instructional materials do not take into account the limited capacity of working memory and may add to the cognitive load already present due to the nature of the material being presented (Sweller, 1993; Sweller, van Merriënboer, & Paas, 1998). The nature of the material and the learner’s characteristics will determine whether design issues will impact on understanding and learning (Sweller, 1993; Sweller, van Merriënboer, & Paas, 1998).

Much of the early research which contributed to the development of the cognitive load theory was in the area of problem-solving in mathematics and physics, looking at the differences between novices and experts (Owen & Sweller, 1985; Sweller & Cooper, 1985; Sweller & Levine, 1982; Tarmizi & Sweller, 1988; Zhu & Simon, 1987). Sweller (1988) used a computational model to calculate cognitive load during problem-solving and then used these results to suggest ways to reduce the cognitive load associated with the presentation of the instructional materials. His research looked into the problem-solving processes of novices. He suggested that problem-solving and learning were two different cognitive processes for novices and that problem-solving used a large amount of working memory capacity leaving little or no capacity left for learning.
Cognitive load is not a new concept, for it is closely related to the constructs of mental load and task difficulty that have been extensively researched with respect to different professions within the field of human factors and psychology (Plass, Moreno, & Brünken, 2010; Salvendy, 2012). Within cognitive load theory, cognitive load is a complex construct that represents the load that performing a particular task imposes on the cognitive processing system (Paas & van Merriënboer, 1994; Sweller, 2010b; Sweller, van Merriënboer, & Paas, 1998). According to cognitive load theory there are three types of cognitive load: a) extraneous cognitive load, b) intrinsic cognitive load, and c) germane cognitive load. Extraneous cognitive load is the cognitive load that is imposed due to the poor presentation and design of the learning materials. Intrinsic cognitive load is the load imposed due to the nature of the material, that is, the intellectual complexity and interconnections between the ideas, which directly affects the number of elements that need to be processed simultaneously in working memory. Germane cognitive load is the load associated with learning the material, i.e., the load imposed on working memory due to schema construction and transfer of information into long-term memory.

The construct of cognitive load involves two sets of factors, causal and assessment factors. Causal factors are seen to affect cognitive load while the assessment factors can be measured to assess cognitive load (Choi, van Merriënboer, & Paas, 2014; Paas & van Merriënboer, 1993, 1994). The causal factors include the learner characteristics (e.g., ability and prior knowledge), the task (environment) characteristics (e.g., novelty and time pressure, noise level and temperature, respectively), and the interactions between them. In a real learning situation, task and learner characteristics always interact as learning cannot happen without a task and conversely, learning from a task cannot happen without a learner. Assessment factors are the indices we use to measure or assess cognitive load which include mental load, mental effort, and performance. Mental load is task based and is constant for a given task in a given environment (e.g., the number of interacting elements in the material to be learned) (Paas & van Merriënboer, 1994; Choi, van Merriënboer, & Paas, 2014). Mental effort is learner based and reflects the amount of controlled, conscious processing the learner is engaged in to meet the task demands, so this factor reflects the interaction
between the learner and the task (Shiffrin & Schneider, 1977; Choi, van Merriënboer, & Paas, 2014). Performance (what the learner does or achieves based on a given task) is influenced by all of the internal and external interactions of the causal and assessment factors. Hence, variability in performance is also an indication of the cognitive load imposed by the task, i.e., a learner who attains high performance for less effort indicates lower cognitive load than a learner who does not achieve as high in performance and had to put in more effort. Examples of performance measures would be test marks, time to completion, number of errors made, or success in problem-solving transfer.

Recently, Choi, van Merriënboer, and Paas, (2014) have proposed an amendment to the causal factors affecting cognitive load in which the task and environment are viewed separately (which were combined in the original model). They also changed the causal factor “environment” to “physical learning environment.” The physical learning environment is the physical aspects of the place where teaching and learning is taking place. These can include learning tools, physical features of the space, other people present, or sensory stimuli, which they argue can also have an effect on the learner and the task. Consequently, the physical learning environment embraces both the learner and the task factors. There are no empirical studies within the cognitive load theory framework to support their suggestions, however, there is evidence from other studies of the impact of the physical environment on behaviour, attitudes and performance, which can all impact on learning (Lan, Wargocki, Wyon, & Lian, 2011; Uline, & Tschannen-Moran, 2008; Vredeveldt, Hitch, & Baddeley, 2011).

Cognitive load theory has been used to develop techniques to facilitate understanding and learning by modifying the presentation of instructional materials. These techniques have been developed within the context of consistent patterns of experimental effects found by research generated from cognitive load theory. There are twelve effects which will be briefly discussed below and can be grouped into three sections according to the primary type of cognitive load on which it has the most impact.
**Extraneous cognitive load.**

As indicated earlier, extraneous cognitive load is the cognitive load that is imposed due to the poor presentation and design of the learning materials. This has been the main focus of cognitive load research from the late 1980’s till the early 2000’s. The experimental effects which have been linked to extraneous cognitive load are the worked example (Renkl, 2005), completion (Paas & van Merriënboer, 1994), split-attention (Chandler and Sweller, 1991, 1992), modality (Low & Sweller, 2005), redundancy (Sweller, 2005), expertise reversal (Kalyuga, 2005), guidance fading (Renkl, 2005) and goal-free (Owen & Sweller, 1985). These effects occur primarily when there is an attempt to employ manipulations which help learners to focus their attention on the critical information. Studies which have documented these effects offer extensive evidence which supports the suggestion that instructional formats which impose a high extraneous (avoidable) cognitive load on working memory can interfere with further cognitive processing, which is critical for facilitating understanding and learning (Sweller, 1989, 1994). Since the main aim of instruction is to transform a novice (simple schemas) into an expert (complex schemas) in a given field, learning (schema acquisition) is a vital part of this transformation. If the instructional materials impose an unnecessary load on working memory then learning could be affected and as a result the intended transformation from novice to expert could also be affected.

**Intrinsic cognitive load.**

As indicated earlier, intrinsic cognitive load is the load imposed due to the nature of the material. This type of cognitive load is not alterable (Pollock, Chandler, & Sweller, 2002; Sweller & Chandler, 1994). The two experimental effects that relate to intrinsic cognitive load are element interactivity (Sweller, 1994) and isolated-elements (Pollock, Chandler, & Sweller, 2002). These indicate that element interactivity dictates whether cognitive load will have an effect on learning. That is, material high in element interactivity (complex information) could impose a high cognitive load, which in turn could interfere with learning – element interactivity effect. In some situations where the learners are novices and the material is very complex (which could potentially overload working memory), it is
more effective to initially present the material as isolated-elements before presenting it with all the interacting elements - isolated-elements effect. Empirical evidence from further studies has confirmed this effect (Ayres, 2006; Blayney, Kalyuga, & Sweller, 2010; Kester, Kirschner, & van Merriënboer, 2006; Lee, Plass, & Homer, 2006).

**Germane cognitive load.**

As indicated earlier, germane cognitive load is the load associated with learning the material. The related effects are the variable examples and imagination effects. The variable examples and imagination effects have been found in studies focused on manipulating germance cognitive load. These effects relate to instructional design which encourages students to use freed up working memory resources as a result of reducing extraneous cognitive load for construction of schemas e.g., using examples with variable surface features and imagining skills and procedures (Leahy & Sweller, 2004; Paas & van Merriënboer, 1994).

By the mid 2000’s, Cognitive Load Theory had shifted it’s focus from reducing extraneous cognitive load to using the techniques and guidelines developed to eliminate or significantly reduce extraneous cognitive load to focusing on strategies to either manage intrinsic cognitive load (Gerjets, Scheiter & Catrambone, 2004; Pollock, Chandler, & Sweller, 2002;) and/or increase germane cognitive load (Renkl, Atkinson, Maier, & Staley, 2002; van Merriënboer, Kester, & Paas, 2006).

In summary, the total cognitive load imposed on working memory is the sum of the three types of cognitive load. If tasks are complex (high element interactivity) then mental load and mental effort may both be high (Halford, Mayberry, & Bain, 1986), and this is the result of high cognitive load (Paas & van Merriënboer, 1994; Pollock, Chandler, & Sweller, 2002). With respect to understanding new information, Marcus, Cooper, and Sweller, (1996) suggest that there are three main factors that affect the total cognitive load imposed by the instructional materials, that is, prior knowledge (the presence or absence of relevant schemas),
the intrinsic nature of the material (the number of elements and their degree of interactivity), and the presentation and organisation of the instructional materials.

More recent discussion and research has focused on a revised framework of the categories of cognitive load based on element interactivity (Sweller, 2005, 2010a). Within this context, there are now two main independent categories of cognitive load (intrinsic and extraneous) and the total cognitive load imposed on working memory for a given task is the addition of intrinsic and extraneous cognitive load (Sweller, 2010a, 2010b). There is still a third type of cognitive load (germane) but this is seen as dependent on intrinsic cognitive load. The three types of cognitive load interact in the following ways.

Intrinsic cognitive load is still a function of the interactivity of the elements in the learning task if understanding is the goal. This load will differ depending on whether the learners are novices or experts with respect to the task. That is, increasing one’s expertise through learning would reduce element interactivity. This occurs because information that is complex and contains many interacting elements for a novice, may for experts, already be combined into schemas which can then be processed in working memory as one element (Sweller, 2010a, 2010b). Extraneous cognitive load is now defined as unnecessary interacting elements that are introduced to the instructional design of the material presented for learning. Germaine cognitive load is still seen as being imposed when learning the material and is a reflection of the working memory resources allocated to processing the interacting elements (Sweller, 2010b, 2012). These have also been referred to as “germane resources” (Leppink et al., 2014). If there are many interacting elements then learning this material will impose a high germane cognitive load for novices.

In summary, intrinsic cognitive load is caused by the number of interacting elements within the to-be-learned information but this only imposes a load on working memory (germane cognitive load) when trying to make sense and learn the information. Germaine and extraneous cognitive load are complementary. This means that for a specific task and learners, if intrinsic cognitive load is high and not changeable, and if extraneous cognitive load increases, then germaine cognitive load will decrease to keep within working memory limits. As intrinsic and
germane cognitive load are closely related and often difficult to separate, when calculating the total cognitive load of learning a particular task, they are effectively treated as one source (Sweller, 2005, 2010a). However, they are only the same when learners put in the maximum effort (assuming maximum motivation) and extraneous cognitive load is eliminated or negligible. In summary, efficient instruction occurs when the total cognitive load, (i.e, the addition of extraneous and intrinsic cognitive load) are within an individual’s working memory limits (Kalyuga, 2010).

**Measuring cognitive load.**

The development of techniques to measure cognitive load have contributed to the success of cognitive load theory as it has provided an empirical base for the theory and will continue to be instrumental in its future development (Paas, Tuovinen, Tabbers, & van Gerven, 2003). However, measuring cognitive load is a very complex task as it is affected by the interactions of the learner, the task, and the physical learning environment (Sweller, van Merriënboer, & Paas, 1998; Choi, van Merriënboer, & Paas, 2014). When measuring cognitive load the main issues of sensitivity, validity, implementation, and intrusion need to be taken into account (Eggemeier, 1988; Salvendy, 2012). Sensitivity is the degree to which the technique can measure differences in cognitive load, validity is the extent to which the technique measures what it is supposed to measure, implementation is what is needed before and during the use of the technique (equipment, subject preparation and supervision are included in this aspect), and intrusion is the extent to which the technique interferes with the completion of the task and therefore performance. There are many and varied techniques discussed in the literature (Wierwille & Eggemeier, 1993). This review will focus on the three main categories of techniques for measuring cognitive load: subjective measures, physiological measures, and performance-based measures, (which includes both primary and secondary task measures). Cognitive load cannot be directly measured but is inferred indirectly from these measures. At present there is no universally accepted method for measuring the individual categories of cognitive load, so the following section will describe the techniques to measure total cognitive load (Moreno and Park, 2010; Sweller, 2010b).
Subjective techniques. These have frequently been used in cognitive load theory research to estimate cognitive load (Ayres, 2006; Paas, Tuovinen, Tabbers, and van Gerven, 2003). These techniques use likert scales to assess the mental effort expended or the difficulty of the task. Subjects are asked to translate either the mental effort expended on completing the task or the perceived difficulty of the task into a number, the range of numbers depend on the scale used, that is for example a 5, 7 or 9 point scale (Ayres, 2006; Hoffman & Schraw, 2010). Most of these scales are an adaptation of the Bratfisch, Dornic, and Borg (1972) symmetrical 9-point rating scale. In this scale 1 is rated as very very easy and 9 is rated as very very difficult. Therefore, number 1 equates to very little mental effort expended or a very very easy task and 9 equates to a very high mental effort or a very very difficult task. The techniques assume that subjects can accurately assess the level of mental effort invested (Paas, van Merriënboer, & Adam, 1994) and that there is a direct relationship between subjective measures and actual cognitive load (Paas, Tuovinen, Tabbers, & van Gerven, 2003). Other scales that have been used successfully are the NASA-TLX which assesses mental effort and other associated variables eg, fatigue, frustration, and pace (Paas, Tuovinen, Tabbers, & van Gerven, 2003). These scales usually provide an overall assessment of cognitive load as they cannot be used repeatedly throughout the learning or they could interfere with learning (Ayres, 2006).

Rating scales need to be explained to students before their use but have been found to be sensitive to small differences in cognitive load, (Kalyuga, Chandler, & Sweller, 1998; Sweller, van Merriënboer, & Paas, 1998) reliable, valid, and less intrusive than physiological methods so are more suitable for use in classrooms (van Gog & Paas, 2008). It is, however, important that these are administered without delays (within 15-30 minutes) as these could cause the loss of critical rating information (Wierwillie & Eggemeier, 1993). In addition, they suggest that the participants should not be engaged in recall of information or performing other tasks which require storage or retrieval of information in and from long-term memory as this can interfere with recall of experiences relevant to the rating scales.
**Physiological techniques.** These use measurement of body physiology in response to the stress of expending mental effort. For example, these methods may look at changes in heart rate, eye movements, or blood glucose levels in response to mental effort invested. These methods assume that there are significant changes in physiological response to differing amounts of mental effort expended and are therefore useful for measuring the changes in cognitive load throughout a learning session (Paas, Tuovinen, Tabbers, & van Gerven, 2003). The first problem with this approach is that responses to stress may differ from person to person and from situation to situation. The second problem is that these measures were found to be unreliable, invalid, and not sensitive to small changes in cognitive load (Paas, van Merriënboer, & Adams, 1994; Sweller, van Merriënboer, & Paas, 1998).

**Performance measures.** Performance measures assess the learning and/or understanding that occurs or occurred during a period of instruction. Measures of performance can include test results, error rates, time for task completion, number of problems solved, and accuracy on tasks. These performance measures relating to studies measuring cognitive load can be divided into primary task and secondary task measures.

**Primary task measures.** Primary tasks are the main task, that is, the focus of the research. Sweller and others have investigated the relationship between cognitive load and instructional design, in which they relied on performance measures alone to estimate cognitive load (Chandler & Sweller, 1991, 1992; Sweller, 1988, 1989; Sweller, Chandler, Tierney, & Cooper, 1990). The difficulty with only using performance is that it is often not a true indicator of the load imposed, as subjects can increase their effort to maintain the same level of performance (Paas, Tuovinen, Tabbers, & van Gerven, 2003; Paas, van Merriënboer, & Adams, 1994). In these cases, mental effort measures are more reliable as an indicator of cognitive load than performance based measures (Paas & van Merriënboer, 1994). Hamilton (1979) suggests that the essence of cognitive load is the mental effort expended and therefore a combination of performance and measures of mental effort have provided the best estimates of cognitive load, this is called instructional efficiency. This view is supported by Paas and van

Within a school setting, using subjective rating scales and performance measures have been found to be non-intrusive, valid, easy to explain to students, administer, and analyse (Wierwille & Eggemeier, 1993; Paas & van Merriënboer, 1994). Paas and van Merriënboer (1993) have described in detail a successful way of using these measures to assess cognitive load and to calculate the efficiency of the instructional conditions which can easily be applied to research in a secondary school setting.

**Secondary task measures.** These assess the capability of the learner to perform a secondary task (e.g., an auditory beep or a coloured letter on a computer screen) while engaged in the target or primary task. The secondary task performance measure (as measured by reaction time, accuracy, error rate, or time on task) gives an indication of spare mental capacity. The assumption is that if the participants can attend to the secondary task then they have spare mental capacity and this influences the cognitive load of the primary task (Brünken, Plass, & Leutner, 2003; Sweller, 1988). An appropriate secondary task is one that is easy to perform, easy to learn, and is possible to complete concurrently with the primary task, e.g., responding to a tone or a visual signal (Brünken, Plass, & Leutner, 2003; Marcus, Cooper, & Sweller, 1996; Paas, Tuovinen, Tabbers, & van Gerven, 2003). Secondary tasks have been found to affect performance on the primary task, usually in a negative direction. The size of the negative effect depends on the nature of the primary task (Meshkati & Loewenthal, 1988) so for this reason they have not been used extensively in cognitive load theory research. Using a secondary task could also be particularly problematic if the primary task is complex and would therefore take most of the learners working memory resources for understanding and learning leaving few or none for attending to the secondary task (Brünken, Plass, & Leutner, 2003; Paas, Tuovinen, Tabbers, & van Gerven, 2003). However, when used appropriately they give a more continuous indication of the changes in cognitive load throughout a task (Brünken, Plass, & Leutner, 2003), as cognitive load is seen to fluctuate during learning (Xie & Salvendy, 2000).
Using cognitive load measures and performance measures to calculate instructional efficiency.

Instructional efficiency in cognitive load research most often involves looking at the instructional outcomes in terms of the mental resources used to produce these i.e., the relationship between maximum performance and minimum effort (Hoffman, 2012). Although there is no consensus about how best to measure instructional efficiency (Hoffman & Schraw, 2010), in cognitive load research an instructional lesson is deemed efficient if learners achieve above average performance for below average mental effort expended to reach this level of performance (Clark, Nguyen, & Sweller, 2011). Instructional efficiency can be used to assess the quality of learning outcomes, i.e., the extent to which cognitive schemas have been acquired, elaborated, and automated (van Gog & Paas, 2008). There are two measures used in the calculation of instructional efficiency: a) performance measure, and b) a mental effort measure. The performance measure is most often the learner’s score on a post test but some studies have used time for completion of a summative task in their calculations of instructional efficiency (Kalyuga & Sweller, 2005). One approach to measuring mental effort involves the use of subjective rating scales which ask questions about the difficulty of performing a particular task, e.g., learning, completing a test, or the difficulty of understanding ideas presented. Other approaches use objective measures of cognitive load, which could include the use of a secondary task, time on task when there is no time limit imposed, eye movements, or heart rate variability.

Research has found that calculating instructional efficiency scores adds another dimension to the analysis of results not captured by purely focusing on performance scores. For example, if two students achieved the same performance score and one student achieved their performance with little effort and the other had to invest a large amount of effort then calculating an instructional efficiency score would capture an important way in which these two students differed which would be missed by only looking at their performance scores (Hoffman & Schraw, 2010). Similarly when looking at the success of an intervention where comparisons are made between different treatment groups receiving variations on the same material to be learned, measuring instructional efficiency gives a better understanding of the effectiveness than by just looking at
performance because it gives an indication of the benefit to learning compared to the mental effort invested (Plaas, Moreno, & Brünken, 2010). Also, if only the mental effort scores were used then this could reflect confidence, whereas combining mental effort and performance scores eliminates this as a possible interpretation (Kayuga, Chandler, & Sweller, 2000). The reason that instructional efficiency is widely used in educational research is that it is easily calculated and has been found to be reliable (Hogg, 2007; Plaas, Moreno, & Brünken, 2010).

Hoffman and Schraw (2010) completed a comparison of learning and problem-solving instructional efficiency across a large number of studies and have categorised the methods used to assess instructional efficiency. They found that there are three main but very diverse models used in the assessment of instructional efficiency which differ with respect to the actual formula used in the calculation, the measurement of the variables, and how they are applied (i.e., when to use each particular model and the justification for doing so).

The first model (deviation model) was originally developed by Paas and Merriënboer (1993) and assesses the difference between the standardised scores for performance and mental effort expended to achieve this performance (Hoffman & Shraw, 2010). This is essentially the effect size for the difference between performance and effort. This model assumes a one to one relationship between performance and effort and so an ideal instructional efficiency is a line with a slope of 1. Individual instructional efficiency scores are therefore calculated as a deviation from this line and hence the name for this model. There are two versions of the deviation model, an original and adapted version. The original one, mentioned above, (Paas & van Merriënboer, 1993) looked at the mental effort expended during the test to produce the test scores (performance) and measures the instructional efficiency of the learning outcomes. In contrast, the adapted version looks at the mental effort during learning and the performance on the test which means that this measures the instructional efficiency of the learning process (van Gog & Paas, 2008). The actual calculation performed while assessing the instructional efficiency using the deviation model is:
Instructional efficiency = \[ \frac{z (\text{Post test scores}) - z (\text{Mental effort score})}{\sqrt{2}} \]

The second model (likelihood model), uses the ratio between performance and effort expended to achieve this performance, to calculate instructional efficiency (Hoffman & Shraw, 2010). It measures the rate of change of effort for performance achieved, i.e., the relative gain of performance with respect to the effort expended to achieve this. In this model, time to completion can be used as a performance measure (Kalyuga & Sweller, 2005). The actual calculation performed while assessing the instructional efficiency using the likelihood model is:

\[
\text{Instructional efficiency} = \frac{\text{Performance scores}}{\text{Mental effort score}} \text{ or } \frac{\text{Performance scores}}{\text{Time to completion}}
\]

The third model (conditional likelihood) is an adaptation of the likelihood model which also takes into account prior learning or experience. This is in effect very similar to using prior knowledge as a covariate in analysis when instructional efficiency is calculated using the likelihood model.

Each of these three models measure a different aspect of learning efficiency and therefore, the most appropriate model of instructional efficiency to be used will depend on the nature of the research being conducted. For example, the deviation model measures the relative instructional efficiency of different conditions and therefore is most appropriately used in intervention studies (Hoffman, 2012; Hoffman & Schraw, 2010; Plaas, Moreno, & Brünken, 2010). As this study is an intervention study which is investigating the effect of pre-training on the learning process relative to three different treatment groups receiving the same information the adapted deviation model is the appropriate method for calculating instructional efficiency.

As indicated earlier, using the deviation model employs two measures, a performance score and a mental effort score related to the targeted performance. In the current study the performance is a measure of the learning that happened
as a result of the teaching and was measured using a post test. Within the current study, several measures of mental effort were employed, firstly, an online subjective cognitive load rating scale which was used twice during the teaching to assess the difficulty of understanding the material being taught and secondly, a summative cognitive load rating scale with five separate questions was used at the end of the teaching, summative Question 1 and 2 were used to assess overall learning and understanding of the material being taught.

**Recent advances in cognitive load theory.**

Recent advances in cognitive load theory compare human cognitive processing and the process of biological evolution. For example, long-term memory is compared to the genetic code as they are both vast stores of information and changes to the genetic code and changes long-term memory happen slowly and incrementally (Schnotz & Kurschner, 2007). Within this framework two types of knowledge have been distinguished, biologically primary and biologically secondary knowledge. Biologically primary knowledge is information that has been acquired over generations and is part of being human e.g., speech, listening and social skills. In contrast, biologically secondary information is acquired consciously and requires mental effort, eg., all institutional learning. Cognitive load theory is concerned with the latter knowledge, i.e., biologically secondary knowledge (Sweller, 2004, 2005). Five principles of human cognitive architecture have been developed as a result of comparing these two biological processes. The recent advances suggest that for effective functioning of the cognitive processing system to acquire biologically secondary knowledge requires all of the five basic principles to be met (Sweller, 2005, 2010a). These principles are not new additions to the theory but a different way of looking at and explaining why the cognitive processing system functions as it does within an evolutionary perspective.

**The five new principles.** The first principle is the information store principle which states that long-term memory is the main storage centre for all information and that this in turn drives the cognitive processing system. The next two principles are concerned with how material gets into long-term memory. The second principle is the borrowing and reorganising principle which suggests that
most of the information in long-term memory is second hand, i.e., it is borrowed from others by listening to, reading, or watching presentations of new information from the long-term memory store of other people. The third principle, the randomness as genesis principle, states that if no previous information is available in long-term memory (no prior knowledge) that random trial and error is the only way for new information to get into long-term memory. This is a very mentally demanding process and not an effective use of the cognitive processing system so the principle of borrowing and reorganising is the preferred way for information to get into long-term memory. The fourth principle is the narrow limits of change principle which explains why the changes to information in long-term memory are incremental and happen slowly over time. This is a protection mechanism so that major changes cannot happen easily to guard against the inadvertent deletion of important information already held in long-term memory. This principle primarily concerns the limited capacity and duration of working memory which determine the changes in long-term memory. Lastly, the fifth principle describes the way information stored in long-term memory can override the limited capacity of working memory for experts and is referred to as the environmental reorganising and linking principle.

These five principles explain the central role of long-term memory in learning from experiences in the external environment, how and why this compensates for a limited working memory, where the information in long-term memory comes from, and why the process of changes to long-term memory happen slowly and progressively over time. These reasons are seen to be analogous to the natural change process in the biological world called evolution.

In summary, the main aim of cognitive load theory is to produce guidelines to maximise the processing capacity of working memory to facilitate understanding and learning (Sweller, 1993; Sweller, van Merriënboer, & Paas, 1998). For this reason the cognitive load theory is most applicable to instructional materials that impose high intrinsic cognitive load (high element interactivity), that is, materials or instruction that are difficult to learn, in situations where the learners are novices (do not have expansive automated schemas in the field of instruction). In these situations the mental effort required to process instructional materials could interfere with understanding and learning.
Critique of cognitive load theory.

This section starts with a discussion and critique of the theory, then evaluates the individual categories of cognitive load before moving on to looking at the concept of total cognitive load and lastly discusses possible problems with the measurement of cognitive load.

Cognitive load theory. Gerjets, Scheiter, and Cierniak (2009) question whether cognitive load theory is actually a theory based on Karl Popper’s critical rationalism definition. That is, a theory should be possible to falsify and the main assumptions should be able to be empirically tested. However, they do concede that it could still be regarded as a theory using the structuralist view of theories by Joseph Sneed (Sneed, 1979). Others acknowledge that cognitive load theory has been used to successfully develop guidelines for the design of learning tasks in environments that would have been difficult or even impossible for novices for over 30 years. However, these researchers also suggest that there are several factors which determine or have an effect on learning besides cognitive load that are not included in cognitive load theory. For example, motivation, metacognitive skills, self-confidence, spatial ability, cognitive ability, learning style, and anxiety (de Jong, 2010; Mayer, 2014b; Moreno, 2006, 2010; Paas, Tuovinen, van Merriënboer, & Darabi, 2005; Valcke, 2002). In response to some of these concerns associated with motivation, the definition of germane cognitive load has been widened and now includes the idea that germane cognitive load only refers to dealing with all the elements associated with intrinsic cognitive load in situations where the learner invests maximum effort, i.e., in situations involving maximum motivation (Kalyuga, 2011; Sweller, 2010b). Kirschner, Paas, and Kirschner (2011) have extended cognitive load theory research into looking at affective reasons why learning in groups (collaboration) is more effective for learning complex information or in complex problem-solving environments. Regarding the cognitive load theory model of working memory, Brünken, Plass, and Moreno (2010) criticise cognitive load theory for the inclusion of only parts of Baddeley's (2000) model of working memory, namely the visuo-spatial sketchpad and phonological loop but not including the central executive and episodic buffer. In addition, Brünken, Plass, & Moreno (2010) also suggest that the modality effect is
wrongly described using these two subsystems. Cognitive load theory uses sensory modes to assign material to the different subsystems but Baddeley uses representations. That is, cognitive load theory suggests that the reason for the modality effect is that written and spoken language is processed in different subsystems. However, according to Baddeley (2000) all verbal information is processed by the phonological loop which contradicts the cognitive load theory explanation for the modality effect (see also Rummer, Schweppes, Furstenberg, Seufert, & Brünken, 2010).

Brünken, Plass, and Moreno (2010) raise issues regarding cognitive load theory notions of mental models produced through working memory processes. They suggest that models could differ with respect to detail so it could be argued that the mental effort required to produce different representations and consequently the mental models could be different. Differential impact of different mental models on cognitive load is not reflected in cognitive load theory. Although this is not specified in cognitive load theory the increased effort required to produce a more comprehensive mental model will result in a higher cognitive load rating and better performance on a transfer test, which is able to be measured. However, cognitive load theory does not account for the ease with which some learners may build a mental model irrespective of prior knowledge, i.e., if they have more efficient processing systems.

Criticism has also been made regarding the lack of detail with respect to processing that occurs in working memory as part of cognitive load theory. Without this detail, it is difficult to describe the nature of working memory resources that are allocated to the different processes of understanding and learning included in cognitive load theory (Mayer, 2005, 2014a; Schnitz & Kurschner, 2007). Although these are separate processes in working memory, the measurement of these separate processes is not possible at this stage. Schnitz and Kurschner (2007) discuss this further stating that there is no clear distinction between understanding and learning in cognitive load theory, i.e., understanding involves processing demands on working memory but this cannot occur until at least some elements have been combined into schemas which is defined in cognitive load theory as learning (a change in long-term memory). This criticism is consistent with earlier versions of cognitive load theory (Sweller, van Merriënboer,
& Paas, 1998) but this has been resolved with the suggestion that all categories of cognitive load are now related to element interactivity and that germane cognitive load is now defined as the actual load of processing the material which results in learning and the concept of understanding and subsequent learning has been included in the revised definition of intrinsic cognitive load (Kalyuga, 2011; Sweller, 2010b, 2012). These revisions also address the concerns by Brünken, Plass, and Moreno (2010) that germane cognitive load did not involve any working memory processing so it was difficult to attribute germane cognitive load to situations where learning occurred. In response to Schnottz and Kurschner (2007), Kalyuga (2011) has begun to align cognitive load theory with the cognitive theory of multimedia learning (Mayer, 2014a)(which differentiates between understanding and learning in terms of processing) by describing active processing as processing which involves the selecting and organising of elements (understanding) and integrating them with knowledge structures from long-term memory in working memory (learning).

The 2005 revision of cognitive load theory introduced the idea that the cognitive processing system is analogous to the biological process of evolution by natural selection (Paas & Sweller, 2014; Sweller, 2005). Moreno and Park (2010) question the value and practical usefulness of the revision to cognitive load theory (Sweller, 2005) within an evolutionary framework, with respect to the generation of new cognitive load theory effects. Cowan (2014) goes further and questions the distinction between biologically primary and secondary knowledge in terms of working memory involvement and claims that some of the examples of biologically primary knowledge used, actually involve situations of limited working memory resources, so technically should be classified as biologically secondary knowledge, e.g. recognising faces (Sweller, 2010a).

Lastly, questions have been raised regarding the narrow range of contexts that cognitive load theory research has covered, the lack of research in authentic settings, the short lengths of instructional time and the predominance of adults as participants (Bannert, 2002; Brünken, Plass, & Moreno, 2010). Since these issues were raised there has been progress made in widening the contexts and the using authentic settings, namely, Youssef, Ayres, and Sweller (2012) in geography, Kyun, Kalyuga, and Sweller (2013) in English, Leslie, Low, Jin, and Sweller (2012)
in primary school science and Lee and Kalyuga (2011) in learning Chinese. This study also extends the context beyond university students learning about physical systems to New Zealand prescribed physics curriculum content using high school students as participants and involving a teaching period lasting 30 minutes.

*Triarchic model of cognitive load.* There are a variety of problematic issues that have been raised concerning the specification of the three types of cognitive load: a) specification of intrinsic cognitive load and the different influences on it, b) differentiation between the three types of cognitive load, c) the nature and appropriateness of creating a total cognitive load from the three different types of load, and, d) measurement of cognitive load.

*Intrinsic load.* De Jong (2010) suggests that experienced difficulty due to issues surrounding intrinsic load could also be due to misconceptions and this is not accounted for in cognitive load theory. I agree as these have particular relevance to learning science especially physics and graphing, as these misconceptions interfere with learning for example, in the topic of speed (the context for this study) students often confuse the graphic representation of speed and think that the graph is a literal picture of the motion being represented. This confusion makes graph interpretation very difficult (Biechner, 1994; Berg & Phillips, 1994; Glazer, 2011; Shah & Hoeffner, 2002). Also, regarding difficulty of learning there is material in science that requires the rote learning of a large number of elements, according to cognitive load theory even though it is difficult to learn this does not pose a high intrinsic cognitive load and therefore does not have working memory consequences as difficulty and element interactivity are two different concepts (Sweller, 2012). However, students can be overloaded with the amount of material (large number of elements) required to be learned even if the material is not high in element interactivity and doesn’t require simultaneous processing. This overload situation is not accounted for in cognitive load theory. Questions have also been asked concerning the fixed nature of intrinsic cognitive load, i.e., is it actually fixed given that there are techniques to control it (de Jong, 2010; Moreno, 2010; Moreno & Park, 2010; Schnotz & Kurschner, 2007). However, if the material is altered (elements reduced) then it is not the exact same material so the altered material has reduced intrinsic cognitive load but the original material does not. Finally, Moreno and Park (2010) note that at present there are
no guidelines for the consistent measurement of element interactivity which make it difficult to assess both intrinsic cognitive load and whether material is actually too difficult for novices and therefore needs to be reduced or at least managed. The specification of intrinsic load is problematic, consequently, in this study material was chosen at a higher curriculum level and in a difficult topic to ensure that it would be complex for the participants. This approach was to ensure high intrinsic load due to the fact that there was no universally accepted method to determine difficulty based solely on a measure of element interactivity.

*Different categories of cognitive load.* There are also many questions surrounding the distinction between the different categories of cognitive load. Looking firstly at the difference between intrinsic and germane cognitive load, suggestions are that if intrinsic cognitive load is to do with the materials and germane cognitive load has to do with the processing of these materials, then without a learner being involved in processing you can’t have cognitive load imposed and therefore intrinsic cognitive load is not actually a load on working memory (de Jong, 2010). Sweller, Ayres, and Kayuga (2011b), have clarified this distinction which follows the redefining of the three categories of cognitive load by Sweller (2010b). Within this new framework intrinsic and extraneous cognitive load both impose a load on working memory but germane cognitive load does not. Working memory resources are allocated to both intrinsic and extraneous cognitive load but germane cognitive load resources are allocated to dealing with intrinsic cognitive load. Kalyuga (2011) also suggests that intrinsic and germane cognitive loads by definition both contribute to schema acquisition, consequently, the terms are indistinguishable at a macro and micro level and so intrinsic and germane cognitive load could be regarded as essentially the same load. These two loads are specifically related to the learner and the learning goal, and therefore the category of germane cognitive load is redundant.

In terms of extraneous cognitive load, there are suggestions that the distinction between extraneous and germane can sometimes be blurred, e.g., when material is simplified and redundant information is removed this may affect the processes of comparison and elaboration which are germane processes, therefore removing extraneous cognitive load has a negative effect on learning which is not consistent with cognitive load theory (de Jong, 2010). In addition, the type of cognitive load
(extraneous or germane) could also be determined by the processing required for the learning outcome, that is, if the load concerns recall or transfer e.g., using concept mapping which facilitates understanding would be germane cognitive load if the goal was transfer of this new knowledge but extraneous if the goal was just recall of facts (Brünken, Plass, & Moreno, 2010). On the other hand, it is possible for there to be too much germane cognitive load which inhibits learning and therefore can be categorised as extraneous cognitive load (bad or negative cognitive load). Kalyuga, (2007), de Jong (2010) and Moreno (2010) see this as contrary to the definitions of extraneous and germane cognitive load. Schnotz and Kurshner (2007) indicate that the difference between extraneous and germane cognitive load is in fact related to the expertise of the learner and the actual learning goal. This contradiction has been clarified in cognitive load theory with the redefinition of the categories of cognitive load. The situation described above where too much germane cognitive load inhibits learning would now within the revised definitions be described as having too much intrinsic cognitive load due to the complexity of the material and would be an example of a situation where intrinsic cognitive load needs to be managed so it is within learner’s cognitive capacity.

Further to the discussion regarding the categories of cognitive load, Moreno and Park (2010) question whether reducing elements to reduce intrinsic cognitive load could in fact also be reducing the number of extraneous elements in which case it would be difficult to conclude if the success of the intervention was due to a reduction in intrinsic or extraneous cognitive load. However, if (as is the case in this study) the intervention was designed to eliminate or minimise intrinsic cognitive load then the success would have to be attributed to intrinsic cognitive load. In support of this, Ayres (2006) argues that by keeping extraneous and germane cognitive loads constant that the resulting subjective measurements would be assessing intrinsic cognitive load.

To conclude this discussion on categories of cognitive load, de Jong (2010) and Gerjets, Scheiter, and Cierniak (2009) suggest that the definitions are circular, i.e., the definitions for the categories of cognitive load determine which load is involved and not the actual measurements of cognitive load. For example, if lower learning is achieved by participants and cognitive load increases then it is determined to be
extraneous. This is consistent with other criticisms already mentioned in that until the measurement of individual categories of cognitive load are developed, these questions and difficulties cannot be addressed (Brünken, Plass, & Moreno, 2010; de Jong, 2010; DeLeeuw & Mayer, 2008; Kirschner, Ayres, & Chandler, 2011; Moreno, 2010; Moreno & Park, 2010).

Total cognitive load. De Jong (2010) questions the additivity of the three types of cognitive load given the different nature of intrinsic cognitive load. The additivity of three cognitive loads to get a total has been somewhat answered in the redefining of the categories of cognitive load but for reasons involving the different nature of germane instead of intrinsic cognitive load. Kalyuga (2011) and Sweller (2010a, 2010b) have removed germane cognitive load from the equation as there is overlap between the definitions of intrinsic and germane cognitive load and for this reason they suggest a revision to the additivity rule from total cognitive load being the addition of all three types of cognitive load to the addition of intrinsic and extraneous cognitive load only. In terms of asking students about the effort required to understand the material which is regarded as intrinsic cognitive load (DeLeeuw & Mayer, 2008), this change makes sense in that what the students are actually thinking about is the processing that they are undertaking to understand the material, which is now included in intrinsic cognitive load as there is now no distinction between intrinsic and germane cognitive load at a macro or micro level (Kalyuga, 2011). However, there are still questions as to whether the addition of intrinsic and extraneous cognitive load to get the total cognitive load is too simplistic in that it does not take into account offloading from one working memory subsystem to another (Brünken, Plass, & Moreno, 2010).

In summary of the discussion of total cognitive load Kalyuga (2011) and Sweller (2010b, 2012) suggest that extraneous cognitive load always needs to be reduced and that intrinsic cognitive load be managed.

Measurement of cognitive load. A major limitation identified by many researchers is that there is no consensus on measurements critical to cognitive load theory, namely the categories of cognitive load, difficulty, mental effort, expertise, and prior knowledge (Brünken, Plass, & Moreno, 2010; de Jong, 2010;
Moreno, 2006, 2010; Moreno & Park, 2010). Techniques for measuring cognitive load have not been universally agreed upon either, especially in authentic settings where the use of assessment techniques is somewhat limited. Sweller (2010a) states that subjective rating scales can be used to determine total cognitive load but this is questioned by others. Cierniak, Scheiter, and Gerjets (2009) used different questions to attempt to access different working memory loads imposed but they concluded that this implies that learners can interpret the different questions asked in the way that the researchers intended, which they questioned in their study. DeLieuw and Mayer (2008) and Leppink et al., (2013) also used different rating scales for the different categories of cognitive load but these have not been widely adopted. Within the revised framework, Kalyuga (2011) and Sweller, Ayres, and Kalyuga (2011b) indicate that finding separate empirical measures of the three different categories of cognitive load is now redundant as germane cognitive load is now incorporated in the construct of intrinsic cognitive load and therefore is not regarded as a separate measure. As there are now only two measurable sources of cognitive load, intrinsic and extraneous which are easily distinguishable, it is therefore possible to measure these by experimental means. For example, measuring intrinsic cognitive load is possible if extraneous cognitive load is kept constant and conversely if intrinsic cognitive load is kept constant then extraneous cognitive load could be measured. Schnotz and Kurschner (2007) do not consider the measurement of individual categories of cognitive load as an issue for current views of cognitive load theory. They agree that total cognitive load as a variable should be measured but see cognitive load theory as a theoretical framework for guiding research and research based practice, as such all the individual variables do not require empirical measurement. They give the example of schema theory which does not have an empirical measurement for a schema but as a theory has been instrumental in guiding research and other theoretical frameworks. Therefore, the reduction to two main types of cognitive load which are fundamentally different, and considering that both can be managed, reduces the need to develop individual measures.

Van Gog, Kester, Nievelstein, Giesbers, and Paas (2009) used eye tracking, concept mapping and verbal reporting as techniques to uncover the processing involved in the different categories of cognitive load, they conclude by suggesting that these be used to extend information gained from performance and cognitive load measures, i.e., concept mapping can give added information on the quality of
cognitive structures. However, allocating the processes to categories of cognitive load was not possible. Moreno (2006, 2010) and de Jong (2010) question the reliability and validity of different versions of subjective questions asked, they give examples of differences in questions asked, number of units on the scale, labels on the scale, and of administering the scale. They are also concerned that researchers have used a different question to the original one used by Paas (1992) and have still quoted the reliability used in this study, for this reason van Gog and Paas (2008) recommend that reliability be calculated for each study.

Moreno (2006, 2010) and de Jong (2010) also question the interchangeable use of mental effort and task difficulty questions and suggest that there are differences in outcome (conclusions) depending on the question used. For example, in extreme cases mental effort does not equate to task difficulty, i.e., if the task or learning materials are very difficult participants may chose not to invest mental effort and so score this relatively low, when this low score (which will be interpreted as easy or little effort needing to be expended for the task) is paired with performance (which will correspondingly be low) in the instructional efficiency calculation, the result will be opposite to what is expected. Schnotz and Kurschner (2007) acknowledge that subjective rating scales are often used but they raise doubts that learners can report on their mental effort expended during or after learning but concede that if they are used carefully they can provide valuable data particularly in authentic settings, which is the case in this study. However, Paas, van Merriënboer, and Adam (1994) found that subjective rating scales were more successful and sensitive to variations in instruction than physiological measures, which are more intrusive and not easily administered. Ayres (2006) also found that by controlling extraneous load that subjective rating scales reflected both the difficulty of problems and were highly correlated with respect to error rates.

With respect to timing of administering the subjective rating scale, van Gog and Paas (2008) suggest if looking at extraneous cognitive load then it should be administered during the learning phase and if looking at intrinsic/germane cognitive load then it should be administered during the testing phase. Kalyuga (2011) disagrees with this and suggests that as cognitive load theory focuses on what cognitive processing occurs during schema acquisition (learning), cognitive load measures should always be administered during the learning. Further recent studies by van Gog, Kirschner, Kester, and Paas (2012) suggest that repeated
measures during the learning are favoured. Sweller, Ayres, and Kalyuga (2011b) acknowledge that although there have been some discrepancies in the consistency of subjective measurement of cognitive load, they suggest that this measure has had a profound influence on cognitive load theory research and has proven to be a useful tool to measure cognitive load in a large number of studies.

Another technique that has been used to measure cognitive load is dual task methodology. Brünken, Plass, and Moreno (2010) raise issues regarding the validity of dual task methodology in that they question whether conclusions can be made regarding the cognitive load of the primary task if motivation and engagement are not considered as part of cognitive load theory. In addition, Schnottz and Kurschner (2007) suggest that it is also possible that the participants try to maintain their secondary task performance at the expense of the primary task performance. In a study which employed dual task methodology, students chose to attend to the secondary task in the first instance and only attended to the primary task between the beeps, and in some instances not at all (Haslam & Hamilton, 2010). In this instance, the secondary task was not measuring what was it was supposed to be measuring. De Jong (2010) favours this dual-task method as it gives an instantaneous and continuous measure of cognitive load throughout the task (See also, Sweller, Ayres, and Kalyuga, 2011b). However, de Jong (2010) does not acknowledge the disadvantages of this method particularly in authentic settings and in this respect his discussion is limited.

Moreno (2010) and de Jong (2010) have suggested that cognitive capacity is rarely measured in cognitive load theory research despite there being a set of well developed intelligence tests that will measure this construct. Consequently, it would be difficult to claim that learners are overloaded given that this prediction cannot be tested without a measure of cognitive capacity. Cognitive load theory uses measures of very high mental effort or difficulty paired with low performance as indicators of overload which would also be the case despite measuring cognitive capacity. However, having access to cognitive capacity would enable the researcher to quantify the instruction in terms of this measure and would then be able to determine roughly the resources that would have to be allocated to meet the demands of the instruction. Kalyuga (2011) also advocates the measurement of cognitive capacity but for different reasons. He suggests that
there is a relationship between cognitive capacity and performance, i.e., high cognitive capacity has been found to correlate with high performance (See also Seufert, Schutze, & Brünken, 2009), and therefore advocates the use of measures of cognitive capacity as an individual difference measure that could impact on the cognitive load experienced during instruction.

Another and alternative method of indirectly measuring cognitive load involves performance measures e.g., test scores (Owen & Sweller, 1985), error rates (Ayres, 2001; Ayres & Sweller, 1990; Sweller & Cooper, 1985), and instructional time (Chandler & Sweller, 1991,1992). Brünken, Plass, and Moreno (2010) also question whether the presentation mode of the instruction is aligned to the performance test i.e., if the presentation is a computer-based animation and the performance measure used is a pen and paper multichoice test, is this appropriate and valid? However, they acknowledge that this is a very new field of research in multimedia learning and suggest that cognitive load theory research do the same. Despite all the research that has been carried out over 30 years, Kirschner, Ayres, and Chandler (2011) state that in their opinion the most problematic issue in cognitive load theory is the measurement of mental load (cognitive load).

In summary, cognitive load theory has been used to guide instructional design particularly in relation to novices learning complex information for decades. This section has summarised criticisms that have been raised regarding many aspects of cognitive load theory, some of these have been addressed by cognitive load theory researchers and have resulted in changes to the theory, others have not. The unresolved issues will need to be addressed if cognitive load theory is to continue to be effective in influencing instructional design especially in regard to authentic learning environments, and in particular, the consistent measurement of cognitive load. As Moreno (2010) states, the measurement of cognitive load both globally and individually is the most important aspect of cognitive load theory that requires addressing in terms of universal acceptance, consistency, reliability, and validity. Scientific theories are never proven (unlike facts) but evolve with evidence and critique, cognitive load theory has reached the stage where it needs some modifications to remain viable.
The next section will discuss the cognitive theory of multimedia learning (Mayer, 2005a, 2014a) which is based on cognitive load theory. This theory concerning the design of multimedia presentations is applicable to this study as information was presented to participants using pictures and words.

**The Cognitive Theory of Multimedia learning (CTML)**

This is one of the most influential cognitive theories on how people learn from instructional messages that contain pictures and words (Eitel, Scheiter, & Schüler, 2013). That is, it is a learner-focused theory not a technology-focused theory (Mayer, 2005a, 2014a) and is based on earlier theories of information processing and cognitive load theory (Sweller, 1988, 1989, 2010a; Sweller, van Merriënboer, & Paas, 1998). The main focus is on the design and presentation of pictures and words (multimedia) to facilitate processing and knowledge construction. The main hypothesis is that learning is improved by incorporating pictures and words as an integral part, in the presentation of new knowledge. The following section will look at the development of the theory, its main ideas, implications for designing multimedia instruction, future developments, and its relationship to another theory of understanding the presentation of pictures and words.

The CTML was proposed following a lengthy and extensive research programme spanning over 15 years. Over the years, the CTML has evolved from a variety of iterations of models and theories concerned with meaningful learning and the influence of verbal and visual materials within this context. Mayer (1989) first introduced the idea of learning from pictures and words within a model of meaningful learning. Soon following this, Mayer and Gallini (1990) when describing illustrations containing pictures and words referred to cognitive conditions for effective illustrations and then in the same year, Mayer and Anderson (1991) discussed the dual coding model. A few years later Mayer, Steinhoff, Bower, and Mars (1995) referred to a generative theory and Mayer (1997) renamed it the generative theory of multimedia learning only to be replaced by the dual processing model of multimedia learning (Mayer & Moreno, 1998). Finally, the CTML was introduced (Mayer, Bove, Bryman, Mars & Tapangco, 1996) but CTML was not officially proposed as the name of this theory until 2003 (Mayer & Moreno, 2003). Although there have been multiple name changes and
some refinement to the theory over the years, the main assumptions of the CTML have remained the same (Mayer, 2014a). The next section describes the human cognitive processing system and its functions related to processing pictures and words according to the CTML before describing the three main assumptions of the CTML in more detail.

Processing pictures and words in CTML.

The cognitive model of multimedia learning explains how people learn from multimedia instructional messages involving the human cognitive processing system, this system was previously described in the cognitive load theory section and is briefly summarized here. This model includes three memory stores (Atkinson & Shiffrin, 1968): sensory, working, and long-term memory. Sensory memory has an unlimited capacity and is where visual and auditory information is stored for a brief length of time. Working memory has a limited capacity and stores information for a longer but still brief amount of time. There are two parts to working memory and two distinct processes (organising and integrating) that take place here. The first part of working memory involves incoming selected visual and auditory information (sounds and images) is held very briefly, before it is organised into verbal and pictorial (spatial) models (representations), which is then integrated with relevant existing information from long-term memory. This results in a mental model which is consequently encoded (transferred) to long-term memory. Long-term memory has unlimited capacity, no limits on its duration and is where learned information is stored permanently.

Main assumptions.

The CTML is based on three assumptions derived from previous cognitive theories: a) dual channel assumption, b) limited capacity assumption and, c) active processing assumption.

Dual channel assumption. The first assumption is that the human cognitive processing system has two channels, a visual/pictorial channel and an auditory/verbal channel which are both located in working memory. This is consistent with dual coding theory (Clark & Paivio, 1991; Paivio, 1991) and
Baddeley's (1992) model of working memory. When considering the two proposed channels (visual/spatial and verbal/auditory) it is important to distinguish between presentation-mode and sensory-mode approaches. The presentation mode approach is the way the information is presented, that is, verbal refers to the presentation of words, this can be in a spoken or written form, whereas, visual presentation refers to the presentation of non verbal information e.g., pictures, videos or animations (Paivio, 1991). The sensory mode approach refers to which sensory memory store (visual or auditory sensory memory) the information is first processed which corresponds to whether the eyes or ears first register the information and consequently where in working memory it is processed initially so e.g., visual information comes in via the eyes so is processed in the visual channel and auditory information (narration or non verbal sounds) comes in via the ears so is processed in the verbal channel initially (Baddeley, 1992). There is a conceptual difference between these two approaches and the channels that initially process this information, particularly for printed text and background sounds, e.g., printed text is processed in the verbal channel in the presentation-mode approach but in the visual channel for the sensory-mode approach. The CTML has overcome these differences by opting for a compromise. That is, Mayer (2005a, 2014a) has adopted a sensory-mode approach for how information enters the cognitive processing system, however, information may enter a particular channel but it can be transferred to another channel for processing. For example, for printed text, the CTML suggests that it enters the cognitive processing system via the visual channel but can be transferred to the verbal channel for processing. This is consistent with the idea of cross-channel representations and with dual coding theory (Paivio, 1991).

*Limited capacity assumption.* The second assumption of the CTML is that each channel has limited capacity for processing new information [also consistent with cognitive load theory (Sweller, 1988, 1989, 2010a; Sweller, van Merriënboer, & Paas, 1998) and (Baddeley, 1992)]. This applies to both the time it is able to be held for and the amount of information that can be processed [cf., Baddeley (1992) and Sweller (2005)]. The capacity limit of each channel is 5-7 chunks of information (Miller, 1956) which impacts on the amount of processing that can occur. Baddeley (2000) posits that a central executive is responsible for what is selected and therefore processed. The central executive serves a
metacognitive function in that it oversees the management and monitoring of processing within working memory. This differs from Sweller (2005) who suggests that schemas perform the function that Baddeley attributes to the central executive, as we attend to information which we recognise (already included in existing schemas) and so this is responsible for the attention we give to incoming information. The consequence of this limited capacity and the fact that a selection process occurs before information reaches working memory is that we don't usually remember exactly all the information that is presented to us in cases where the information is complex or there are a large number of elements.

**Active processing assumption.** The third assumption is that active learning requires active processing of the information presented in working memory and subsequent storage in long-term memory (Mayer, 1989; Wittrock, 1989). Included in the CTML is a description of the steps involved in the active processing of pictures and words in the human cognitive processing system to effect learning. This active processing involves the three cognitive processes of selecting, organising, and integrating new information. Selecting information is attending to relevant information and transferring it to working memory, organising refers to making internal connections between the ideas in the selected information and lastly, integrating is the process of making external connections between the selected material and existing information in long-term memory.

When describing how both pictures and words are processed there are five processes involved: selecting pictures, selecting words, organising pictures, organising words, and finally integrating pictures and words. The process can be described simply by using examples of images and words. Firstly, for spoken words, these are auditory and enter the sensory memory via the ears. The words are then selected and transferred to the verbal channel in working memory and are organised into a verbal model (connections between the selected words) and integrated with relevant existing knowledge from long-term memory and then, finally, encoded in long-term memory. Secondly, for an image e.g., a photo or magazine picture, this is visual information and consequently enters the sensory memory via the eyes, it is then transferred to the visual channel in working memory and organised into a pictorial model (connections between the selected visual information) then integrated with relevant existing information from long-
term memory and then, finally, encoded in long-term memory. In reality however, and in particular with multimedia learning, information presented contains both visual and auditory information, in this instance the mental model produced will be the result of the integration of both the verbal and pictorial models and information retrieved from long-term memory (prior knowledge). This is what is referred to as integrating pictures and words. Lastly, a slightly different process occurs for printed words in that these are considered visual so they enter sensory memory via the eyes and words are selected and transferred to the visual channel in working memory. They are then transferred to the verbal channel for organising into a verbal model and integrating with relevant existing knowledge in long-term memory. The processes detailed above are prerequisite for meaningful learning (Mayer, 2009, 2014a).

Meaningful learning in CTML.

Meaningful learning fosters a deep understanding of the target material and performance resulting from this type of learning can be measured by the transfer of this learning to new situations. The end result of meaningful learning is the construction of coherent knowledge structures (mental models) in long-term memory (Mayer, 2005a). These knowledge structures are different depending on the type of processing, e.g., a knowledge structure constructed after the process of comparison of two ideas will be different from the structure constructed after classifying different objects into groups (Mayer, 2014a). It is important when presenting multimedia instructional messages that the information is both well organised and provides guidance on how to construct a coherent knowledge structure that matches the type of processing required to build it (Mayer, 2005a).

As described above processing pictures and words requires extensive cognitive resources, which are limited, so the CTML categorises this processing according to the allocation of cognitive resources and the outcome of this processing.

Categories of cognitive processing and basic multimedia principles.

Initially the three categories of cognitive processing which represent the demands on cognitive capacity during multimedia learning identified in the CTML were
essential, incidental and representational holding (Mayer & Moreno, 2003). These
three categories were changed to intrinsic, extraneous and germane processing
which is aligned with the three types of cognitive load described in cognitive load
theory (DeLeeuw & Mayer, 2008). Currently the categories of cognitive
processing are referred to as essential, extraneous and generative processing to
more accurately reflect the cognitive processing that is occurring. As cognitive
resources in working memory are limited it is important to keep the amount of
processing within the learners’ working memory capacity so as not to cause
cognitive overload (which is also an important construct in cognitive load theory)

After extensive research involving over 100 studies based on the three categories
of cognitive processing, the CTML has developed evidence-based techniques or
principles to improve multimedia learning. The current 15 principles will be
discussed in the following paragraphs along with the relevant category of cognitive
processing according to the latest version of the CTML (Mayer, 2014a).

**Essential Processing.** Essential processing is the processing required to
make sense of or understand new information. This type of processing involves all
the activities that take place in working memory that are required to produce a
pictorial or verbal model. That is, it includes the processes of selecting the
relevant images and words and organising the selected words and images into the
two different models. Essential processing is more demanding if the material is
complex and so this concept is closely aligned with intrinsic cognitive load in
cognitive load theory. In situations where the material is complex, and the
learners are novices, the CTML principles for managing essential processing are
segmenting, pre-training, and modality. Segmenting involves breaking the
presentation into segments that the learner can process separately, and can also
involve learner pacing of the segments to allow time for processing cognitively
demanding information. Pre-training involves using a preinstructional strategy to
provide relevant prior knowledge to prepare the learners for when the new
complex to-be-learned information is presented. Modality involves balancing the
use of the separate channels in working memory so one channel is not
overloaded. With respect to the modality principle, Schüler, Scheiter, Rummer, &
Gerjets (2012) question the CTML assertion that the modality effect is due to
overloading and suggest that it is not always better to present visuals with spoken text as they found evidence that when presenting complex visual and written information in multimedia learning (same sensory channel) that the visual channel does not always become overloaded. It appears that in situations of simultaneous presentation of pictures and written text where overloading could occur, that learners ignore the pictures and concentrate on the written text.

**Extraneous processing.** Extraneous processing is any processing that is required of learners that does not contribute to understanding and learning. It is often caused by the poor design of instructional materials and is synonymous with extraneous cognitive load in cognitive load theory. If material to be learned is complex and extraneous processing demands are high, then learners may have little or no cognitive resources available for essential processing that is required for learning. The CTML principles aimed at reducing extraneous processing include coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. The coherence principle is overarching and states that people learn best when extraneous material is eliminated. Signaling involves using cues to highlight important ideas that are essential to understanding and learning the new information. Redundancy involves not using narration in addition to printed text in multimedia presentations. Spatial contiguity involves placing words and diagrams or pictures close to each other in a space while temporal contiguity involves presenting words and pictures that are related to each other at the same time.

**Generative processing.** Generative processing is the processing which directly results in changes in the knowledge structures in long-term memory. These include the processes of reorganising the pictorial and verbal models into a mental model and integrating it with relevant existing knowledge from long-term memory. This term aligns closely with germane cognitive load in cognitive load theory. The CTML principles aimed at promoting changes in long-term memory include the multimedia, personalization, voice, embodiment, guided discovery, self-explanation, and the drawing principles. The multimedia principle involves using words and pictures instead of just words and is considered the guiding principle for the CTML. Personalization involves using a conversational or speaking style when presenting verbal information rather than a more formal style, while the voice principle involves using the human voice for narration rather than a
computer generated voice. Embodiment involves making on-screen characters human-like in their actions. Guided discovery involves giving guidance and feedback when learners are problem-solving. Self-explanation involves asking learners to explain the essential learning to themselves, so essentially talking to themselves about the learning that they are involved in (reflecting on their learning). Lastly, the drawing principle involves learners drawing diagrams to summarise their learning (Mayer, 2014a).

Advanced multimedia learning principles.

Along with the basic multimedia learning principles already described there are a further eleven advanced principles which come under the umbrella of Multimedia Learning rather than exclusively under the CTML. The advanced principles of multimedia learning are guided discovery learning, worked examples, self explanation, generative drawing, feedback, multiple representation, learner control, animation, collaboration, expertise reversal, and individual differences in working memory capacity. Two of these have already been mentioned in the previous section, guided discovery and self explanation, and another two, worked examples and expertise reversal are further elaborations on cognitive load theory effects, so they will not be included in this discussion. The generative drawing principle involves learner support to draw diagrams while processing text to improve generative processing (meaningful learning). The feedback principle involves giving learners explanatory feedback to support their learning and guide further learning. The multiple representation feedback principle refers to support given to learners to help them integrate multiple specialized graphical representations in multimedia learning. The learner control principle refers to support given to learners enabling them to control the pace, sequencing or selection of information in multimedia presentations and is referred to as interactivity of learners with the information presented. The animation principle involves the use of strategies to guide the design of animation in multimedia environments. The collaboration principle refers to the use of group work to support learning in multimedia environments. Lastly, the individual differences in working memory capacity principle refers to specific individual differences that affect working memory processing and hence learning (Sorden, 2012). All the basic and advanced principles mentioned are effective if certain conditions are met and these
conditions could be related to the characteristics of the learner or the multimedia materials presented, and these are referred to as boundary conditions.

**Boundary conditions.**

Mayer (2009) introduced boundary conditions pertaining to the multimedia learning principles and cited instances where the principles may not be effective e.g., a condition which impacts the individual differences principle is that some instructional methods may be effective for novices but not for experts. This boundary condition aligns with the expertise-reversal effect in cognitive load theory (Kalyuga, Ayres, Chandler, & Sweller, 2003). Also, the modality effect can be found for system or teacher paced instruction but when the instruction is learner paced this modality effect disappears (Mayer, 2014b; Stiller, Freitag, Zinnbauer, & Freitag, 2009; Tabbers, 2002). De Jong (2010) has criticized these principles for these “exceptions to application” and sees these as a weakness of the theory and an excuse for when the principles do not work. Mayer (2010a), however, suggests that the CTML is not just a set of rules but it is a theory that is dynamic and these boundaries conditions refine the theory and its application to instructional settings.

**Motivation, metacognitive strategies, and the CTML.**

There is a difference between instructional methods aimed at reducing extraneous processing and managing essential processing, and ones that foster generative processing. Instructional methods to reduce extraneous and manage essential processing rely on ways of manipulating the material that is presented whereas instructional methods that foster generative processing involve motivating the learner to invest working memory resources in learning (to engage in the learning process). This awareness of the influence of motivation (increasing engagement) and metacognitive strategies (regulating processing) in the overall processes of learning led to the development of the cognitive-affective theory of learning with media (CATLM) (Moreno, 2006, 2009). This theory integrates the cognitive and affective aspects that impact learning from multimedia (Park, Plass, & Brünken, 2013). A limited number of studies have looked into techniques to foster generative processing by adding appealing graphics to motivate students (Magner, Schwonke, Aleven, Popescu, & Renkl, 2013; Plass, Heidig, Hayward, Homer, &
Um, 2013), or by adding challenging and contradictory scenarios to incite confusion and stimulate deeper engagement (D'Mello, Lehman, Pekrun, & Graesser, 2013). Although, initial findings are promising they are not convincing (Mayer, 2014b). The results are consistent with evidence that confusion in the initial stages of learning material can be beneficial for learning (Große & Renkl, 2007; Mayer, 2008). However, these motivational strategies only work if learners are not already overloaded processing essential and at times, extraneous material (Mayer, 2014b).

Overall the CTML suggests that to enhance meaningful learning the instruction should manage essential processing, eliminate extraneous processing, and foster generative processing (Mayer, 2009, 2014a). Essential and generative processing contribute to understanding and learning and can be viewed as constructive processing, whereas, extraneous processing diverts cognitive resources away from understanding and learning and therefore is destructive (Mayer, 2001, 2005a, 2014a). De Jong (2010) questions this triarchic model and suggests that essential and generative processing are the same with respect to learning and therefore difficult to separate. Kalyuga (2011) agrees with this suggestion and in relation to cognitive load theory suggests that it is difficult to separate intrinsic and germane cognitive load in terms of learning and that germane cognitive load is included in the definition for intrinsic cognitive load, therefore germane cognitive load is a redundant concept, this has already been discussed in the previous section on cognitive load theory. However, DeLeeuw and Mayer (2008) found some evidence for distinguishing and measuring the three distinct constructs and in so doing validating the triarchic cognitive load model in the CTML and cognitive load theory. In a reply to de Jong, Mayer (2010a) indicated that describing the three demands and their corresponding processing with respect to learning makes the theory more instructionally useful. In accordance with de Jong (2010) and Kalyuga (2011) it would be important to ascertain if both students and teachers (the target for these theories of instruction) could easily distinguish between both these processes and loads with respect to learning and if they could not, then defining them as essentially the same makes sense.
Future Directions.

Mayer (2014a) acknowledges that there are still areas that need to be incorporated into the CTML, as outlined above, i.e., metacognition and motivation. In addition, the theory needs to be widened to include new multimedia presentations e.g., games and mobile learning environments. A recurring theme throughout the development of the theory and subsequent refinements has been the need to move research out of university laboratories with university undergraduate students into real classroom settings with school students learning curriculum content with longer learning periods compared to the relatively short periods of less than five minutes that have been used for large numbers of the CTML studies (de Jong, 2010; van Merriënboer & Ayres, 2005). Mayer (2010) suggests that there have been some studies in actual settings in CTML and cognitive load theory. Harskamp, Mayer, and Suhre (2007) found that the modality principle in the CTML applies to multimedia learning in school settings in the context of science when time was a limiting factor. Tindal-Ford, Chandler and Sweller (1997) also confirmed the modality principle in cognitive load theory with trade apprentices in a training program of a large company when material was complex. Ayres (2006, 2013) and Pollock, Chandler, and Sweller (2002) confirmed the isolated-elements strategy in cognitive load theory with secondary school students in mathematics and with first year industrial trade students respectively. Mayer (2010a) also suggests that there are strengths in using multiple research methods, that both lab-based and school-based studies each have their own strengths and weaknesses and so it is both useful and beneficial to combine the results of the two environments.

Final Critique.

The evolving nature of this theory means that it has continuously addressed critique, looked forward and aligned with new findings. However, one aspect which has not yet been resolved is the measurement of cognitive load (de Jong, 2010; Schnotz & Kurshner, 2007), i.e., very few studies in the CTML programme have measured cognitive load. Mayer (2005a, 2014a) accepts this critique and has indicated that this is a future development for the theory but acknowledges that the strength of the CTML is in its ability to help instructional designers to
develop teaching and learning materials, so the research programme aims to make causal claims about different techniques using experimental comparisons and controls. That is, the studies aim to achieve an effect size ($d$) greater than 0.8 and learning is measured by transfer and recall tests (Mayer, 2009).

**Similarities and differences with the integrated text and picture comprehension model (Schnotz & Bannert, 2003).**

The CTML is not the only model for describing how people process and make sense of incoming verbal and pictorial information. Schnotz and Bannert (2003) have developed an integrated model of text and picture comprehension (ITPC) which consists of three memories, 2 channels, and two methods of sensory input via the auditory register (ear) and the visual register (eye) as in Mayer's (2014a) CTML. The end product of processing within ITPC also results in the construction of mental models and involves the integration of existing prior knowledge from long-term memory. However, there are differences between the two models. The ITPC consists of four types of processing (perpetual surface structure, semantic deep structure, descriptive, and depictive processing) in contrast to the CTML which consists of five processes (selecting pictures, selecting words, organizing pictures, organizing words, and integrating pictures and words). In ITPC, mental models and propositional representations (a forerunner to the final construction of a mental model) also link to information from long-term memory for integration as in the CTML. ITPC distinguishes between listening comprehension, reading comprehension, visual picture comprehension, and auditory picture comprehension. The CTML differs in that it only looks at processing of pictures and words. Both models account for processing visual and verbal information in the human cognitive processing system, although ITPC is a more complex model than the CTML.

In summary, the CTML is a successful theory that is supported and has been validated by a large body of research. It has been flexible and has accepted critique which has been reflected in on-going changes to the theory. This has meant that the theory has stood the test of time despite changes in the nature of multimedia presentations as technology has advanced at a rapid rate. Some areas are problematic and there are some unanswered questions that demand
new research methods or methods of measurement but with a strong foundation this theory looks likely to continue to influence multimedia teaching and learning.

The next section will discuss and compare the strategies of pre-training as one of the basic principles of the CTML and the isolated-elements strategy which is one of the cognitive load effects situated in cognitive load theory.

**Pre-training and Isolated-elements strategies**

Mayer (1975) called advance organisers a form of pre-training, that is, preparing a learner for learning target information. However, the term “pre-training” as a specific strategy to reduce the complexity of learning materials wasn’t suggested until around 2002. Pollock, Chandler, and Sweller (2002) called their strategy the isolated-elements approach and Mayer, Mathias, and Wetzell (2002) and Mayer, Mautone, and Prothero (2002) referred to a very similar strategy as pre-training. As described in cognitive load theory, by definition complex information has high element interactivity, i.e., lots of elements related to each other that need to be held in working memory simultaneously to be understood (high intrinsic cognitive load).

**Pre-training for reducing essential processing (intrinsic cognitive load).**

When complex material is presented at a fast pace to novice learners using a multimedia presentation the likelihood of essential overload (cognitive overload) is high. Essential overload occurs when the working memory capacity for essential processing (intrinsic cognitive load) is exceeded and this hinders further understanding and learning. Mayer (2005b) and Mayer and Pilegard (2014) distinguish between type 1 and type 2 essential overload. Type 1 overload occurs when both channels in working memory (verbal and pictorial) are overloaded whereas, type 2 occurs when only one channel is overloaded. One of the three strategies suggested by the CTML for managing essential processing so essential overload (type 1) is avoided is the strategy of pre-training. Pre-training is a specific preinstructional strategy that splits the presentation of essential material
into two stages, initially (stage one) only the names and characteristics of the main concepts or the main ideas and the states that they can exist in, for example, the main concepts and their definitions are presented to give the learners relevant prior knowledge for the next step, which is when all the essential information needed for understanding and learning is presented (stage two). In stage one there is no relationship between the elements, i.e., the elements acts as separate units while being processed in working memory rendering the material as low interactivity, therefore, making it easier to learn than the original learning materials. The provision of relevant prior knowledge in the first stage equips learners with preliminary schemass, which reduce the amount of cognitive processing during the final presentation (second stage) of all the essential material. So having preliminary schemas already in long-term memory makes it easier to learn the new information as it reduces the processing load on working memory (Mayer, 2005b). The reason for this is that it is easier to add onto existing schemas than to construct new schemas (Mayer & Pilegard, 2014). For example, when learning how a system works, e.g., the braking system on a car, the initial stage helps learners build a component model (learning the names of the parts of the system and how they can change) so when all the material is presented they can more easily build a causal model (how the whole process of braking occurs) of how the whole system works (Mayer & Pilegard, 2014; Mayer, Mathias, & Wetzell, 2002). This is referred to as a two-stage theory of mental model construction. So if learners were required to learn essential information about the physics topic of motion which includes graphing motion, (as was the case in this study) they would receive definitions, SI units regarding the concepts of time, distance, and speed and in addition basic graph information e.g., axes, titles, line, and bar graph examples.

The principle of pre-training has the support of a large number of studies whose effect sizes have been averaged to $d = 0.75$, this is over the threshold of $d = 0.4$ which is quoted by Hattie (2009) as being educationally important (Mayer & Pilegard, 2014). The three initial studies into the strategy of pre-training date back to 2002. Pollock, Chandler, and Sweller (2002) found that by reducing complexity of electrical engineering information relating to resistance testing in stage one, apprentices were able to learn more about how to perform resistance testing when all the information was presented in stage two. Mayer, Mathias, and Wetzell
(2002) found that by providing preinstructional material relating to the individual components in a car braking system or a bicycle tyre pump in stage one, participants who knew little about car systems or bicycle tyre pumps learned more about these systems when all this information was presented in the second stage of learning and were able to transfer this new knowledge to unfamiliar situations. Similarly, Mayer, Mautone, and Prothero (2002) found that providing a pictorial sheet (pictorial scaffolding) showing the major geological features that participants were asked to locate, before a simulation game lesson in the context of geology, helped them to solve more authentic geology problems than participants who did not receive the pre-training.

Subsequently, Clarke, Ayres, and Sweller (2005) found that students with low spreadsheet skills performed at a higher level after receiving spreadsheet instruction before receiving mathematics content in the area of graphic representations of linear functions using spreadsheets. Similarly, Kester, Kirshner, and van Merriënboer (2006) found that a form of pre-training that involved separating procedural and declarative information, helped students solve electrical circuit troubleshooting problems. Finally, Eitel, Scheiter, and Schüler (2013) found that pre-exposure to a picture provided a scaffold for students when learning from text in the context of physical pulley systems.

Although these studies have been identified by Mayer and Pilegard (2014) as being in support of pre-training, the strategies used are not actually referred to as pre-training by the authors of the articles except Mayer, Mautone, and Prothero (2002) and, Mayer, Mathias, and Wetzel (2002). Pollock, Chandler, and Sweller (2002) refer to their strategy as isolated-elements strategy and Eitel, Scheiter, and Schüler (2013) refer to the strategy as scaffolding instead of using the term pre-training. It is interesting to note that Mayer and Pilegard (2014) include Kester, Kirshner, and van Merriënboer (2006) as an example of pre-training but Ayres (2013) associates this study with the isolated-elements strategy. The following paragraphs describe the difference between these three strategies and additional research studies in support of these strategies.
Similarities between the pre-training and the isolated-elements strategies.

Pre-training and Isolated-elements are both preinstructional strategies in which specific essential material is presented in stage one, this material is processed serially and so is not considered complex (low element interactivity) and not likely to overload either channel with essential processing, the result of both strategies after stage one is that some preliminary schemas are constructed but understanding is limited (Pollock, Chandler, & Sweller, 2002). Then in stage two, for both strategies all the material is presented which is considered complex or high element interactivity, this material is required to be processed simultaneously and if all cognitive processing occurs then this results in full understanding. Both these strategies are beneficial for novices who could easily be overloaded with the cognitive processing required for understanding complex information.

Differences between the pre-training and the isolated-elements strategies.

The difference between pre-training and isolated-elements is that in pre-training, the specific essential information is the names and characteristics of the main concepts, so includes definitions, states and behaviours of the main concepts or ideas. This is in contrast to the isolated-elements strategy where the specific essential information is a simplified version of the main content, a reduction in elements of the full version (which is presented in stage two). If this is put into a problem-solving or training context, stage one involves partial worked examples or part tasks whereas, stage two involves full worked examples or whole tasks [See Ayres (2006, 2013) in the context of mathematics and Pollock, Chandler, and Sweller (2002) with electrical apprentices]. Blayney, Kalyuga, and Sweller (2010) showed that the isolated-elements strategy was most effective with low prior knowledge learners. Lee, Plass, and Homer (2006) used the isolated-elements strategy to reduce visual complexity to help students learn from computer-based science simulations. Gerjets, Scheiter, and Catrambone (2006) found that reducing high element interactivity materials by splitting the material into smaller successive units with lower element interactivity enhanced learning of problem-solving. Finally, the classification of the study undertaken by Clark, Ayres, and
Sweller (2005) as a pre-training study (See Ayres, 2013) is not consistent with the definition of this strategy. Although Ayres refers to this as a pre-training effect and places it in a section titled “Increasing prior knowledge (pre-training)” (Ayres, 2013, p116) in which Mayer, Mautone and Prothero (2002) and Mayer, Mathias, and Wetzell (2002) are also included. This study actually reduced the number of interacting elements by providing prior training on a skill required to be used simultaneously when learning mathematical concepts. Consequently, this was not a case of presenting the names and characteristics of the main concepts before presenting all the information to the students, so according to the definition (Mayer, 2005b; Mayer & Pilegard, 2014) it is not an example of employing the pre-training strategy.

The isolated-elements strategy is a direct approach to reducing element interactivity and is always situated in the CLT framework whereas pre-training can be associated with both cognitive load theory and the CTML. It is also interesting to note that Mayer and Chandler (2001) describe a strategy to assist learners to build a component model and then a causal model as the strategy of pre-training and they go on to use this as confirmation of a two stage theory of mental model construction. This is consistent with the strategy of pre-training but subsequently they changed the name to segmenting (Mayer, 2005a, 2014a). They then attribute the benefit of this strategy of segmenting to the fact that the multimedia presentation was divided into learner-paced segments for the group who performed the best in transfer tests and these learner-controlled chunks could be processed in working memory in their own time. In terms of reclassifying this strategy, this is appropriate as there was not a dedicated preinstructional period in which the names and characteristics of the main parts or concepts were administered as would be the case if the strategy was pre-training. Bos, Terlouw, and Pilot (2009) employed a novel approach to pre-training as they used a pre test as pre-training which was embedded in an introductory model before teaching science concepts in the disciplines of physics, chemistry, biology, applied mathematics, and computer sciences. They found this strategy produced higher learning gains for students involved in these introductory models. However, this strategy is not altogether consistent with the definition of pre-training (Mayer & Pilegard, 2014) as the material does not contain names and characteristics of the main concepts. Finally, Eitel, Scheiter, and Schüler (2013) employed the strategy
of presenting a picture (for different amounts of time) before reading a text, so the picture had the effect of facilitating the building of a spatial pictorial model which acted as prior knowledge for understanding the text. Although Eitel, Scheiter, and Schüler (2013) refer to this strategy as “scaffolding”, this approach is consistent with a pre-training strategy in that it was used preinstructionally with specific information which enabled the learners to build preliminary schemas to aid understanding of the text, similar to Mayer, Mautone, and Prothero (2002). The strategy of scaffolding is discussed in more detail in the section discussing the 4C/ID model for learning complex tasks.

In summary, the pre-training and isolated-elements strategies have proven effective in reducing essential processing (intrinsic cognitive load) in a number of studies across a wide range of contexts, these studies and the contexts in which they are used are summarised in Table 2.
### Table 2

*Studies investigating the Pre-training and the Isolated-elements strategies*

<table>
<thead>
<tr>
<th>Study</th>
<th>Effective for learning about</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-training</strong></td>
<td></td>
</tr>
<tr>
<td>Mayer, Mathias, and Wetzell (2002)</td>
<td>braking systems</td>
</tr>
<tr>
<td>Mayer, Mautone, and Prother (2002)</td>
<td>how a tyre pump works</td>
</tr>
<tr>
<td>Clarke, Ayres, and Sweller (2005)</td>
<td>spreadsheets in mathematics</td>
</tr>
<tr>
<td>Kester, Kirshner, and van Merriënboer, (2006)</td>
<td>trouble shooting in electrical circuits</td>
</tr>
<tr>
<td>Bos, Terlouw, and Pilot (2009)</td>
<td>science concepts</td>
</tr>
<tr>
<td><strong>Isolated-elements</strong></td>
<td></td>
</tr>
<tr>
<td>Pollock, Chandler, and Sweller (2002)</td>
<td>electrical resistance testing</td>
</tr>
<tr>
<td>Ayres (2006)</td>
<td>problem-solving in mathematics</td>
</tr>
<tr>
<td>Lee, Plass, and Homer (2006)</td>
<td>science in computer based simulations</td>
</tr>
<tr>
<td>Blayney, Kalyuga, and Sweller (2010)</td>
<td>accountancy</td>
</tr>
<tr>
<td>Ayres (2013)</td>
<td>mathematics problem-solving</td>
</tr>
<tr>
<td>Gerjets, Scheiter, and Catrambone (2006)</td>
<td>probability problem-solving</td>
</tr>
</tbody>
</table>

Another recent model that has been developed which is consistent with cognitive load theory and the CTML with respect to its focus on the learning and understanding of complex information is the 4C/ID model. This model embeds components and strategies which are similar to the isolated-elements and pre-training strategies and the strategies focus on making it easier to learn and
understand complex information. The next section will describe the actual model and discuss similarities between this model and pre-training.

**4C/ID model**

Within the cognitive load theory framework there exists a model for learning complex, authentic tasks which include the use of multiple skills in real life situations. Cognitive load theory suggests that when novices are learning new information that extraneous cognitive load should be reduced and germane cognitive load optimised within the limits of working memory. However, when tasks are complex even if extraneous cognitive load is eliminated, the demands of the task may exceed working memory capacity due to high intrinsic cognitive load. In these situations cognitive load theory recommends only using germane inducing strategies with simple tasks. Van Merriënboer, Kirschner, and Kester (2003) oppose this approach and refer to a model designed for teaching and learning complex tasks within a cognitive load theory framework. This is called the four-component instructional design (4C/ID) model and it includes scaffolding practice and just-in-time information to control extraneous and intrinsic cognitive load in contexts where overload could easily occur. This model is based on and is consistent with cognitive load theory but is also consistent with the CTML. It was developed for use in educational settings using multimedia (computer-based) learning environments when learning complex skills in real-life contexts (van Merriënboer & Kester, 2005).

In real life-situations performing complex tasks includes not only a number of different tasks that need to be coordinated but also the integration of theory knowledge and appropriate use of each (i.e., when to apply each component) (Kester, Paas, & van Merriënboer, 2010). Another problem is that for some complex tasks the skills are sometimes used in a set order (e.g., operating an electrical circuit) but for other tasks there is not a set order (e.g., troubleshooting in the context of electrical circuits), these non recurrent tasks (no set order) impose a heavy load on working memory as novices do not have cognitive schemas to assist them in deciding which skills to use and when to use them. Experts on the other hand not only have cognitive schemas which guide their actions but some of the tasks are
automated which further reduces the load on working memory of completing these complex tasks (Kester, Paas, & van Merriënboer, 2010).

The 4C/ID model offers up a useful approach to support novices when faced with learning to perform complex tasks. This model developed by van Merriënboer (1997) includes four components. Firstly, the model is based on the use of learning tasks that are relevant, real-life, and complete whole tasks. Secondly, learners are given support information or scaffolding while completing the complex tasks. Scaffolding is a general term that refers to any support or guidance given to learners to assist learning (Reiser, 2004; Rosenshine & Meister, 1992). Traditionally, scaffolding referred to situations where a knowledgeable other (e.g., peer, parent, teacher, or expert in a field) helped a learner to achieve a task that would otherwise have been too difficult for them to perform on their own (Vygotsky, 1978). An important feature of scaffolding (which is particularly important for learning complex authentic tasks) is the guidance for not only what to do but also why to do it, i.e., it guides the learners regarding when to use the new knowledge they have acquired and why they are using it in this situation. Scaffolding is essential for the teaching of higher level thinking, complex information and learning complex skills (Rosenshine & Meister, 1992) for which the 4C/ID model is designed. An important feature of scaffolding is fading, i.e., reducing the guidance as learner's expertise increases (Rosenshine and Meister, 1992). Fading guidance is also an important feature in the 4C/ID model (van Merriënboer, Kirschner, & Kester, 2003). Scaffolding in computer-based learning (and utilised within the 4ICD model) can include reducing the options available when performing tasks so that the whole tasks become more manageable (Hmelo-Silver, Duncan & Chinn, 2007). Scaffolding is important when material is complex and learners are novices, this strategy can be viewed using cognitive load theory as a way of reducing the need for using search strategies which induce a high cognitive load (Sweller, 2010a). This has the effect of reducing overall processing (reducing cognitive load) to a manageable level within the learners working memory capacity. However, it is important that this scaffolding (guidance) be integrated with the tasks or split-attention can occur. The third component of the 4C/ID model is that procedural information is presented in chunks as needed (just-in-time)(van Merriënboer & Kester, 2005). Lastly, the fourth
component is part-task practice, where tasks that are required to be automated are practiced using exercises.

The 4C/ID model is consistent with cognitive load theory and concerned with instructional design control, using a variety of strategies (isolated-elements, sequencing from simple to complex, integrated guidance, worked examples, and part-task practice) to reduce intrinsic and extraneous cognitive load when learning to perform complex real life tasks, e.g. project and problem-based learning. Management of cognitive load is essential, as learning to coordinate the many and varied tasks and skills involved in complex learning environments could easily overload working memory capacity.

A new principle which has been developed for use with the 4C/ID model is the training wheels principle (van Merriënboer & Kester, 2014). This principle involves the reduction of elements initially to those that are absolutely necessary, as part of the whole task for novices. The 4C/ID model is similar to the isolated-elements strategy associated with cognitive load theory and the pre-training strategy associated with the CTML, in that all three are used in the teaching of complex information where it is essential to reduce the intrinsic cognitive load of the material so learning can occur, in situations where cognitive overload is likely, and for use in authentic educational settings.

Preinstructional strategies embedded in cognitive load theory and the CTML, and the 4C/ID model, whose purpose is to make it easier to learn complex information are consistent with earlier research which focused on the influence of advance organisers as a way of briefly introducing information relevant to upcoming instruction in advance, to support, and guide new learning (Ausubel, 1968; Mayer, 1983, 1979; Novak, 1980).

**Advance organisers, Graphic organisers, and Pre-training**

This section will discuss the nature, use and similarities of advance organisers, graphic organisers and the strategy of pre-training. The sections on advance organisers and graphic organiser will be followed by a comparison of the two
strategies and finally, after the section on pre-training, an overall comparison of the three preinstructional strategies will occur.

**Advance organisers.**

The concept of advance organisers was first introduced by Ausubel (1960) as an adjunct aid to learning from text (Ausubel, Novak, & Hanesian, 1978). Advance organisers are an abstract written statement read (sometimes twice through) by learners before reading the to-be-learned text and were designed to enhance meaningful learning or understanding of the main ideas in the target (to-be-learned text). Ausubel (1960) identified two types of advance organisers, expository and comparative advance organisers. Expository advance organisers are used when the learner has no prior knowledge and so the text in this advance organiser uses information that is related to the target material, providing links to relevant prior knowledge. Comparative advance organisers are used when learners have some prior knowledge of the target material (Corkill, 1992). The advance organiser in this case identifies similarities and differences between new and existing knowledge related to the target text (Corkill, 1992; Shihusa & Keraro, 2009).

Advance organisers were first used as a preinstructional reading strategy which introduced higher level concepts in advance of the target learning. In this way, they could also be described as providing scaffolding before the reading of new information in text (Robinson, 1998). Ausubel (1960) defined meaningful learning as the correct linking of existing and new information in long-term memory, which facilitates understanding. This is in direct contrast to rote learning in which material is transferred to long-term memory in verbatim form, and therefore, not connected to appropriate prior knowledge so not understood (Ausubel, Novak, and Hanesian, 1978; Mayer, 1979; Novak, 1980). According to Ausubel's theory of cognitive subsumption, if the pre-reading passage was at a "higher level of abstraction, generality, and inclusiveness" (Ausubel, 1968, p. 148), then it acts as both a framework and a bridge to anchoring these new ideas to existing ideas held in hierarchical structures in long-term memory (Corkill, 1992). These cognitive structures are now called schemas (Rumelhart, 1984; Sweller, 2010a). So in effect, advance organisers activate prior knowledge and assist the learner to attach the new knowledge to the correct existing knowledge in long-term memory.
As the concepts in the advance organiser and the to-be-learned material are not the same, that is, the material in the advance organiser is at a higher level, this helps the learner to build a framework in advance of learning to anchor or incorporate the new knowledge. This is consistent with Mayer (2005a, 2009, 2014a) who as previously described, lists three processes which must happen before meaningful learning or understanding can occur: selecting information, organising information, and lastly, integrating information with existing knowledge in long-term memory. It is within the last process of integrating new and existing knowledge that advance organisers have the greatest effect.

Further research by Mayer (1979) and Derry (1984) refined Ausubel’s definition of advance organisers to include any vehicle that acts in assisting the reader to access prior knowledge relating to the new information in the target text. In this expanded definition, the advance organisers provide anchors for the main ideas from the to-be-learned material. As a result, this prior knowledge would also help the reader to make sense of and learn the new information and could even assist in elaborating or modifying their existing schemas. This new definition helped explain some instances when advance organisers were found to be ineffective: a) if the advance organiser is poorly written and not easy to read or understand, b) when the learning is predominantly technical details or prose, c) when the material was not meaningful, d) when the learners were not given enough time to read the advance organiser or the target information, e) if learners were already able to activate relevant prior knowledge themselves and so the advance organiser was redundant or extra information and f), if learners had difficulty reading, i.e., their reading ability is low, presenting more reading in addition to the target material may not prove to be beneficial for learning (Corkill, 1992). Overall advance organisers have been found to be an effective strategy for helping learners understand and recall main ideas from text. In a review of advance organisers and their effectiveness Corkill (1992) reports that 24 out of 29 experiments which conform to Ausubel’s strict guidelines were effective for learning from unfamiliar text. A selection of these studies report that advance organisers are effective for facilitating the recall of main ideas (Mayer & Bromage, 1980), recall of badly organised text (Mayer, 1979), making inferences from the text (Mayer & Bromage, 1980) particularly for lower proficiency students (Tudor, 1986), and problem-solving (Mayer, 1984).
Graphic Organisers.

Graphic organisers are graphic or visual displays of information with a spatial component which take the form of a hierarchical structure. This is intended to help learners to organise information so it can be successfully stored (encoded) within appropriate familiar structures (schemas) in long-term memory. The suggestion is that graphic organisers are effective because schemas are also hierarchical structures and they enhance the effective understanding and encoding of the target information (Horton, Lovitt, & Bergerud, 1990). Some examples of graphic organisers are concept maps, venn diagrams, tables, charts, and flow charts (Robinson, 1997). Graphic organisers, therefore, show the relationship between ideas in the target information and they highlight the main ideas to be learned and important vocabulary (Robinson, 1997). Research on graphic organisers suggest that they are effective in, promoting higher level thinking, metacognitive skills, and critical thinking (Alshati, Waiters & Kidman, 2011; Bellanca, 2007; Parry & Gregory, 2003), motivating students to learn and developing problem-solving skills (Rosenshine, 1995), encouraging independent thinking and active learning (Billmeyer, 2003), facilitating active and meaningful learning and assisting students to recognise patterns, and relationships especially in e-learning (Amin, 2005).

According to Mayer (2005, 2009, 2014a), graphic organisers which contain verbal and visual information, help students build internal connections between ideas in the target information, that is, organise the information in preparation for storage in long-term memory, as material that is organised is easier to incorporate into existing schemas.

It appears that many of the earlier studies which looked at advance organisers were not strictly adhering to the original definition proposed by Ausubel (1960) and Ausubel, Novak, & Hanesian (1978). This appeared to be due to the fact that the original definition was vague and did not provide guidelines for their construction or use in education (Corkill, 1992). So many of the reported studies (Barnes & Clawson, 1975) used what we currently refer to as graphic organisers (Mayer, 1984) and therefore were not strictly advanced organisers as described by Ausubel (1960). This led to confusion as to whether these adjunct aids (based on their interpretation of the original construct) were beneficial or not as the results differed widely on their actual effectiveness in promoting meaningful learning.
A comparison of advance organisers and graphic organisers.

The following discussion will focus on a comparison of advance organisers and graphic organisers with respect to intention for use, how they affect the human cognitive system, what information is contained in each and when you would typically use them to promote learning. The intention of both advance organisers and graphic organisers are to enhance understanding of target material and both can be described as adjunct aids to learning new information but there are distinct differences between the two types of organiser. Firstly, according to Mayer (2005a), advance organisers and graphic organisers involve different parts of the human cognitive processing system, advance organisers involve making external connections (between working memory and long-term memory) and graphic organisers involves making internal connections (in working memory). Secondly, advance organisers do not contain ideas from the target information but graphic organisers do. Thirdly, advance organisers are only written text and graphic organisers are organized, structured visual representations of the information. Lastly, advance organisers are always prepared and administered by the teacher in advance of learning the target information but graphic organisers can be generated by the teacher or the student and can be used before, during or after the learning of the target material.

Pre-training.

Pre-training as already described in the previous section is a two stage process used for the learning of complex information which attempts to deal with the possibility that, for novices, the new to-be-learned information could potentially overload their cognitive processing system (Mayer, 2005b; Mayer & Pilegard, 2014b). In stage one of the pre-training strategy, the learners are presented with the main concepts and their characteristics. New preliminary schemas are constructed in long-term memory in stage one, and these preliminary schemas are then used in stage two when all the complex information is presented to help the learner make sense of it.
A comparison of advanced organisers, graphic organisers and pre-training.

The aim of all three strategies is to promote learning of new information, but advance organisers and pre-training both do not lead to understanding of the target material immediately after their use, that is, learners who were involved in pre-training or who used an advance organiser and then sat a test of transfer (understanding) which only included the target material would not be advantaged in any way. The reason for this is that advance organisers do not contain any of the target material and pre-training only contains unrelated ideas or elements of the target material so no connections are made between the ideas and therefore this will not facilitate understanding. In contrast, graphic organisers do promote understanding of the target information while they are being used by the learner, so if these learners were tested on understanding of the target material after the presentation of the graphic organisers, these learners would be advantaged. All three strategies are intended to promote understanding of the target material and therefore are not beneficial for encouraging rote learning.

When comparing the cognitive processes that each promote, advance organisers promote appropriate encoding in long-term memory, graphic organisers promote organisation of the information in working memory and pre-training promotes the construction of preliminary schemas in long-term memory when no schemas exist so it is not concerned with relating new information to existing schemas. Pre-training is also used with novices when the to-be-learned information is complex whereas advance organisers and graphic organisers involve new learning but not necessarily complex information that could potentially overload the cognitive processing system. Also, there are two categories of advance organiser, expository and comparative, which are used for different types of learners but there are not these same distinctions between graphic organisers and pre-training. All three strategies assist learners by making the learning of target information easier but as an actual strategy for learners, pre-training is not cognitively demanding, whereas, advance organisers and graphic organisers require active processing of information and building connections which are potentially demanding. Lastly, looking at the mode of delivery, advance organisers are in
written text only format but graphic organisers and pre-training can include spatial, visual and textual formats.

In summary all three strategies promote meaningful learning of target information but the effect they have on the cognitive processing system, how and when they are used and their mode of delivery are quite different. Pre-training is beneficial for facilitating understanding of complex information (visual and verbal) and is relatively easy to prepare but always by the instructor. Advance organisers are used specifically with written text, however, these are more difficult to prepare and also have to be provided by the instructor. Graphic organisers can be used with visual and verbal information and can be constructed by learners or instructors.

Given the central role that spatial ability plays in creating and interpreting graphs, the next section will discuss the impact of spatial ability on complex learning involving graphing and science achievement as well as the different patterns of performance and behaviour of male and female students within and related to science courses as influenced by spatial ability skills.

Spatial Ability

This section will discuss different aspects of spatial ability, how it is related to learning STEM (Science, Technology Engineering and Mathematics) subjects, in particular learning physics concepts which involve the use of graphs, e.g., the topic of motion, and lastly, factors which affect individual differences with respect to spatial ability.

Aspects of spatial ability.

Spatial ability is “the ability to generate, retain, retrieve, and transform mental images” (Mayer, 2001 p.172) or the ability to spatially transform, whole or component parts of mental images (Kozhevnikov, Motes, & Hegarty, 2007). Spatial ability is an important skill that is used everyday in tasks, e.g., reading maps, driving a car, and assembling kit set furniture (Feng, Spence, & Pratt, 2007). Spatial ability is not one single ability but has been separated into a range of different but related abilities or factors. There is, however, not a consensus on
the categories used or one universal set of tests that can be employed for identifying these aspects of spatial ability. This section will look specifically at the categories or factors of spatial orientation and spatial visualization.

According to Ekstrom, French and Harman (1976) spatial orientation is the ability to mentally rotate whole 2D and 3D images whereas spatial visualization is the ability to mentally transform component parts of an image. They suggest that the latter ability is a more complex skill, but not all researchers in this field agree with this. Tests for both of these factors involve speeded rotation tests, that is, they are timed and the object is to finish the tests in the time allocated. As is common in research there is not universal agreement on the factors contributing to spatial ability or the definitions of these factors, in particular relating to spatial orientation. In contrast to Ekstrom, French and Harman’s (1976) definition for spatial orientation, Kozhevnikov and Hegarty (2001), who have contributed widely to this field, define it as the ability to imagine how objects look as an observer from different perspectives or physical positions of the observer i.e., spatial ability in relation to one’s own position. However, they do agree that spatial visualization and spatial orientation are different spatial abilities but closely related. For the purposes of this study, we have adopted Ekstrom, French, and Harman’s (1976) definition of spatial orientation and visualization and have used their spatial ability tests developed specifically for the factors of spatial orientation and spatial visualization. Spatial ability has been found to be important in learning physics and in particular graphing motion which is the context for Study 2 (Kozhevnikov, Motes, & Hegarty, 2007).

**Spatial ability and learning STEM subjects.**

Educational research has investigated spatial ability and how it affects learning STEM (Science, Technology, Engineering and Mathematics) subjects, the link between the two is that both learning and spatial ability involve generating, maintaining and manipulating mental images in working memory (Mayer, 2001). As suggested earlier, working memory is made up of a central executive and 2 separate limited capacity processing subsystems called the phonological loop and the visuo-spatial sketchpad (Baddeley, 1992). Spatial ability tests of mental rotation reflect an individual’s memory capacity for visual and spatial processing
(Just & Carpenter, 1985). However, more recent studies undertaken due to the emergence and importance of multimedia and hypermedia learning environments have provided evidence that the visuo-spatial sketchpad may itself consist of two subsystems which work independently (i.e., spatial and visual imagery subsystems) (Logie, 1995; Kozhevnikov, Hegarty, & Mayer, 2002). A revised classification describes people as verbalizers (people who prefer to process verbal information) and, high and low spatial visualizers (people who prefer to process visual information). This suggests that spatial manipulation competes for space in working memory with visual images and therefore, if the information is complex the visuo-spatial sketchpad may become overloaded. High spatial visualizers may concentrate on spatial manipulation at the expense of detail on the visual image and in this way they reduce the load on working memory and therefore avoid possible overload. In contrast, low spatial visualizers may concentrate on the detail in the image and may not have the capacity to spatially transform the image (Kozhevnikov, Hegarty, & Mayer, 2002). These categories would have an impact on learning most STEM subjects as these rely heavily on both visual and verbal information, and spatial ability particularly where graphs and diagrams are concerned (Kozhevnikov, Motes, & Hegarty, 2007). Spatial visualization has been found to be an important skill in solving physics problems especially those relating to the interpretation of kinematics (motion) graphs as these involve building and manipulating mental images of the motion that is represented by lines on a graph (both speed-time and distance-time graphs)(Kozhevnikov, Motes, & Hegarty, 2007; Larkin, 1982). There is evidence that competing visual/spatial processing demands increase as information becomes more complex in physics, especially in the topic of motion, and so this could potentially cause cognitive overload (Kozhevnikov, Hegarty, & Mayer, 2002). Students with low spatial ability have difficulty visualising the motion represented on a graph and interpret it as the actual movement of the object (Clements, 1983; Kozhevnikov, Motes, & Hegarty, 2007; Kozhevnikov & Thornton, 2006).

As physics learning involves the integration of visual and verbal information, multimedia presentations are an appropriate vehicle for delivering this subject matter. Some kinds of multimedia presentations benefit some learners more than others. Research has been conducted on the impact of multimedia presentations on individuals differing in terms of their spatial capabilities. Learners with higher
Spatial ability (greater visual/spatial processing capacity) were found to not only be better at creating and manipulating mental images from multimedia presentations but it was less likely that their working memory would become overloaded (Mayer, 2001; Mayer & Simms, 1994).

Spatial ability has also been found to be an important indicator of future study and employment in STEM related work. In a longitudinal study, a large percentage of students who scored highly on spatial ability tests at age 13 went on to enter STEM careers (Shea, Lubinski, & Benbow, 2001; Wai, Lubinski, & Benbow, 2007, 2009). In the area of learning physics, in particular, there is a high correlation between successful learning of physics concepts and spatial ability but not between learning physics concepts and verbal ability (Kozhevnikov & Thornton, 2006).

**Correlates of spatial ability.**

There are many proposed environmental and genetic factors that could contribute to an individual’s spatial ability. In this section, two of these will be discussed: gender and practice or training specifically targeted at improving or involving spatial ability skills.

**Gender.** A common finding in many studies involving spatial ability is a gender difference where males outperform females (Fennema & Sherman, 1977; Vandenberg & Kuse 1978; Linn & Petersen, 1985; Casey, Nuttall, Pezaris, & Benbow, 1995). However, there is also evidence that the magnitude of the gender differences are decreasing (Voyer, Voyer, & Bryden, 1995) and non-existent for spatial visualization (Casey, Nuttall, & Pezaris, 2005). In fact, even earlier studies found no difference in performance of males and females (including spatial visualization) except on mental rotation tests of spatial orientation (Fennema & Sherman, 1977; Benbow & Stanley, 1984; Linn & Petersen, 1985). An alternative explanation for gender differences on mental rotation tests is that these tests are timed and speed is a factor in scoring highly, however, because females consider alternative answers, they do not complete all problems on the test (Kail, Carter, & Pellegrino, 1979; Peters, Chisholm, & Laeng, 1995). On the other hand, some more recent studies have found that females outperformed males on mental
rotation tests but the difference was not significant (Nuttall, Casey, & Pezaris, 2005).

Gender differences have been found in adolescents but not in children younger than 12 years old. This may be due to the fact that traditional spatial ability tests are not appropriate for younger children and not that these differences do not exist (Kerns & Berenbaum, 1991). Titze, Jansen, and Heil (2010) support this and also found that traditional mental rotation tests were not suitable for children under the age of 13 years, as they did not understand what the test involved, the task was too abstract for them, and that it required too greater level of attention (Titze, Jansen, & Heil, 2010). Studies involving younger children using specially modified tests have found that boys outperformed girls only at ages 11-12 years (Hoyek, Collet, Fargier, & Guillot, 2012) that gender differences only appear from age 9 or 10. Other studies have found that gender differences increase with the onset of puberty (Kerns & Berenbaum, 1991; Vandenberg & Kuse 1978).

**Practise.** Another factor that could have an effect on spatial ability is time spent practising actual spatial ability tasks or skills that enhance spatial ability, that is, e.g., drawing diagrams or visualising situations. Battista (1990) found that performance on spatial ability tasks was greater for students that were exposed to prior teaching that required drawing diagrams to assist in the solving of mathematics problems. Sorby and Baartmans (1996) found that teaching high school students engineering drawing improved spatial ability, and another finding is that concrete experiences with drawing 2D and 3D geometric shapes also improved high school students spatial ability (Olkun, 2013). Repeat performance on spatial ability tests has been found to improve student performance on the second attempt at the same test (Casey & Brabeck, 1989). Also, Baenninger and Newcombe (1989) found a weak but reliable relationship between spatial ability and previous experience or prior participation in activities that promote spatial ability skills e.g., playing with lego blocks. They found a stronger relationship between actual training and performance on spatial ability tests.

Micro computer-based laboratories (MBL) use computer graphics in the teaching of physics concepts, e.g., using probes to graph motion in real-time so students can see the motion of the object and the construction of the graph of this motion at the same time. The use of these computer graphics within this context has been
found to improve high school students' performance on spatial ability tests. These computer programmes are in effect training students to develop their spatial ability skills, as MBL environments provide graphics to help students to improve spatial visualization, it is therefore these tests of spatial ability where improvements were found (Barnea & Dori, 1999; Kozhevnikov & Thornton, 2006). Finally, research at the University of Toronto found a marked improvement in adult’s spatial skills after playing video games which require these same spatial skills (Feng, Spence, & Pratt, 2007).

In summary, spatial ability can be divided into different aspects or skills and is tested with separate tests. Spatial ability is an important indicator of achievement in STEM subjects, i.e., learners with high spatial ability have higher achievement in STEM subjects both at secondary and tertiary study and then are more likely to go on to work in STEM related careers. Individual differences in spatial ability can be found relating to gender and prior experience, learning, or training in spatial ability skills. That is, males traditionally outperform females on spatial ability tests involving mental rotation, and experience, learning or training significantly improves spatial ability. In addition to finding a gender effect relating to spatial ability, there is much documented evidence for a gender effect relating to achievement in science. The next section will explore this issue.

**Gender and Achievement in Science**

The issue of gender differences in science has been debated since the 1970’s when equal opportunities for women also became an issue and consequently, this generated a considerable amount of research on the topic. There is widespread concern that compared to boys, girls are falling behind in many areas of science. “A science education must be available to all New Zealand students”, “all “ in this sentence refers to gender, race, ethnicity, abilities and disabilities (Ministry of Education, 1997, p11). The next section will look at the issue of gender differences in science with respect to achievement, class participation and teacher influences.
History of gender differences in science achievement.

The second IEA (International Association for the Evaluation of Educational Achievement) report of 1988 established a continuing gap in achievement between boys and girls, which increased with age and was more pronounced in the physical sciences (IEA, 1988). This was confirmed in the USA by the NAEP (National assessment and educational progress)(1978) report and worldwide in the TIMSS 1999 report on gender differences (Mullis, Martin, Fierros, Goldberg, & Stemler, 2000). Further to this, in the UK, boys outnumbered girls by 4 to 1 in A level physics, statistical data for GCSE (General Certificate of Secondary Education) confirm that boys outperform girls in physics and chemistry (Bell, 2001), and in high schools in the USA, Willingham and Cole (1997) found that there were more high achieving males than females in mathematics and science. At degree level in the UK in physical sciences, males dominated females by 81% and in engineering the figure is higher at 83% (Murphy, 2000). In New Zealand, the figures for 1987 show that at degree level only 16% enrolled in physics courses and 8.3% in engineering courses were female, in 1989 only 16% of female Year 13 students took bursary physics compared to 42% of male students, and in 1990 only 13% of ESR (Institute of Environmental Science and Research Ltd) scientists were women (Ministry of Education, 1991). In addition, it was found that girls were less likely to study science when it was no longer a compulsory subject (Baker, 1998).

Suggested reasons for gender differences in science achievement and participation.

No consistent evidence has been found that there are consistent cognitive differences between boys and girls (Murphy, 2000). However, Whyte (1986) found that girls scored lower on visual spatial ability tests and suggests that this is important in understanding why girls’ achievement is lower in the physical sciences. National surveys of science achievement in England reported that at ages 11, 13 and, 15 there was a marked difference between girls and boys, with boys outperforming girls in the application of physics concepts but girls outperforming boys in practical tasks (Black, Harlen, Johnson, & Palacio, 1988). Murphy (2000) suggests that as achievement is related to success, and boys have
been found to have a more positive attitude to physical sciences and a higher level of interest, this could account for these differences in this particular area of science. This could also be a reason why boys more often choose to study physical sciences at higher levels compared to girls (Tobin, Kahle, & Fraser, 1990).

There are interest differences in areas of science, which are gender linked i.e., girls prefer health, human biology and plants whereas boys preferred cars, motors, machinery and space travel (Jones, Howe, & Rua, 2000; Murphy, 1991, 2000). Weinberg (1995) also confirmed after a meta-analysis of students from 1970 – 1991 into gender differences in attitude, interest, and achievement that boys had a more positive attitude to science and that positive attitudes result in higher achievement. These findings are consistent with Keeves and Kotte (1992) who looked at students from ten countries aged 10, 14, and 18 and found that females reported that science was harder to learn, males showed more interest in studying science, that the magnitude of the difference between male and female student’s attitudes increases from elementary school to high school, and that males have higher achievement in physics, chemistry, and earth science topics.

The Secondary Science Curriculum Review (SSCR) (1987) suggests that differences in the everyday “out of school” experiences between boys and girls, could account for gender differences in achievement, participation, and interest found. For example, in relation to making models from a kit, 6% of girls reported that they had engaged in this activity quite often compared to 42% for boys, and with respect to knitting or sewing, 46% of girls reported that they had engaged in this activity quite often compared to 5% for boys. This could account for the differences found in practical work and complex problem-solving. Baker and Leary (1995) suggest that science experiences and attitudes could affect career selection, they found that overall girls were not as interested in a physics career as it did not involve helping people. Jones, Howe, & Rua (2000) agree as a result of surveying 6th graders on their attitudes, out of school experiences related to science and science topics of interest.

**Class interest and participation in science.** Given the previously described difference in patterns of outside of school interests as related to science
and participation in science, it is not surprising that girls show less interest in science as they often see little relevance, particularly in the physical sciences (Ministry of Education, 1997). This section will look at the possible reasons for the differential participation of males and females in science, e.g., writing focus/skills, the use of models/illustrations to portray science as a male domain in textbooks, risk-taking, and activity preferences. The lack of participation by girls (Tobin, Kahle, & Fraser, 1990), could be due to the ways girls prefer to write which is undervalued in science (i.e., looking at different perspectives and writing extensively and reflectively) whereas the way boys write is valued in science (i.e., focused and factually) (Murphy, 2000). Analysis of illustrations in science texts, show a large percentage of male illustrations (e.g., pictures and photos showing males in action) compared to female ones. In Australia, UK, and USA, 78% of illustrations in science texts were male (Tobin, Kahle, & Fraser, 1990). In a New Zealand analysis of the most commonly used junior science text, 82% of the illustrations were male (Ministry of Education, 1991). This gender bias could reinforce the idea that science is more of a male oriented subject and influence the participation levels of girls in a negative way (Baker, 1998). In addition the pronoun “he” is most often used when referring to scientists (Zacks, 1999). The idea that science is more of a male orientated subject is supported by studies that have found that only a very small percentage of students drew female scientists when asked to draw a “scientist” (Kahle, 1987; Zacks, 1999).

In a classroom setting, girls are less likely than boys to take risks (i.e., boys are not afraid to ask or give wrong answers to questions), which leads to girls asking questions less frequently (Kelly, 1987; SSCR, 1987). Girls generally prefer activities which include drawing, colouring and strive for neatness, which are qualities which are not highly valued in science, whereas, boys prefer activities which include construction, which is valued in science. These preferences could negatively influence girls’ perceptions of their competence in science (Murphy, 2000).

**Teacher Influences.** Teachers are an important influence on students, however, for girls this can be negative. Teachers can influence students by the expectations they have for them, the way that they interact with them, the teaching strategies that they use, and the unspoken messages they give (Brophy, 1985).
Although there is some controversy over the direct effects of teacher expectation on student achievement and development (Spitz, 1999), there is clear evidence that teacher expectations influence teacher behaviours within the classroom and their interactions with students (Rubie-Davies, 2009; Turner, Christensen, & Meyer, 2009; Weinstein & McKown, 1998). That is, teachers have been shown to moderate their interactions with students based on their expectations for different levels of achievement or performance (Ennis, 1998). This is important as differential teacher behaviour has been found to be related to differences in student performance and classroom environments (Brophy, 1985). Consequently, in the worst possible instance, teacher behaviours may create a negative snowball effect by virtue of their reactions and feedback to students.

It has been found that teacher’s expectations can affect students’ beliefs of their competence and self-confidence (Murphy, 2000). Teachers generally expect boys to answer higher level questions and therefore a greater number of higher level cognitive interactions occur for boys than for girls (Crossman, 1987; Tobin, 1987; Tobin, Kahle, & Fraser, 1990). Overall, teachers have been found to give more attention to boys (i.e., praise, criticism, one-to-one assistance, and answering questions) (Howe, 1999; Sadker & Sadker, 1985; Tobin, Kahle, & Fraser, 1990). In addition teachers have been found to overrate the work of boys compared to girls, which could (in combination with giving girls less attention) contribute to the lower motivational levels of girls (Spear, 1987). Teachers often reinforce stereotypical images of males versus females without realising it (e.g., reinforcing the relevance of biology for girls and physical sciences for boys) (Kelly, 1987; Tobin, Kahle, & Fraser, 1990). Teacher behaviours and expectations can encourage girls to actively participate in science activities and provide a classroom climate which is favourable to the achievement of girls as well as boys (Tobin, Kahle, & Fraser, 1990). On the other hand, they can increase the differences in self-confidence, skills development, subject anxiety, and participation between boys and girls which could impact on achievement as they progress through secondary school in science (Kelly, 1987).

Achievement is related to success and ultimately, career choices. On the evidence presented so far, and confirmed by Campbell, Hombo, & Mazzeo (2000) who found that over the previous 30 years from 1969 -1999, little progress had
been made in closing the male-female gender gap in science favouring males, it appears that there is just reason to be concerned about gender differences in achievement in science. However, the situation appears to be changing and the next sections will look at the more recent evidence of gender influences in science interest, participation and achievement in schooling and at university, and the effect this has on science related careers.

**Gender differences and science achievement in the 21st century.**

When looking at recent international results the evidence is not convincing for the persistence of male supremacy in science achievement, in fact the opposite has been reported in many studies and the participation of females in advanced science courses at high school and at Universities has also increased (Machin & McNally, 2005). Despite the nature and direction of the difference, it has been found that gender differences in achievement and interest in science first appears at elementary school and continues into high school and culminates in gender differences in degree attainment in STEM subjects (Amelink, 2009; Miller, Slawinski Blessing, & Schwartz, 2006). In the UK at all levels (age 7, 11, and 16) girls outperform boys in schooling for reading, writing and mathematics which are important skills in science achievement (Machin & McNally, 2005; Snow, 2010; Snow & Ucceli, 2009). In the US differences between male and female performance in science based on results from standardised tests are not significant but girls confidence in their ability to do science still shows a significant gap in favour of males (Miller, Slawinski Blessing, & Schwartz, 2007). With these inconsistencies Baker (2002) still sees gender as an important issue in science education. The following section will look at data from the UK, US, New Zealand, and internationally and discuss possible reasons for these results regarding science achievement in the 21st Century.

**Boys still outperforming girls in science.** In TIMMS 2003 boys outperformed girls in 65% of the countries involved (Nosek et.al., 2009) and in TIMMS 2011, New Zealand boys significantly outperformed girls as they also did in 9 other countries (Martin, Mullins, Foy, & Stanco, 2011). In PISA 2012 boys performed better than girls in 37 out of 65 countries (57%) (Kelly, Nord, Jenkins, Chan, & Kastberg, 2013). In the UK boys’ mean performance on national tests
was marginally higher than girls at age 11 years and boys were slightly more represented in the highest scoring groups (Calvin, Fernandes, Smith, Visscher, & Dreary, 2010). In the US in TIMMS 2007 grade 8 (14 years old) boys outperformed girls (Miller, Sen, Malley, & Burns, 2009). Gender differences are smaller in the mid range of science scores for standardised tests than for the highest and lowest achievers (Halpern, Benbow, Geary, Gur, Hyde, & Gernbacher, 2007). A longitudinal study “The National Educational Longitudinal Study” (NELS) 1988-2004 found that in standardised tests in 1988 there was a significant difference in achievement between males and females in favour of males but by 2004 gender had a limited influence on achievement (Amelink, 2009).

The National Centre for Educational Statistics (NCES) reports that National Assessment for Educational Progress (NAEP) performance for 12th grade students (18 years) indicates that overall, males scored higher in science in 2009 and males significantly outperformed females in advanced biology, chemistry and physics (Cunningham, Hoyer, & Sparks, 2015). At University, these male high school graduates outperform their female counterparts in all science related courses except engineering but numbers of females enrolled in engineering at undergraduate level is significantly lower than males (Amelink, 2009; Cunningham, Hoyer, & Sparks, 2015; Maltese & Tai, 2011). In PISA 2012 males outperformed females by 2 score points on average and males performed significantly better than females in 13 countries (Thomson, de Bortoli, & Buckley, 2013). In New Zealand, NEMP results for Year 4 (10 years old) and Year 9 (14 years old) indicate boys perform slightly better than girls (Crooks, Smith, & Flockton, 2007).

Possible reasons suggested for this continued gender gap in favour of males are not that males are more mathematically skilled but that they are more proficient at complex problem-solving and are more spatially able and these are important components of achievement in science (Hyde, Lindberg, Linn, Ellis, & Williams, 2008) particularly in standardized exams and physics topics (Halpern, Benbow, Geary, Gur, Hyde, & Gernbacher (2007). Spelke (2005) and Hyde (2005) agree that males are not more mathematically able than females and suggest that cognitive capacities develop equally and therefore women and men should be equally represented in science courses, achievement, and careers. Miller, Slawinski Blessing, and Schwartz (2007) and Halpern, Benbow, Geary, Gur, Hyde,
and Gernbacher (2007) suggest that females have a lower interest in science not a lower ability and suggest that males have higher variability in achievement and so are more represented in the higher and lower achieving scores making them appear to have higher achievement. In addition, girls are less likely to take advanced science courses at high school so the effect of having lower numbers of students is that if some of these students don't perform well it will be more noticeable. In addition female underachievement in science could be due to the lack of female role models in STEM subjects and careers (US Department of Commerce, 2011; Lindberg, Hyde, Petersen, & Linn, 2010). Cunningham, Hoyer, and Sparks (2015) reported that 70% of 12th grade males like science compared to 59% of females. Jones, Howe, and Rua (2000) found that males reported more science related interests e.g., microscopes, pulleys and electric toys and they had a more favourable attitude to learning science and scientists.

Girls outperform boys in science. In the US Freeman (2004) reported a trend toward educational equality for girls and women with both making strong gains in science achievement, course taking, and entering science careers and that the number of female high school graduates enrolled in advanced science courses at universities between 1982 and 2004 increased [in 2004 female enrolment (37%) exceeded male enrolment (30%)]. In addition female senior high school students outperformed males in overall GPA's. Britner (2008) found when looking at the self efficacy, motivation, and achievement in a US study that females had higher self efficacy, motivation, and achievement in earth sciences, higher grades but not higher self efficacy in life sciences and overall had higher levels of reported anxiety. In the UK, Dreary, Strand, Smith, and Fernandes (2007) found that there was no difference in intelligence at age 11 and by age 16 girls outperformed boys in all 25 General Certificate of Secondary Education (GCSE) subjects tested except physics, and when considering the individual science subjects (e.g., double science, single science, biology, chemistry or physics) girls performed significantly better than boys in all science subjects except in physics. In PISA 2012, females performed significantly better than males in 9 countries (Thomson, de Bortoli, & Buckley, 2013). In TIMMS 2011, girls significantly outperformed boys in 15 countries (Martin, Mullins, Foy, & Stanco, 2011).
In New Zealand attainment in the National Certificate of Educational Achievement (NCEA) which includes the subject of science, has improved for both males and females but a greater percentage of females achieved NCEA level one in 2011 (81% of females and 71% of males) (NZQA, 2012). In 2013 over all NCEA levels and in all age groups females attained the highest percentage of national qualifications (NZQA, 2013).

Possible reasons that have been suggested for the success of girls in science include that educational testing methods may favour girls as they may have a possible verbal advantage in written exams, girls are more motivated and therefore more likely to compete coursework and submit work for internal assessment (Amelink, 2009; Calvin, Fernandes, Smith, Visscher, & Dreary, 2010; Dreary, Strand, Smith, & Fernandes, 2007). On the other hand boys lack of achievement in the UK could be due to boys watching too much football or because there appears to be an anti-learning culture among boys (Machin & McNally, 2005).

**No difference in science achievement between boys and girls.** In the US, Hyde (2005, 2014) reported no difference after meta-analysis of 46 studies and suggested a “similarity hypothesis” i.e., that males and females are highly similar in relation to STEM subjects. Quinn and Cooc (2015) looked at science achievement in early grades and found very small differences which disappear when controlling for maths ability. Britner (2008) found no difference in self-efficacy or achievement but higher anxiety in chemistry and physics for females. Lindberg, Hyde, Petersen and Linn (2010) found that the gender gap in course taking and STEM subject performance has closed except in physics. Hyde, Lindberg, Linn, Ellis, and Williams (2008) and Hyde and Linn (2006) found that on standardised tests in grade four girls score just as well as boys. Amelink (2009) reported that in the national American College Test (ACT) for admission to university there was a minimal difference between the performance of males and females.

In PISA 2006, science was the domain with the smallest difference between male and female performance (Coll, Dahnah, & Faikhamta, 2010) and in PISA 2012, 8% of males were in the highest 2 proficiency levels compared to 7% for females.
(Thomson, de Bortoli, & Buckley, 2013). In New Zealand, the PISA 2006 and 2012 results show very little difference between boys and girls (Breakspear, 2012; Telford & Caygill, 2007). Britner (2008) suggests that the gender gap is decreasing because studies look at the overall performance in science and not by individual science subjects where there are still gender gaps in e.g., physics and computer science.

**Gender differences in degree level science subjects and STEM careers.**

Despite the reported improvement in the achievement of girls in science worldwide, women are still under-represented in STEM disciplines particularly physics and engineering, in degrees awarded in these subjects, and in science related careers (US Department of Commerce, 2011; Eccles, 2007; Lindberg, Hyde, Petersen & Linn, 2010; Millar, Slawinski Blessing, & Schwartz, 2007; Miyake, Kost-Smith, Finkelstein, Pollock, Cohen & Ito, 2010). Females are still outnumbered by males in physics courses at undergraduate level and only 16% of women are enrolled in engineering PhD’s in the US (Hyde, Lindberg, Linn, Ellis, & Williams, 2008). In STEM careers woman hold less than 25% of STEM jobs (US Department of Commerce, 2011). Benbow, Lubinski, Shea, and Eftekhari-Sanjani (2000) assessed gifted students at age 12 and 14 and found that gender differences in mathematical reasoning were accurate predictors of careers 20 years later, males were more likely to be engineers, computer scientists or academics. Many reasons have been suggested for this disparity including the life of a scientist does not appeal to women as it is perceived as a male dominated environment, childhood stereotypes that girls don't have mathematical ability which affects subject choices including science and physics at high school, and societal barriers (Eccles, 2007; Millar, Slawinski Blessing, & Schwartz, 2007). Also, the value placed on these occupations as appropriate careers for women (Eccles, 2007), and the difference that exists in complex problem-solving ability which favours males, which is an important component in science achievement in higher education (Hyde, Lindberg, Linn, Ellis & Williams, 2008). Halpern et al., (2007) suggest that early experiences, genetic factors, education, and cultural experiences affect the choice of science related careers.
There is evidence that female achievement in science is improving and in some cases surpassing males, however there are still specific areas within the science field that a gap in interest, participation and achievement still exists. This is particularly true regarding achievement in the science discipline of physics. Study 2 is set in the context of physics and since research indicates that gender differences can exist when learning physics it is important to determine if pre-training in a physics context affects learning differently for males and females.

The next section will look at the science achievement of Māori and Pasifika students in New Zealand secondary schools. In addition to the influence of gender on achievement in science, within the New Zealand context, another individual difference variable i.e., ethnicity, appears to affect achievement. The nature and extent of these effects will be discussed in relation to international and national studies and national exam statistics. Reasons for this ethnicity influence on science achievement and New Zealand government initiatives will also be explored.

**Māori and Pasifika Achievement in Science**

In New Zealand due to our reliance on agriculture, horticulture, and our environment, the development of scientific knowledge is deemed as critical to our economy and our future (Ministry of Education, 2009). Consequently, the learning of science at school is seen as vital in terms of scientific literacy. Scientific literacy in this context has a two-fold purpose equipping students with adequate science knowledge to make everyday decisions involving science and for laying the knowledge foundation for those students who will become future scientists. However, achievement in science is not equal for all students in New Zealand and for many years there has been a gap between the achievement of the main ethnic groups (New Zealand/European, Māori, Pasifika, and Asian). Māori and Pasifika achievement in science is significantly below that of the two other main ethnicities in New Zealand schools (Baker & Jones, 2005; Bishop, Berryman, Cavanagh, & Teddy, 2009; Chamberlain & Caygill, 2012; Coll, Dahsah, & Faikhamta, 2010; Cowie, Jones, & Otrel-Cass, 2010; Hodson, 1993; Telford, 2010; Telford & Caygill,
Māori are the indigenous people of New Zealand and Pasifika people have been immigrating to New Zealand since the mid 1940s, in search of a better life which included a good education for their children. The collective term Pasifika refers to people from Samoa, Tonga, Niue, Tokelau, Cook Islands, Fiji, Kiribati, Solomon Islands, Vanuatu, and Papua New Guinea. These island groups are all found in the Pacific Ocean. The largest four groups are Samoan, Tongan, Cook Islands, and Niuean.

**International and national studies of science achievement.**

Many of the studies which identified a gap in science achievement between New Zealand/European students and those of Māori and Pasifika descent have looked at international and national studies involving the assessment of scientific literacy. The three most quoted are, the Trends in International Mathematics and Science Study (TIMMS), the Programme for International Student Assessment (PISA), and the National Educational Monitoring Project (NEMP). TIMMS and PISA include secondary assessment at Year 9 and 10 respectively and are pencil and paper tests. NEMP includes practical and oral questions but is only at primary school level. These three studies and their findings for the latest rounds or the last where science was the focus (PISA and NEMP) will be discussed in the next sections.

**TIMMS.** Involvement in International studies is important as a benchmark for New Zealand students in order to compare their progress with students of the same age in countries around the world. Firstly, TIMMS is administered by the Evaluation of Educational Achievement (IEA) and is run on a four yearly cycle which measures trends in student achievement (understanding and cognitive processing) in the areas of mathematics and science. In 2010/2011, 60 countries were involved, 5300 students from New Zealand participated which represented the involvement of 158 schools. In Year 5, (9-10 years old) New Zealand students performed significantly lower than 29 countries but were above the international mean and physical science was the weakest area. In Year 9 (13-14 years old) the New Zealand mean (512) was significantly lower than 10 countries
and significantly higher than 25 countries but was above the international mean of 500. The strongest content area was earth science and the weakest was chemistry and physics. When analysing the results using the four main ethnicities, New Zealand/European, Māori, Pasifika, and Asian, the highest to lowest means were for New Zealand/European, Asian, Māori, and Pasifika, respectively. The means for Māori and Pasifika were significantly lower than New Zealand/European and Asian. Only 15% of Māori and 6% of Pasifika were in the higher benchmarks compared to 42% for New Zealand/European and 43% for Asian. The achievement levels for the period 2002/2003 to 2010/2011 has remained fairly constant and the pattern of significantly lower achievement for Māori and Pasifika students compared to their New Zealand/European and Asian counterparts has been repeated throughout this time period (Chamberlain & Caygill, 2012). The means for the four main ethnic groups in New Zealand from 2002/2003 to 2010/2011 are shown in Table 3.

Table 3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ/European</td>
<td>532</td>
<td>534</td>
<td>540</td>
<td>533</td>
</tr>
<tr>
<td>Māori</td>
<td>472</td>
<td>472</td>
<td>488</td>
<td>466</td>
</tr>
<tr>
<td>Pasifika</td>
<td>430</td>
<td>430</td>
<td>465</td>
<td>439</td>
</tr>
<tr>
<td>Asian</td>
<td>500</td>
<td>516</td>
<td>543</td>
<td>533</td>
</tr>
</tbody>
</table>

Note: New Zealand did not take part in TIMMS in 2006. Adapted from information (Chamberlain & Caygill, 2012)

**PISA.** PISA was initiated by the Organisation for Economic Co-operation and Development (OECD) in 2000 and is administered to 15 year old students on a three year cycle. Each administration of the tests (3 year cycle) focuses on a different set of skills and knowledge (reading, mathematical, or scientific literacy). The most recent assessment of scientific literacy occurred in 2006. In 2006,
400,000 students from 57 countries participated in the study, 4824 students from 170 schools in New Zealand participated. Scientific literacy scores are the combination of competency scores from three sub scales, identifying scientific issues, using scientific evidence, and explaining scientific phenomena. Overall, New Zealand students scored well above the OECD mean of 500 points, New Zealand's mean score was 530 points, the third highest in the study. Both New Zealand and Finland had the highest number of students in the top proficiency level but New Zealand had one of the widest distributions of scores across all the countries. When looking at the scores by ethnicity the achievement is positively skewed towards New Zealand/European students. The mean scores for the four main ethnic groups overall and for physical systems (one of the explaining phenomena scientifically categories) are shown in Table 4 (Telford, 2010). Māori students performed 25% lower than New Zealand/European and Asian students, and Pasifika 32% lower. Māori and Pasifika students are clearly over represented in the lower achievement percentiles (Coll, Dahsah, & Faikhamta, 2010).

Table 4
*Means in PISA 2006 for NZ students by ethnicity*

<table>
<thead>
<tr>
<th>Ethnic group</th>
<th>Overall Mean</th>
<th>Mean for Physical systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ/European</td>
<td>554</td>
<td>536</td>
</tr>
<tr>
<td>Māori</td>
<td>480</td>
<td>473</td>
</tr>
<tr>
<td>Pasifika</td>
<td>454</td>
<td>449</td>
</tr>
<tr>
<td>Asian</td>
<td>542</td>
<td>528</td>
</tr>
</tbody>
</table>

*Note:* This table focuses on scores related to Physical systems as it is the context for this PhD study. Adapted from information (Telford, 2010)

Of the six proficiency levels with 1 being low and 6 being high, Māori and Pasifika students are more likely to be in the lowest proficiency levels and least likely to be in the highest proficiency levels (see Table 5).
Table 5
Percentage of New Zealand students in the highest and lowest proficiency levels in PISA 2006 by ethnicity

<table>
<thead>
<tr>
<th>Ethnic group</th>
<th>Level 1 or below</th>
<th>Level 5 and 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ/European</td>
<td>7%</td>
<td>22%</td>
</tr>
<tr>
<td>Māori</td>
<td>25%</td>
<td>7%</td>
</tr>
<tr>
<td>Pasifika</td>
<td>33%</td>
<td>5%</td>
</tr>
<tr>
<td>Asian</td>
<td>15%</td>
<td>23%</td>
</tr>
</tbody>
</table>

Note: Adapted from information (Telford, 2010)

**NEMP.** NEMP is a national study undertaken by the University of Otago at primary school level in New Zealand. In 2007, looking at Year 4 and Year 8 children in the three main ethnic groups, New Zealand/European children outperformed Māori and Pasifika children on all tasks and Māori and Pasifika children performed significantly lower than New Zealand/European children on 42/49 tasks in Year 4 and 46/57 tasks in Year 8. In Year 8, however, Māori and Pasifika children achieved at the same level as New Zealand/European children on practical tasks and group work (Crooks, Smith, and Flockton, 2008). Overall, in New Zealand, Year 9 and 10 New Zealand students are performing above the mean for both TIMMS and PISA, however, the means for Māori and Pasifika students are consistently below the overall New Zealand mean. In NEMP assessment Māori and Pasifika students are also significantly below the mean for New Zealand/European and Asian students except for group and practical tasks. In addition to these three studies, the New Zealand Qualifications Authority (NZQA) which administers all qualifications in New Zealand, collects data from national exams at secondary school level. The next paragraphs will discuss evidence from this national qualification at Level 1 in science.
National Certificate of Educational Achievement - Level 1 external science achievement standards.

The National Certificate of Educational Achievement (NCEA) the New Zealand national qualification for secondary school students is divided into three levels each of which correspond to one of the final three years at secondary school (high school), which in New Zealand are called Year 11, 12, and 13. In each subject e.g., science or mathematics, there are a number of achievement standards (science has a possible 16 achievement standards) which schools offer and these make up a year long course in this subject, most science courses in secondary schools are made up or 5 or 6 achievement standards. Most achievement standards in science are worth 4 credits, so a whole year's science course would be worth 20 or 24 credits. To earn credits for an achievement standard a student must demonstrate the knowledge and skills specified in the criteria for this particular standard. Some achievement standards (thirteen in science) are internally assessed by teachers in schools and some are externally assessed (three in science) by NZQA. Internally assessed achievement standards are taught and marked during the year by classroom teachers and samples sent to NZQA for moderation. Externally assessed achievement standards are taught during the year but are assessed in national exams at the end of the year, which are administered and marked by NZQA. There are four possible grades that can be awarded for achievement standards in NCEA, a student will be awarded ‘not achieved’ if the criteria for the achievement standard is not met, ‘achieved’ if it is met to a satisfactory standard, ‘merit’ if it is met to a very good standard, and ‘excellence’ for outstanding performance against the criteria for the achievement standard. A student needs to achieve a minimum of 80 credits at Level 1 across all their subjects to be awarded NCEA level 1, and the same for levels 2 and 3.

Within the present context, the three achievement standards that are most relevant are: Science (90940) - Demonstrate understanding of aspects of mechanics, Science (90944) - Demonstrate understanding of aspects of acids and bases, and Science (90948) - Demonstrate understanding of biological ideas relating to genetic variation. The target information used in this study was taken from Achievement Standard Science 90940. This achievement standard will be
looked at separately and compared to achievement for all three combined. Tables 6 - 8 include these three achievement standards organised by the four main ethnic groups. When considering the means over the four years from 2011 to 2014 for the four main ethnic groups, Māori and Pasifika have the highest percentage of students who did not achieve the science external achievement standard (90940). In the lowest two grades (not achieved and achieved) Māori and Pasifika also have the highest combined percentages of students (both groups have a total for these two grades over 80%). This means that less than 20% of students in these two ethnic groups were in the highest two grades (merit and excellence). In the highest two grades Asians have a combined total of 50% which is more than double the percentage of Māori students and three times the percentage of Pasifika students. Regarding excellence grades, Asian students have almost five times the percentage of Māori and over seven times the percentage of Pasifika students. When analysing the overall results for the three external achievement standards, Pasifika students have the largest percentage of students who do not achieve, nearly half of the students who sit (47.7%) while Māori have 43.3% who do not achieve. New Zealand/European have the largest percentage of achieved grades (37.4%) and Asians have the highest number of merit (29.9%) and excellence grades (18.9%) across all the years from 2011 to 2014. This information (percentage scores) is summarised in Tables 6, 7, and 8.

As demonstrated in the previous discussion and the presented tables, the trend of Māori and Pasifika underachievement in science is consistent across International data (TIMMS and PISA) and National data (NEMP and NCEA). Based on all this evidence the New Zealand government has taken action to try to reverse this trend in the underachievement of Māori and Pasifika students.
Table 6
Achievement Standard 90940 Results for Māori, Pasifika, NZ European and Asian

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ/Eur</td>
<td>26.3</td>
<td>21.3</td>
<td>27.1</td>
<td>21.7</td>
<td>36.2</td>
<td>42.8</td>
<td>34.5</td>
<td>37.6</td>
<td>25.9</td>
<td>25.1</td>
<td>25.6</td>
<td>28.4</td>
</tr>
<tr>
<td>Māori</td>
<td>47.5</td>
<td>39.9</td>
<td>48.0</td>
<td>40.4</td>
<td>33.5</td>
<td>42.2</td>
<td>31.5</td>
<td>38.1</td>
<td>14.6</td>
<td>14.5</td>
<td>15.1</td>
<td>17.5</td>
</tr>
<tr>
<td>Pasifika</td>
<td>50.2</td>
<td>43.8</td>
<td>54.4</td>
<td>45.9</td>
<td>33.3</td>
<td>41.5</td>
<td>31.1</td>
<td>36.4</td>
<td>13.0</td>
<td>12.6</td>
<td>12.6</td>
<td>14.9</td>
</tr>
<tr>
<td>Asian</td>
<td>18.3</td>
<td>14.8</td>
<td>17.3</td>
<td>16.9</td>
<td>31.2</td>
<td>37.0</td>
<td>32.1</td>
<td>32.4</td>
<td>31.4</td>
<td>29.0</td>
<td>29.7</td>
<td>31.9</td>
</tr>
</tbody>
</table>

Note: Adapted from data received from NZQA
Table 7

*Achievement Standard 90944 Results for Māori, Pasifika, NZ European and Asian*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ/Eur</td>
<td>23.7</td>
<td>26.7</td>
<td>26.0</td>
<td>19.2</td>
<td>39.1</td>
<td>36.6</td>
<td>28.6</td>
<td>32.9</td>
<td>24.8</td>
<td>26.5</td>
<td>32.3</td>
<td>31.4</td>
<td>12.5</td>
<td>10.3</td>
<td>13.1</td>
<td>16.5</td>
</tr>
<tr>
<td>Māori</td>
<td>42.8</td>
<td>47.6</td>
<td>47.4</td>
<td>38.5</td>
<td>38.1</td>
<td>34.1</td>
<td>27.7</td>
<td>34.0</td>
<td>14.3</td>
<td>14.6</td>
<td>20.6</td>
<td>21.3</td>
<td>4.9</td>
<td>3.6</td>
<td>4.4</td>
<td>6.2</td>
</tr>
<tr>
<td>Pasifika</td>
<td>46.9</td>
<td>50.8</td>
<td>53.2</td>
<td>39.3</td>
<td>37.6</td>
<td>33.0</td>
<td>25.4</td>
<td>31.8</td>
<td>11.6</td>
<td>13.1</td>
<td>16.9</td>
<td>21.8</td>
<td>3.9</td>
<td>3.1</td>
<td>4.5</td>
<td>7.1</td>
</tr>
<tr>
<td>Asian</td>
<td>16.6</td>
<td>19.9</td>
<td>22.6</td>
<td>15.4</td>
<td>36.0</td>
<td>30.3</td>
<td>22.8</td>
<td>25.8</td>
<td>27.2</td>
<td>31.6</td>
<td>32.8</td>
<td>33.8</td>
<td>20.2</td>
<td>18.1</td>
<td>21.8</td>
<td>24.9</td>
</tr>
</tbody>
</table>

*Note: Adapted from data received from NZQA*
### Table 8

**Achievement Standard 90948 Results for Māori, Pasifika, NZ European and Asian**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ/Eur</td>
<td>27.6</td>
<td>27.3</td>
<td>20.5</td>
<td>20.8</td>
<td>42.1</td>
<td>39.6</td>
<td>41.0</td>
<td>38.5</td>
<td>22.9</td>
<td>29.1</td>
<td>27.1</td>
<td>29.4</td>
<td>7.3</td>
<td>9.0</td>
<td>11.5</td>
<td>11.4</td>
</tr>
<tr>
<td>Māori</td>
<td>48.5</td>
<td>40.8</td>
<td>39.4</td>
<td>38.6</td>
<td>36.8</td>
<td>38.4</td>
<td>39.9</td>
<td>38.1</td>
<td>12.3</td>
<td>17.4</td>
<td>16.5</td>
<td>18.7</td>
<td>2.4</td>
<td>3.3</td>
<td>4.2</td>
<td>4.6</td>
</tr>
<tr>
<td>Pasifika</td>
<td>56.0</td>
<td>41.6</td>
<td>45.2</td>
<td>44.8</td>
<td>31.5</td>
<td>38.0</td>
<td>38.5</td>
<td>36.6</td>
<td>10.9</td>
<td>17.3</td>
<td>13.2</td>
<td>15.1</td>
<td>1.6</td>
<td>3.1</td>
<td>3.1</td>
<td>3.4</td>
</tr>
<tr>
<td>Asian</td>
<td>24.0</td>
<td>20.0</td>
<td>19.4</td>
<td>20.6</td>
<td>37.7</td>
<td>32.0</td>
<td>36.9</td>
<td>33.7</td>
<td>24.8</td>
<td>31.5</td>
<td>27.1</td>
<td>28.0</td>
<td>13.5</td>
<td>16.6</td>
<td>16.6</td>
<td>17.6</td>
</tr>
</tbody>
</table>

*Note: Adapted from data received from NZQA*
New Zealand Government initiatives.

This persistent underachievement by Māori and Pasifika students has prompted the New Zealand Government to set up initiatives to lift the achievement of these two ethnic groups. These initiatives are not specifically for the curriculum area of science but science is included in the aims to improve overall achievement in primary and secondary schools for Māori and Pasifika students. The Māori initiatives are Ka Hikitia - Managing for Success: The Māori Education Strategy (2008 - 2012) and Accelerating for Success (2013 -2017) which is an updated strategy of the earlier version, and Te Kotahitangi Phase 1 - 4. The Pasifika initiative is The Pasifika Education Plan (2009 - 2012), which has also been renewed (2013 - 2017). The Best Evidence Synthesis (BES) was a programme initiated in 2003 to collect evidence for what works for diverse learners considering the poor outcomes identified for some learners in the New Zealand school system (Alton-Lee, 2008). In addition to separate targeted initiatives for Māori and Pasifika students the New Zealand Government has funded TeachFirst NZ to increase the number of quality secondary teacher education graduates working in low decile schools and has established five new partnership schools Kura Hourua in Auckland and Northland to raise Māori and Pasifika achievement.

These initiatives have been successful in retaining Māori and Pasifika students in school and lifting the number of students who achieve the required number of literacy and numeracy credits for NCEA Level 1 and 2. Māori and Pasifika Level 2 achievement has improved 6.2 and 5.9 percentage points respectively, more than any other group. In addition, more Māori and Pasifika students are completing level 4 or higher qualifications within 5 years of tertiary study (Bishop, Berryman, Cavanagh, & Teddy, 2009; Ministry of Education, 2009).

The issue of Māori and Pasifika achievement has prompted many studies which also offer potential reasons for this underachievement. These studies will be discussed in the next section.
Reasons for Māori and Pasifika underachievement.

The PISA study includes information on students’ attitudes to science and confidence in science, teacher-student relationships, access to educational resources, school decile rating, highest level of parental education, language spoken at home, and socio-economic status (SES) data. Analysis of the achievement data was correlated with these additional factors. The PISA (2006) data showed that Māori students had the least positive attitudes about their science learning which included enjoyment and motivation to learn and were the least likely to report that they were good at science. Māori and Pasifika students were less confident in science and Māori students in particular were least likely to think that science was important and consider a science related career. Of all the factors SES was the most powerful factor influencing performance in PISA (2006). New Zealand/European and Asian were over represented in the highest SES groupings and Māori and Pasifika were over represented in the lowest SES groupings and there was a high correlation between SES and achievement. In addition to these patterns, PISA (2006) results also showed that students’ achievement improved with the education level of the parents, access to educational resources, attending higher decile schools, and a positive school climate. More Māori and Pasifika were in the lowest groupings for all of these additional factors (Song, Perry, & McConney, 2014; Telford, 2010).

In addition to the PISA information, a number of other studies have looked at possible reasons for the underachievement of Māori and Pasifika students in New Zealand. Whereas these are not all specifically for science the reasons apply to all subjects including science. Hattie (2003, 2005) disagrees with the importance of SES affecting the achievement of Māori students and suggests that the cultural relationships between teachers and students, and excellence in teaching is the major factor in all student achievement (See also, Morris and Patterson, 2013). Inequitable education for Māori students is another reason suggested, i.e., low inclusion of topics relevant to Māori and ineffective teaching strategies for Māori students (Alton-Lee, 2008; Biddulph, 2003; Waiti & Hipkins, 2002). Cultural beliefs have also been suggested, e.g., teaching the topic “evolution” is inconsistent with
Māori cultural beliefs but is part of the New Zealand Curriculum (Cowie, Jones, & Otrel-Cass, 2010; Harker, 2006). Also, family stress, health problems, and a lack of resources (linked closely to SES) have been advocated by Harker (2006).

Teacher expectations which are lower for Māori and Pasifika students in reading, which is used in science has also been suggested as being linked with lower level of achievement (Rubie-Davies, Hattie, & Hamilton, 2005). A factor particularly relevant to Pasifika student achievement in science is the level of English used at home, i.e., students who speak English at home outperform students who do not speak English at home (Bishop, Berryman, Cavanagh, & Teddy, 2009; Ministry of Education, 2009; Telford 2010). Finally, the use of assessment procedures which do not include contexts relevant to Māori and Pasifika students could adversely affect achievement (Harker, 2006). For example, using the context of skiing for exam questions about motion. Most Māori and Pasifika students due to the financial restraints of living in low socio-economic status families may not have experienced skiing.

In summary, the issue of whether Māori and Pasifika are underachieving in science is not in question but the reasons and effective cognitive strategies to reverse the trend are still open to discussion and investigation. This study will use a cognitive strategy (pre-training) in an attempt to enhance the achievement of Māori and Pasifika students in science. With respect to Māori, McKinley and Gan (2014) stress the importance of supporting participation, academic achievement, and equity in science classrooms and advocate for an integrated approach. Most of the initiatives undertaken so far have suggested practices involving culturally responsive pedagogies and motivational strategies would be beneficial at ameliorating the cultural gap in achievement.

**Summary and Research Questions**

The literature discussed in this chapter covered a variety of research and theory surrounding the learning of complex science concepts within a secondary context. Given the focus of the current study, the majority of literature and research discussed concerned models and theories of cognitive load, how to address cognitive load in order to facilitate learning and understanding, and how best to
measure cognitive load. In addition, in order to better understand the impact of cognitive load and the pre-training strategy, there was also a need to discuss literature on gender differences in science achievement and in spatial ability, and, within a New Zealand context, the impact of ethnicity and a school’s decile rating on student achievement.

Cognitive load theory suggests that understanding and learning complex information puts a heavy load on our cognitive processing system and in some situations i.e., when novices are learning new information, can cause cognitive overload which can hinder further understanding and learning. Cognitive load theory and the CTML developed preinstructional strategies to reduce the cognitive load of learning complex information. This study will employ the CTML strategy of pre-training in the context of understanding and learning physics content which includes graphs in whole class teaching in secondary science classrooms with novices.

As discussed in the literature review there are many ways cognitive load can be measured. Within the context of this Study 2 different subjective cognitive load rating scales will be developed for assessing understanding and learning during and after the teaching. Subjective cognitive load measures (i.e., asking participants to rate cognitive load on a rating scale) will be used as they have been found to be sensitive to small changes in cognitive load, reliable, valid, and are less intrusive than other measures e.g., physiological techniques. The subjective cognitive load measure employed during the teaching phase will focus on assessing participant load while they are learning the target information. Participants will be asked to assess the cognitive load at two different times during teaching in order to gauge their cognitive load over different aspects of the teaching. That is, understanding the conceptual physics knowledge and graphing skills related to motion. The measure used after teaching will assess overall cognitive load and associated variables and will be administered immediately after the teaching to prevent loss of critical rating information by the participants. The actual cognitive load of the pre-training material will also be assessed in order to better understand its affect and role within this context. The focus on the impact of pre-training on cognitive load while learning and attempting to understand complex
science concepts within an authentic setting led to the development of the first research question.

Research Question 1

Does pre-training reduce the cognitive load associated with learning complex science information involving graphs as measured during and after learning?

According to cognitive load theory and the CTML if the cognitive load of learning complex information is reduced then learners will have more mental resources available for understanding and learning. Therefore in this study if pre-training is employed and cognitive load consequently was reduced then it would be expected to have a positive effect on overall learning. In addition, given that the pre-training instructional strategy is aimed at reducing processing within working memory and related cognitive load during learning, and that one might expect variations in the complexity or difficulty of to-be-learned knowledge and skills to differentially influence processing and load within working memory, two additional sets of comparisons were planned. One set of comparisons assessed the impact of pre-training on understanding of the conceptual physics knowledge and graphing skills related to motion, while the other focused on the impact of pre-training on higher and lower order learning outcomes.

Assessing performance in relation to overall understanding and learning and with respect to the different categories of outcomes (graphing versus calculation) and different levels of outcome (high and low order) as a result of employing the strategy of pre-training led to the development of the second research question.

Research Question 2.

Within the context of learning complex information involving graphs, what is the effect of pre-training on secondary science students’ overall learning, performance on different levels of question (high versus low order), and performance on different content questions (calculation versus graphing)?
As discussed earlier within this chapter, when cognitive load and performance is assessed it is possible to combine the two to give a more in depth knowledge of the instructional efficiency of the intervention which in this study was the use of the strategy of pre-training to understand and learn complex science information. This study will use the cognitive load of understanding and learning the target information, and the difference between participant’s pre and post test scores in the instructional efficiency calculation using the adapted version of the deviation model. The deviation model assesses the difference between the standardised scores for performance and mental effort expended to achieve this performance. The adapted model uses the mental effort expended while learning the information. This model has been found to be the most appropriate method when assessing the instructional efficiency in intervention studies as it measures the instructional efficiency of the learning process which would be expected to differ in different conditions (treatments).

Therefore, instructional efficiency will be calculated to give a better indication of the extent to which schemas have been acquired, elaborated and automated as a result of pre-training. This led to the development of the third research question.

Research Question 3.

Does pre-training increase instructional efficiency for secondary science students when learning complex information involving graphs?

Conducting research in authentic settings is a complex task and one which few cognitive load and CTML researchers have undertaken. Within the area of science, participant individual difference variables could potentially affect learning and understanding. Within a New Zealand context there are potentially at least three variables which could significantly impact on understanding and learning within a science classroom, i.e., gender, decile, and ethnicity. The target science content within the present study is the physics topic of motion which includes graphs. Previously discussed literature indicates that understanding and learning this physics topic could be more difficult for females and they will have to put in more effort to learn the information than males. Literature also suggests that students in low decile schools are more likely to underachieve compared to those
in higher decile schools. Consequently, one might expect that their performance will be lower and their effort to learn the complex information will be higher. In addition, with respect to ethnicity the literature suggests that Māori and Pasifika students will find understanding and learning the complex information more difficult and will have to put in more effort to learn the information.

Given the previous research on the potential moderating influences of student characteristics on science learning and achievement, this study will look at the influence of gender, decile and ethnicity on cognitive load, performance, and instructional efficiency. In addition it will also investigate whether pre-training is a robust strategy which overcomes potential differences in gender, decile and ethnicity. This is reflected in the fourth research question.

Research Question 4.

Are there differential effects on cognitive load, learning (difference scores) and instructional efficiency with respect to gender, decile and ethnicity when using pre-training in a secondary science context involving graphs?

Finally, the literature on graphing and on spatial ability suggest an important link between these two, i.e., spatial ability has been found to influence graphing skills. In addition, there appears to be some literature that suggests that there are differences between spatial ability across males and females. Given this, spatial ability measures were included in this study in order to assess whether there were significant gender differences and to potentially employ these measures as covariates in relevant analyses.

The next chapter will detail the methodology developed as a result of the literature reviewed in Chapter 3 with the guidance of the research questions.
Chapter 4

Method

Research Design

The overall purpose of this study was to test the effectiveness of the strategy of pre-training on the learning of complex information in the science discipline of physics. The pre-training materials were designed to reduce the cognitive load of understanding and learning complex ideas. This research is grounded in cognitive load theory (Sweller, 1988, 1989, 2010a; Sweller, van Merriënboer, & Paas, 1998) and the CTML (Mayer, 2005a, 2014a). The procedures employed within the present study are modelled on those used in Mayer, Mathias, and Wetzell (2002), Mayer, Mautone, and Prothero (2002) and Pollock, Chandler, and Sweller (2002).

There are three main issues with conducting research in educational settings as is the case with this study. The first is that it is very difficult to conduct well controlled research as there are so many variables to control. Consequently, uncontrolled extraneous or confounding variables could compromise the internal validity of the experiment, i.e., differences in results may not be able to be conclusively attributed to the manipulation of the independent variable(s). Secondly, if the experiment is so controlled that these conditions are not what is usually experienced in educational settings then the external validity is compromised as the results may not be able to be generalised or applied to real classrooms. Lastly, if working with existing groupings i.e., school classes, then it is not possible to randomly allocate participants to the treatment groups. These three issues had to be taken into consideration in this study and the following paragraphs will look at the experimental design and how these issues were addressed.

With respect to controlling variables, secondary school classrooms are unpredictable environments which differ within and between schools. Therefore, controlling all the potential confounding variables that could have had an effect on the performance and cognitive load measures was not only impractical but impossible. So as many of these variables as were practical were controlled to
give overall consistency between the different treatment groups in the eight schools where the study was conducted. The initial phase instruction and the data collection phase teaching was all done by the researcher, however, personal attributes of the researcher i.e., the style of teaching and its affect on individual participants could not be controlled. Experimenter bias was controlled by using the same PowerPoint presentations for the teaching of all treatment groups and all the slides also contained notes (the script) that was closely adhered to by the researcher. This was closely monitored by the teachers of the classes concerned. They were given a copy of the script and asked to annotate where the verbal instruction (teaching) deviated from the printed script. In addition, the teachers were also able to monitor any interruptions or events (history) that could have an impacted on the learning of the participants e.g., if a teacher or student from the school entered the room to give a message to the teacher or request a participant to leave the room. Unplanned interruptions are a real concern for educational research in ecological settings as it could potentially happen at any time, consequently, these cannot be controlled but can be minimised. Therefore, other staff at the schools involved were alerted to the timetable of the study, and this was intended to try to minimised interruptions.

All materials given to participants were identical and all participants were tested before and at the immediate conclusion of the teaching. This eliminates any pre test-treatment interaction. The pace of the teaching had to be carefully managed and was kept as consistent as practically possible, two of the three treatments required a minimum of 55 minutes to complete and the option of going over time did not exist. Laminated worksheets were used with all the treatment groups during the teaching and these reinforced the main teaching ideas and ensured that all participants had access to important information even if they missed the verbal instruction. Electronic timers were used to ensure that all participants had exactly the same amount of time to complete the pre and post test. The main extraneous variables that could not be controlled were the physical classroom environment and the classroom climate. In some classrooms students sat in groups and in others in rows which affects student interaction.

The behaviour of the participants was controlled as much as possible by the researcher employing proven, effective classroom management techniques e.g.
using the word “thanks” after instructions to imply that it was expected that they would comply with the instructions. In addition, the time of the day, the day of the week and the timing of the research in the school term was also not able to be controlled. For all participants the content that they were being taught was not the topic that they were learning at the time that the study was conducted so the interruption to science teaching was consistent among all treatment groups and all the schools involved. However, the time lapsed between the initial phase and the data collection phase was not able to be controlled. All of these confounding variables could potentially have had an impact on learning of the target material.

To maintain the status quo as much as possible for the participants so that the results could be applied to other classrooms, the study was conducted with the individual classes in their usual classroom and their teacher was also present. This means that they were being taught science in a school period when they were expecting to receive science instruction. To increase the external validity of this study a large number of participants from eight different schools were involved, the school’s differed with respect to gender, decile, geographical location in Auckland, ethnic make-up, and type (co-educational, single-sexed, state, and private schools were involved). The large number of participants also assisted in dealing with attrition which can also be a real concern in schools especially if the study is conducted over more than one period. Participants who were absent for one of the periods and therefore do not have a full set of data were not included in all the analyses. Participants were also taught material that was assessed for NCEA Level 1, which meant that the learning was appropriate for these participants as it was aligned to the New Zealand Curriculum (Ministry of Education, 2007).

A quasi-experimental treatment design was used for the method, as it was not possible to completely randomly assign the subjects to the three treatment groups. Quasi-experimental designs are a valuable tool for applied researchers but they do have disadvantages. Quasi-experiments maximise internal and external validity when it is impossible to completely control all variables. This means that they are easier to set up than fully controlled experiments and more applicable and realistic for educational settings. However, without random assignment of participants to treatments, this will impact the generalizability of the results to other populations of students.
The type of quasi-experimental treatment design used was the non-equivalent control group design as the three treatment groups were given a pre and a post test at the beginning and the end of the different treatments. Intact existing classes (pre-formed groups) were randomly assigned to the different treatment groups but this means that not all participants had an equal chance of being in the experimental treatment group (Treatment group 1) and this opens up the possibility of the experimenter choosing the most favourable classes for the experimental treatment groups. In this study this did not occur as the researcher did not know any of the participants or the characteristics of the classes and they were randomly assigned to treatment groups by putting the class names in a container and drawing them out. The first class was Treatment 1, the second Treatment 2 and the one left was Treatment 3. The three classes provided by the school were chosen by the school from the mid range of ability i.e., no classes were selected from the extended (high ability) or remedial (low ability) bands or streams. Ability differences between the three treatment groups were identified using PAT Mathematics and Reading Comprehension scores provided by the schools as a baseline measure of ability. These scores were used as covariates in all the analyses performed. A disadvantage of using existing groups is that participant recent experiences can affect the whole class (treatment group) e.g., if the class had just all performed badly in a science test, they may not put in the effort to learn the material in the study whereas if only some of the participants in a treatment group had this bad experience it may not have such a large effect on performance measures. Another disadvantage is that certain participant misbehaviour can affect the whole group e.g., bullying by some participants of other participants in their class, whereas if the students were not all from the same group these individuals may not misbehave and therefore performance would not be compromised. However, having pre existing groups removes the ethical considerations of placing people in groups and means that the participants are comfortable being together and the study is not a change to school routine (which can have an negative impact on participant behaviour). In addition, the teacher present will know the participants and this can have a positive effect on their behaviour which can also potentially affect performance on dependent measures.
Participants

The participants were six hundred and six Year 9 and 10 secondary school science students recruited from eight different secondary schools in Auckland, New Zealand. Three classes from each school were involved. There were two schools where all classes involved were from Year 9, and 6 schools where all classes were from Year 10. The schools were chosen to represent Auckland geographically and to obtain a spread of decile ratings. Decile ratings (See Table 9) are ratings from 1-10 allocated to schools by the New Zealand Ministry of Education based on Statistics New Zealand information. These ratings take into account the socio-economic status of the parents of the children at the school and they determine the Operations Funding given by the Ministry of Education for the running of the school. Higher decile ratings indicate a higher socio-economic status of the parents and therefore lower funding given to the school. The average decile rating in this study was 4.50. The whole class participated as part of their regular classroom activities and classes were randomly assigned to the three treatment groups. The total number of subjects in Treatment group 1 was 176, Treatment group 2 was 172, and Treatment group 3 was 178. None of the students had studied the physics topic of motion at the time that the study was undertaken, as part of their classroom science programme in Year 9 or 10. That is, they were novices for this motion topic. Discussion with the HOD science at all schools confirmed that participants had not studied the topic of speed or motion. There was a slight imbalance of students with respect to gender (Male, N = 285; Female, N = 321). There were two single sex schools involved in the study and the girl’s school had higher numbers of students per class than the boy’s school and this accounts for the slightly higher participation of female students in this study. The other six schools were co-educational secondary schools. Table 9 gives a summary of the schools involved.
### Table 9

**Summary of Secondary Schools involved in the experimental treatment**

<table>
<thead>
<tr>
<th>School</th>
<th>Gender and Year</th>
<th>Total No. of participants</th>
<th>No. in Treatment 1</th>
<th>No. in Treatment 2</th>
<th>No. in Treatment 3</th>
<th>Decile rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Girls Year 10</td>
<td>76</td>
<td>27</td>
<td>24</td>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>Boys/Girls Year 10</td>
<td>67</td>
<td>24</td>
<td>21</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Boys/Girls Year 9</td>
<td>71</td>
<td>26</td>
<td>23</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Boys/Girls Year 10</td>
<td>51</td>
<td>17</td>
<td>16</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Boys/Girls Year 10</td>
<td>68</td>
<td>19</td>
<td>24</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Boys/Girls Year 10</td>
<td>71</td>
<td>24</td>
<td>24</td>
<td>23</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>Boys Year 10</td>
<td>55</td>
<td>15</td>
<td>20</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>Boys/Girls Year 9</td>
<td>67</td>
<td>24</td>
<td>20</td>
<td>23</td>
<td>10</td>
</tr>
</tbody>
</table>

### Materials and Equipment

This study was conducted in schools over three school periods (1 hour each). The first period is referred to as the initial phase, the second period referred to as the data collection phase, was when the actual study was conducted, and the third period was when the spatial ability testing occurred.

**Initial phase.**

This phase included the use of a PowerPoint presentation during which the participants were introduced to the researcher and given an overview of the initial phase and the data collection phase procedures and were guided through a set of practice cognitive load questions. This phase concluded with the participants completing a pre test.
**Initial phase PowerPoint.** This was an 18 slide PowerPoint presentation with relevant visuals and text at an appropriate language level for Year 9 or 10 students (See Appendix E). The slides had between 2 and 46 words per slide over the whole presentation. The presentation started with an introduction for the participants regarding the researcher and the study, an overview of this initial phase, an overview of the actual study, and indicated to participants the date the researcher was returning to conduct the data collection phase with their class (See Slides 1 to 5). The next set of slides (See Slides 6 to 13) were concerned with introducing the participants to the subjective cognitive load rating questions that they would be required to complete during the data collection phase. The slides described the rating scales to be employed and initially asked the participants to practice using the scale on a relevant personal experience, i.e., enjoyment of a movie or DVD. The participants were then asked to apply the rating scale to a recent science lesson when they learned something new with respect to how much effort they had to invest to, understand the science ideas, learn and remember the science ideas, their success on a test of the science ideas, their frustration level while learning the new science ideas, and lastly, about the pace of the science lesson for them (See Figure 5 below for a copy of Slide 9). Students were given an opportunity at this point to ask questions regarding the use of the rating scale.

The final part of the Initial phase consisted of giving students instructions for the pre test (See Slides 15 and 16). Participants were then given the pre test to complete. For consistency, all slides included notes which the researcher could access when in presenter view using an Apple MacBook.

![THINK ABOUT THE LAST SCIENCE LESSON YOU HAD WHERE YOU LEARNED SOMETHING NEW](image)

**Figure 5.** Slide 9 A Sample slide from Initial Phase PowerPoint
**Practice cognitive load question worksheets.** During the initial phase participants were given an A4 practice worksheet on which they wrote their ratings to the six questions embedded in the practice using the previously described rating scale (See Appendix F). The questions on this sheet were exactly the same as those used in the Initial Phase PowerPoint presentation (See Appendix E) and questions two to six were identical to the questions they would be required to complete in the data collection phase. The sheet did not have a space for them to write their name as it was not collected.

**Pre test materials.** As indicated earlier, at the conclusion of the initial phase the participants completed a pre test. The pre test was on two A4 pages and had spaces for the participants to write their name, school, and class at the top of the first page (See Appendix G). The pre test included six short answer questions that assessed their knowledge of the physics topic of motion. The Structure of the Learning Outcomes (SOLO) taxonomy was used to classify the questions included in the pre test. SOLO taxonomy is used to classify levels of complexity in student’s understanding of content (Biggs & Collis, 1982). The system classifies learning using five levels of understanding, prestructural (ideas are not grasped), unstructural (one idea is understood), multistructural (many ideas are understood), relational (many related ideas are understood), and extended abstract (many related ideas and content beyond that presented is linked and understood). The following assignment of questions in the pre test to SOLO levels of classification was verified by an expert in the field of assessment from The University of Auckland. Question 1 and 2 were unistructural questions and worth one mark each. For example, Question 1 reads, “Give an example of a unit that is used to measure speed.” Question 5c was classified as extended abstract in SOLO, for example, “Compare the sketch you drew and the actual graph of her fall. Give one reason why the actual graph of her fall may not look exactly like the graph that you drew above in part b.” Table 10 shows the SOLO taxonomy classification for this pre test. In addition, the questions were also categorised into high and low order questions and questions related to graphing and calculation for use in subsequent analyses. The high order questions were 3b, 4, 5c, 6c, 6d and 5c, the low order questions were 1, 2, 3a, 5a, 5b, 6a, and 6b. The questions related to graphing were 4, 5b, 5c, and 6, those related to calculation were 1, 2, 3, and 5a. Participants wrote their answers on the sheet in
the spaces provided. The maximum marks a participant could achieve for this pre test was 20 marks. The maximum time allowed to complete the pre test was determined after times were recorded for all three classes to complete the pre and post test during the pilot study. No participants in the pilot study took longer than ten minutes to complete either the pre or post test so participants in the actual study were allocated ten minutes to complete the pre and post test. The internal consistency of the pre test measure (Cronbach’s alpha) was 0.83.

Table 10
SOLO taxonomy classification of the pre test questions

<table>
<thead>
<tr>
<th>Question number</th>
<th>Question description</th>
<th>SOLO taxonomy classification</th>
<th>Mark allocated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Recall unit of speed</td>
<td>Unistructural</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Recall definition of constant speed</td>
<td>Unistructural</td>
<td>1</td>
</tr>
<tr>
<td>3a</td>
<td>Calculate distance using speed formula</td>
<td>Multistructural</td>
<td>2</td>
</tr>
<tr>
<td>3b</td>
<td>Calculate time using speed formula</td>
<td>Relational</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Graph a journey</td>
<td>Relational</td>
<td>3</td>
</tr>
<tr>
<td>5a</td>
<td>Calculate average speed</td>
<td>Unistructural</td>
<td>2</td>
</tr>
<tr>
<td>5b</td>
<td>Graph constant speed</td>
<td>Multistructural</td>
<td>2</td>
</tr>
<tr>
<td>5c</td>
<td>Compare a graph of constant speed and a graph of a real journey</td>
<td>Extended abstract</td>
<td>3</td>
</tr>
<tr>
<td>6a</td>
<td>Interpreting a distance-time graph</td>
<td>Unistructural</td>
<td>½</td>
</tr>
<tr>
<td>6b</td>
<td>Reasons for their answer</td>
<td>Multistructural</td>
<td>1</td>
</tr>
<tr>
<td>6c</td>
<td>Interpreting a distance-time graph</td>
<td>Multistructural</td>
<td>½</td>
</tr>
<tr>
<td>6d</td>
<td>Reason for their answer</td>
<td>Relational</td>
<td>2</td>
</tr>
</tbody>
</table>

Data collection phase.

The second phase of the study involved a pre-training PowerPoint presentation (only for the participants in the experimental group), a teaching PowerPoint presentation either once or twice through depending on which Treatment group the participants were assigned to, cognitive load questionnaires, materials for monitoring the teaching, and a post test. The cognitive theory of multimedia learning was used to guide the development of the two PowerPoint presentations (Mayer 2005a, 2014a).
**Instructional materials.** The instructional materials included two PowerPoint presentations which were used to both prepare the participants for the teaching session (experimental Treatment group only) and to teach the participants science ideas about the topic of motion which included graphing (all Treatment groups). The science ideas in the topic of motion in all the instructional materials were verified by a specialist physics teacher with over 20 years experience in secondary schools.

**Pre-training PowerPoint.** The pre-training PowerPoint consisted of 14 slides with relevant visuals and text at an appropriate language level for Year 9 or 10 students (See Appendix H). There were between 2 and 42 words per slide. The presentation began with an introduction of the strategy of pre-training (See Slides 1 and 2), then introduced the important concepts of distance, time, speed, constant speed, average speed and line graphs. Each concept was introduced on a separate slide using its definition, the units it could be measured in and included an example using an illustration (See Slides 3 to 13). See Figure 6 for a copy of Slide 3 and 9. Twice during this presentation the participants were asked to answer a cognitive load question (See Slides 9 and 14). These two slides were that the context was the pre-training information not a previous science lesson. For consistency, all slides included notes which the researcher could access when in presenter view using an Apple MacBook.
Teaching PowerPoint. The teaching PowerPoint consisted of 19 slides with relevant visuals and text at the appropriate language level for Year 9 or 10 students (See Appendix I). There were between 4 and 60 words or numbers per slide. The teaching PowerPoint was divided into two parts. Part one was concerned with teaching the concept of speed, the units used, and calculations involving distance, time, and speed (See Slides 1 to 7). Part one also included watching an embedded video of Usain Bolt running a 100m race. (See Figure 7 for a copy of Slide 5). Part two was concerned with teaching about graphing speed (See Slides 9 to 15). (See Figure 7 for a copy of Slide 14). Participants were asked to answer a cognitive load question at the end of part one and the end of part two (See Slides 8 and 16). Finally, at the end of the Teaching PowerPoint presentation, participants were asked to answer a summative cognitive questionnaire (See Slide 17). For consistency, all slides included notes which the researcher could access when in presenter view using an Apple MacBook.
Student worksheet. The participants were given a worksheet just before the teaching PowerPoint presentation (See Appendix J). The worksheet was headed "The need for Speed", it consisted of two A4 sheets backed and laminated. The worksheet had two functions. Firstly, it acted as a set of notes (the main ideas from the teaching PowerPoint presentation) and secondly, as a space for students to practice calculations and graphing questions. The first side was concerned with part one of the teaching PowerPoint presentation, i.e., the main ideas about the topic of speed, a triangle with the symbols d, s, and t drawn inside it, and two questions requiring the participants to calculate speed, distance, and time using the speed formula (speed = distance/time). The second side included copies of Slides 9 to 15 (part two of the teaching PowerPoint presentation). Participants were given an overhead transparency pen to write on the laminated worksheet during the teaching.
**Cognitive load questionnaires.** This study employed two cognitive load questionnaires. One was a cognitive load question which was completed repeatedly during the study while the participants were learning (online). The second was a summative cognitive load questionnaire which was completed at the conclusion of the teaching.

**Repeated cognitive load questions.** During both types of PowerPoint presentations (pre-training and teaching) the participants were asked to complete the same cognitive load question twice. Participants in the experimental Treatment group and the Treatment group who received the teaching PowerPoint presentation twice answered this question four times, participants who only received the teaching PowerPoint presentation once answered this cognitive load question twice. The cognitive load question read "How hard was it to understand what I was just teaching you?" (See Appendix K). The scale was a line with the numbers 0, 50 and 100 evenly spaced along the line from left to right. Under the number 0 the label read "very very easy" and under the number 100 the label read "very very hard". Under this scale were four grey boxes labelled time 1, 2, 3, and 4 for participants to write their answers. The sheet was A5, with spaces for participants to write their "name, school, and class" at the top of the sheet. Under this were instructions for the participants regarding the completion of this sheet which read "At several times during the teaching today you will be asked to think about how difficult it is for you to understand what I am teaching at that time. You need to choose a number from between 0 and 100 and write it in the box below."

**Summative Cognitive Load Questionnaire.** The summative cognitive load questionnaire was completed by all participants at the conclusion of the teaching (See Appendix L). This cognitive load questionnaire had five different questions. All the questions included a scale from 0 to 100 with the half way mark labelled as 50. Question 1 asked the participants to rate the effort they invested to understand the science ideas taught. The labels on the scale read "very very easy" at the zero end of the scale and "very very hard" at the one hundred end of the scale. Question 2 asked the participants about the effort required to learn and remember the science ideas taught, for example, this question read "Question 2: Effort to learn. How hard was it for you to learn and remember the science ideas? Think: Would I have to use the worksheet to remember the science ideas?"
Exactly the same scale and labels as used for Question 1 were also used on Question 2. Question 3 asked participants to rate their perceived success at understanding, remembering, and learning the science ideas with respect to the percentage they thought they would achieve if they sat a test on the science ideas they had just been taught, the labels on the scale read "not at all successful" at the zero end of the scale and "very successful" at the 100 end of the scale. Question 4 asked about their frustration level while learning the new science ideas, the labels on the scale read "Not at all frustrated" at the zero end of the scale and "very frustrated" at the 100 end of the scale. Question 5 asked the participants to indicate their feeling on the pace of the lesson, the question read "Question 5: Timing How rushed did you feel the whole lesson was today? Think: How was the pace of the teaching and the time given to do the tasks?" The labels on the scale read "far too slow" at the zero end and "far too rushed" at the 100 end of the scale. In addition, on this question only, a label was included above the number 50, the label read “about right,” this clarified for the students which number would indicate that the teaching was at the right pace for them. The participants indicated their score by writing their chosen number in a grey box to the immediate right of the scale for each question. The sheet had spaces for participants to write their name, school, and class at the top of this questionnaire. Under this were instructions for completing this sheet, the instructions read "For all questions below think about the whole teaching time today. Give your answer as a number between 0 and 100 and write it in the grey box at the end of the scale." (All sheets completed by the participants were in greyscale, unlike the examples supplied electronically in the Appendices).

**Monitoring of the presentations and participant behaviour.** Prior to both the pre-training PowerPoint and the teaching PowerPoint the classroom teacher was given a copy of the PowerPoint Slides and presenter notes. During the instruction the teacher annotated the script to indicate any deviations from the script. This was important so teaching consistency was maintained for all the teaching of each class involved in all of the eight schools and to establish the integrity of the different treatments. In addition to the script, the teacher was also given a small sheet (6 x 21cm) which had a space for the teacher to write their name, school, and class at the top of this sheet. This sheet was for describing the behaviour of the participants during the data collection phase (See Appendix M).
The instruction read “In my experience a teacher new to teaching this class would find the management of this class”. Under this were five descriptors the descriptors read from left to right “very difficult, difficult, reasonable, straightforward and extremely straightforward”. Teachers indicated their choice by ticking a small box above the descriptor. The reasons for this will be discussed in the next section on procedure (See pp. 132).

**Post test.** This test was identical to the pre test except that it was headed “Post test.” Participants completed this immediately after completing the Summative Cognitive Load Questionnaire (See Appendix N). The time limit for the post test was also 10 minutes. The internal consistency of the post test measure (Cronbach’s alpha) was 0.83.

**Ability testing.**

Within the present study both Progressive Achievement Test (PAT) and e-asTTle scores for both Reading Comprehension and Mathematics will be used as a standardised baseline achievement measure. These scores were supplied by the school for the participants involved in this study.

**Progressive Achievement Test (PAT).** PAT are standardised tests used in New Zealand primary and secondary schools to compare subjects with national norms and for longitudinal monitoring of individual student achievement. The main PAT tests used are Mathematics, Reading Comprehension, Reading Vocabulary, and Listening. These tests are usually administered to subjects early in the year and are often used to determine ability grouping, remedial assistance needed by individual subjects and to inform and review current practice. The New Zealand Government sees these tests as an important source of standardised data for improving practice particularly in the area of numeracy and literacy. The Kuder-Richardson Formula 20 method was used to measure the reliability of the PAT Mathematics and Reading Comprehension tests when norming the test. The KR20 values reported for the PAT Reading Comprehension test ranged from 0.87-0.94 while for the PAT Mathematics test ranged from 0.90-0.93. With respect to validity, the PAT Mathematics and Reading Comprehension test are reported to have high content and concurrent validity (Reid & Elley, 1991; Reid, 1993).
the present study the PAT Mathematics and Reading Comprehension age percentile scores will be used as a standardised, base line achievement measure.

**e-asTTle tests.** e-asTTle tests are web-based assessment tools aligned to the New Zealand Curriculum (Ministry of Education, 2007) that can be accessed and used by schools and teachers to assess students' achievement and progress against the national average in the following curriculum subjects - Reading Comprehension, Mathematics, and Writing. The Reading Comprehension and Mathematics scores used in this study were developed for use with students from Year 5-10 (ages 10-15 years old). These web-based tests can be used by the school at any time during the year. e-asTTle test results can be used to inform planning and teaching to maximise individual student learning but can also be used by students' for self assessment and by teachers to inform parents of their child's achievement. This website is administered by the New Zealand Ministry of Education. Reliability and validity data is not available for e-asTTle due to the dynamic nature of the testing. Items are selected by teachers for students from a pool of several thousand items. This effectively prevents the collection and presentation of simple validation summary tables for e-asTTle. An alternative approach to validation was developed for e-asTTle and established the validity of e-asTTle for the construct of Reading Comprehension (Hattie et al., 2003)

**Spatial ability testing.**

Between one and three months after the data collection phase, the classroom teacher administered two spatial ability tests which the participants completed during a scheduled science class. A file box was delivered to the school by the researcher which contained all the materials needed for the spatial ability testing. The kit included the instructions for the teacher and the participants to follow, these were provided in a PowerPoint presentation on a USB data stick, a hard copy of the PowerPoint presentation with annotated notes for the teacher, 30 copies of each test, and answer sheets for the participants. Both tests for spatial orientation and for spatial visualization were in two parts. The total time allowed for participants to complete the two tests was 18 minutes. The tests were developed by Ekstrom, French & Harman (1976) and the researcher was granted permission by the Education Testing Service (ETS) to use the two tests specified
from the "Manual for Kit of Factor-Referenced Cognitive Tests" which contains 72 tests for 23 aptitude factors. Both the tests were on laminated A4 sheets but the participants wrote their answers on an answer sheet which was provided for them.

**Spatial ability test for spatial orientation.** The spatial ability test for spatial orientation was referred to as "S2 - Cube Comparison Test" (See Appendix O). It was divided into two parts, Part one included questions one to twenty one and Part two included questions twenty two to forty two, participants had 3 minutes to complete each part of this test. Instructions at the bottom of the first page informed participants not to go onto the next part till instructed to do so. This cube comparison test was a double-sided, laminated, A4 sheet.

**Spatial ability test for visualization.** The spatial ability test for spatial visualization was referred to as "VZ3 - Paper Folding Test" (See Appendix P). It was also divided into two parts, Part one included questions one to six and Part two included questions seven to twelve, participants had 6 minutes to complete each part of this test. Instructions at the bottom of the first part informed participants not to go onto the next part till instructed to do so. This Paper Folding Test was two, doubled sided, laminated, A4 sheets (4 sides in total).

**Spatial ability PowerPoint.** The spatial ability PowerPoint consisted of 14 slides with relevant visuals and text at an appropriate language level for Year 9 or 10 students. (See Appendix Q). There were between 22 and 64 words per slide. The first set of slides included information on the nature of the tests and instructions for completing the tests (See Slides 1 to 4). The next set of slides introduced the Cube Comparison Test, gave two examples of Cube Comparison Test questions with answers, provided the participants with two similar questions to those in the test as a chance to practice this skill before attempting the actual test, and instructions for completing the Cube Comparison Test part one and two respectively (See Slides 5 to 9). The next set of slides introduced the Paper Folding Test, gave them an example of how to correctly answer the questions in this test, and included instructions for completing Part one and two of this test (See Slides 10 to 13). (See Figure 8 for a copy of Slides 7 and 11). The final slide thanked participants for their participation (See Slide 14). For consistency, all
slides included notes which the researcher could access when in presenter view using an Apple MacBook.

*Figure 8. Slide 7 and 11 from the Spatial ability PowerPoint*

**Spatial Ability Tests Answer Sheet.** The answer sheet consisted of one A4 sheet with spaces at the top for the participants to write their name, school, and class (See Appendix R). The Cube Comparison Test answer section was on the top half of the A4 sheet. Part one and Part two were separated by a bold horizontal line. The Paper Folding Test answer section was divided into two columns, column one had Part one questions 1-6 and column 2 had Part two questions 7-12. Each question had five answers, these answers were formatted into a 6x2 table (See Figure 9 for Paper Folding Test Question 1).
**Procedure**

This section is divided into two main parts outlining the procedures for the Pilot and Main Studies. Together these two studies comprise Study 2, both studies contain an initial phase and a data collection phase. The data collection phase in the Main Study is also referred to in this Chapter as the experimental treatment. In the Pilot Study subsection, the procedure followed and the changes made to the Main Study procedure as a result of completing the Pilot Study are outlined in detail. In addition, other issues that arose as a result of this study being conducted in an ecologically valid setting and how these would be managed in the Main Study are also outlined. Finally, the results are summarised briefly. Regarding the procedures for the Main Study, each phase (initial and data collection) are outlined including Ethics procedures and justification for the cognitive load measures used.

**Pilot Study.**

A Pilot Study was conducted using three Year 10 (14-15 years old) classes from one secondary school in Auckland and involved 70 participants over two science class periods. The three classes were randomly assigned to each of the three treatment groups that were to be used in the present study. Treatment group 1

---

**Paper Folding Test**

**Part 1**

1.

<table>
<thead>
<tr>
<th>Side</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 9. Paper Folding Test Question 1 on Spatial Ability Tests Answer Sheet*
was the intervention Treatment group. This was the only Treatment group to be taught using the pre-training strategy which occurred in advance of being taught the science ideas in the topic of motion. Treatment group 2 received teaching of the science ideas once through, similar to the teaching participants would receive in a school setting. It could be argued that the success or otherwise of the intervention Treatment group 1 was because this group received more time being taught the science ideas. In order to address this potential criticism Treatment group 3 was established as a group which received more teaching time than the other two treatment groups. Treatment group 3 were given two opportunities to learn the science ideas whereas the other two treatment groups had only one opportunity. See Table 11 for a summary of the three treatment groups.

Table 11  
*Treatment groups used in Study 2*

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Description of Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pre-training PowerPoint presentation followed by the teaching</td>
</tr>
<tr>
<td></td>
<td>PowerPoint</td>
</tr>
<tr>
<td>2</td>
<td>Teaching PowerPoint once through</td>
</tr>
<tr>
<td>3</td>
<td>Teaching PowerPoint twice through</td>
</tr>
</tbody>
</table>

The purpose of conducting this Pilot Study was three-fold. Firstly, to trial all the materials (PowerPoint presentations, cognitive load questionnaires, pre and post tests, and the student worksheet) and to get participant feedback. Secondly, to determine time allocation for teaching using the PowerPoint presentations, pre and post testing, and completing the cognitive load questionnaires and thirdly, to determine if there was a pre-training effect within the Pilot Study context.

The Pilot Study was useful in identifying changes that were required to both the procedure and materials employed in the initial and data collection phase. These changes will be outlined in more detail in the next few paragraphs.

**Changes to the materials.** *Initial phase PowerPoint.* One change occurred to this material due to the Pilot testing. A grey box was added to Slides 7 to 10 so that the questions on the PowerPoint were exactly the same as the questions on the practice cognitive load question sheet so the participants
recognised the PowerPoint slides as matching their sheet so as to minimise any confusion.

*Practice cognitive load question sheet.* Originally this question sheet had only one question on it matching Slide 7 of the initial phase PowerPoint regarding rating their enjoyment of a recently viewed movie or DVD. Due to the results of the Pilot Study, all the questions from Slides 7 to 10 were added so that firstly, participants could practice answering all the questions which would give the participants more confidence in answering them in the data collection phase and it was hoped that as a result this would speed up the time taken to complete these questions in the data collection phase. Secondly, participants’ understanding of the cognitive load questions as the researcher intended them to be understood has been found to be a potential confounding issue with subjective rating scales (Cierniak, Scheiter, & Gerjets, 2009; Schnitz & Kurschner, 2007). In order to address this potential problem in this study, the questions were fully explained to participants, they had an opportunity to ask questions and could practice them in a context similar to the one used in the data collection phase.

*Pre and post test.* Participant feedback and results from the marking of the pre and post test in the Pilot Study suggested a need for multiple changes to the nature and procedures surrounding the implementation of the pre and post test. Participant feedback indicated that question six was too difficult for them, especially after the limited amount of teaching possible and based on the fact that they were novices for this topic. Therefore, this question was deleted and replaced with a question that related to Slide 14 regarding the graphing of different speeds on the same graph. In addition, the units used in question 3b and 4 for speed (ms⁻¹) were changed to m/s for two reasons, firstly, as the participants were familiar with km/h as a unit of speed then as m/s is in the same format, feedback suggested that they would feel more comfortable with this notation, and secondly, explaining the difficult mathematical concept of s⁻¹ being equivalent to one divided by s was not possible in the time allocated to the teaching in the data collection phase. Also, due to the frequency of questions relating to question 5c, a sentence was added to the instructions to clarify this question, the sentence added read "Compare the sketch graph you drew and the actual graph of her fall... in part 5b."Finally, personal access to a set of calculators proved to be difficult and
therefore a set of 30 calculators and rulers were taken to each school and given to participants for use during the data collection phase of the Main Study.

*Teaching PowerPoint.* Results of the Pilot Study suggested a need for multiple changes to the Teaching PowerPoint presentation. Slide 3 contained a hyperlink to a video of Usain Bolt running a 100m race in Berlin, however, due to technology incompatibility and a problem with internet access this hyperlink did not work during the presentation. This video was an important feature of this presentation and so the video was embedded in the presentation so it did not require a connection to the internet. In addition, as a result of participant feedback and questions asked during the PowerPoint presentation changes were made to Slides 7, 9, and 10. Numbers along the horizontal line in Slide 7 were deleted as they were too small to be seen from the back of the classrooms used in the Pilot Study and this was likely to be the same in other classrooms in schools in Auckland. The words "slope = speed" were animated in Slide 9 so that the definition of slope could be explained in this context before the concept of slope being equivalent to speed on a distance–time graph was introduced as the sequence of presenting these two ideas is important in understanding the main ideas of graphing speed. Finally, the order of the cars in the key for Slide 10 were changed so they were in the same order as the lines on the graph. This made answering questions verbally in the teaching of Part two and in the post test regarding information presented on this slide easier.

*Other issues.* In addition to procedural and hard copy materials amendments, there were three issues regarding participant behaviour, timing of Treatment groups 1 and 3, and the post test that were identified during the Pilot Study that required some attention. All three of these issues could potentially have had an impact on the performance and cognitive load measures collected in the data collection phase and so required careful consideration. Participant behaviour impacts teaching and student learning, i.e., time on task and engagement (Beadle & Murphy, 2013; Shulman, 1991). The consequence of bad behaviour by only a few participants could impact negatively on the teaching and hence the learning of the whole class. It was decided to ask the teacher to monitor the teaching and note when any interruptions occurred to the teaching schedule (i.e., participants entering or leaving the class or interruptions to the
teaching initiated by questions or negative participant behaviour requiring discipline). In addition to this monitoring, the teacher would also be asked to rate student classroom behaviour at the end of the teaching by completing a 5-point rating scale (very difficult – extremely straightforward).

The second issue was timing of Treatment 1 and 3. Class periods are a maximum of one hour but allowing for participants to travel to the class from elsewhere in the school reduces this time to under an hour. In the Pilot Study, the class that received the teaching PowerPoint presentation twice were rushed and did not finish in the time allocated. When the bell went for the end of the lesson the participants wanted to leave and this was problematic. Some participants did not have enough time to adequately complete the summative cognitive load questionnaire and the post test. In the Pilot Study the participants completed the online cognitive load question during the teaching three times for each PowerPoint presentation, that is, six times in Treatment 1 and 3 and three times in Treatment 2. This was reduced to twice for each presentation, i.e, four times for Treatment 1 and 3 and twice for Treatment 2. By making these changes to the frequency of this question, the two online cognitive load measures would now occur at the end of part one (speed definitions and calculations) and part two (graphing speed) which represents the cognitive load imposed for parts one and two, respectively. Another change that was implemented to reduce the timing of the experimental procedure was to impose a time limit of 10 minutes for participants to complete the post test. This meant that all classes would receive the same amount of time to complete this test, an important consideration for consistency between the schools. Therefore the instruction "You have 10 mins" was added to the end of Slide 19 of the teaching PowerPoint. Another change that was made regarding the issue of timing was the deletion of a chart on the student worksheet that students completed regarding units of distance, time and speed. This chart was removed as participant feedback suggested that it was redundant and it required additional time for teacher instructions regarding chart completion and for participants to complete it. After the chart was removed from the worksheet and slides were amended the final worksheet was laminated so that the participants could write answers on it with overhead transparency markers and then the sheets were immersed in water and dried ready for the next class. As already mentioned, practicing the cognitive load questions in the initial phase was introduced to the
procedure to potentially reduce the time taken to complete the online cognitive load question during the teaching and the summative cognitive load questionnaire at the end of the teaching. Lastly, it was decided that time could also be saved if all the materials required were already set out in advance on the desks before participants entered the room.

The last potential problematic issue identified in the Pilot Study was that of participants not taking the post test seriously. It was stressed to the participants before and during the teaching that this post test was not part of their school assessment but that it was important for the results of this study that they demonstrated what they had learned during the teaching. It was decided to inform the participants in the initial phase that they would be required to answer questions about what they had learned after the teaching. These instructions were added to the initial phase PowerPoint on Slide 4. In addition, a verbal explanation reinforced this written instruction and it was stressed that it was important to try your best even though it was not part of your school assessment as this was material that they would be required to learn next year for NCEA science and this was a good opportunity to learn this in advance of Year 11.

Results. Results from the Pilot Study were analysed by calculating means and standard deviations of each treatment and it was found that the participants in Treatment group 1 (pre-training Treatment group) reported lower cognitive load scores and that the differences between pre and post test scores were greater for this group also. This indicated a pre-training effect and suggested that further research into this strategy was warranted.

Main Study.

Two months before the study, Ethics sheets were sent to the Principal of the nine schools selected to be involved in the study. Once their approval was given for the school to be involved in the study the Head of the Science Department (HOD Science) was contacted and a meeting arranged. At this meeting the study procedures were fully explained and an information pack was given for the HOD Science and the teachers involved in the study, this included Ethics sheets for both the HOD Science and the teachers of the classes involved, all sheets that the
participants would be required to complete and a one page overview of the
different treatments. Four weeks before the study the procedure was explained to
all the science teachers from the school during a science department meeting.
Following this meeting, teachers involved were given the added opportunity to ask
any further questions and a schedule agreed upon for the initial phase and the
data collection phase of the Main Study. (See Appendix D for all Ethics sheets)

**Initial phase.** Within 10 days of the data collection phase the participants
were addressed and shown the organisation of the experimental treatment using a
PowerPoint presentation. Participants were made familiar with the cognitive load
materials they would be required to complete and given an opportunity to practice
answering these questions in the context of a recent science lesson. Each
question was carefully explained and participants were given an opportunity to ask
questions. This was to ensure that they fully understood what was being asked of
them, that they interpreted the questions as the researcher intended them to and
that they understood how to answer the cognitive load questions. The order of
events in the experimental treatment was then explained. Ethics sheets were then
distributed to the participants, these included a participant information sheet, a
participant consent form and a parental information sheet. No parental consent
was required given that the teaching methods were similar to ones used by
teachers in New Zealand schools and the content was appropriate for these
participants, non controversial and aligned to the New Zealand Curriculum
(Ministry of Education, 2007). Participants consented for their data from the Main
Study to be used in any analysis performed. Participant consent forms were
collected from those who wished to return them in this period while others were
collected by the teacher in subsequent science classes before the data collection
phase. Participants who did not consent to their data being used were still
involved in the data collection phase but their results were not included in the
study. Participants were informed that in the experimental treatment they were not
required to bring anything except a pen or pencil, i.e., this was not a bring your
own device lesson, calculators and rulers would be supplied. After ascertaining
that there were no further questions the participants were asked to complete a pre
test. They were given 10 minutes to complete this pre test. Participants were then
thanked and reminded of the date of the data collection phase.
**Data collection phase.** The experimental treatment was conducted over one science class period in the timetabled classroom for each of the classes involved. Classes were randomly assigned to the three different treatment groups before the data collection phase. All materials participants required were set up in the classroom in advance of them entering the room, these included the laminated student worksheet, online cognitive load question sheet, an overhead transparency marker, a ruler, and a calculator. The data show was set up and the title Slide was projected onto the screen. Spare pens were provided for those participants who did not bring one. The teacher was given the monitoring materials to complete during the teaching. The teaching began when the participants had all arrived and were seated.

The procedure which followed depended on which Treatment group the participants were involved in. Participants in Treatment group 1 were taught using the Pre-training PowerPoint which was then followed immediately by the Teaching PowerPoint. Participants in Treatment group 2 and 3 were taught using the Teaching PowerPoint once (Treatment group 2) and twice (Treatment group 3). The initial procedure for Treatment 1 which differed from the other two treatment groups will be explained first followed by the procedure used for the Teaching PowerPoint presentation which was identical for all treatment groups. The final part of the procedure in the data collection phase involving the summative cognitive load questionnaire and administration of the post test was identical for all treatment groups.

At the beginning of the Pre-training PowerPoint presentation Treatment group 1 were instructed to leave all materials already on their desks alone and that they would need these for the second PowerPoint presentation after a short special introduction called pre-training to the topic of speed. The strategy of pre-training was explained and participants watched the Pre-training PowerPoint presentation and listened to the accompanying verbal explanations. The Pre-training PowerPoint presentation was also divided into part one and part two, these were designed to match part one and two on the Teaching PowerPoint presentation. Part one concerned the brief definitions and units of time, distance and speed and part two concerned details pertaining to line graphs. At the conclusion of both part one and part two the participants were asked to turn to the little sheet (half size)
with the question "How hard was it to understand what I was just teaching you?" and four little grey boxes labelled Time 1 to Time 4. They were then instructed verbally and this was also written on the accompanying PowerPoint Slide as to which grey box they were to write their answer in. The sheet was not given a title so that the question could be read out to the participants to help them identify the correct sheet and also as some of the participants in each class were English Second Language learners and this would assist them if their reading level was low. See Table 12 for how the Times 1 to 4 measurement points relate to each treatment group.

Table 12
*Online Cognitive load question and Times 1 to 4 relating to treatment groups*

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pre-training part one</td>
<td>Pre-training part two</td>
<td>Teaching part one</td>
<td>Teaching part two</td>
</tr>
<tr>
<td>2</td>
<td>Teaching part one</td>
<td>Teaching part two</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Teaching part one</td>
<td>Teaching part two</td>
<td>Teaching part one</td>
<td>Teaching part two</td>
</tr>
</tbody>
</table>

*Note.*
1. The online cognitive load question was asked at the conclusion of the parts mentioned
2. Treatment two did not require Time 3 and 4 so at the beginning of the Teaching ppt presentation the participants were instructed to cross out Time 3 and Time 4 as they would not need to fill these in.

The Teaching PowerPoint presentation began with a Title Slide that had a large photo of a cheetah running on it and a verbal question was asked as to why this photo was chosen, this was to access the participants prior knowledge about speed. Part one of the Teaching PowerPoint presentation started with a detailed explanation of speed and then participants watched a You tube video of Usain Bolt winning a 100m race in Berlin. This video was used as the context for teaching about speed, calculating speed, and drawing distance-time graphs of speed. This video was chosen as most participants would be familiar with Usain Bolt. This would mean that using his distance and time to calculate his speed and drawing a graph of his running of the 100m race would make the learning relevant and could motivate them to learn about speed. This is consistent with current teaching practice in New Zealand as outlined in the New Zealand Curriculum (Ministry of
Education, 2007, pp. 34) which states "students learn best when teachers enhance the relevance of new learning" and is aligned to the Learning Objectives which state "Demonstrate an understanding of physical phenomena and concepts by explaining and solving questions and problems that relate to straightforward situations" (Science: Level Six, Physical World - Physical enquiry and physics concepts). At the conclusion of both part one and part two the participants were asked to turn to the little sheet (half size) with the question "How hard was it to understand what I was just teaching you?" and four little grey boxes labelled “Time 1 to Time 4.” They were then instructed verbally and this was also written on the PowerPoint Slide as to which grey box they were to write their answer in. Treatment group 3 repeated this Teaching PowerPoint presentation exactly as just described.

Immediately after the Teaching PowerPoint presentation (second presentation for Treatment 3), the Summative Cognitive Load Questionnaire was distributed to the participants and they were reminded that these were the same questions that they had already practised but this time the questions were about today’s teaching about speed. As there was the possibility that there would be some participants that had not been present in the initial phase, they were all instructed that these questions were about your understanding and learning today. They were also instructed to read the question carefully and the importance of answering these questions honestly was emphasised. Immediately after this was completed the post test was distributed and participants were instructed to wait for a signal to start. The two large Salter timers were started and they were instructed to start answering the post test questions. The timers were placed at the front on either side of the room so all participants could see them and these timers counted down from 10 minutes so it was possible for them to check how much time was remaining whenever they needed to. When all the sheets had been collected and materials were all accounted for, the participants were thanked for their participation in this study.

**Cognitive load measures justification.**

As suggested in the previous chapter, there are a variety of ways in which one can measure cognitive load and each approach has identified advantages and
disadvantages. This section will look at the nature, placement, and justification for the use of the specific cognitive load measures employed in this study. As this study looked into the effects of an intervention on cognitive load, the cognitive load measures employed in this study were direct measures reported by participants during and after the learning. These measures can provide valuable information on the effect of differential instructional formats as used in this intervention study (van Gog, Kester, Nievelstein, Gisbers, & Paas, 2009). Two subjective measures were used, both were rating scales, one question was asked repeatedly during periods of instruction (online) and a summative cognitive load questionnaire was used at the conclusion of the teaching (off-line) for all treatments. The online measure was used to assess fluctuations in cognitive load during the learning and the off-line measure was used to assess overall (total) cognitive load (van Gog & Paas, 2008). Subjective measures are sensitive to changes in task complexity (Paas, van Merriënboer, & Adam, 1994), easy for participants to understand (Yeo & Neal, 2008), less intrusive than other measures e.g., secondary tasks, reliable, and easy to administer. These characteristics are all important considerations for the use of these subjective measures in authentic settings, which is why they were employed in this study which took place in high school classrooms (Ayres & Paas, 2012; van Gog & Paas, 2008). Concurrent cognitive load reporting is not optimal for use in complex learning environments with novices as it can interrupt the learning. For this reason when asking the participants to answer the repeated cognitive load question retrospective reporting was employed. This occurred immediately after short periods of learning (part one and two) consisting of about 12 minutes each. The aim of this was to ensure that the information was likely to be still partly activated in working memory (recency effect) whereas questions asked at the conclusion of longer periods of learning may require retrieval from long-term memory which may not be as accurate (van Gog, Kirschner, Kester, & Paas, 2012). Employing one retrospective cognitive rating question at the conclusion of learning does not assist in identifying which part of the learning participants found the most cognitively demanding (van Gog, Kirschner, Kester, & Paas, 2012). As this study was looking to determine whether speed calculations and definitions (part one) or speed graphing (part two) was the most mentally demanding information to understand and learn, the cognitive load question was asked after each part (one and two). Participants were informed during the initial phase that they would be required to answer the same questions as were being
practised in this phase during and at the conclusion of the teaching in the data collection phase. It is possible that in these situations, as is the case in this study, that when participants are expecting to answer cognitive load questions that they devote some cognitive resources to monitoring their cognitive load and therefore their subjective rating score would be more accurate (van Gog, Kirschner, Kester, & Paas, 2012).

The rating scales that were used in the two cognitive load measures were the same in both the online and off-line measures and went from 0 to 100. This was to ensure consistency between the two measures, to ensure ease of use by participants, and to reduce the possibility of confusion which could have occurred if all the scales were different. However, the labels on the scales were different for most of the questions between and within the two measures as will be further described below.

The online rating scale (one question used repeatedly) was based on the question and scale labels used by Paas (1992) and asked participants to rate their perceived mental effort to understand the science ideas on a scale of 0 to 100, with 0 being "very very easy" and 100 being "very very hard". As the aim of this study was to use the strategy of pre-training to reduce the cognitive load of learning complex ideas in physics, in this instance assessing understanding during the learning was a good indicator of the load imposed which results in meaningful learning (Mayer, 2005, 2014a).

The rating scale and questions used in the off-line measure were both a modification of the NASA-TLX (Hart & Staveland, 1988), and the versions employed by Gerjets, Scheiter, and Catrambone (2004, 2006), Paas (1992), and Pollock, Chandler, and Sweller (2002). The first two questions asked about the mental effort imposed when trying to understand (Question 1) and learn (Question 2) the physics ideas presented. Question 1 was an exact repeat of the online cognitive load measure. The next three questions were based on the NASA-TLX which measures different factors associated with a task (in this study the understanding and learning of the physics ideas in the topic of motion). The factors used were performance, frustration, and temporal demand (Questions 3 - 5 respectively). These three questions were asked as these are factors that impact
on cognitive load and learning and so this allowed for more detailed analysis of the data, e.g., investigating if there was an inverse correlation between the effort of learning and performance ratings. In the initial phase explanations given to participants regarding the three additional NASA-TLX factors were, for performance - what percentage do you think you would get if you had a test on the ideas you learned right now, for frustration - sometimes when we find understanding and learning difficult we feel frustrated or angry with ourselves or our teacher, and for temporal demand (labelled as “timing” on the summative cognitive load questionnaire) - when learning is difficult sometimes we feel like it is all happening too fast, we would like to slow it all down. These explanations were to ensure that participants understood what was being asked of them and interpreted these questions as the researcher intended (Cierniak, Scheiter, and Gerjets, 2009).

Finally, van Gog, Kirshner, Kester, and Paas (2012) suggest that one subjective rating score at the end of learning could possibly measure one of three things: a) it could measure what participants remember as the average cognitive load over the learning task, b) it could measure primarily the cognitive load of the last task that they remember the best (recency effect) or, c) it could measure the most difficult task that they remember over the learning period, i.e., maximum cognitive load during the whole period of learning (Xie & Salvendy, 2000). In order to address this potential confounding variable, participants were told in the initial phase and in the data collection phase both verbally and in written form on the sheet and the accompanying PowerPoint Slide 9 (Initial phase PowerPoint) and Slide 17 (Teaching PowerPoint) that this was about the “whole teaching time today”. The next section will give details of the analyses planned for Study 2.

Analysis

The analysis of data collected in this study is divided into five parts. The first section of analyses focuses on participant ability and prior knowledge measures, and behaviour rating scale measures. The analyses within the remaining sections are linked to the four research questions: an analysis of cognitive load measures (Research Question 1); an analysis of performance measures (Research Question 2); an analysis of instructional efficiency (Research Question 3); a series of
analyses on the influence of gender, decile, and ethnicity with respect to cognitive load, learning, and learning efficiency measures (Research Question 4). The statistical programme that was used for all the analyses is SPSS version 23.

In addition to the specific analyses employed within this study, the following sections will also include the rationale for choosing some of the specific measures employed and the expected results of all of the intended analyses.

**Participant ability (mathematics, reading and spatial ability), behaviour (class level), and prior knowledge measures analyses.**

Participant ability data was collected from all the schools involved in this study. Participant mathematics and reading ability scores were provided by each school. Some schools provided PAT scores and other schools provided e-asTTle scores. Within the New Zealand context, schools are free to choose how best to assess students’ reading and mathematics ability. In order to standardise these measures, the e-asTTle scores were converted into PAT scores for analysis [See Technical note 3, The University of Auckland, The Quantitative Data Analysis and Research (Quant-DARE)]. Spatial ability data was collected in the third session with the participants after the initial and data collection phase. The tests that were used were developed by Ekstrom, French, & Harman (1976), involved mental manipulation of objects, and all four tests were subject to time limits. Prior knowledge was measured using a pre test. The pre test included a range of questions developed to assess the learning after teaching of the content (motion in physics). The three analyses in this section were only assessing the influence of the independent variable of treatment.

The participant ability measures analysis involved two multivariate analyses of variance for treatment to determine if there was a significant difference in ability (PAT scores and spatial ability scores) between the participants in the three different treatment groups. As the groups were from the same school ability groupings (mid-range students) and three classes were randomly assigned to the three different treatment groups it was expected that there would be no significant difference between the ability of the treatment groups across all the schools involved. However, because significant treatment effects were found on PAT
scores and spatial orientation scores, these were integrated as covariates in the analyses performed to address the four research questions in order to moderate for these differences.

As indicated in the Methods section, participant classroom behaviour information was collected by asking the teachers of the classes involved in this study to rate the overall behaviour of the students within the classroom during treatment. Overall classroom behaviour was rated on a 5-point scale for each class at the end of the experimental treatment (1= very difficult, 5 = extremely straightforward). A univariate analysis of variance for treatment was employed to determine if there were significant differences in behavior between the three different treatment groups across schools. It was expected that overall there would be no difference in the behavior between the three different treatment groups across schools.

Finally, analysis of prior knowledge data involved a univariate analysis of variance for treatment to determine if there were significant differences in prior knowledge of the participants in the three different treatment groups. As these participants were all novices for this physics topic it was expected that there would be no significant difference between the pre test scores for participants in the three different treatment groups.

**Research Question 1: Cognitive load analysis.**

Online and summative cognitive load data was collected during the data collection phase of Study 2. In order to address Research Question 1 and to establish the integrity of treatment, four analyses were performed. Analysis one was a multivariate analysis of covariance to determine if there was a significant difference in self-reported cognitive load scores (online) between learning the pre-training materials only (Treatment 1) and learning the teaching information for the first time (Treatment 2 and 3) i.e., the difference in cognitive load between learning the two different sets of materials. It was expected that since there was a deliberate attempt to reduce the intrinsic cognitive load of the pretraining material compared to the teaching materials, that the self-reported cognitive load scores for learning the pretraining materials would be lower than for learning the teaching materials.
The second analysis involved a multivariate analysis of covariance to determine if there was a significant difference in self-reported cognitive load scores (online and summative) between the three different treatments when learning the teaching information. The online cognitive load measures consisted of two scores: cognitive load during the graphing section and cognitive load during the calculation section of the teaching. The summative cognitive load measures consisted of five different questions which were completed by the participants immediately after the conclusion of the teaching. The five questions focused on effort to understand, effort to learn, self-predicted performance, frustration, and pace of teaching, respectively. It was expected that if pre-training makes it easier to learn complex information then the online cognitive load scores for the pre-training treatment group (Treatment 1) would be lower, than the other two treatment groups. In addition, for the summative cognitive load questions, if the pre-training made it easier to learn the information presented then it would be expected that the cognitive load scores for understanding and learning the information (Question 1 and 2) and the frustration scores (Question 4) would be lower for the pre-training treatment group.

The third analysis was a paired sample T-test to determine if cognitive load scores differ for part one and part two of the teaching (calculation and graphing respectively) across all the treatments. As reported in the literature it is expected that participants will find the graphing more difficult than the calculations in the physics topic of motion, so the cognitive load scores for the calculation (part one of the teaching) will be lower than the cognitive load scores for the graphing (part two of the teaching).

The fourth analysis was a univariate analysis of covariance to determine if the pretraining Treatment group’s online cognitive load scores during teaching were different from the online cognitive load scores for Treatment 3 during the second time they received the teaching PowerPoint. Assuming that the pretraining did reduce cognitive load during teaching, that a similar reduction in cognitive load would also occur the second time that participants in Treatment 3 received the teaching.
Research Question 2: Performance measures analysis.

The performance measures that were used in this analysis were the learning of target concepts (difference between pre and post test scores), scores on high and low order questions, and scores on the calculation and graphing questions. This section involved three separate analyses. Firstly, a multivariate analysis of covariance determined if there was a significant difference between the difference scores of the pre-training Treatment group and Treatment group 2 and 3. If pretraining made it easier to learn complex information then it would be expected that the pre-training Treatment group would have higher difference scores than Treatment groups 2 and 3. The second and third analysis looked at differences between high and low order questions and calculation and graphing respectively, between the pre and post test. This involved two multivariate analyses of covariance that determined if there was a significant difference between the three treatment groups for differences between pre and post tests, high and low order questions, and calculation and graphing questions. If pretraining made it easier to learn complex information then it would be expected that the pretraining Treatment group (Treatment 1) would have greater difference scores for high order and graphing questions respectively.

Research Question 3: Instructional efficiency measures analyses.

Instructional efficiency scores were calculated using mental effort scores and performance scores. Given the use of two sets of mental effort measures (online and summative), the following section will identify which measures were used to calculate the efficiency scores and discuss the rationale for choosing the particular measures employed in the efficiency calculation.

In terms of generating the instructional efficiency scores within the current study, the current study calculated an average mental effort score using four of the cognitive load measures, the two online cognitive load measures for the comparable teaching of all groups (difficulty of understanding question) and the summative cognitive load Questions 1 and 2 (overall difficulty of understanding and overall difficulty of learning questions).
This combination of these measures was chosen to calculate the instructional efficiency scores for the following reasons. Firstly, the overall aim of this study was to enhance learning of complex information. Learning is described as involving both remembering and understanding, material that is remembered is used for recall of knowledge and material that is understood is used for the transfer of knowledge, that is, using the knowledge in new situations (Mayer, 2001). Given that the post test was focused on remembering and understanding (contained 10% recall and 90% transfer questions), all mental effort questions (online and summative) which include understanding and remembering were included in the mental effort measures. In addition, a focus on the learning of new material requires looking at all the processes which occur in working memory to facilitate this learning. These are organising the new material (making internal connections between the ideas) and then integrating these ideas with existing knowledge in LTM (making external connections between the new ideas and existing schemas). These processes are demanding and impose a load on working memory and therefore require mental effort (Mayer 2001, 2005; Paas, Tuovinen, Tabbers, & Van Gerven 2003). Therefore the mental effort questions as described above which are involved in important processes of learning (understanding and remembering) were included in the mental effort measures to reflect the load that these processes impose on working memory.

Secondly, integrating subjective rating scale questions about learning and understanding in the mental effort measure is consistent with mental effort measures used in previous research (Kalyuga, Chandler, & Sweller, 2000; Pollock, Chandler, & Sweller, 2002). These studies support using all the cognitive load measures involved in learning and understanding the material taught, in the calculation of instructional efficiency scores. Finally, Paas and van Merrienboer (1993) suggest that it is important to consider all the effort required to produce an outcome, i.e., both the effort to learn and the effort to understand. In summary, prior research and theory supports the approach used in this study to calculate the efficiency score. That is, calculating an average mental effort score using four of the cognitive load measures, the two online cognitive load measures for the comparable teaching of all groups (difficulty of understanding question) and the two summative cognitive load questions (overall difficulty of understanding and overall difficulty of learning questions).
With respect to instructional efficiency two analyses were employed. Firstly, a univariate analysis of covariance was used to determine if the participants in the pretraining Treatment group were more efficient in their learning. It is expected that if their performance was higher and their cognitive load scores lower then they will have higher efficiency scores than Treatment group 2 or 3. Secondly, a univariate analysis of covariance was used to determine if the pretraining group were more efficient in learning during the first time they received the teaching compared to the second time that Treatment 3 received the teaching. It is expected that pretraining will make it easier for Treatment 1 participants to learn information that they are taught (for the first and only time) in comparison to participants who had already received the teaching and were receiving it for the second time (Treatment 3), i.e., Treatment 1 participants will be more efficient in their learning.

**Research Question 4: The influence of gender, decile, and ethnicity on dependent measures analyses.**

This section involved three parts, Part A involved treatment and gender, Part B involved treatment and decile, and Part C involved treatment and ethnicity. For these three analyses the four dependent measures were, online cognitive load, summative cognitive load, difference between pre and post test scores (learning), and instructional efficiency.

**Part A Treatment and gender analysis.** This part involved a multivariate analysis of covariance to determine if gender influenced the four dependent measures for participants in the three treatment groups. It was expected that: a) if gender affected online cognitive load and summative cognitive load questions 1, 2, and 4, then the scores for males would be lower than females, b) if gender affected summative cognitive load question 3, learning, and instructional efficiency then the scores for males would be higher than for females, and, c) if gender affected summative cognitive load question 5 that males would have scores closer to 50 (just the right pace). These predictions are based on previous research into gender differences and achievement in science. That is, despite evidence relating to achievement over all science subjects that the gender gap in favour of males is
closing or in some instances reversing, this same trend has not been found in studies looking into physics achievement (Britner, 2008; Dreary, Strand, Smith, and Fernandes, 2007; Lindberg, Hyde, Petersen and Linn, 2010).

Females have also reported higher anxiety when learning physics (Britner, 2008), less positive attitudes to learning science, and a lower interest in science as a subject (Jones, Howe, & Rua, 2000). These findings could translate into males being more motivated, interested, and less anxious, therefore rating scores for cognitive load (online cognitive load, summative question 1, 2 and 4) could be lower than females. For summative cognitive load question 5 females could find the pace of presenting the ideas too fast and therefore their rating scores for this question could be further from 50 compared to those of males. Higher anxiety, less interest, and less positive attitudes towards physics could affect female’s confidence and therefore their self-predicted performance could be lower than for males (summative question 3). Finally, achievement of males in physics is reported to be higher so in combination with their lower cognitive load scores, instructional efficiency scores could also be higher.

The focus of these analyses was on the interaction of treatments and gender, and the impact of gender on the target outcomes. It was expected that any treatment main effects found in these analyses would mirror those found in the initial set of analyses which focused on the impact of treatment on the individual target outcomes.

**Part B Treatment and decile analysis.** As there were only very small numbers of schools in each decile category (one or two) and the schools were all very different and some were atypical for their decile rating, it would not be appropriate to put decile 1 and 2 together to make a low category, decile 4 and 5 to make a mid category and decile 9 and 10 to make a high category. Consequently, in the analyses all six decile categories were included.

The analyses within this part involved a multivariate analysis of covariance to determine if decile influenced the four dependent measures for participants in the three treatment groups. It was expected that: a) if decile affected online cognitive load and summative cognitive load questions 1, 2, and 4, then these scores for
high decile schools would be lower than lower decile schools, b) if decile affected summative cognitive load question 3, learning, and instructional efficiency then these scores for high decile schools would be higher than for low decile schools, and, c) if decile affected summative cognitive load question 5 that high decile schools would have scores closer to 50 (just the right pace). These predictions are based on research findings in New Zealand looking into ethnicity differences between NZ/European and Asian achievement in science and that of Maori and Pasifika. TIMMS, PISA, and NEMP reported findings, and NZQA data confirm that Maori and Pasifika achievement in science is below that of NZ/European and Asian ethnicities (Chamberlain & Caygill, 2012; Coll, Darsah, & Faikhamta, 2010; Crooks, Smith, and Flockton, 2008). In Auckland, lower decile schools have larger numbers of Maori and Pasifika students compared to higher decile schools. If Maori and Pasifika students have lower achievement and find learning in science more difficult then these participants could have to put in more effort to learn the science ideas and so their cognitive load scores could be higher (online cognitive load, summative cognitive load questions 1, 2, and 4). Correspondingly, their confidence in self-predicted performance scores and instructional efficiency scores could be lower than NZ/European and Asian participants. Finally, if learning was more difficult then these participants could find the pace of presenting the ideas too fast and therefore their scores for summative cognitive load question 5 could be further from 50 than those of NZ/European and Asian participants.

The focus of these analyses was on the interaction of treatments and decile and the impact of decile on the target outcomes. It was expected that any treatment main effects found in these analyses would mirror those found in the initial set of analyses which focused on the impact of treatment on the individual target outcomes.

**Part C Treatment and ethnicity analysis.** This part involved a multivariate analysis of covariance to determine if ethnicity influenced the four dependent measures for participants in the three treatment groups. It was expected that: a) if ethnicity affected online cognitive load and summative cognitive load questions 1, 2, and 4, then the scores for NZ/European and Asian participants would be lower than for Maori and Pasifika participants, b) if ethnicity affected summative cognitive load question 3, learning, and instructional efficiency
then the scores for NZ-European and Asian participants would be higher than for Maori and Pasifika participants, and, c) if ethnicity affected summative cognitive load question 5 that NZ-European and Asian participants would have scores closer to 50 (just the right pace) than Maori and Pasifika participants scores. These predictions are based on previous studies and results from national and international data (See Chapter 3 – Maori and Pasifika Achievement in Science).

The focus of these analyses was on the interaction of treatments and ethnicity and the impact of ethnicity on the target outcomes. It was expected that any treatment main effects found in these analyses would mirror those found in the initial set of analyses which focused on the impact of treatment on the individual target outcomes.

The next chapter discusses the results of the analyses detailed in this section.
Chapter 5

Results

The current study was concerned with the effect of a pre-training intervention on cognitive load imposed when novices learn complex physics information including graphs in science. The research questions for this study were:

1. Does pre-training reduce the cognitive load associated with learning complex science information involving graphs as measured during and after learning?

2. Within the context of learning complex information involving graphs, what is the effect of pre-training on secondary science students’ overall learning, performance on different levels of question (high versus low order), and performance on different content questions (calculation versus graphing)?

3. Does pre-training increase instructional efficiency for secondary science students when learning complex information involving graphs?

4. Are there differential effects on cognitive load, learning (difference scores) and instructional efficiency with respect to gender, decile, and ethnicity when using pre-training in a secondary science context involving graphs?

Performance was measured by using identical pre and post tests, and then calculating the difference between the two scores. This is referred to as “difference” or “learning” in all consecutive reporting. Participants’ prior knowledge with respect to graphing and the physics topic of motion was integrated into the design of the study using pre test measures which gave an index of what participants knew before treatment. In addition participant ability was measured and focused on spatial ability, reading and mathematics skills, as it is likely that these could have an impact on participants’ performance in this study.
To assess the extent to which behaviour may have impacted on performance, both the teacher of the class and the researcher gave the class a behaviour rating on a 5-point scale, from 1 being very poor to 5 being excellent behaviour. A paired sample T-test found no significant difference between the teacher and researcher rating of behaviour of the participants (Teacher behaviour rating $M = 3.34, SD = 0.62$; Researcher behaviour rating $M = 3.36, SD = 0.86$).

As indicated in the Methods section, in order to assess the cognitive load which was imposed during the learning, subjective ratings of the mental effort required to understand and learn the material were gathered during and at the end of the teaching. It was expected that the participants in the pre-training treatment group would attain greater differences between pre and post test scores and would also evidence lower cognitive load as demonstrated by lower cognitive load ratings. In addition, as a result of this lower cognitive load, it was expected that their instructional efficiency would be higher.

Before specifically addressing the results pertaining to the research questions the following section will summarise differences found in the participant ability measures, (spatial ability, reading and mathematics ability), prior knowledge scores, and behaviour ratings. Participant ability, behaviour, and prior knowledge differences could have an impact on performance in this study and therefore it was important to establish whether any differences existed between treatments so that they could be accounted for in the analyses.

**Participant ability (reading, mathematics, and spatial ability) and prior knowledge measures**

In this study the PAT Mathematics and Reading Comprehension scores were used as baseline ability measures. As already described in the Analysis section some schools provided PAT scores and other schools provided e-asTTle scores. These are not compatible so a method was used to convert the e-asTTle scores into PAT scores for analysis [see The University of Auckland, The Quantitative Data Analysis and Research (Quant-DARE) Technical Note 3]. PAT tests are standardised tests used throughout New Zealand to assess ability and are used to compare students to national norms and for longitudinal monitoring.
Spatial ability was measured using two tests, one for spatial orientation (SO) and one for spatial visualization (SV). Each spatial ability test was divided into two parts. The tests were developed by Ekstrom, French, & Harman (1976) and involved mental manipulation of objects within time limits. Prior knowledge was measured using a pre-test. The pre-test included a range of questions developed to assess the learning after teaching of the content (motion in physics). The results in this section will only be reported for comparisons between the three different treatment groups for the factors being investigated.

**Mathematics and reading ability.**

A multivariate analysis of variance comparing the participants’ mathematics and reading ability in each treatment group found significant main effects for treatment (Treatment: Wilks λ, $F = 7.04$, $p < .001$, partial $\eta^2 = .03$). Tests of Between-Subjects effects found significant effects for both of the PAT measures [Treatment: PAT Mathematics, $F (2,535) = 6.06$, $p = .002$, partial $\eta^2 = .02$; PAT Reading Comprehension, $F (2,535) = 12.40$, $p < .001$, partial $\eta^2 = .05$]. Multiple Comparisons using the Bonferroni method found that Treatment 1 (pre-training) scored significantly lower on PAT Mathematics tests than Treatment 2 (teaching once) ($p = .002$) and significantly lower on PAT Reading Comprehension tests than Treatment 2 ($p < .001$) and Treatment 3 (teaching twice) ($p = .001$). In order to assist with the interpretation of these results, additional information has been included. The overall mean for PAT Mathematics was ($M = 4.88$), and for PAT Reading Comprehension ($M = 5.14$). Scores of 4, 5, and 6 represent average ability for these Year levels. See Table 13 for the PAT means for Mathematics and Reading Comprehension for Treatments 1, 2, and 3.

**Spatial ability.**

A multivariate analysis of variance comparing the participants’ spatial ability in each treatment group found significant main effects for treatment (Treatment: Wilks λ, $F = 2.34$, $p = .017$, partial $\eta^2 = .02$). Tests of Between-Subjects effects found significant effects for both of the spatial orientation measures [Treatment: spatial orientation 1, $F (2,473) = 6.70$, $p = .001$, partial $\eta^2 = .03$; spatial orientation
Multiple Comparisons using the Bonferroni method found that the participants in Treatment 1 scored significantly lower on spatial orientation 1 than Treatment 2 ($p = .026$) and participants in Treatment 3 scored significantly lower than Treatment 2 for both spatial orientation 1 ($p = .001$) and 2 ($p = .022$). The following information has been added in order to better interpret the nature and magnitude of the scores within each treatment group. The range of scores for each test across the different treatments was almost identical for spatial orientation 1 and 2 (0 or 1 to 20 or 21) and identical for spatial visualization 1 and 2 (0 to 30). The best possible score for spatial orientation 1 and 2 was 21 and for spatial visualization 1 and 2 was 30. See Table 13 for the means for all four spatial ability tests for Treatments 1, 2, and 3.

Table 13

Means and Standard deviations for, PAT Mathematics and Reading, Spatial ability and Pre test for all Treatment groups.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Treatment 1 (Pre-training)</th>
<th>Treatment 2</th>
<th>Treatment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>PAT Mathematics</td>
<td>9</td>
<td>4.59 1.80</td>
<td>5.21</td>
</tr>
<tr>
<td>PAT Reading</td>
<td>9</td>
<td>4.65 1.54</td>
<td>5.47</td>
</tr>
<tr>
<td>Spatial orientation 1</td>
<td>21</td>
<td>11.24 3.80</td>
<td>12.32</td>
</tr>
<tr>
<td>Spatial orientation 2</td>
<td>21</td>
<td>11.54 3.33</td>
<td>11.80</td>
</tr>
<tr>
<td>Spatial vizualisation 1</td>
<td>30</td>
<td>9.95 7.33</td>
<td>9.71</td>
</tr>
<tr>
<td>Spatial vizualisation 2</td>
<td>30</td>
<td>11.97 8.65</td>
<td>10.70</td>
</tr>
<tr>
<td>Pretest</td>
<td>20</td>
<td>2.71 2.93</td>
<td>3.30</td>
</tr>
</tbody>
</table>

Note.

a N for PAT scores: Treatment 1=181, Treatment 2=183, Treatment 3=171
b N for Spatial ability: Treatment 1=168, Treatment 2=141, Treatment 3=164
c N for Pretest: Treatment 1=178, Treatment 2=186, Treatment 3=186
Indications from previous research (see Chapter 3) into gender differences in spatial ability suggest that there may be differences between the scores of males and females, this was therefore investigated. A multivariate analysis of variance comparing the spatial ability of males and females found no significant main effects for gender. See Table 14 for the means and standard deviations for the four spatial ability tests by gender.

Table 14

Means and Standard deviations for Spatial ability and Gender.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Max score</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Spatial orientation 1</td>
<td>21</td>
<td>11.29</td>
<td>3.74</td>
</tr>
<tr>
<td>Spatial orientation 2</td>
<td>21</td>
<td>11.51</td>
<td>3.74</td>
</tr>
<tr>
<td>Spatial visualisation 1</td>
<td>30</td>
<td>9.05</td>
<td>6.44</td>
</tr>
<tr>
<td>Spatial visualisation 2</td>
<td>30</td>
<td>10.69</td>
<td>8.58</td>
</tr>
</tbody>
</table>

Note. N for males = 222, N for females = 251

Pre test measures.

A univariate analysis of variance comparing the pre test scores of participants in each treatment group found no significant main effect for treatment. See Table 13 for the means for the pre test scores for treatment.

Summary.

Given the differences in PAT scores between treatments, and in order to control for ability differences in mathematics and reading which can have an influence on learning in science, PAT Mathematics and Reading Comprehension scores were used as covariates in all between treatment analyses. In addition, given the
differences in spatial orientation across treatments which can influence learning in physics, it was decided to also include these measures as covariates in all between treatment analyses.

**Behaviour Rating Scale measures**

A univariate analysis of variance was conducted in order to examine group differences on the behavior rating measures. There was no significant main effect for treatment. The means for each treatment were Treatment 1 \((M = 3.32, \ SD = 0.50)\), Treatment 2 \((M = 3.33, \ SD = 0.65)\), and Treatment 3 \((M = 3.43, \ SD = 0.69)\).

The following sections will report on results of analyses related to each research question in turn. Research question 1 focuses on the effects of treatment on cognitive load, research question 2 focuses on treatment effects on overall learning, and performance on questions of different levels and content, research question 3 on treatment effects on instructional efficiency, and lastly, research question 4 which focuses on gender, decile, and ethnic differences in cognitive load, learning, and instructional efficiency and the moderating influence of these characteristics on the impact of pre-training.

**Cognitive load measures**

**Research Question 1.** Does pre-training reduce the cognitive load associated with learning complex science information involving graphs as measured during and after learning?

In addressing this research question, four comparisons (involving four different analyses) were of interest: a) were the pre-training materials less complex (lower intrinsic cognitive load) compared to the teaching materials, b) did the pre-training reduce the cognitive load of learning the target information compared to the other two treatment groups, c) was the cognitive load of learning calculation and graphing information different, d) were the cognitive load scores for the pre-training group for their first time of teaching different to Treatment 3 for their second time of teaching?
This study employed two sets of subjective cognitive load measures. The first set included a repeated online cognitive load question which was completed by participants immediately after the calculation and the graphing portion of the first time each treatment group received the teaching. The second set of subjective measure of cognitive load employed was a summative cognitive load questionnaire containing five different questions, completed at the conclusion of the teaching. The five questions focused on effort to understand, effort to learn, self-predicted performance, frustration, and pace of teaching, respectively.

Firstly, the assumption in using the strategy of pre-training to manage essential processing of complex information to avoid cognitive overload was that it reduced intrinsic cognitive load (essential processing) (Mayer 2014b). Therefore in this study, it was expected that the participants who were taught using the pre-training teaching materials (Treatment 1) would report lower online cognitive load scores during pre-training than participants who were taught with materials that contained all the essential information required to understand the content (Treatment 2 and 3). A multivariate analysis of covariance (with PAT Mathematics, PAT Reading Comprehension, and spatial orientation as covariates) was conducted. In this analysis, the two measures of online cognitive load (for learning calculation and learning graphing respectively) collected during pre-training in Treatment 1 were compared to the two measures of online cognitive load for learning calculation and learning graphing collected during the first presentation of the teaching PowerPoint module for Treatment 2 and 3. A significant main effect was found for treatment (Treatment: Wilks $\lambda$, $F = 44.52, p < .001$, partial $\eta^2 = 0.19$). Tests of Between-Subjects effects found significant main effects for online cognitive load scores for learning calculation and learning graphing [Treatment: calculation, $F (2,483) = 94.96$, $p < .001$, partial $\eta^2 = 0.33$; graphing, $F (2,483) = 58.75$, $p < .001$, partial $\eta^2 = 0.24$]. Pairwise Comparisons using the Bonferroni method found that for the cognitive load of learning, Treatment 1 had significantly lower online cognitive load scores than Treatment 2 ($p < .001$) and 3 ($p < .001$) during their first presentation of the teaching PowerPoint module. This indicates that the pre-training materials were lower in cognitive load than the teaching materials.

Secondly, a multivariate analysis of covariance (with PAT Mathematics, PAT Reading Comprehension, and spatial orientation as covariates) focusing on the
two online cognitive load measures and the five summative cognitive load measures found a significant main effect for treatment (Treatment: Wilks $\lambda$, $F = 5.31$, $p < .001$, partial $\eta^2 = 0.09$). Tests of Between-Subjects effects found significant main effects for the two online cognitive load measures and summative question 1 (effort to understand) and 2 (effort to learn) [Treatment: calculation online cognitive load, $F(2,375) = 21.17$, $p < .001$, partial $\eta^2 = 0.10$; graphing online cognitive load, $F(2,375) = 8.76$, $p < .001$, partial $\eta^2 = 0.05$; summative question 1 (effort to understand), $F(2,375) = 17.80$, $p < .001$, partial $\eta^2 = 0.09$; summative question 2 (effort to learn), $F(2,375) = 5.97$, $p < .001$, partial $\eta^2 = 0.03$]. For the online cognitive load measures (calculation and graphing) and summative question 1 (effort to understand), Pairwise Comparisons using the Bonferroni method found that the Treatment 1 had significantly lower cognitive load scores than Treatment 2 ($p < .001$) and 3 ($p < .001$). For summative question 2 (effort to learn), Pairwise Comparisons using the Bonferroni method found that Treatment 1 had significantly lower cognitive load scores than Treatment 2 ($p = .020$) and Treatment 3 ($p = .006$). That is, pre-training reduced the cognitive load of learning the complex ideas in science. See Table 15 for the means for online and summative cognitive load scores for Treatments 1, 2, and 3.

Due to the reported difficulty that students experience learning graphing skills it was expected that students would find learning the graphing portion of the teaching more difficult than the calculation portion. The third comparison involved a T-test to determine the difference in cognitive load scores reported immediately after learning the graphing and calculation portion of the teaching. A paired sample T-test found a significant difference between the online cognitive load reported for the graphing and calculation part of the teaching (the mean difference $M = - 4.10$, $SD = 19.77$, $t = - 4.85$, $df = 545$, $p < .001$). Previous research into the complexity of learning graphing skills would suggest that learning graphing skills is a more complex task than learning to perform calculations in physics. The results from this study also suggest that there is a difference in complexity between learning graphing and calculation skills.

Lastly, a univariate analysis of covariance (with PAT Mathematics, PAT Reading Comprehension, and spatial orientation as covariates) comparing cognitive load scores for Treatment 1 (first time of teaching) and Treatment 3 (second time of
teaching) found no significant difference between the cognitive load scores for Treatment 1 and Treatment 3 as evidenced by the similarity in means, [Treatment 1 ($M = 33.94$, $SD = 21.24$); Treatment 3 ($M = 34.41$, $SD = 20.37$)].

### Table 15

*Means and Standard deviations for the Online and Summative cognitive load measures for all treatment groups*

<table>
<thead>
<tr>
<th></th>
<th>Treatment 1 (Pre-training)</th>
<th>Treatment 2</th>
<th>Treatment 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Online cognitive load</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>30.11</td>
<td>21.29</td>
<td>38.68</td>
<td>27.29</td>
</tr>
<tr>
<td>G</td>
<td>37.61</td>
<td>23.70</td>
<td>42.51</td>
<td>24.56</td>
</tr>
<tr>
<td>Summative cognitive load</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>33.41</td>
<td>20.39</td>
<td>42.07</td>
<td>23.38</td>
</tr>
<tr>
<td>2</td>
<td>39.13</td>
<td>22.30</td>
<td>42.33</td>
<td>18.98</td>
</tr>
<tr>
<td>3</td>
<td>63.77</td>
<td>22.79</td>
<td>67.90</td>
<td>22.66</td>
</tr>
<tr>
<td>4</td>
<td>30.98</td>
<td>27.15</td>
<td>21.49</td>
<td>24.60</td>
</tr>
<tr>
<td>5</td>
<td>51.26</td>
<td>16.90</td>
<td>53.33</td>
<td>15.75</td>
</tr>
</tbody>
</table>

| $N$ | 143 | 108 | 131 |

*Note: C = calculation, G = graphing*

### Summary.

The main cognitive load findings were that: a) the cognitive load reported for learning the pre-training materials was significantly lower than the cognitive load reported for learning the teaching materials, b) the cognitive load scores for Treatment 1 were significantly lower than for Treatment 2 and 3 for the first time each group received the teaching, c) the cognitive load scores for learning graphing was significantly higher than for learning calculation, and d) the cognitive load scores for Treatment 1 (first time of teaching) and Treatment 3 (second time of teaching) were not significantly different.
Performance measures

Research Question 2. Within the context of learning complex information involving graphs, what is the effect of pre-training on secondary science students’ overall learning, performance on different levels of question (high versus low order), and performance on different content questions (calculation versus graphing)?

In addition to reporting the difference between overall pre and post test scores this section will look at the difference between high and low order questions as categorised using SOLO, and the difference between scores on calculation and graphing questions between the pre and post tests.

Learning.

A univariate analysis of covariance (with PAT Mathematics, PAT Reading Comprehension, and spatial orientation as covariates) comparing the difference in performance between the pre and post test for participants in each treatment group found a significant main effect for treatment [Treatment: difference, $F(2,378) = 35.34, p < .001$, partial $\eta^2 = .16$]. Pairwise Comparisons using the Bonferroni method found that Treatment 1) had a significantly higher difference between pre and post test scores than Treatment 2 ($p < .001$) and Treatment 3 ($p < .001$). That is, pre-training had a positive effect on overall learning. See Table 16 for means for learning for Treatment 1, 2, and 3.
Table 16

Means and Standard deviations for the Performance measures for all treatment groups

<table>
<thead>
<tr>
<th></th>
<th>Treatment 1 (Pre-training)</th>
<th>Treatment 2</th>
<th>Treatment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Learning</td>
<td>5.38</td>
<td>3.00</td>
<td>3.51</td>
</tr>
<tr>
<td>Level of question</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High order</td>
<td>1.72</td>
<td>1.61</td>
<td>1.15</td>
</tr>
<tr>
<td>Low order</td>
<td>3.22</td>
<td>2.06</td>
<td>2.34</td>
</tr>
<tr>
<td>Context</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cal</td>
<td>2.48</td>
<td>1.96</td>
<td>1.67</td>
</tr>
<tr>
<td>Gra</td>
<td>2.84</td>
<td>2.06</td>
<td>1.82</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>133</td>
<td>113</td>
<td>132</td>
</tr>
</tbody>
</table>

Note. Cal=Calculation, Gra = Graphing

High versus low order questions.

A multivariate analysis of covariance (with PAT Mathematics, PAT Reading Comprehension, and spatial orientation as covariates) comparing the difference in performance by participants on high and low order questions between the pre and post test found a significant main effect for treatment (Treatment: Wilks λ, $F = 17.78, p < .001$, partial $\eta^2 = .07$). Tests of Between-Subjects effects found significant main effects for high order and low order questions [Treatment: high order questions, $F (2,378) = 19.29, p < .001$, partial $\eta^2 = .09$; low order questions, $F (2,378) = 15.06, p < .001$, partial $\eta^2 = .08$]. Pairwise Comparisons using the Bonferroni method found for high and low order questions that Treatment 1 had a significantly higher difference between pre and post test scores than Treatment 2 ($p < .001$) and Treatment 3 ($p < .001$). Pre-training had a positive effect on learning both high and low order questions. See Table 16 for means for high order and low order question differences for Treatment 1, 2, and 3. In order to put these results in context the following additional information is provided. The best
possible score for high order questions was 10.5 marks and for low order questions was 9.5 marks.

**Calculation versus graphing questions.**

A multivariate analysis of covariance (with PAT Mathematics, PAT Reading Comprehension, and spatial orientation as covariates) comparing the difference in performance by participants on calculation and graphing questions between the pre and post test found a significant main effect for treatment (Treatment: Wilks λ, $F = 18.36, p < .001$, partial $\eta^2 = .08$). Tests of Between-Subjects effects found significant main effects for calculation and graphing questions [Treatment: calculation questions, $F (2,378) = 13.73, p < .001$, partial $\eta^2 = .06$; graphing questions, $F (2,378) = 25.83, p < .001$, partial $\eta^2 = .12$]. Pairwise Comparisons using the Bonferroni method found for calculation and graphing questions that Treatment 1 had significantly higher difference scores between the pre and post tests scores than Treatment 2 ($p < .001$) and Treatment 3 ($p < .001$). As mentioned previously, graphing is a complex skill to learn, however, the pre-training group participants performed better on both graphing and calculation questions as evidenced by their larger difference scores between the pre and post test. See Table 16 for means for calculation and graphing question differences for Treatment 1, 2, and 3. In order to put these results in context the following additional information is provided. The best possible score for calculation questions was 8 marks and for graphing questions was 12 marks.

**Summary.**

The main findings related to learning in this study were that difference scores for Treatment 1 were significantly higher than Treatment 2 and 3 for overall learning, learning high and low order learning outcomes, and graphing and calculation learning outcomes.
Instructional Efficiency.

Research Question 3. Does pre-training increase instructional efficiency for secondary science students when learning complex information involving graphs?

This study used learning (difference between the pre and post test scores) and cognitive load measures (the average of the two online cognitive load measures for the first time of teaching and summative questions 1 and 2) to calculate instructional efficiency. Two analyses were employed to compare: a) the instructional efficiency of the first time all treatment groups received the teaching and b) the first time the pre-training group received the teaching with the second time Treatment 3 received the teaching.

Firstly, a univariate analysis of covariance (with PAT Mathematics, PAT Reading Comprehension, and spatial orientation as covariates) comparing the instructional efficiency of participants in each treatment group for the first time they received the teaching found a significant main effect for treatment [Treatment: efficiency, $F(2,381) = 31.22, p < .001$, partial $\eta^2 = .14$]. Pairwise Comparisons using the Bonferroni method found that Treatment 1 had a significantly higher instructional efficiency scores than Treatment 2 ($p < .001$) and Treatment 3 ($p < .001$). That is, pre-training increased the instructional efficiency of learning complex science information. See Table 17 for the means for the instructional efficiency scores for Treatments 1, 2, and 3. Instructional efficiency combines the performance and mental effort ratings to give an indication of the relative efficiency of different conditions (treatments). Typical ranges for efficiency scores would be between -1.0 and 1.0. A positive score (above 0) indicates that performance was higher based on the level of effort expended to achieve this performance i.e., more efficient learning. Hence, the greater the efficiency value the more efficient the experimental condition (lower effort for greater performance). A negative score indicates that the learning is not efficient and a score of 0 indicates that learning and mental effort are as expected for the performance.
Table 17

Means and Standard deviations for the Instructional Efficiency measures for all treatment groups

<table>
<thead>
<tr>
<th></th>
<th>Treatment 1 (Pre-training)</th>
<th>Treatment 2</th>
<th>Treatment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.43</td>
<td>1.24</td>
<td>0.06</td>
</tr>
<tr>
<td>N</td>
<td>142</td>
<td></td>
<td>108</td>
</tr>
</tbody>
</table>

Secondly, a univariate analysis of covariance (with PAT Mathematics, PAT Reading Comprehension, and spatial orientation as covariates) comparing the instructional efficiency scores of Treatment 1 (first time of teaching) and Treatment 3 (second time of teaching) found a significant main effect for treatment [Treatment: efficiency, $F(1,275) = 92.17$, $p < .001$, partial $\eta^2 = .41$]. Pairwise Comparisons using the Bonferroni method found that Treatment 1 had a significantly higher instructional efficiency scores for their first time of teaching than Treatment 3 ($p < .004$) for their second time of teaching. That is, using pre-training is more effective for increasing the instructional efficiency of learning compared to receiving teaching twice through the same information. See Table 18 for the means for the instructional efficiency scores for treatments 1 and 3

Table 18

Means and Standard deviations for the Instructional Efficiency measures for Treatment 1 and 3 comparing the first and second times of teaching respectively.

<table>
<thead>
<tr>
<th></th>
<th>Treatment 1 (Pre-training)</th>
<th>Treatment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.43</td>
<td>1.25</td>
</tr>
<tr>
<td>N</td>
<td>144</td>
<td></td>
</tr>
</tbody>
</table>
Summary.

The main findings for instructional efficiency were that: a) the instructional efficiency of Treatment 1 was significantly higher than Treatment 2 and 3, and b) the instructional efficiency of Treatment 1 (first time of teaching) was significantly higher than Treatment 3 (second time of teaching).

The influence of gender, decile, and ethnicity on dependent measures

Research Question 4. Are there differential effects on cognitive load, learning (difference scores), and instructional efficiency with respect to gender, decile, and ethnicity when using pre-training in a secondary science context involving graphs?

This section involves three separate analyses which focus on the influence of gender, decile, and ethnicity, respectively. For all three analyses, the four dependent measures are online cognitive load, summative cognitive load, difference between pre and post test scores (learning), and instructional efficiency, with PAT Mathematics, PAT Reading Comprehension, and spatial orientation as covariates. In contrast to the previous analyses of online cognitive load, the analyses within this section employed an average of the two separate online cognitive load measures (graphing and calculation) as a composite online cognitive measure

Gender.

A multivariate analysis of covariance comparing cognitive load, performance and instructional efficiency for males and females in each treatment group found a significant main effect for treatment and gender (Treatment: Wilks λ, F = 6.67, p < .001, partial η² = .14; Gender: Wilks λ, F = 4.47, p < .001, partial η² = .10). There was no significant interaction between treatment and gender. The outcomes of interest in these analyses are the main effect of gender and the interaction between gender and treatment. Previous analyses have already focused on assessing the main effects of treatment on these outcomes.
Tests of Between-Subjects effects found significant main effects for gender on online cognitive load, summative cognitive load questions 1 and 3, and difference scores [online cognitive load $F(1,355) = 10.51$, $p = .001$, partial $\eta^2 = .03$; summative cognitive load question 1 (effort to understand), $F(1,355) = 5.99$, $p = .015$, partial $\eta^2 = .02$; summative cognitive load question 3 (self-predicted performance), $F(1,355) = 10.38$, $p = .001$, partial $\eta^2 = .03$; difference (learning), $F(1,355) = 5.71$, $p = .017$, partial $\eta^2 = .02$].

Pairwise Comparisons using the Bonferroni method found that, for online and summative question 1 cognitive load scores, and difference scores males scored significantly lower than females ($p = .001$), ($p = .015$) and ($p = .017$) respectively. That is, females reported higher cognitive load scores for learning and understanding the information but their performance was also higher. For summative cognitive load question 3 (self-predicted performance) males scored significantly higher than females ($p = .001$). That is, males predicted higher performance on the post test. See Table 19 for the means for treatment and gender.
Table 19
Means and Standard deviations for Treatment and Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Treatment 1 (Pre-training)</th>
<th>Treatment 2</th>
<th>Treatment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Online cognitive load</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>35.75</td>
<td>23.67</td>
<td>33.71</td>
</tr>
<tr>
<td>Female</td>
<td>34.33</td>
<td>19.96</td>
<td>46.16</td>
</tr>
<tr>
<td>Summative question 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>34.58</td>
<td>22.12</td>
<td>36.80</td>
</tr>
<tr>
<td>Female</td>
<td>31.44</td>
<td>18.46</td>
<td>46.38</td>
</tr>
<tr>
<td>Summative question 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>39.77</td>
<td>23.48</td>
<td>37.95</td>
</tr>
<tr>
<td>Female</td>
<td>37.93</td>
<td>19.57</td>
<td>45.58</td>
</tr>
<tr>
<td>Summative question 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>64.39</td>
<td>24.04</td>
<td>71.70</td>
</tr>
<tr>
<td>Female</td>
<td>62.97</td>
<td>20.45</td>
<td>64.82</td>
</tr>
<tr>
<td>Difference (learning)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>33.35</td>
<td>29.66</td>
<td>20.87</td>
</tr>
<tr>
<td>Female</td>
<td>29.70</td>
<td>24.50</td>
<td>23.31</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>.23</td>
<td>1.21</td>
<td>.20</td>
</tr>
<tr>
<td>Female</td>
<td>.52</td>
<td>1.28</td>
<td>.01</td>
</tr>
<tr>
<td>N</td>
<td>129</td>
<td>100</td>
<td>126</td>
</tr>
</tbody>
</table>

Decile.

The categories described for this factor (designated school socio-economic index number) are 1 = low and 10 = high. Only one school from each decile was involved in this study except for decile 1 and 4 (2 schools involved). All six decile
categories are used in this and subsequent analyses. There was no attempt to collapse the schools into low, medium, and high categories. The rationale for not doing this was that in looking at the characteristics of the different schools in similar decile categories it was clear that important differences existed between these schools which suggested that collapsing schools into the three categories (e.g., high, medium and low) would be inappropriate.

A multivariate analysis of covariance comparing cognitive load, performance and instructional efficiency for different decile schools in each treatment group found a significant main effect for treatment and decile (Treatment: Wilks $\lambda$, $F = 6.39$, $p < .001$, partial $\eta^2 = .14$; Decile: Wilks $\lambda$, $F = 1.97$, $p < .001$, partial $\eta^2 = .05$). There was also significant interaction between the two factors (Treatment x Decile: Wilks $\lambda$, $F = 1.56$, $p = .002$, partial $\eta^2 = .05$). The outcomes of interest in these analyses are the main effect of decile and the interaction between decile and treatment. Previous analyses have already focused on assessing the main effects of treatment on these outcomes.

Tests of Between-Subjects effects found significant main effects for decile on online cognitive load and instructional efficiency [online cognitive load, $F (5,355) = 2.44$, $p = .034$, partial $\eta^2 = .04$; efficiency, $F (5,355) = 3.07$, $p = .010$, partial $\eta^2 = .02$]. Pairwise Comparisons using the Bonferroni method found that for decile, for online cognitive load scores decile 1 was significantly higher than decile 5 ($p = .001$), and for instructional efficiency scores decile 1 scored significantly lower than decile 9 ($p = .022$). That is, decile 1 schools: a) reported higher online cognitive load for understanding the information than the decile 5 school, and b) were less efficient in their learning compared to the decile 9 school. Tables 20 and 21 display the means for treatment and decile.

There was a significant treatment x decile interaction for summative cognitive load question 2 (effort to learn), [$F (10,355) = 1.99$, $p = .034$, partial $\eta^2 = .06$]; summative cognitive load question 5 (pace), [$F (10,355) = 1.94$, $p = .039$, partial $\eta^2 = .06$]; difference, [$F (10,355) = 2.19$, $p = .018$, partial $\eta^2 = .06$]; and efficiency, [$F (10,355) = 2.35$, $p = .011$, partial $\eta^2 = .07$]. These interactions are displayed visually in Figures 10 -13. When examining these, it appears that for the decile 2 school, the pattern of means across treatments is different to other deciles. That
is, Treatment 2 has lower cognitive load than Treatment 1 and higher learning and instructional efficiency scores than Treatment 1. Also, for the decile 2 school, the pre-training group had the highest cognitive load scores and the lowest learning and instructional efficiency scores compared to all other deciles. The combination of these different performance patterns across treatments for the decile 2 school in comparison to the other schools, could be responsible for the significant interaction. In addition, for summative question 2 in the decile 4 schools, scores for Treatment group 2 do not follow the same pattern as the other deciles except decile 2, i.e., their scores are lower than decile 5, 9, and 10. For summative question 5, the interaction appears to be caused by the pattern of performance within the decile 9 and 10 schools whose scores for Treatment 1 do not follow the same pattern as the other deciles. These schools report the scores furthest from 50 which is the target score for this question.
<table>
<thead>
<tr>
<th>Decile</th>
<th>Online cognitive load</th>
<th>Treatment (Pre-training)</th>
<th>Treatment 2</th>
<th>Treatment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>50.24</td>
<td>22.80</td>
<td>56.47</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>44.05</td>
<td>19.07</td>
<td>25.28</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>35.25</td>
<td>21.75</td>
<td>48.22</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>27.61</td>
<td>19.45</td>
<td>40.50</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>20.87</td>
<td>12.90</td>
<td>31.28</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>23.95</td>
<td>13.24</td>
<td>46.90</td>
</tr>
<tr>
<td></td>
<td>Summative question 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>42.82</td>
<td>20.42</td>
<td>29.41</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>31.88</td>
<td>22.19</td>
<td>46.00</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>29.31</td>
<td>22.73</td>
<td>38.07</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>21.04</td>
<td>13.02</td>
<td>35.30</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>29.09</td>
<td>20.31</td>
<td>51.92</td>
</tr>
<tr>
<td></td>
<td>Summative question 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>49.10</td>
<td>19.86</td>
<td>49.96</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>47.50</td>
<td>22.62</td>
<td>30.68</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>37.96</td>
<td>21.89</td>
<td>40.40</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>28.67</td>
<td>23.07</td>
<td>48.79</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>31.78</td>
<td>19.53</td>
<td>43.00</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>37.09</td>
<td>20.48</td>
<td>50.85</td>
</tr>
<tr>
<td></td>
<td>Summative question 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>58.28</td>
<td>14.82</td>
<td>57.81</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>47.10</td>
<td>14.91</td>
<td>50.00</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>47.44</td>
<td>15.92</td>
<td>56.61</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>46.67</td>
<td>17.91</td>
<td>55.23</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>51.04</td>
<td>16.21</td>
<td>45.11</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>59.73</td>
<td>12.32</td>
<td>53.50</td>
</tr>
<tr>
<td>Decile</td>
<td>Treatment 1 (Pre training)</td>
<td>Treatment 2</td>
<td>Treatment 3</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( M )</td>
<td>SD</td>
<td>( M )</td>
<td>SD</td>
</tr>
<tr>
<td>1</td>
<td>4.64</td>
<td>3.12</td>
<td>3.50</td>
<td>2.53</td>
</tr>
<tr>
<td>2</td>
<td>3.86</td>
<td>2.67</td>
<td>5.20</td>
<td>2.55</td>
</tr>
<tr>
<td>4</td>
<td>5.41</td>
<td>2.77</td>
<td>3.83</td>
<td>2.18</td>
</tr>
<tr>
<td>5</td>
<td>6.28</td>
<td>2.90</td>
<td>2.58</td>
<td>2.52</td>
</tr>
<tr>
<td>9</td>
<td>6.91</td>
<td>3.21</td>
<td>3.56</td>
<td>2.71</td>
</tr>
<tr>
<td>10</td>
<td>5.77</td>
<td>1.85</td>
<td>3.00</td>
<td>2.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Difference (learning)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(- .53 )</td>
<td>1.13</td>
<td>(- .86 )</td>
</tr>
<tr>
<td>2</td>
<td>(- .32 )</td>
<td>1.03</td>
<td>1.18</td>
</tr>
<tr>
<td>4</td>
<td>.51</td>
<td>.94</td>
<td>(- .39 )</td>
</tr>
<tr>
<td>5</td>
<td>.77</td>
<td>1.19</td>
<td>(- .26 )</td>
</tr>
<tr>
<td>9</td>
<td>1.64</td>
<td>.93</td>
<td>.97</td>
</tr>
<tr>
<td>10</td>
<td>.87</td>
<td>.86</td>
<td>(- .40 )</td>
</tr>
</tbody>
</table>

| N       | 129      | 100       | 126       |

Note.

\( N \) for Decile 1 = 69, Decile 2 = 60, Decile 4 = 80, Decile 5 = 41, Decile 9 = 63, and Decile 10 = 42
**Figure 10.** SPSS generated interaction plot for summative question 2 (effort to learn) for decile x treatment

**Figure 11.** SPSS generated interaction plot for summative question 5 (pace) for decile x treatment
Figure 12. SPSS generated interaction plot for difference (learning for treatment x decile)

Figure 13. SPSS generated interaction plot for efficiency for treatment x decile
The four main ethnicities included in these analyses are New Zealand/European, Māori, Pasifika, and Asian (which included Indian). Data for other ethnicities was collected but low numbers precluded their inclusion.

A multivariate analysis of covariance comparing cognitive load, performance and instructional efficiency for the four different ethnicities in each treatment group found a significant main effect for treatment and ethnicity (Treatment: Wilks λ, $F = 5.56, p < .001$, partial $\eta^2 = .12$; Ethnicity: Wilks λ, $F = 1.57, p = .04$, partial $\eta^2 = .04$). There was no significant interaction between treatment and ethnicity. The outcomes of interest in these analyses is the main effect of ethnicity and the interaction between ethnicity and treatment. Previous analyses have already focused on assessing the main effects of treatment on the outcomes.

Tests of Between-Subjects effects found significant main effects for ethnicity on online cognitive load, summative question 1, and efficiency [Ethnicity: online cognitive load $F (2,338) = 3.68, p = .012$, partial $\eta^2 = .03$; summative cognitive load question 1 (effort to understand) $F (2,338) = 3.04, p = .029$, partial $\eta^2 = .03$; efficiency $F (2,338) = 3.77, p = .011$, partial $\eta^2 = .03$]. Pairwise Comparisons using the Bonferroni method found that for online and summative question 1 cognitive load scores Pasifika scored significantly higher than Asian ($p = .015$) and ($p = .028$) respectively, and for efficiency scores Pasifika scored significantly lower than Asian ($p = .005$). That is, the significant ethnicity differences are associated with the cognitive load for understanding and for learning where Pasifika found this more demanding than Asian, and instructional efficiency where Asians were more efficient in their learning than Pasifika. See Table 22 for the means for treatment and ethnicity.
### Table 22

*Means and Standard deviations for Treatment and Ethnicity.*

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>N</th>
<th>Treatment (Pre-training)</th>
<th>Treatment 2</th>
<th>Treatment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>NZ/Eur</td>
<td>101</td>
<td>24.62</td>
<td>17.19</td>
<td>44.08</td>
</tr>
<tr>
<td>Māori</td>
<td>43</td>
<td>45.97</td>
<td>17.88</td>
<td>42.21</td>
</tr>
<tr>
<td>Pasifika</td>
<td>129</td>
<td>44.54</td>
<td>21.47</td>
<td>49.40</td>
</tr>
<tr>
<td>Asian</td>
<td>65</td>
<td>24.65</td>
<td>19.74</td>
<td>23.94</td>
</tr>
</tbody>
</table>

#### Online cognitive load

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ/Eur</td>
<td>101</td>
<td>24.34</td>
<td>17.25</td>
<td>45.38</td>
<td>22.55</td>
<td>38.50</td>
<td>21.08</td>
</tr>
<tr>
<td>Māori</td>
<td>43</td>
<td>42.00</td>
<td>17.64</td>
<td>44.58</td>
<td>13.12</td>
<td>46.38</td>
<td>21.61</td>
</tr>
<tr>
<td>Pasifika</td>
<td>129</td>
<td>41.88</td>
<td>19.25</td>
<td>51.27</td>
<td>24.15</td>
<td>51.02</td>
<td>21.30</td>
</tr>
<tr>
<td>Asian</td>
<td>65</td>
<td>22.95</td>
<td>17.06</td>
<td>28.58</td>
<td>17.77</td>
<td>32.29</td>
<td>17.01</td>
</tr>
</tbody>
</table>

#### Summative question 1

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ/Eur</td>
<td>101</td>
<td>32.48</td>
<td>19.31</td>
<td>45.08</td>
<td>19.70</td>
<td>42.95</td>
<td>22.15</td>
</tr>
<tr>
<td>Māori</td>
<td>43</td>
<td>46.22</td>
<td>18.65</td>
<td>40.00</td>
<td>17.11</td>
<td>52.54</td>
<td>23.03</td>
</tr>
<tr>
<td>Pasifika</td>
<td>129</td>
<td>46.49</td>
<td>20.08</td>
<td>48.44</td>
<td>17.47</td>
<td>50.12</td>
<td>17.42</td>
</tr>
<tr>
<td>Asian</td>
<td>65</td>
<td>30.80</td>
<td>20.91</td>
<td>33.29</td>
<td>15.32</td>
<td>32.85</td>
<td>19.72</td>
</tr>
</tbody>
</table>

#### Summative question 2

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ/Eur</td>
<td>101</td>
<td>5.26</td>
<td>2.69</td>
<td>2.75</td>
<td>2.47</td>
<td>3.73</td>
<td>3.15</td>
</tr>
<tr>
<td>Māori</td>
<td>43</td>
<td>4.73</td>
<td>3.17</td>
<td>3.41</td>
<td>2.66</td>
<td>2.42</td>
<td>2.21</td>
</tr>
<tr>
<td>Pasifika</td>
<td>129</td>
<td>5.49</td>
<td>3.21</td>
<td>4.19</td>
<td>2.50</td>
<td>2.13</td>
<td>2.67</td>
</tr>
<tr>
<td>Asian</td>
<td>65</td>
<td>6.00</td>
<td>3.08</td>
<td>4.33</td>
<td>2.57</td>
<td>3.98</td>
<td>2.61</td>
</tr>
</tbody>
</table>

#### Difference (learning)

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ/Eur</td>
<td>101</td>
<td>23.54</td>
<td>25.53</td>
<td>28.79</td>
<td>29.25</td>
<td>27.25</td>
<td>25.05</td>
</tr>
<tr>
<td>Māori</td>
<td>43</td>
<td>34.86</td>
<td>28.82</td>
<td>20.00</td>
<td>25.66</td>
<td>34.87</td>
<td>30.99</td>
</tr>
<tr>
<td>Pasifika</td>
<td>129</td>
<td>39.75</td>
<td>27.00</td>
<td>27.21</td>
<td>24.06</td>
<td>34.43</td>
<td>31.49</td>
</tr>
<tr>
<td>Asian</td>
<td>65</td>
<td>24.50</td>
<td>23.28</td>
<td>8.35</td>
<td>13.34</td>
<td>25.00</td>
<td>25.23</td>
</tr>
</tbody>
</table>

*Note: N is the same for all measures as is reported for online cognitive load*
Summary.

Main findings for gender were that males scored significantly lower on all cognitive load measures related to learning and understanding, and performance measures than females. Males scored higher on self-predicted performance scores. There was no interaction between treatment and gender. Decile findings indicate that decile 1 schools scored significantly lower than the decile 5 school for online cognitive load and significantly lower than the decile 9 school for instructional efficiency. There was a significant decile x treatment interaction. Main findings for ethnicity were related to cognitive load measures for learning and understanding, and efficiency. For cognitive load measures, Pasifika scored higher than Asian and for efficiency Asian scored higher than Pasifika. There was no interaction between treatment and ethnicity.
Chapter 6

Discussion

This chapter will consist of: a) discussion and interpretation of the results related to each research question separately, b) overall discussion of the theoretical and practical implications of the results, and, c) recommendations for future research.

Cognitive load theory and the CTML have independently investigated the presentation of materials to maximise understanding and learning, in many varied contexts with different instructional materials. This study is unique in that it investigated the use of the strategy of pre-training to teach content which included complex representations (graphs) in an ecologically valid setting, i.e., a secondary school science classroom. It involved a large number of participants in a teaching period of 30 minutes in which cognitive load during and after teaching was also measured.

The overall aim of this study was to investigate the effectiveness of the strategy of pre-training, to reduce the cognitive load related to the understanding and learning of complex information. The instructional context was the physics topic of motion (which includes graphs) and the participants were novices. Treatment 1 received pre-training (names and characteristics of the main concepts) and then teaching (full version of the information), Treatment 2 received the teaching once and Treatment 3 received the teaching twice.

Research Question 1

There are three related areas of interest in addressing this research question with respect to cognitive load measures. Firstly, were the pre-training materials less complex (lower intrinsic cognitive load) compared to the teaching materials, secondly, did pre-training reduce the cognitive load of learning the target information compared to the other two treatment groups, and lastly, was the cognitive load of learning about calculation lower than learning graphing skills?
As suggested in previous research, one possible explanation for the difficulty students have when learning complex science ideas is that the nature of information overloads working memory (Marcus, Cooper, & Sweller, 1996; Mayer, 1984; Mayer & Pilegardi, 2014; Pollock, Chandler, & Sweller, 2002). When working memory is overloaded this hinders further learning and therefore, in order to facilitate learning, a reduction in intrinsic cognitive load related to the target information is required in order to enhance the potential effectiveness of the instructional procedure. Previous studies investigating the isolated-elements and pre-training strategies were predicated on the notion that novices would be overloaded learning complex concepts and skills. These studies focused on a variety of domains ranging from algebraic calculations (Ayres, 2006), spreadsheet and mathematical skills concurrently (Clarke, Ayres, & Sweller, 2005), how mechanical systems work (Mayer, Mathias, & Wetzell, 2002), to apply new geological information in a simulation game (Mayer, Mautone, & Prothero, 2002) solving accounting problems (Blayney, Kalyuga, & Sweller, 2010), chemistry concepts using computer simulations (Lee, Plass, and Homer, 2006) and performing electrical tests (Pollock, Chandler & Sweller, 2002).

These previous studies have attributed a reduction in complexity of the target information to a reduction in intrinsic cognitive load. These studies used a variety of methods to reduce the complexity of the materials. Ayres (2006, 2013) used part tasks and worked examples respectively. Lee, Plass, and Homer (2006) reduced visual complexity by separating the display of information onto 2 screens. Mayer, Mathias, and Wetzell (2002) included main parts and their states pertaining to a mechanical system, whereas, Mayer, Mautone, and Prothero (2002) provided visuals of geological features students were required to identify in a simulation game. Finally, Pollock, Chandler, and Sweller (2002) only incuded the main steps in an electrical test, removing the reasons for the different procedural steps. In the present study, only the names and definitions of the main concepts were presented using a multimedia PowerPoint presentation to reduce the complexity of the physics ideas related to motion and graphing.

Two of the previous studies included indirect measures of the intrinsic cognitive load instructional materials within stage one. Pollock, Chandler, and Sweller (2002) calculated the number of interacting elements and reduced this in stage
one (equivalent to pre-training) and Ayres (2006, 2013) calculated the number of individual calculations to solve algebraic problems and reduced these in stage one. In contrast, in the present study, rather than assessing the number of interacting elements or mental manipulations embedded in the pre-instructional material, the complexity of the two sets of materials (pre-training and teaching) were compared using the participants’ cognitive load scores. The pre-training materials were found to be significantly less complex (lower intrinsic cognitive load) than the teaching materials. Therefore, this study was able to confirm using the analysis of cognitive load measures, the reduction in intrinsic cognitive load between the pre-training and teaching materials, a finding which has not been demonstrated previously.

Regarding the effect of pre-training on the cognitive load of learning the target information, the pre-training group reported significantly lower cognitive load and a small – medium effect size for the first time of teaching (learning) compared to the other two treatment groups. These results suggest that pre-training as a strategy reduced cognitive load and may have made it easier for novices to learn the target information, i.e., made the task of learning less mentally demanding. This is consistent with results for the isolated-elements strategy (which is similar to pre-training) found by Ayres (2006) for low ability students in a mathematical context, Ayres (2013) in a mathematical context but paired with practice on areas of high cognitive load, Clarke, Ayres, and Sweller (2005) for students learning spreadsheet skills and mathematical concepts, and Pollock, Chandler, and Sweller (2002) for electrical apprentices. These studies found that cognitive load measures showed significant main effects for the isolated-elements strategy which supported hypotheses that this strategy reduced the cognitive load of learning high element interactivity information. Other pre-training or isolated-elements studies (Blayney, Kalyuga, & Sweller, 2010; Lee, Plass, & Homer, 2006; Mayer, Mathias, & Wetzel, 2002; Mayer, Mautone, & Prothero, 2002) did not calculate cognitive load or estimate element interactivity to establish the reduced intrinsic cognitive load of the materials or the effectiveness of these strategies for reducing essential processing. However, they attribute the success of the strategy used to a reduction in intrinsic cognitive load.
However, it could be said that when considering the first time all groups received the teaching, that the pre-training group had already been exposed to the information albeit in a very concise form with reduced complexity. This could be the reason that they reported lower cognitive load and not that the pre-training reduced the cognitive load of learning. This issue of extended exposure to the materials was also identified by Mayer, Mathias and Wetzel (2002) in Experiments 1 and 2, and was addressed in Experiment 3 by including pre-training after the teaching. Their results indicated that pre-training after the teaching did not have the same effect on learning as pre-training before the teaching. In the present study Treatment 3 was used to counter the greater time exposure to the complex science ideas between Treatment 1 and 2, Treatment 3 received the teaching twice in contrast to Treatment 1 and 2 who received it once. The lack of significant difference between the cognitive load scores for Treatment 1 (1st time of teaching) and Treatment 3 (2nd time of teaching) indicate that even though Treatment 3 had 2 opportunities to learn the information their cognitive load scores are the same as Treatment 1 who only had one opportunity.

In all previous studies investigating the effectiveness of the isolated-elements and pre-training strategies the cognitive load measures relate to overall cognitive load for the learning or the task. In contrast, this study looked at the cognitive load of two different aspects of the learning, that of graphing and calculation. Findings indicate that learning graphing skills is a more mentally demanding task than learning to perform physics calculations. This further confirms findings within a New Zealand context which identified learning graphing skills as both complex and difficult (Hipkins, 2011). This is also consistent with researchers in the field of learning graphing which also describe this as a complex skill which students find difficult (Bowen & Roth, 1998; Culbertson & Powers, 1959; Mautone & Mayer, 2007; Shah & Hoeffner, 2002; Woolnough, 2000). However, these studies do not confirm the difficulty of learning graphing skills by analysing cognitive load measures in contrast to this study.

The next section which is related to Research Question 2, will explore if pre-training had a positive effect on learning including the learning of complex and less complex ideas, and calculation versus graphing ideas.
Research Question 2

Learning in this study was measured using a post test which contained high and low order questions (SOLO taxonomy) and questions related to the calculation and the graphing content of the target information presented in the teaching. High and low order questions are similar to transfer and retention questions (Mayer, Mathias, & Wetzel, 2002) and high and low element interactivity questions (Pollock, Chandler, & Sweller, 2002).

Regarding the learning (difference between pre and post test scores), the results indicate that the pre-training strategy was superior to receiving no pre-training (Treatment 2 and 3). The effect size was large (partial $\eta^2 = 0.15$). The difference between the three treatment groups does not appear to be due to a prior knowledge effect as the pre test mean for the pre-training group was numerically lower than Treatment 2 and 3 but the difference was not statistically significant. This result also cannot be attributed to the extended exposure of the pre-training treatment group to the information when compared to treatment 2, as Treatment 3 participants did not outperform the pre-training treatment group despite having two opportunities to learn the information. This positive impact of pre-training on learning within the present study is consistent with results of Ayres (2006, 2013) and Blayney, Kalyuga, and Sweller (2010) who focused on solving algebraic problems and accounting problems respectively; Mayer, Mathias, and Wetzel (2002) who tested students on information regarding mechanical systems, Mayer, Mautone, and Prothero, 2002 who assessed the number of problems students solved in a simulation game, and Pollock, Chandler, & Sweller, 2002 who focused on procedural knowledge and conceptual knowledge related to electrical tests. In all these related studies participants in the isolated-elements or pre-training treatment groups performed significantly better than those in other treatment groups. In contrast, Lee, Plass, and Homer (2006) who manipulated intrinsic cognitive load by splitting the information onto two screens to reduce complexity, found that the benefit to novices of reducing intrinsic cognitive load was offset by the increase in extraneous cognitive load of intergrating the information from two screens in order to understand the complex information.
Pollock, Chandler and Sweller (2002) attribute the success of the isolated-elements strategy on understanding and learning to reducing the number of elements to be assimilated thereby reducing the load on working memory initially. Therefore, in the initial stages of schema construction when learning complex information, only a manageable number of elements are required to be incorporated. In the present study there was a reduction in the number of interacting elements between the complex information (teaching) and the pre-training, therefore, this could also be the explanation for the success of the pre-training strategy in this context. Pollock, Chandler, and Sweller (2002) conducted the first study looking into the effects of the isolated-elements strategy on cognitive load and learning. The findings and rationale for the success of the strategy for all following isolated-elements studies except Lee, Plass, and Homer (2006) concur with this initial study. Pre-training studies situated in the CTML, similarly attribute the success of the strategy to reducing essential processing and building preliminary schemas supporting a two stage theory of mental model construction. That is, building a component model before building a causal model (Mayer, Mathias, & Wetzell, 2002). The explanations for the success of these two strategies are closely aligned and both are consistent with the findings of the present study as both lead to increased understanding and learning of the complex information.

Regarding high and low order questions, the pre-training group outperformed Treatment 2 and 3 for both high and low order questions. This is not consistent with Mayer, Mathias, and Wetzell (2002) or Pollock, Chandler, and Sweller (2002) as these studies only found a significant difference for high element interactivity (high order or transfer) questions. The present study was unique in testing for differences between the two content areas of calculation and graphing and found that the pre-training treatment group outperformed the participants in Treatment 2 and 3 for both calculation and graphing questions. The difference was significant for both content areas but the effect size for the difference was greater for graphing which is the more complex (difficult) skill (see, Bowen and Roth, 1998; Friel, Curcio, and Bright, 2001; Mautone and Mayer, 2007; and Shah and Hoeffner, 2002). The significant impact of pre-training on graphing outcomes within the present study, further confirm the findings of Mautone and Mayer (2007) that using cognitive aids (signalling and concrete graphic organisers which have similarities
to pre-training, as described in Chapter 3) enhance the understanding and learning of information presented in a graph.

In this study participants in Treatment 1 outperformed participants in Treatment 2 and 3 for high order questions indicating that their understanding was superior. Mautone and Mayer (2007) attribute this increased understanding on transfer or higher order outcomes to the influence of cognitive aids on the learning processes of organising and integrating. In this study increased understanding and learning could also be due to the positive influence of pre-training on the working memory processes of organising and integrating the new complex information.

Ayres (2006) found that reducing cognitive load does not automatically facilitate learning if germane resources were not directed to the task of learning the information. In this study the reduction of intrinsic cognitive load may have made it easier to understand the complex information and consequently the participants in the pre-training group invested more germane resources to the task of learning.

One way in which the current study is unique is its focus on the integrated learning and understanding of conceptual information and skills which occurred within a whole class instructional setting, that is, learning a series of interrelated concepts and skills within a science topic in a classroom situation. Other studies which focused on the impact of pre-training or the isolated-elements strategy used individualised learning i.e., problem-solving in mathematics (Ayres, 2006, 2013) and accounting (Blayney, Kalyuga, & Sweller, 2010), conducting electrical resistance tests (Pollock, Chandler, & Sweller, 2002), how a physical system works (Mayer, Mathias, & 2002), and problem-solving using a simulation game (Mayer, Mautone, & Prothero, 2002). Two exceptions to this pattern relating to the learning of integrated concepts and skills occur. In Experiment 3 of Pollock, Chandler, and Sweller (2002) focused on conceptual knowledge and skills related to electrical principles and Lee, Plass and Homer (2006) used science simulations to teach chemistry concepts, however the learning in both these studies was individual. Learning science concepts and skills within whole class instruction is more relevant to everyday teaching and learning science in high schools and is therefore more directly useful for application to teaching in a classroom situation.
Overall, results suggest that the pre-training strategy was superior for all types of learning for novices, especially when information is complex. The next section related to Research Question 3 will explore if pre-training had a positive effect on instructional efficiency.

**Research Question 3**

Instructional efficiency was measured using online and summative cognitive load measures and performance measures using the method outlined in Paas and van Merriënboer (1993) and using the same combination of measures (mental effort of learning and performance on a test) as Pollock, Chandler, and Sweller (2002) and Kester, Kirschner, and van Merriënboer (2006). These measures are most useful in situations when the aim is reducing the cognitive load during learning as is the case in this study (van Gog & Paas, 2008). Instructional efficiency can be used to measure the effects of different conditions on learning (van Gog & Paas, 2008). In this study, the effects of pre-training were compared to no pre-training. Efficient instructional conditions would be characterised by lower intrinsic cognitive load and increased performance, resulting in a higher instructional efficiency score which would indicate the higher quality of the learning process or acquired cognitive schemas (van Gog & Paas, 2008).

In this study the pre-training treatment group reported lower cognitive load when learning the complex information, and their performance was superior to the other two treatment groups. Therefore, when instructional efficiency was calculated the pre-training treatment group was found to be more efficient, as indicated by a positive efficiency score and a large effect size. This is consistent with previous findings using the isolated-elements strategy (Ayres, 2006, 2013; Pollock, Chandler, & Sweller, 2002) and the pre-training strategy (Clarke, Ayres, & Sweller, 2005). This suggests that in the present study the lower cognitive load experienced by the pre-training treatment group resulted in more efficient schema acquisition i.e., the pre-training group were more able to direct their mental resources toward learning the information presented. In addition, they were able to build preliminary schemas during the pre-training which they were consequently able to modify (add information to) during the teaching.
Clarke, Ayres, & Sweller (2005) without calculating instructional efficiency also reported that students who received training in spreadsheets before learning mathematical skills performed better than those who learned spreadsheet skills concurrently with learning mathematical skills. They attributed this to the students building preliminary schemas in using spreadsheets which helped them to chunk the information (which otherwise would have been several elements) when learning the mathematical information. In contrast, students in the concurrent group were not able to do this. Chunking therefore appeared to reduce the load on working memory. In another study focused on the impact of pre-training which did not directly measure/calculate instructional efficiency, the researcher (Mayer, Mathias, and Wetzel, 2002) suggest that pre-training helps build preliminary schemas which the researchers refer to as a component model (individual components). This model is seen to assists learners to build a full causal model of how a mechanical system works when all essential information for understanding this system is presented.

The second analysis looking at the instructional efficiency of the pre-training group (Treatment 1) for their first time of teaching and Treatment 3 for their second time of teaching found a significant difference with a very large effect size. This confirms the success of using the pre-training strategy paired with teaching compared to teaching the material twice, even though there was no significant difference between the cognitive load scores, the benefit to learning and instructional efficiency was considerable. Another benefit to consider is a reduction in teaching time, as the time it takes to use the strategy of pre-training and then teach complex information is less than the time it would take to teach the material twice. The current study extends our understanding of the impact of pre-training by further supporting the idea that its positive impact is not due to increased exposure to target concepts and information. Finally, no other studies have analysed the difference in instructional efficiency of treatments that are more equivalent in time.

The next section will discuss the influence of gender, decile, and ethnicity on cognitive load (online and summative), learning (difference between pre and post test scores), and instructional efficiency as related to Research Question 4.
Research Question 4

The effects of pre-training on the dependent variables of cognitive load, learning, and instructional efficiency have been reported in many studies. This study is unique in that it also looked at the influence of gender, decile, and ethnicity on cognitive load, learning, and instructional efficiency. As reported in the results chapter each dependent variable was examined separately. Within this study, the decile rating of a school represents the designated school socio-economic index, numbers can range from 1 (low) to 10 (high). With respect to ethnicity, the four main ethnicities employed within this study are New Zealand/European, Māori, Pasifika, and Asian (which includes Indian). This is consistent with reporting of NCEA results by the New Zealand Qualifications Authority and National Educational Monitoring Project reports for science.

Gender findings.

There was a significant main effect for gender on the online cognitive load, summative cognitive load question 1 and 3, and learning (difference between pre and post test) scores. These differences were in the same direction, i.e., females were significantly higher than males. Females reported higher online cognitive load which is less desirable in a study aiming to lower cognitive load, however, females outperformed males in one of the main performance measures (learning). This indicates that the pre-training strategy had a beneficial effect on females when learning physics information which is contrary to most other gender and achievement studies in science (see Chapter 3, subsection Gender and Achievement in science). It is unclear why females reported higher online cognitive load, however it is possible that females and males actually self-report cognitive load (mental effort expended) differently. Alternatively, the males could have found the information easier or on the other hand could just not have put in the same effort to learn the information as females i.e., the females were more motivated to learn the information. Higher levels of motivation for females has been suggested as a possible reason for the improvement in science achievement by girls in the last 15 years (Amelink, 2009; Calvin, Fernandes, Smith, Visscher, & Dreary, 2010; Dreary, Strand, Smith, & Fernandes, 2007). There is a need to
integrate motivational influences into models of learning which assess impact of cognitive load.

Mayer (2014a) suggests that there is a need to incorporate motivational constructs into the CTML (Mayer, 2014a). A limited number of studies have investigated the influence of motivation on learning within a cognitive load framework but findings at this stage are not convincing and results indicate that not all motivational aids are effective in promoting learning (Mayer, 2014b). Plass, Heidig, Hayward, Homer, and Um (2013) found that by using appealing graphics (colour and shape) in a multimedia presentation about immunization induced positive emotions that increased comprehension but not transfer of information (meaningful learning). Magner, Schwonke, Aleven, Popescu, and Renkl (2013) used decorative (irrelevant) illustrations in a multimedia presentation about geometry and found that this had a positive effect on interest but a negative effect on learning, particularly for novices. Paas, Tuovinen, van Merriënboer, and Darabi (2005) looked at using efficiency measures to calculate learner motivation. Their findings indicate that there are issues (conceptual and methodological) with using the efficiency measure to calculate motivation and in addition, there are other factors which impact on mental effort expended which are not accounted for in using this measure e.g., fatigue and anxiety. In the present study measures of motivation were not assessed, therefore it is not possible to conclude with any level of certainty what part motivation might have played in any learning and understanding differences which occurred.

There was no significant gender main effect for the instructional efficiency scores. That is, male and female performance was the same on this measure. With respect to the lack of effect of gender on the instructional efficiency measure, this result is likely due to the differential influence of gender on the two measures which are employed to calculate instructional efficiency. As indicated earlier, females had higher cognitive load and higher learning scores than males. As instructional efficiency is calculated using the difference between performance and cognitive load scores and higher efficiency scores will occur when low cognitive load is paired with high levels of learning, the differences in the results for males (low cognitive load and low learning) and females (high cognitive load and high levels of learning) would have generated similar efficiency scores. In addition,
there was no significant interaction between gender and treatment for both of these measures indicating that the pre-training strategy did not influence males and females differently on any of the outcome measures.

**Decile findings.**

Findings indicate that there were significant main effects for online cognitive load and instructional efficiency. These differences focused on comparisons between decile 1 schools and other higher decile schools. There is documented evidence that lower decile schools underachieve compared to higher decile schools in Auckland (Hawk, Hill, 1996; McNaughton, Phillips, & MacDonald, 2003). Therefore the results for learning and instructional efficiency were not unexpected.

There was also a significant interaction between decile and treatment for online and summative cognitive load, and instructional efficiency, indicating that decile had a differential effect on participants across treatment groups from different schools for these measures. This implies that pre-training results would be different depending on the decile of the school. However, in this case most of the interactions involve the decile 2 school. This suggests that a possible explanation could be that at the decile 2 school the performance was atypical for the decile rating or that small sample sizes impacted on these results and therefore, the interaction was a school effect and not related to decile. Without further investigation it is not possible to come to a definite conclusion, however, these findings related to decile are unique and could provide an interesting area for further research.

**Ethnicity findings.**

There was no significant main effect for ethnicity on summative cognitive load questions 2, 3, 4, and 5 and on learning. Given the research evidence for the underachievement of Māori and Pasifika students in science (see Chapter 3, subsection Māori and Pasifika achievement in science) this result is particularly pleasing especially when compared to the results for National exam performance in Achievement Standard 90940 (see Table 6), which includes the same content. The results for National exam performance on this achievement standard indicates
that Māori and Pasifika students are more than twice as likely than New Zealand/European and almost three times that of Asian students to be awarded a not achieved for this achievement standard. This study contained large numbers of students in each ethnicity from many different schools, particularly in the Pasifika group, so the results do not lack power. The significant main effect for instructional efficiency was between Pasifika, and Asian students. Even though the effect size for this difference was small (0.03), this result holds promise as Māori and Pasifika were performing with the same instructional efficiency as New Zealand/European students and Māori were not significantly different from Asians. When examining all science data for achievement in NCEA this result (equivalence between Māori, Pasifika and New Zealand/European and Māori and Asians) is non-existent. In addition, the lack of interaction on any of the measures indicates that pre-training is a robust strategy that does not differentially influence students from different ethnic groups.

**Spatial ability**

An unexpected finding within this study was the lack of gender differences in spatial ability scores. Previous research has consistently found males score higher on spatial ability tests (in particular the factors of spatial orientation and spatial visualization) (Fennema & Sherman, 1977; Vandenberg & Kuse 1978; Linn & Peterson, 1985; Casey, Nuttal, Pezaris & Benbow, 1995; Peters, Laeng, Latham, Jackson, Zaiyouna & Richardson, 1995). There is more recent evidence that gender differences in spatial ability are decreasing or in some cases non-existent particularly for spatial visualisation (Voyer, Voyer & Bryden, 1995; Casey, Nuttal & Pezaris, 2005). The results of this study are therefore consistent with recent trends indicating no difference between males and females on spatial ability tests for visualization, but inconsistent with recent results for spatial orientation.

Overall the participants in this study scored similarly for the Cube comparison test (spatial orientation) and higher for the Paper folding test (spatial visualization) when compared to previous studies using these same tests. These two spatial ability tests are suggested for use with normal adolescents or adults from ninth grade or higher (Ekstrom, French & Harman, 1976) which would be equivalent to Year 10 in New Zealand.
When comparing the results for participants in previous studies who were in higher grades (equivalent to Year 12 and 13 in New Zealand) with participants in this study it was found that their scores were similar. Mean scores for the Cube Comparison test (spatial orientation) for previous studies was 23.5 for males and 21.5 for females compared to this study, males had a mean scores of 22.80 and females 22.71. For the spatial vizualisation test males in previous studies had a mean of 11.5 and females 10.4 whereas in this study the males had a mean of 19.74 and females 21.61. These results support the validity of using these tests with this population.

**Theoretical and educational significance**

The results gained in this study can be applied to both theory and practice. All the practice and teaching materials were developed specifically for this study by the researcher and are therefore unique to this study. The treatment results for the effect of pre-training implies that the researcher was successful in developing these instruments from the guidelines and research examples provided by Mayer, Mathias, and Wetzell (2002), Mayer, Mautone, and Prothero (2002), and Pollock, Chandler, and Sweller (2002) and that these were successfully transferred and administered in a classroom setting. As a result, there are implications for cognitive load theory, the CTML and for the use of this strategy in classroom situations. These will be outlined in the following sections.

**Theoretical Implications.**

The learning of complex information by novices is often problematic. If all the elements needed for the material to be understood and learned are required to be processed simultaneously, then this may be an impossible task for learners who have no prior knowledge. The strategies of isolated-elements (cognitive load theory) and pre-training (cognitive theory of multimedia learning) offer a potential solution. Most previous pre-training and isolated-elements studies have been trialled using small samples sizes, e.g., $N = 33$ for Experiment 2, (Mayer, Mathias, & Wetzell, 2002) and $N = 18$, Experiment 3 (Pollock, Chandler, & Sweller 2002) respectively, involving undergraduate or technical students, where testing has
been individual and in a laboratory setting. However, the isolated-elements strategy has been trialled with higher numbers of undergraduate students, e.g., $N = 505$ (Blayney, Ayres, & Sweller, 2010). Other isolated-elements studies have involved mathematics high schools students, however, the largest sample size was 54 participants (Ayres, 2013). Therefore, results from previous studies and related implications have not always been directly transferable to a classroom setting. This study expanded the settings where pre-training has been found to have an impact, as it was trialled with large numbers of secondary students ($N = 606$) in a variety of different science classrooms.

The previous studies involving pre-training did not employ measures of cognitive load, whereas, the current study employed two measures of cognitive load (online and summative). When novices learn complex information which is high in element interactivity and therefore high in intrinsic cognitive load, the potential for cognitive overload is also high. If this occurs then further learning will be compromised. In these situations it is important to reduce element interactivity (intrinsic cognitive load) to reduce the likelihood of cognitive overload to enable learning in these contexts. This is consistent with both cognitive load theory - isolated-elements effect, (Pollock, Chandler, & Sweller, 2002; Sweller, 1994) and the CTML - pre-training effect, for managing essential processing (Mayer, 2005b; Mayer & Pilegard, 2014).

This study was able to confirm that pre-training reduced intrinsic cognitive load for students who were novices, but was also able to confirm that pre-training materials were lower in intrinsic cognitive load than the teaching materials. This has not been established in previous studies.

The present study has also provided valuable data on measuring cognitive load in an ecologically valid setting and will therefore contribute to the body of research on measuring cognitive load. All cognitive load instruments were developed specifically for use in this study and were modifications of several previously employed instruments. The use of a 1-100 point scale and using a combination of online and summative cognitive load measures were unique. Including practice of the cognitive load questions (online and summative) in a separate session to the data collection phase was not only unique to this study but it proved to be
successful. These instruments were able to be understood and used by participants and supplied useful data on mental effort expended during and after the teaching. In addition, cognitive load information related to confidence, frustration and timing was collected. These provided valuable information on the teaching and learning that occurred and could be used in future research.

Given that understanding graphic representations is a complex process (Friel, Curcio, & Bright, 2001; Gerber, Boulton-Lewis, & Bruce, 1995; Hipkins, 2011), several strategies have been suggested in previous research into the development of graphing skills, to assist student learning of this complex skill. The results of some studies have suggested that reducing the cognitive load of learning graphing skills by using computers has had a positive effect on learning these complex skills (Adams & Shrum, 1990; Berg & Phillips, 1994; Brasell, 1987; Mokros & Tinker, 1987). The results of this study into the effects of pre-training to reduce the cognitive load of learning these complex skills is consistent with these previous studies. Hence this strategy could be added as an alternative way of learning complex graphing skills through the reduction of cognitive load.

In summary, participants in the pre-training group outperformed the other two treatment groups further providing evidence that pre-training is effective for the teaching of complex information to novices by reducing cognitive load.

**Educational implications.**

**Overall implications.** The positive impact of pre-training in an ecologically valid setting supports the use of pre-training by teachers in science classrooms. This study has revealed some of the complexities of implementing pre-training within a classroom setting and will enable the writing of guidelines to assist teachers. As part of the agreement with the schools who took part in this study, these guidelines will be developed so that science teachers can confidently use this strategy for the teaching and learning of complex information in science. This could ultimately be used by other schools who were not involved in the study. As previous studies were all performed in laboratory settings the procedures employed in previous studies do not lend themselves easily to the development of specific guidelines for actual use of this strategy in science classrooms. This
makes the research findings from this study directly relevant for teachers and readily accessible for them to incorporate into their teaching to enhance student learning and understanding of conceptually difficult information. In this study pre-training reduced the level of frustration experienced by students when learning complex material. This may be especially useful when students are tired and cognitive overload is a threat to further learning, which often occurs at the end of a school term or at the end of a school year. The summative cognitive load measures of frustration, confidence in learning and pace of the lesson could also easily be used by teachers in their own classrooms to give quick but meaningful feedback at the end of a lesson. One teacher requested a summary of data collected in the introductory phase (practice cognitive load questions), as this was related to their teaching and they indicated that this provided valuable feedback. Lastly, the strategy of pre-training could now also be included as one that is beneficial when teaching graphing skills which is an important component of learning science in New Zealand (Hipkins, 2011).

**Specific implications.** In addition to overall implications of employing the strategy of pre-training in this study, it has also been found to be beneficial in more specific circumstances, with respect to gender and ethnicity and in certain contexts. Firstly, in terms of gender implications, males often outperform females in science especially in physics, but in this study females outperformed males in the pre-training group. This suggests that this strategy may be useful to improve the achievement of females in physics topics. Secondly, in terms of ethnicity, in New Zealand, statistics reported on achievement in national exams and international science studies indicate that Māori and Pasifika students underachieve compared to New Zealand/ European and Asian. However, this was not the case in this study indicating that pre-training may be useful for improving the achievement of Māori and Pasifika students in science. The New Zealand government has implemented long-term strategies and plans for improving the achievement of Māori and Pasifika students in all curriculum areas of secondary schooling including science. This strategy could work alongside existing initiatives and strategies as it is relatively easy to implement and does not require any additional resources that would not be present in a typical New Zealand secondary school. Given the documented underachievement this could be one way to reduce the achievement gap between Māori and Pasifika students and other
ethnic groups in science. Consequently this strategy could be employed successfully in schools with high percentages of Māori and/ or Pasifika populations to improve science achievement.

Lastly, as pre-training had a positive effect on complex learning which included learning graphing skills and the answering of high order questions, it may be a useful strategy to employ when teaching material related to excellence questions in national science exams. Improving the success of students in answering excellence questions in national exams would have a positive effect on overall achievement. In addition, as this strategy worked equally well for all ethnicities the positive impact could span all ethnic groups which employed this strategy. With respect to instructional efficiency, in this study pre-training had a positive effect on improving the quality of schema acquisition with less time exposure to the complex ideas and as such would be useful for teachers to employ especially when time restraints for learning exist as it would allow more time for elaboration (Kalyuga, 2009).

Future Research

In this section future research will be discussed as related to a) target content areas, b) specific methodological changes, c) comparison with other methods, d) different target populations, e) participant characteristics, and f) improving overall performance.

Target content areas.

This study is set in the context of physics and may be applicable to other areas of physics, however, further study would be needed to support the value of pre-training for use in other topics and disciplines of science e.g., chemistry or geology. As pre-training has been found to be beneficial for the understanding and learning of complex information within the current study, it may be a useful strategy to employ when learning other complex science topics in the senior school. These could be addressed using pre-training before actual teaching to reduce the cognitive load of learning this complex information. Three topics that come to mind as appropriate in the senior school that would offer the same
complexity and richness of words and their meanings in biology are osmosis, photosynthesis and respiration. These are level 2 NCEA topics (Year 12). This study investigated the effect of pre-training in a science context, future research could ascertain if this strategy was beneficial for use in other curriculum subjects, e.g., history or geography.

**Specific methodological changes.**

There are a variety of potential methodological changes and evaluations that could be pursued in order to further enhance our knowledge of the impact of pre-training and extend the results of the current study. These include looking at: a) characteristics of teachers, b) an integration of focus topics into overall curriculum, c) including immediate and long-term measures of learning and understanding, d) assessing the impact of standardisation of procedures, and, e) including measures of interactivity of pre-training materials.

In terms of teacher characteristics, there is a possibility that the results obtained within the present study are due to the specific characteristics of the teacher (researcher) employed across all classrooms within the present study. There are at least two ways in which this may have influenced the results. One could be that the teacher (researcher) employed within this study is a very experienced science teacher and pre-training may require this level of skill to have impact. In addition, not being taught by one’s regular teacher may have had an impact on the participants’ motivation to learn and to take the lessons seriously. Therefore, it would be important to replicate the current results using the student’s own teacher implementing the strategy of pre-training as part of their science programme.

Assessing the impact of integration of the target lessons into the overall curriculum would be a useful avenue for future research, i.e., in a topic that is part of their school scheme and at the time that they would be learning this in their science class. In this study the teaching was out of context with regards to their science learning and the New Zealand Curriculum level. Therefore, there were no consequences for not trying (putting in effort) to learn the information, however, if it was part of their science course this could be different.

Expanding the measures employed to assess learning and understanding when employing a pre-training strategy would expand our knowledge of the impact of
this strategy. Mayer, Mathias and Wetzell (2002) suggest that it would be useful to test the pre-training treatment group immediately after the pre-training. This would assess whether the participants have indeed learned what the main components are before the relationships between these components are presented in the main teaching session. This may be able to be incorporated into future research but for the current study, time constraints within the class period did not allow for this. In addition, it could be beneficial to establish through further investigation if this strategy improves learning long-term, that is by conducting immediate (straight after the teaching) and delayed testing (several weeks or months later).

Research in classroom settings will always present many variables that are difficult to control, i.e., time of day, day of the week, time during the school term. These factors could potentially affect student behaviour and motivation. In future studies it may be possible to be more consistent with some of these factors. Natural attrition in this study was partially due to the split days in which the initial introductory phase and the teaching phase took place. Consequently, it would be an advantage to run the whole study on one day so students are present for both phases and they are less likely to forget what they are expected to do in the teaching phase as the time lapse will not be more than four hours.

This study ensured complexity of the information by using physics concepts which have been found to be more conceptually difficult than other science disciplines, incorporating graphs in the content, increasing the curriculum level of the information for the participants involved, and using novices. In future research, it could be beneficial to calculate the number of interacting elements to give a measure of complexity and to ensure that the potential for overload is a reality for the students concerned. If this was paired with the addition of measuring the cognitive capacity of participants (as suggested earlier and also discussed later), one could also assess whether high cognitive capacity correlated with high performance and/or interacted with the nature of intrinsic load embedded within the pre-training materials.

**Comparison with other methods.**

Pre-training is only one way suggested in cognitive load research to help students learn complex information. Kirschner, Paas and Kirschner (2009) looked at the
benefits of group work and collaboration to help students learn complex information with reported success. Future research could compare both these strategies and their effect on learning, instructional efficiency and cognitive load.

**Different target populations.**

This study was trialled with Year 9 and 10 students but it would be an advantage to also trial this with other Year levels particularly senior students (Year 12 and 13) where the material is significantly more complex.

In addition, given the promise provided by this strategy for targeting the underachievement of Māori and Pasifika in science, further research in this area is warranted. Future investigation could include larger numbers of high and low decile schools to increase the numbers of Māori and Pasifika students for comparison with in particular, New Zealand/ European and Asian students, not only in Auckland but across New Zealand. In New Zealand there is a separate Māori Curriculum and future research could investigate if this strategy worked equally well in Māori immersion schools (where the medium is the Māori language).

Another investigation which would assist Māori and Pasika students but not exclusively, would be to involve using the strategy of pre-training on a regular basis throughout a topic in the teaching of difficult concepts. This would ascertain if this strategy could be used more than once during extended learning to improve the understanding and learning of a whole topic or achievement standard as compared to a 30 minute teaching session as was used in this study, this could have a positive impact on Māori and Pasifika achievement in NCEA science (national exams). Another suggestion which could contribute to better understanding the impact of this strategy would be to include interviews after the teaching as part of the data collection, this could establish what aspects of pre-training Māori and Pasifika students found beneficial regarding their understanding and learning, this could further improve the strategy particularly for these two ethnic groups.

**Participant characteristics.**

In terms of student characteristics, this study was conducted with participants who were in the middle range of ability at the schools involved, further studies would be
needed to look at whether this strategy was also effective for low and high ability students. Another student characteristic of interest would be to measure the cognitive capacity of participants to ascertain if high cognitive capacity correlated with high performance as found by Seufert, Schutze, and Brünken (2009). Measuring the cognitive capacity could also assist in determining if participants are indeed overloaded as recommended by de Jong (2010) and Moreno (2010).

A common finding in many studies involving spatial ability is a gender difference where males outperform females (Fennema & Sherman, 1977; Vandenb & Kuse 1978; Linn & Petersen, 1985; Casey, Nuttall, Pezaris, & Benbow, 1995). Given the unexpected results in this study for spatial ability with respect to gender it would be recommended that these tests (spatial orientation and visualization) be employed in further research into the effectiveness of pre-training, especially in the context of teaching about graphs within the physics. This would be a useful next step in terms of research in order to further investigate if the results of the present study were a unique occurrence or if these results (no gender difference in spatial ability) can be replicated in the same context. In addition, given the link between learning physics and spatial ability (Mayer, 2001) it would be important to further investigate if there is a correlation between spatial ability and performance in physics as suggested by Kozhevnikov, Motes, and Hegarty (2007) and Larkin (1982).

**Improving overall performance.**

Despite finding robust significant effects in terms of significance and effect sizes in this study, the level of performance overall was relatively low. This could have been due to the fact that the topic was physics and was at a higher curriculum level, for the participants involved. Another explanation could be that the teaching session was relatively short (25 minutes). However, further research would be warranted to investigate lifting this relatively low level of performance. There is clearly value in using a pre-training approach for preparing students to study, learn, and understand complex concepts, but it is clear that it is just the beginning of a process that should be supplemented by more instruction and practice with the target concepts.
Conclusion

In summary, this study proved beneficial for the understanding and learning of complex information by novices in science in a classroom setting. Findings suggest that pre-training enabled the novices to construct preliminary schemas when the main concepts and their definitions were presented, and then they were able to build on these schemas when all the complex information was presented. The data also suggests that this occurred because the pre-training strategy reduced the intrinsic cognitive load of learning complex ideas, therefore also reducing the potential for cognitive overload. In addition, this strategy proved beneficial for increasing the instructional efficiency of learning for all learners.

The findings of the influence of gender, decile, and ethnicity on cognitive load, learning, and instructional efficiency are unique to this study. This strategy had a positive effect on females and their learning in science despite reporting slightly higher cognitive load. Regarding decile, the cognitive load induced by the complex information was higher and the performance was lower for the lowest decile school and learning was correspondingly less efficient for these schools. There were no differences between all of the four main ethnic groups in New Zealand on all measures except instructional efficiency, where Māori and Pasifika students were only lower than Asian students. Despite the influences of gender, decile and ethnicity the pre-training strategy was superior to receiving teaching once (as would be expected in a classroom situation) or receiving two opportunities to learn the information, on all measures (online and summative cognitive load, learning and instructional efficiency). Effect sizes for the main effects of treatment overall were medium to large. In this study pre-training reduced the cognitive load of learning complex information and increased understanding and learning in a science classroom, indicating that this strategy works in an ecologically valid setting when students are learning curriculum appropriate information. In conclusion, this study has provided powerful evidence that pre-training can be used to reduce cognitive load and successfully enhance learning for all students in a New Zealand science classroom.
Most cognitive load theory and CTML studies into the strategy of pre-training or isolated-elements have involved undergraduate University students as participants, often in laboratory situations. By contrast, this study involved over 600 students from secondary schools learning in their own classrooms. Evaluating instructional treatments is not an easy task when carried out in authentic settings as there are many factors that singly or in combination can influence the impact of the instructional treatments, which in turn makes it difficult to isolate and test these factors (Kirschner, Ayres, & Chandler, 2001). Nevertheless, Kirschner, Ayres, & Chandler (2001) and Mayer (2005, 2014a) suggest that it is important for cognitive load theory research to be carried out in authentic settings despite the difficulties that this presents. This chapter outlines the challenges of doing research in school classrooms, where classroom dynamics and student interactions create a complex set of factors that are not easily controlled compared to involving undergraduate students in controlled laboratory settings where participants are tested individually. It will describe the processes and techniques used in this study to address these challenges and justification for their use. It could prove useful for researchers intending to carry out research in schools who may not be familiar with educational settings and the nature of the research challenges in these contexts to incorporate the processes outlined here in their planning and execution of research in schools. These suggestions if implemented may result in a more positive outcome in relation to approaching schools, be a more beneficial and useful experience for school personnel involved, and result in a richer set of data from a research perspective. There are specific challenges in firstly, making contact with schools and getting permission, and secondly, working with participants in schools. This chapter will look at these two aspects separately.
The challenges of making contact and getting consent

This section will outline the challenges a researcher faces: a) before making contact i.e., getting Ethics approval to work with human participants, b) when making contact with schools and explaining the research, c) gaining permission (consent) from the Principal and staff involved to work with their students.

**The challenges of getting Ethics approval for research in schools.**

Most undergraduate students tested individually would only have to consent to take part in the research and as such gaining Ethics approval is a relatively straightforward procedure. However, the process of getting Human Participants Ethics approval is a more complicated process when doing research in schools. The approach I adopted in getting Ethics Committee approval for my research required taking into consideration two major issues, firstly, involving the whole class in the research as part of their regular science classes and secondly, not requiring parental consent. This approach was aimed at reducing the workload on participating teachers and streamlining the process to reduce the impact on teaching and learning of the New Zealand Curriculum (Ministry of Education, 2007). The next few paragraphs will look at how these two issues were resolved in my study.

If not all students in the class consent to participate in a study then this puts extra strain on teachers at the school, as those who do not consent to participate have to be catered for in another classroom. This option can be particularly appealing for some students and one that they will readily choose as it means not having to do regular class work. The responsibility for the organisation of this will fall on the teachers of the classes concerned. In order to reduce the teacher’s workload and retain their goodwill for future research, it was decided to focus the consent process on asking students to consent to the use of their data in the research. In this way, the whole class was involved in the research as part of their regular science instruction, but they had a choice as to whether the information and data collected during the research and instructional activity could be used as part of the study. This meant less disruption to class routines and not having to provide alternative activities and extra staff to look after students in another setting if they
chose not to be involved. I was able to employ this course of action in my study primarily because the focus of the study involved instructionally relevant and useful content for the students involved.

The second issue concerned the age of the students and the need to obtain parental consent. My study involved classes of Year 9 and 10 participants (ages 13 and 14) and as such the University Ethics process requires parental consent. However, this would have added yet another task for the teachers, i.e., the teacher would have been responsible for the collection of these consent sheets as I was not present in the school, this can be quite an onerous task for teachers (as it typically involves chasing students for the return of the sheets) and one which speaking from experience teachers do not enjoy. So I looked at a way of eliminating this task and decided that if the students could consent then that would solve this issue. However, in order for the Ethics committee to consider this solution it was necessary for me to ensure that the participants were being taught using methods similar to those used by their regular teachers, that the instructions would be easily understood by the participants, that the instructional material was not controversial or beyond the capabilities of the participants and that they would not be harmed or made to feel uncomfortable in any way. Taking this course of action however, did involve providing additional supporting evidence - letter from the Principal of the two schools who potentially could be involved in the study stating that they were supportive of both using whole classes in the study and also of giving participants the right to consent even though they were under 16 years of age as what was being proposed was fully understandable by the students and methods traditionally used by teachers at their schools would be employed in the study. This required preliminary visits to the Principals to fully explain the study and the issues surrounding the consent process (i.e., parental consent and whole class participation) to gain their support and agreement to write the letters detailed above.

An added benefit of not having to obtain parental consent and having the participants consent in class during the initial preparation phase and hand the sheets in at the end of the session was that the return rate was relatively high. In this study about 95% of the consent sheets were returned by the students on the day of the initial phase instruction.
Another challenge regarding gaining Ethics Committee approval for my study was the composition of the Committee. When applying for Ethics approval I found that the Committee reviewing the Ethics application did not include members who were knowledgeable about the structure and nature of schooling within New Zealand. For example, I was including Year 9 or 10 participants so it was necessary to include an explanation accompanying the Ethics proposal that Year 9 and 10 students in New Zealand schools are working from the same curriculum level so despite the slight age difference the difficulty level is the same. To further assist the Ethics Committee in making an informed decision regarding this application given their limited knowledge of the New Zealand Curriculum (Ministry of Education, 2007), I provided evidence that I had aligned what I was teaching the participants to this curriculum, so in my study material presented was aligned to Science Level 6, Physical World strand and the Physical inquiry and physics concepts objectives, and also from the criteria for Achievement Standard 90940. Aligning this study to the New Zealand Curriculum (Ministry of Education, 2007) assisted with justifying the benefit of this research to the participants, the teachers, the HOD science and the Principal of the schools and likely contributed to their agreeing to be involved.

Lastly, as the participants were in Year 9 and 10, I wrote the Participant Information Sheet in student-friendly language in contrast to the other sheets. This ensured that it was easy for the participants to understand what I was asking them to be involved in and what they were required to do. However, I still found it necessary to explain to the participants that these sheets were a formal part of my research study and given that the consent form was long and used some specialised language to also explain the content of the form.

The challenges of making initial contact with schools and explaining the research.

The first challenge when working with outside agencies rather than with students from your own institution is about making contact with the school. This challenge is complicated by the fact that the primary purpose of schooling is teaching and learning [in New Zealand - the New Zealand Curriculum (Ministry of Education, 2007)] and involvement in research will nearly always impact on this primary goal.
So on making contact, it is important to acknowledge the involvement of their time, and potential benefits, even at this preliminary stage. In order for this contact and consent process to be successful it is important to work through a hierarchical process. The specific order I followed was the Principal, Head of Department (in this study, the Head of Department (HOD) Science), teachers in the Department and then the teachers of the classes I would be working with. This order is consistent with the hierarchical levels of authority which operate in a school. In order for the school to be receptive to a request for working with their students I undertook the procedures described in the subsequent paragraphs.

Firstly, I approached schools in which the University of Auckland or myself had an existing positive relationship with the school, the Principal or teachers at the school. So in this study, I included two University of Auckland Partnership schools (these schools regularly accept large groups of preservice teachers on practicum and their staff are regularly in contact with the University), one school where the researcher had previously taught and so had a close working relationship with the science staff and the Principal, three schools where the researcher was on a Science Teacher's Association committee with the HOD science (these teachers had indicated after informal discussions that they would be interested in being involved. After these informal conversations the teachers spoke to their respective Principals to gauge their support for this study before I approached the school. These relationships proved to be very important as only one school declined to be involved.

At the meetings with school personnel I explained the research purposes and procedures. A challenge that a researcher faces when dealing with personnel in schools is being able to clearly and concisely explain the research. This is due to: a) research is not a school’s primary function, b) they are not familiar with research procedures and terminology, and c) students are even less likely to be able to understand what they will be involved in if it cannot be explained simply and concisely. In order to overcome this challenge, I developed a one page visual summary of my research proposal and made this accessible during all meetings (See Appendix S). This concise summary omitted the unfamiliar research terms and presented the procedure as a flowchart that was easy for the Principal and teachers to understand. For the students, I broke up the procedure into
manageable chunks which were presented on separate slides of the initial phase PowerPoint.

I met with all but two of the Principals of the schools concerned when making initial contact, the remaining two were keen to be involved based on the recommendation from the HOD Science and did not deem a visit as a necessary part of the process. This was important in terms of making a good first impression with the school and this could be the reason I achieved positive outcomes to my requests. Thirdly, after gaining consent from the Principal I organised a time to meet the HOD Science to explain the study and what this would involve for the teachers and their students. At this meeting I also asked about their student’s prior knowledge with respect to my study, i.e., had they studied the topic of motion at Year 9 or 10 at their school, as I needed students who had no prior knowledge of this topic. I also enquired about the technology available to teachers as I was planning to use a PowerPoint presentation in the teaching.

Lastly, if the HOD Science agreed to participate I asked about suitable times to conduct the research at their school and a time for me to come and explain the research to all the science staff. It was important to negotiate suitable times to minimize impact on junior classes. Losing valuable time to be involved in research could impact negatively on preparation for common tests or exams. The purpose of meeting all the science staff was twofold; firstly, it was after this meeting that the HOD Science called for volunteers and so explaining the study in person and giving the teachers the opportunity to ask questions meant that they were making an informed decision to participate. This element of choice in their involvement in my study meant that they were enthusiastic and fully supportive of the research. Secondly, the reason for addressing the science staff was so that they were all fully informed regarding the research and this meant that when I came to do the study the other science teachers, who were not directly involved recognised me and knew why I was there. On a number of occasions teachers who were not involved in the research offered their assistance and showed an interest in how the data collection was going. Maintaining a positive relationship with the whole department is important for potential future research at the school. I then set up a separate meeting with the teachers who had chosen to be involved to finalise details and this gave them another opportunity to ask questions.
The challenges of gaining consent to work with students.

Once the proposal was clearly explained an issue that schools also had to consider was if the research was within the capabilities of the students concerned and if the material was at an appropriate level. To assist the schools in assessing these issues I put together a pack containing all the sheets that students would be required to complete during the study, including the Ethics sheets, so that the school could make an informed decision regarding the suitability of their students for this research. In addition, as time constraints put pressure on teachers to complete courses of study with their classes (as prescribed in their school curriculum), it was important that my request reflected the actual time that students and teachers would be required to participate in the study, i.e., the amount of class time that they will lose from their regular class work, how many classes I needed access to and also any school data that I required as this would also incur time for teachers to access this information and in turn make it accessible to me. If schools were discouraged due to excess time taken by the research over and above that which was first indicated this could mean that the potential for further research may have been lost.

Another issue which schools have to consider is the possible impact on student achievement of loss of class time. This was mitigated by explaining the potential benefits of my study to both the students and teachers at the school. For example, in this study I explained that the student participants were going to be introduced to material that was part of the criteria for the national science external exams (NCEA) which they would be taking the following year so learning it during the study could give them an advantage when required to learn it again. In addition, I offered the teachers electronic copies of my teaching materials and a professional development session in which I explained the strategy of pre-training and the results of the study, and also offered to produce a set of guidelines that they could follow in order to implement this strategy in their own classrooms when teaching complex material should the intervention prove to be successful.

As teachers are busy people they also need to consider the impact of being involved in the research on their own time, so it was necessary to clearly indicate the role of the teachers during the study, e.g., in this study teachers were required
to annotate a script during the teaching. It was important that I signal this role as it is usual practice for teachers of the classes involved in practicum visits by lecturers from the University is to be released from their duties during such a visit. By outlining their role I ensured that teachers were prepared for their tasks and were not surprised on the day(s) of the study. If they were expecting to be able to do their own work during the time of the study, this may have created ill feeling towards this study, which could have potentially had an impact on the teacher’s willingness to participate in future research. I did encounter this on one occasion in my study when a new teacher who had replaced a teacher in the initial meetings had not been informed of their role in this study. This teacher had brought marking to do during the lesson and was understandably disappointed that they could not complete this task.

Lastly, from a school’s perspective when giving consent it is important that they have confidence in the researcher being involved in teaching their students, in this study I was able to outline my extensive teaching experience and show evidence of my Teacher Registration.

The challenges of conducting research

A classroom is an unpredictable, dynamic, and challenging place where around 30 students with different needs, abilities, learning styles, cultures, life experiences, emotional states, motivation for the subject, and level of maturity have to co-exist. In junior secondary school, students do not usually choose to take core curriculum subjects (science is a core curriculum subject in New Zealand schools) and they do not choose their classmates or their teacher. In addition to these differences and the issue of choice, there are also factors involving the classroom climate, classroom student-student interactions, school ethos, and discipline strategies at the school, which can all have an impact on classroom behaviour, motivation to learn, and the ability to provide effective opportunities for learning. Bearing all this in mind I had to develop a rigorous plan and be very well organised when teaching students whom I did not know and did not have a relationship with. The following paragraphs detail my approach and the strategies I used in an attempt to control as many features of this complex environment as possible in order to collect meaningful data.
Being organised.

Given that doing large scale research in a number of different schools consecutively is organisationally challenging, I set up an exercise book with pages dedicated to each school to assist in managing this task. Schools are highly organised places and it was important that I also showed this high level of organisation. I took this exercise book to every meeting including the visits to the classes when the research was being conducted. In this book I detailed all meeting times and places pertaining to the school, questions to ask the HOD, questions to ask the teachers and added notes including documented evidence for what was discussed during the visits and what happened during the study which was unique to the class or school. I also used it as the place where I stored all contact details, timetables, maps of the school and specific information about each school, e.g., school gate codes or where to meet on the first visit. This enabled me to meet all my obligations to the school and assisted in the data collection stage of the study running very smoothly despite all the things that potentially could have gone wrong e.g. if I had turned up on the wrong day or time or if I was late.

Working in schools requires a high level of planning and organisation which is consistent with what schools expect from their teaching staff. So in this study, in order to demonstrate this same level of organisation, I made up a check-list of all the equipment and resources I would need for each stage and had this laminated so I could refer to it before each visit to the school, thus ensuring that I did not forget any equipment or materials. In addition to this, I also set up a file box with extra equipment for participants e.g. sets of 30 calculators, 30 rulers, 30 OHT pens and spare ballpoint pens. A copy of the contents of the box was taped onto the lid, so that when materials were returned they could be accounted for against the list to ensure I maintained a full set of equipment. I also put a copy of all the paper resources and the PowerPoint presentations to be used for both the initial and data collection phases on a USB stick should I need more sheets or the PowerPoint presentations became corrupted. In addition, I took an extension lead in case there was not a spare power socket at the front of the room (I needed one for my laptop) and a laser pointer in case the screen was higher up than was comfortable when pointing to important ideas [especially important for English for
Speakers of Other Languages (ESOL) participants]. Another aspect of how I maintained an organised approach is that at the end of the data collection phase I instructed the participants to put all their sheets together in a pile and wait for them to be collected as a set. I asked the classroom teacher to collect these after first checking that the names on the top sheet were correct before taking them off the participants. This was to ensure that no data were lost due to participants using pseudonyms, as even though the participants data were anonymous for reporting purposes, I needed to know the names of the participants for matching school assessment data and data collected in the research.

**Being consistent.**

As the classroom is such a dynamic place it was important that my teaching was as consistent as possible, the tasks were easily understood, and able to be undertaken by the participants without assistance as it was not possible to explain or answer questions in the data collection phase due to time constraints and for consistency purposes. In order to better understand the dynamics of implementing my study within a classroom, I undertook a pilot study, as it was here that I was able to trial all the materials, equipment, and procedures in an authentic setting with real students. In addition, I also used the pilot study as an opportunity to get feedback from teachers and participants. I asked for written feedback from participants on the sheets they completed but also met with a group of participants at a separate time to discuss the whole procedure. I found that in a group situation the participants felt less threatened and were more willing to share ideas, also participants can be reminded of ideas by listening to others and therefore the feedback was richer and more valuable. The participants were very honest and said that they felt valued when their opinions were asked for and they mentioned potential changes that I had not considered e.g., removing a chart from the worksheet to practice SI (Standard International) units in physics as students felt that this was redundant. The information gained from the pilots was invaluable in terms of modifying the procedures in order to balance my need to collect usable and useful data and the school’s focus on student needs and their learning.
Maximising the data collection.

In order to facilitate the collection of useful data and create the least disruption to the school schedule and student learning, I separated the two visits to the classes when conducting the research, the first one was an introduction to the study and the second the data collection. The reasons for this were so that during the introduction session I could explain the study and the participants could practice tasks that they would need to complete in the data collection session. This would allow them to gain confidence in completing these sheets and it also gave them the opportunity to ask questions to ensure their understanding of what they would be required to do in the data collection phase so that the data would be as accurate as possible. In addition the introduction session served as a way of introducing myself and in working with the students I hoped that they would feel more comfortable with me as their teacher in the data collection phase. From the school’s perspective it was also an opportunity to assess my suitability to teach their students in the data collection phase.

Another way that I facilitated data collection was to create a comfortable environment for the students. One method I employed to do this was that I asked the teachers of the classes involved to inform their classes that they were going to participate in this study and when this would occur. This ensured that my arrival for the initial phase visit was not a surprise for the participants when they were expecting to be involved in their regular science lesson science. An example of this occurred within my study, when one teacher forgot to inform the class and so this class was expecting to dissect flowers and many students had made an effort to bring flowers from home with them and they were understandably disappointed that they were not able to complete this science activity. In addition to avoiding disappointing the students, forewarning them when this is happening will be more likely to help them to perceive this as just another part of their science programme and it also indicates to them that their teacher supports this research. In the schools that followed these suggestions the participants were more cooperative, interested and took the study more seriously. I also asked the teacher to introduce me to the class at the start of the initial phase. This reinforced to the participants that their teacher supported this research and when this happened I found that overall it resulted in a greater percentage of students consenting to participate.
Recognising their contribution.

With the large number of participants involved in my study it was not possible to give them a monetary reward. Because I was not a teacher at the school and not all classes at the particular Year level in each school were involved it was not possible to give course credit. These are both ways that are used when involving undergraduate students from Universities in laboratory studies. However, it was important to recognise the contribution of both the teachers and the participants involved in my study and maintain goodwill so that future research at the school could be possible. Therefore in my study the participants were given a small chocolate frog after the study was complete, and the teachers were also given a small gift and a Thank You card.

In the next sections specific issues will be discussed pertaining to the challenges of an increased time commitment and timetabling the visits to schools for research purposes (collection of data), using technology in schools, classroom management, and teaching ESOL students.

The challenges of increased time commitments and timetabling visits to collect data.

The challenge of visiting schools to conduct research imposes both time and mileage costs. A related issue is the limited number of opportunities a week to access these students,.

I found that organising visits to schools for my research required a lot of forward planning, particularly when involving participants from secondary schools where a rigid timetabling structure exists. Most schools were happy to provide the timetables for the classes involved, indicate when in the term and which periods would be suitable and then leave the scheduling to me. This reduced the workload on the teachers and helped retain their goodwill. However, this took a considerable amount of time, e.g., becoming familiar with these timetables, planning the schedule of visits, allowing time for emails confirming the schedule with the teachers and the relevant HOD concerned. Another challenge I had to address was taking into consideration the time allocated to class periods at the
schools concerned. I found that there was a predominance of the use of 60
minute periods in secondary schools but that this was not universal. I encountered
examples of 45, 50, 55 minutes periods, and also examples where the length of
periods changed throughout the day. In my study I needed at least 60 minutes as
one of the treatment groups required 55 minutes and so a 50 or 55 minute period
would not be enough time. From my experience in a secondary school, classes
that directly follow other classes and are not after lunchtime or interval do not have
an allowance for student travelling time and so a 55 minute period is realistically
50 or 45 minutes of actual teaching time, by the time students get to class, get
themselves organised and are ready to start learning. I encountered an additional
challenge in two schools that were keen to be involved but these schools had only
50 minutes periods and so this required me to come up with a creative solution. I
met with the HOD Science and in consultation we decided to add form period or
tutor time in one school and Sustained Silent Reading (SSR) time in the other
school to the total time allocated to the data collection phase of my study. This
also influenced the time required for me to organise the visits to these schools and
it also impacted on the periods that I could use for the study, so in these two
schools the researcher had to return on three separate days to involve the three
classes as these extra time periods only occurred once a day.

Related to this timetabling challenge is the fact that Year 9 and 10 science periods
only occur three or four times a week. As three classes from each school were
involved in my study I only once managed to complete within one day each phase
of the study (initial preparation or data collection). Typically I had to make six
separate visits.

Another factor which I had to take into consideration which contributed to
increased mileage and time commitments is that when involving large numbers of
classes, teachers, and students in research it is inevitable that despite careful
planning some school visits had to be rescheduled. For example I had to
reschedule two class visits, one due to a fire emergency which occurred just prior
to starting the research and another one where the connection between the TV
and the data show was faulty and could not be repaired before the class was due
to start.
Another challenge due to the differences in purpose and needs between the researcher and the school manifests in the process of getting to the classroom which may take longer than anticipated e.g., in my study I found that visitor car parks were often full and parking was only available outside the school grounds. Further, signing in at the office also takes extra time if office staff also have to attend to student and staff requests. In one of the schools involved in this study office staff were preoccupied with a crisis when pre-booked buses had not arrived to take students to a competition. As a result I had to wait for this to be dealt with as the office staff were responsible for signing visitors in.

Another good example of the tensions that exist between the school’s needs and the needs of doing research in schools is played out in the scheduling of research opportunities within the calendar year. In addition to the time pressures of completing courses of study with their classes, teachers need to make maximum use of the times when their students are keen and eager to learn which is typically at the beginning of the day, the beginning of the week, and towards the beginning of a school term. These are also the optimal conditions for the data collection that was the basis of this study. In order to avoid the tension between teaching need and research need impacting negatively on the study I accepted times that suited the teachers and their classes. I did observe a difference in participant attitude when they were tired which happened when I involved participants at the end of the term or the end of the day. There were also higher numbers of participants absent from classes which occurred later in the day or late in the term. However, I also found that in secondary schools when I did have the opportunity to do my study in period one, that in schools where large numbers of students rely on school buses or parental transport it was equally disruptive if traffic issues occurred and participants were late to the beginning of period one. These were other important factors to consider when scheduling data collection within schools.

Lastly, a final challenge which can contribute to increased time and mileage when doing research in schools is natural attrition. When I planned this study I included more participants and classes than I needed to, to counter the loss of participants, i.e., participants present for the pre test but not the post test or vice versa, and invalid data provided by participants. Also, the variable class sizes I encountered at the different schools meant that the number of participants in the study was
reduced for the same investment of time when small classes were involved. I needed to sample a range of different schools and I did not want to exclude schools where class sizes were small as they had already agreed to be involved and had invested time in preliminary meetings with me.

**Using technology in schools.**

A difference between doing research in a laboratory and in an authentic school setting is that in a laboratory all the technology can be set up in advance and can be quite sophisticated whereas in a school setting a researcher is limited to the technology available e.g. the data show system, both for practical reasons but also to ensure that the intervention is transferable to multiple classroom settings. To avoid asking the teacher to give up time to set up the equipment I used the school’s data show system but supplied my own laptop. This also meant that I was not plugging a foreign USB stick into their laptops, a practice which schools actively discourage. It did mean however, that I had to take time to check when meeting with the HOD science that the laptop which I proposed to bring was compatible with their system. Another challenge when using your own devices is access to the internet. Schools have internet access for their teachers but getting access as a visitor is a very involved process in most schools and one whose responsibility would fall, once again on the teachers involved in the research or the relevant HOD. It was important that I avoid this so as not to increase the time commitment I required from the teachers or the HOD science, for this reason I embedded a video in the PowerPoint presentation instead of using a hyperlink to the internet.

I found that careful planning and trialling is necessary but that this does not necessarily reduce the stress that there could be technological issues. In a laboratory situation there would likely be technical staff available if issues arose but in a school this is not readily available, especially for visitors. So in preparation for this study I trialled all presentations on technology similar to what was used in the study in a classroom situation before the Pilot study. I included a total practice run of both phases using all equipment planned including laser pointers, embedded videos and demonstrations. Despite all my careful planning I did have a technology incident in this study aside from the one already mentioned,
as the TV monitor did not work in one of the scheduled classroom and so the class was relocated to another classroom.

**The challenges of managing the class.**

When adult undergraduate students are tested individually the issue of classroom management does not arise. However, in a classroom setting this can be a major challenge if teaching is involved in the research and learning is the performance outcome. One of the main factors in preventative classroom management is building a relationship with the class. However, in research situations it is impossible to have this relationship at the beginning of the research and this puts the researcher at a distinct disadvantage. In my study I faced this with every class as I had not met them before and therefore no relationship existed. Another compounding factor was that my study was not part of the participants curriculum course for that year and therefore was not part of their school assessment, so taking part in this research and learning the content was additional to their course of study and therefore could easily have been deemed not important to these participants. This could have meant that they did not put any effort into learning the material presented, they could have misbehaved and in doing so distracted other participants. For this reason, to encourage them to see the importance of putting in the effort required to learn the material I was careful to point out the benefits of taking part in the study to the participants. In this study I told them that they would be introduced to information they would need in Year 11 and learning it now would make it easier when they needed to learn it for NCEA.

From my teaching experience I know that the misbehaviour of even a few participants can have a negative effect on the learning of the whole class, for this reason it was important for me to try to reduce the likelihood of this occurring. Therefore it was very important in the initial phase for me to make a good first impression and start building a relationship with the class immediately. In this study I did this by introducing myself, briefly indicating why I was there, letting the participants know that they had been chosen for this study and thanking them in advance for their involvement. Ensuring that the participants felt comfortable with me as their teacher was also an important factor in managing the class. One strategy that I used which can assist with this aspect is addressing the classes
using their class name (usually their form class teacher’s initials, e.g., 10Pi). This may seem trivial but it is a convention that students are used to, that is, teachers referring to them collectively by their form class name. It is likely that this assisted them in perceiving me as just another teacher at the school.

Another factor that can contribute to increased misbehaviour is if participants do not understand the material or what they are required to do. For this reason, I know that the format and presentation of teaching resources is an important consideration when managing a class as students are less likely to misbehave to mask not understanding the task or the instructions. I therefore ensured that all the materials were student friendly, well organised, professionally formatted, and that the instructions and explanation were very clear (both in verbal and written form) and that the language level was appropriate for their Year level. In this study I also used colour and boldface to highlight important points, I didn't use capital letters for Headings and Titles (as they find these difficult to read, especially ESOL students), I kept words per PowerPoint slide to a minimum and included relevant visuals to complement the text. These strategies assisted in improving clarity of instruction and reflected for the participants the effort I had made to ensure that the material was appropriate and designed specifically for them.

I also know that a teacher’s response is integral to the successful management of student behavior. For this reason it was important for me to display confidence, approachability and encourage the participants to ask questions. So in this study, I used "I" messages or "thanks" in instructions, e.g., "I" need you to write your name at the top of the pre test or "Write your name at the top of the post test, thanks" to portray the expectation that the participants would comply.

The classroom teacher was present for this study and therefore it is reasonable to expect the classroom teacher to assist in dealing with misbehaviour. However, if the classroom teacher cannot manage the class then this makes doing research in classrooms a really challenging task. There were two classes where this situation arose in my study. Some of the behaviours that arose were participants talking in the pre and post test, getting up and walking around during class, difficulty settling the class at the beginning and during transitions and participants refusing to cooperate and complete written responses. In all instances when I had reason to
interact with these participants my verbal requests and interactions were delivered in a calm, assertive voice and without showing emotion. When participants were out of their seats I directed them back to their seats and indicated what they were supposed to be doing. With regards to talking during the tests I reminded them that it was important for this study that they wrote their own answers and not the answers of anyone else, I then moved to stand in near proximity to these participants in the hope that my presence would assist them in making the right decision to comply. Another strategy I used was to ask if they needed help. When students were slow to settle I waited for quiet and then explained that this lesson takes an hour and said “I don’t want to go into your lunch time”.

A final challenge when doing research in schools is that classrooms are very diverse and it is possible to encounter participants with learning difficulties. These participants have the potential to severely disrupt classes, especially if they do not have a teacher aide with them. In this study a particularly disruptive ADHD participant in the data collection phase (who was not present in the initial preparation phase) was causing some concern and so the classroom teacher went to sit next to him, this calmed him down and the lesson went smoothly after this. If this was not effective then I may have asked the teacher to remove this participant. It may also have been possible to have alternate arrangements made for these participants on the days when the research is taking place e.g., working with their teacher aide in the library.

The challenges of teaching ESOL students.

In a diverse classroom setting it is common to encounter ESOL participants present in any class and this is another example of managing the diversity in classrooms when doing research in authentic settings. Any learning situation is more difficult for ESOL students and a research study is no exception. Most of the classes involved in this study included ESOL students. I ensured that I used complementary visual and verbal instructions so ESOL participants had two opportunities to comprehend. I was also careful to pronounce words clearly, read all questions and instructions on sheets to the participants and held up the sheet I was referring to so they could easily identify it visually and did not have to rely on just my verbal instructions. I also rephrased questions that were asked during the
teaching so that these participants also had two opportunities to comprehend the questions and pointed to important words on the PowerPoint presentation as they were spoken to assist them in identifying and learning these words. When designing the teaching materials I also ensured that the visuals used in these materials were relevant to the text so they assisted ESOL learners to understand the ideas presented using visuals also and not just relying on the text. Finally, the last PowerPoint presentation slide said "Thank you" in many different languages so ESOL participants felt included in my show of appreciation for their involvement.

Working with participants in schools is both challenging and rewarding, but requires careful forward planning and a high level of organisation to accommodate the needs and purposes of both researchers and schools. This chapter has offered insights into the challenge of working in authentic school settings with early adolescents and has provided examples of how I addressed these challenges when conducting research in schools in New Zealand.
## Appendices

### Table of Contents

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>Ethics sheets - Study One</td>
<td>224</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Study One : Graphing study questionnaire</td>
<td>230</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Study One : Teacher responses to Question 4 and 7</td>
<td>234</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Ethics sheets - Study Two</td>
<td>239</td>
</tr>
<tr>
<td>Appendix E</td>
<td>Initial phase PowerPoint presentation</td>
<td>251</td>
</tr>
<tr>
<td>Appendix F</td>
<td>Practice cognitive load questions</td>
<td>259</td>
</tr>
<tr>
<td>Appendix G</td>
<td>Pretest</td>
<td>261</td>
</tr>
<tr>
<td>Appendix H</td>
<td>Pretraining PowerPoint presentation</td>
<td>263</td>
</tr>
<tr>
<td>Appendix I</td>
<td>Teaching PowerPoint presentation</td>
<td>269</td>
</tr>
<tr>
<td>Appendix J</td>
<td>Student worksheet</td>
<td>279</td>
</tr>
<tr>
<td>Appendix K</td>
<td>Online cognitive load question</td>
<td>281</td>
</tr>
<tr>
<td>Appendix L</td>
<td>Summative cognitive load questionnaire</td>
<td>282</td>
</tr>
<tr>
<td>Appendix M</td>
<td>Monitoring of participant behaviour sheet</td>
<td>283</td>
</tr>
<tr>
<td>Appendix N</td>
<td>Post test</td>
<td>284</td>
</tr>
<tr>
<td>Appendix O</td>
<td>Spatial ability : Cube Comparison Test S2</td>
<td>286</td>
</tr>
<tr>
<td>Appendix P</td>
<td>Spatial ability : Paper Folding Test VZ3</td>
<td>288</td>
</tr>
<tr>
<td>Appendix Q</td>
<td>Spatial ability PowerPoint presentation</td>
<td>292</td>
</tr>
<tr>
<td>Appendix R</td>
<td>Spatial ability Answer sheet</td>
<td>299</td>
</tr>
<tr>
<td>Appendix S</td>
<td>Overview of Thesis</td>
<td>300</td>
</tr>
</tbody>
</table>
Appendix A  Ethics sheets Study one

School of Science, Mathematics and Technology Education
Faculty of Education

PARTICIPANT INFORMATION SHEET
Principal/ Board of Trustees

Project Title: Graphing in Science Years 9-11
Researcher: Carolyn Haslam

To (Principal's name)

My name is Carolyn Haslam and I am both a staff member and PhD student at The University of Auckland. I am employed by The University of Auckland at the Faculty of Education and I am involved in Teacher Education in the curriculum area of Science. My aim for this Preliminary study is to determine the reported time spent teaching and practising graphing skills in science classrooms from Years 9 – 11. Graphing skills refers to both construction and interpretation of graphs in science.

Graphing is an important skill in science which is used extensively by scientists and it is an important aspect of developing scientific literacy. It is my aim to find out if these skills are being taught and practised in science classrooms from Years 9-11 to the same extent.

I request your permission to approach your Head of Department: Science to complete a questionnaire which will take approximately 20 minutes to complete. The questionnaire asks the Head of Department: Science to estimate the number of periods that they spend teaching and practicing graphing skills in science classrooms from Years 9 to 11, how these align with curriculum knowledge strands, how easy or difficult students find graphing instruction, whether students benefit from graphing instruction, and if so, how do they benefit. In addition, you will be asked to identify which Achievement standards within your Year 11 core science course you are offering this year. If you consent to this request the collated data from 50 schools in Auckland and Northland will be made available to your school so that your science department can compare data with other schools, regardless of the decision to participate or not made by your Head of Department: Science.

The Head of Department: Science decision to participate is voluntary and they will in no way be obligated to complete the questionnaire based on your agreement to let me approach them. I would like your assurance that participation or non participation by the Head of Department: Science will not affect his/her standing in the school. If they consent to complete the questionnaire they can choose to withdraw their data up until the 31st December 2011. In addition, if you choose to allow your school to participate in this research, you can withdraw this approval as well as any data collected within your school at any time up until the 31st May 2012.

The collated data will not be reported in ways that identify your school or Head of Department: Science

If you agree to allow me to approach your Head of Science, please sign and return the Consent Form enclosed to me in the self addressed envelope.
The completed questionnaires and consent forms will be kept separately for a period of 6
years in a secure storage space at The University of Auckland. After this period they will be
shredded.
My contact details should you wish to discuss this further are email
c.haslam@auckland.ac.nz, telephone 6238899 extn. 83918, The Faculty of Education, The
University of Auckland, Private Bag 92601, Symonds St, Auckland 1150.
My main supervisor, Richard Hamilton can be contacted on rj.hamilton@auckland.ac.nz,
telephone 3737599 extn 85619, The Faculty of Education, The University of Auckland,
Private Bag 92601, Symonds St, Auckland 1150.
My Head of Department, Gregor Lomas can be contacted on email g.lomas@auckland.ac.nz,
telephone 6238899 extn. 48517, The Faculty of Education, The University of Auckland,
Private Bag 92601, Symonds St, Auckland 1150.
For any queries regarding ethical concerns you may contact The Chair, The University of
Auckland Human Participants Ethics Committee, The University of Auckland, Office of the
Vice Chancellor, Private Bag 92019, Auckland 1142. Telephone 093737599 extn. 83711

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE
ON.....FOR (3) YEARS, Reference Number ..2011../...438..
My name is Carolyn Haslam and I am both a staff member at The University of Auckland and a PhD student. I am employed by The University of Auckland at the Faculty of Education and I am involved in Teacher Education in the curriculum area of Science. My aim for this preliminary study is to determine the time spent teaching and practising graphing skills in science classrooms from Years 9 – 11. Graphing skills refers to both construction and interpretation of graphs in science. Graphing is an important skill in science which is used extensively by scientists and it is an important aspect of developing scientific literacy. It is my aim to find out if these skills are being taught and practised in science classrooms from Years 9-11 to the same extent. I would like to invite you to complete the attached questionnaire which should take approximately 20 minutes of your time. The questionnaire asks you to estimate the number of periods that you spend teaching and practising graphing skills in science classroom from Years 9 to 11, how these align with curriculum knowledge strands, how easy or difficulty students find graphing instruction, whether students benefit from graphing instruction, and if so, how do they benefit. In addition, you will be asked to identify which Achievement standards within your Year 11 core science course you are offering this year.

Your decision to participate is voluntary and if you consent to complete the questionnaire you can choose to withdraw your data up until the 31st May 2012. The principal has given their assurance that your participation or non partipation will not affect your standing in the school. Your principal will not be advised of your decision to participate or not by the researcher. The collated data will not be reported in ways that identify your school or Head of Department: Science.

If you agree to participate please complete both the questionnaire and the consent form and return them to me in the self addressed envelope enclosed. The completed questionnaires and Consent Forms will be kept separately for a period of 6 years in a secure storage space at The University of Auckland. After this period they will be shredded.
The collated data from 50 schools in Auckland and Northland will be made available to your school so that your science department can compare data with other schools.

My contact details should you wish to discuss this further are
Email c.haslam@auckland.ac.nz, telephone 6238899 extn. 83918, mobile 021 709146, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland
My main supervisor, Richard Hamilton can be contacted on rj.hamilton@auckland.ac.nz, telephone 3737599 extn 85619, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150.
My Head of Department, Gregor Lomas can be contacted on email g.lomas@auckland.ac.nz, telephone 6238899 extn. 48517, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150.
For any queries regarding ethical concerns you may contact The Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142. Telephone 093737599 extn. 83711
CONSENT FORM
Principal/ Board of Trustees
THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS.

Project Title: Graphing in Science Years 9-11
Researcher: Carolyn Haslam

I have read the participant Information Sheet, have understood the nature of the research and why I have been selected and that my involvement is voluntary. I have had the opportunity to ask questions and had them answered to my satisfaction.

• I agree to my school taking part in this research.
• I give consent for an approach to be made to the Head of Science to request participation in this research.
• I understand that it will take approximately 20 minutes to answer the questionnaire
• I give my assurance that participation or non participation by the Head of Science will not affect his/her standing in the school.
• I understand that participation in this research is voluntary and that I may withdraw my permission for you to approach the Head of Department: Science as well as withdraw any data collected at my school up until the 31st May 2012.
• I understand that data will be kept for 6 years after which they will be destroyed
• I understand that the data will remain confidential, they will be collated with other school’s data and my school will not be identified in any publication.

Name ___________________________
Signature _____________________________           Date __________________

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE
ON.....FOR (3) YEARS, Reference Number .2011.../....438.
CONSENT FORM
Participant
THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS.

Project Title: Graphing in Science Years 9-11
Researcher: Carolyn Haslam

I have read the Participant Information Sheet, have understood the nature of the research and why I have been selected and that my involvement is voluntary. I have had the opportunity to ask questions and had them answered to my satisfaction.

• I agree to take part in this research
• I understand that it will take approximately 20 minutes to answer the questionnaire
• I understand that my data may be withdrawn up until the 31st May 2012.
• I understand that data will be kept for 6 years after which they will be destroyed
• I understand that the data will remain confidential, they will be collated with other school’s data and my school will not be identified in any publication.
• I understand that the collated results from this preliminary study will be made available to me.

Name ___________________________
Signature _____________________________        Date __________________

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE
ON.....FOR (3) YEARS, Reference Number .2011.../.438...
Appendix B  Graphing study questionnaire

Graphing Skills Questionnaire

I am interested in the amount of timetabled time science teachers give to teaching **graphing skills** in Years 9 - 11. This includes explicit teaching and student practice of graphing skills.

<table>
<thead>
<tr>
<th>Note: for this questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Graphing skills</strong> = graph construction and graph interpretation.</td>
</tr>
<tr>
<td><strong>Time</strong> = class periods at your school (please enter all time data in periods, you can have ½ or ¼ periods)</td>
</tr>
</tbody>
</table>

An average period at your school = _______ mins

1. **Looking at Areas of science where the most time is spent on graphing skills**

**Q:** Which 3 Units/Topics do you spend the most amount of time on graphing skills?

<table>
<thead>
<tr>
<th>Year</th>
<th>Unit/Topic Name</th>
<th>This unit aligns with which NZC knowledge strand? (circle your answer)</th>
<th>Approximate time (periods)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Interpretation</td>
<td>Construction</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Construction</strong></td>
<td><strong>Interpretation</strong></td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>LW, MW, PW, PE</td>
<td>LW, MW, PW, PE</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>LW, MW, PW, PE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>LW, MW, PW, PE</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>LW, MW, PW, PE</td>
<td>LW, MW, PW, PE</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>LW, MW, PW, PE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>LW, MW, PW, PE</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>LW, MW, PW, PE</td>
<td>LW, MW, PW, PE</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>LW, MW, PW, PE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>LW, MW, PW, PE</td>
<td></td>
</tr>
</tbody>
</table>

Please use the data from the above chart to help you fill in the chart below

1 LW – Living World, MW – Material World, PW-Physical World, PE- Planet Earth and Beyond

2. **Time (periods) spent on graphing skills**

<table>
<thead>
<tr>
<th>Year</th>
<th>Construction</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>teaching</td>
<td>practise</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Total time (periods) allocated to science per week at your school

<table>
<thead>
<tr>
<th>Year</th>
<th>Total allocated number of science periods per week</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

4. Questions: (circle the answer of your choice, think of your average students over all three years)

   a. Do you consider graphing to be an important scientific skill? Yes/No

   b. Overall do your students find Graph Construction?

      Very very easy    very easy    easy    not hard or easy    difficult    very    difficult    very very difficult

   c. Overall do your students find Graph interpretation?

      Very very easy    very easy    easy    not hard or easy    difficult    very    difficult    very very difficult

   d. Do you think that your students would benefit from more time spent on graph construction? Yes / No Why?

   e. Do you think that your students would benefit from more time spent on graph interpretation? Yes / No Why?

5. Referring to your Year 11 core Science course, circle which Achievement standards you are offering in 2011
<table>
<thead>
<tr>
<th>Physical World</th>
<th>Material World</th>
<th>Living World</th>
<th>Planet Earth &amp; Beyond</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS90940 Science 1.1</td>
<td>AS90944 Science 1.5</td>
<td>AS90948 Science 1.9</td>
<td>AS90952 Science 1.13</td>
</tr>
<tr>
<td>Demonstrate understanding of aspects of mechanics</td>
<td>Demonstrate understanding of aspects of acids and bases</td>
<td>Demonstrate understanding of biological ideas relating to genetic variation</td>
<td>Demonstrate understanding of the formation of surface features in New Zealand</td>
</tr>
<tr>
<td>4 credits</td>
<td>4 credits</td>
<td>4 credits</td>
<td>4 credits</td>
</tr>
<tr>
<td>External</td>
<td>External</td>
<td>External</td>
<td>Internal</td>
</tr>
<tr>
<td>AS90941 Science 1.2</td>
<td>AS90945 Science 1.6</td>
<td>AS90949 Science 1.10</td>
<td>AS90953 Science 1.14</td>
</tr>
<tr>
<td>Investigate the implications of electricity and magnetism for everyday life</td>
<td>Investigate the implications of the use of carbon compounds as fuels</td>
<td>Investigate life processes and environmental factors that affect them</td>
<td>Demonstrate understanding of carbon cycling</td>
</tr>
<tr>
<td>4 credits</td>
<td>4 credits</td>
<td>4 credits</td>
<td>4 credits</td>
</tr>
<tr>
<td>Internal</td>
<td>Internal</td>
<td>Internal</td>
<td>Internal</td>
</tr>
<tr>
<td>AS90942 Science 1.3</td>
<td>AS90946 Science 1.7</td>
<td>AS90950 Science 1.11</td>
<td>AS90954 Science 1.15</td>
</tr>
<tr>
<td>Investigate the implications of wave behaviour for everyday life</td>
<td>Investigate the implications of the properties of metals for their use in society</td>
<td>Investigate biological ideas relating to interactions between humans and micro-organisms</td>
<td>Demonstrate understanding of the effects of astronomical cycles on planet Earth</td>
</tr>
<tr>
<td>4 credits</td>
<td>4 credits</td>
<td>4 credits</td>
<td>4 credits</td>
</tr>
<tr>
<td>Internal</td>
<td>Internal</td>
<td>Internal</td>
<td>Internal</td>
</tr>
<tr>
<td>AS90943 Science 1.4</td>
<td>AS90947 Science 1.8</td>
<td>AS90951 Science 1.12</td>
<td>AS90955 Science 1.16</td>
</tr>
<tr>
<td>Investigate the implications of heat for everyday life</td>
<td>Investigate selected chemical reactions</td>
<td>Investigate the biological impact of an event on a New Zealand ecosystem</td>
<td>Investigate an astronomical or Earth science event.</td>
</tr>
<tr>
<td>4 credits</td>
<td>4 credits</td>
<td>4 credits</td>
<td>4 credits</td>
</tr>
<tr>
<td>Internal</td>
<td>Internal</td>
<td>Internal</td>
<td>Internal</td>
</tr>
</tbody>
</table>

6. What methods do you use in teaching graphing skills (Year 9-11). Circle all the methods you currently use.

a. Whiteboard
b. Smart board
c. Powerpoint presentation
d. Computers/i-pads
e. Textbook
f. Worksheet
f. Other (Please specify) ________________________________
7. Comment – please add any extra comments, they could be regarding this questionnaire, graphing skills in general, the data you supplied or any other ideas related to this topic.

Thank you for taking the time to complete this question
Appendix C  Teacher responses to Question 4 of graphing questionnaire

Preliminary Graphing Questionnaire typed answers to Q’s 4 d and e and Q 7
Blue answers are where the school said No to more time spent on graphing in question 4 (d and e)

<table>
<thead>
<tr>
<th>School #</th>
<th>Question 4 d</th>
<th>Question 4 e</th>
<th>Question 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Weaker students find the axes difficult to work out</td>
<td>Weaker students require more time to understand concepts</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Same issues/mistakes arise regularly. Eg irregular intervals on scales</td>
<td>Simple graph reading they seem to understand OK but they overlook things on complex graphs</td>
<td>One issue that students regularly have with drawing line graphs is determining an appropriate interval/scale for the dependent variable. Often the I.V (Independent variable) has a regular interval so numbers are placed even distances apart along the horizontal axis. However, the D.V. does not always have such a regular pattern but students still place these numbers at regular intervals on the vertical axis. Also selecting when to draw bar vs line graphs is a regular issue. Bar graphs seem to be their default. Eg.</td>
</tr>
<tr>
<td>11?</td>
<td>Because they still struggle to get axis evenly spaced and recognise the difference between types of graphs and which should be used for each type of data</td>
<td>Because they still struggle with graphing especially interpretation and it is a vital skill that is needed for science particularly investigations and NOS, as well as being common across many assessments</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td>Would love to teach more graphing skills in science but don’t have time with everything else. Have relied on maths to teach graphing in the past but have found recently that they teach less graphing and the wrong types of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Often done on computer now, not as difficult as interpretation</td>
<td>Students do find this difficult and a new graph can present as a whole new problem. Needs practice and reinforcement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>This is a general view. Talking with my department physics teachers spend more time on this skill with junior classes. They often teach the same class for Maths. Teachers of low ability classes spend a lot more time on this skill.</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Because some need more time</td>
<td>Often students have difficulty transferring learning from maths to science. Also students often choose to do bar graphs even when asked to do line graphs</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>We teach graphing by hand and graphing in excel</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>At the maximum 2 days more as other parts of the curriculum is equally important</td>
<td>At the maximum 2 days more as other parts of the curriculum is equally important</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Depending on the ability – slower kids yes</td>
<td>Particularly as to the slopes of graphs and its meaning</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>-Scales -Interpreting We have a very wide range of ability re graph skills with our students. They can struggle with the : a) scale construction b) interpreting what the graph shows and for Y11 line of best fit and mathematical equations. However many find graphs straight forward after training at least in construction.</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>They need to do a lot more questions within the unit of study relating to graphs</td>
<td>This is becoming part of thinking power for NCEA questions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Graphing skills are very important from our point of view. It represents relationships and is vital part of our science teaching.</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>The time allocated is not enough for some of our students</td>
<td>The time allocated is not enough for some of our students as they have a lower level of numeracy. We hope their time spent in maths</td>
<td></td>
</tr>
<tr>
<td>Page</td>
<td>Text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>236</td>
<td>will also consolidate their learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>By the time they reach senior sciences they are still struggling with appropriate scale, best fit and other basic skills. Graphing is such an incredibly important tool in science and I am realising we don’t spend nearly enough time on it.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>See above (left) this is a vital mechanism for processing data and the ability to see relationships and trends is enormously important</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Students working in digital media – I have had a class in 2011 working via notebooks. They have frequently used the spreadsheets in Excel or Open office to construct graphs which are meaningless and seem to lack the tools to identify why.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• They (students in the digital class) would frequently confuse bar or column graphs with histograms or line scatter plots.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• There seems to be much confusion over what they are taught to do in maths vs science vs social studies.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• The data in this survey is based on rough estimates based on my own teaching of the 3 year levels. It may vary from class to class.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Students always groan when I say &quot; and now lets graph it&quot;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• I have become very conscious as I completed this survey that for something so important we don’t seem to spend a lot of time on it!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>It will develop better understanding of what the data represents</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>More time would mean more practice and this would help them develop their understanding and the process involved in interpreting graphs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I am of the opinion that graphing (construction and interpretation) should be given more emphasis in our teaching programmes. Students will be increasingly confronted by graphical representations of data.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Variety of abilities and don’t seem to retain understanding</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Need to practice but it has to be relevant to the topic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 different courses are run at Year 11 so different combinations of AS Difficult to put accurate hours on this as depends on class especially when comparing top and mid stream classes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>They rarely need to use this skill – graphical calculators and excel more widely used</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Useful to gain and understanding of patterns and relationships. Useful outside</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Some students come to college with v good graphing skills – some with none whatsoever. Could be something intermediate schools focus on so that interpretation could be a college focus.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Page</td>
<td>Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the classroom</td>
<td>We use CUTLASS to help students remember crosses, units, title, line, axes, smooth, scale – from Y9 and continue this throughout. Graphs can easily be peer marked once students are familiar with this.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Graphing and statistics are valuable tools for everyday life. Y9 restructured due to drop in teaching hours Can make the initial interpretation but struggle to justify. Leads into Physics 1.1 and senior sciences – struggle with equations</td>
<td>Y12/13 students rely on a lot of graphing skills for internal assessments. Y11 Physics 1.1 added this year for numeracy credits as well as graphing skills.</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>We cover this well as we are a Y7-13 school and the Year 8 programme spends quite a bit of time on it This is a more difficult aspect and is essential for NCEA</td>
<td>Graphing skills are an essential technique. Students complete introductory skills booklet at the start of each year, to refresh the techniques. This involves teaching practise and interpretation. Graphing is then utilised within each topic as appropriate. However, I think the emphasis is shifting more towards interpretation rather than construction –as there are many technological devices that draw the graphs – the greater skill is interpretation.</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>It is a basic skill that needs to be instilled as it is widely used</td>
<td>They seem to understand this better</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Students individual needs are met already</td>
<td>Students individual needs are met already</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>The more they do it the easier/quicker they get It brings out the point of analysing data with graphs AND the science idea/concept AND literacy skills etc…</td>
<td>Graphing skills get drip fed in Yr9 and 10 in science. But we have worked closely with the Maths department to know what yr 9 maths and yr 10 maths do. So that in year 11 things go a lot quicker but kids still find graphing tricky and time consuming 9 making graphs and interpreting them) especially if they have to do it in test conditions.</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Our students have low numeracy skills which is the biggest barrier to their ability in constructing</td>
<td>Our assessment data still shows poor graph interpretation skills, even in the senior school I have revisited teaching graphing(construction)using spredsheets, etc , as I believe being able to create a graph on paper is important in understanding /interpreting what a graph shows, generally line graphs.</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>There are a number of skills such as even spacing that students need a great deal of practice on graphs.</td>
<td>We have spent a lot of time on sentence structure needed to interpret graphs for ESOL students.</td>
<td>We have had some good help from Liggins institute in drawing and interpreting graphs in a biological context. We use the an acronym (sic) to help the students. The physics advisors through the Physics Teachers have provided good leadership for activities such as Do Nows to help students construct good graphs. Team solutions have been very supportive also.</td>
</tr>
<tr>
<td>46</td>
<td>Probably about right in relation to other skills needed.</td>
<td>Probably about right in relation to other skills needed.</td>
<td>6. Other : Overhead and group work</td>
</tr>
<tr>
<td>48</td>
<td>During the time spent the majority grasp the principles. More time would be detrimental to other aspects of science.</td>
<td>Our general science students who will not take Y12 need to be able to interpret data in newspapers etc.</td>
<td>The major hurdle for our students is the concept of scale. This is not a science issue solely. Many teachers expect to launch into line graphs when students do not have an understanding of relative area/height/distance using a bar graph or pie chart. If they have no comprehension of scale then the graph word is very difficult.</td>
</tr>
<tr>
<td>50</td>
<td>Results show good understanding overall.</td>
<td>Possibly one of 2 lessons of practice.</td>
<td>Mostly taught in context. We no longer teach graphing skills as a stand alone module.</td>
</tr>
</tbody>
</table>
Appendix D  Ethics sheets Study two

PARTICIPANT INFORMATION SHEET
Principal/ Board of Trustees

Project Title:  Can pretraining help science students learn complex information?
Researcher:  Carolyn Haslam

To (Principal’s name)

My name is Carolyn Haslam and I am both a staff member and PhD student at The University of Auckland. I am employed by The University of Auckland at the Faculty of Education and I am involved in Teacher Education in the curriculum area of Science. Your school has been chosen as it complements the other schools that have been invited to participate.

My aim for this study is to investigate whether a strategy called pretraining can be used in secondary science classrooms to help students learn complex information which could potentially cause cognitive overload and therefore interfere with learning and understanding in science. Much of the information presented in senior science is complex and often students are overwhelmed by this complexity and this can interfere with learning and understanding which is important for achievement in school and national assessments. I am planning to trial a teaching strategy that presents complex information in 2 stages to enhance learning and understanding. I request your permission to approach your Head of Department: Science to request the possibility of using three Year 10 classes in science for this study at your school. I am planning to use Year 11 physics material with Year 10 students to ensure that it is appropriately complex. The study will involve one period of teaching and I would like to request permission that I be responsible for teaching the three classes. There will be one class that receives the pretraining strategy and a PowerPoint presentation, the other two classes will receive the same PowerPoint presentation, one class once through the presentation and the other twice through. If the new teaching strategy should prove to be successful then a summary of the research findings will be made available to your school so that your science department can benefit from the results of this study and I will develop guidelines for your teachers so they could use this strategy with their classes to enhance the learning of complex information. I will offer this as professional development for your science teachers regardless of the school’s decision to participate or not made by your Head of Department: Science. All students in the allocated classes will take part in the instructional activities. I will be asking them for permission to use their data in my study. I will ask them to sign a Consent Form to confirm that they understand the research that they are being asked to be involved in. In addition, I will ask the students to give a parental information sheet to their parents/guardians in order to inform their parents/guardians of the nature and value of their participation in the targeted instructional activities.

I will visit the classes prior to the study to introduce myself, to explain the method, purpose and use of the data collected. In addition I will answer any questions the students may have and give them the opportunity to view the materials that will be used. This will take about 15 minutes. In a period before the teaching I will also ask the classroom science teacher to administer a pre-test, which will take about 10 minutes. During the teaching I
would be asking the science teacher of the class to be present in the room and to record what is happening as a running log in order to allow me to document the integrity of the treatment. I will administer a post-test at the end of the instructional activity. In addition I will be asking the classroom science teachers to administer the same post-test about 3 weeks after the study to assess delayed recall and understanding. This will take 10 – 15 minutes. Total time for this study will be about one and a half hours. I also ask for permission to access PAT test results for the individual students who agree to participate, in the classes involved in this study to assess the moderating influences of these abilities on performance on the post-test. The timing of the data collection will be in consultation with the classroom teachers so as not to disadvantage the students from their regular science learning. The benefit to the students will be that they will be introduced to material that they will likely be required to learn for NCEA level one science from AS 90940.

The Head of Department: Science and the classroom science teacher's decision to participate is voluntary and they will in no way be obligated to be involved based on your agreement to let me approach them. I would like your assurance that participation or non-participation by the Head of Department: Science or the classroom science teacher will not affect his/her standing in the school. If they consent to being involved in this study they can choose to withdraw their data up until the 30th June 2014. In addition, if you choose to allow your school to participate in this research, you can withdraw this approval as well as any data collected within your school at any time up until the 30th June 2014. The collated data will not be reported in ways that identify your school, students, teachers or Head of Department: Science.

If you agree to allow me to approach your Head of Science, please sign and return the Consent Form enclosed, to me, in the self addressed envelope.

The data collected and Consent Forms will be kept separately for a period of 6 years in a secure storage space at The University of Auckland. Electronic data will be stored on a password-protected computer. After this period the data will be shredded and electronic copies erased.

My contact details should you wish to discuss this further are email c.haslam@auckland.ac.nz, telephone 6238899 extn. 83918, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150. My main supervisor, Richard Hamilton can be contacted on rj.hamilton@auckland.ac.nz, telephone 3737599 extn. 85619, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150. My Head of School, Judy Parr can be contacted on email jm.parr@auckland.ac.nz , telephone 6238899 extn. 88998, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150. For any queries regarding ethical concerns you may contact The Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142. Telephone 093737599 extn. 83711 APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE ON 7 Nov 2013 FOR (3) YEARS, Reference Number2013/ 9814
My name is Carolyn Haslam and I am both a staff member and PhD student at The University of Auckland. I am employed by The University of Auckland at the Faculty of Education and I am involved in Teacher Education in the curriculum area of Science. Your school has been chosen as it complements the other schools that have been invited to participate.

My aim for this study is to investigate whether a strategy called pretraining can be used in secondary science classrooms to help students learn complex information which could potentially cause cognitive overload and therefore interfere with learning and understanding in science. Much of the information presented in senior science is complex and often students are overwhelmed by this complexity and this can interfere with learning and understanding which is important for achievement in school and national assessments. I am planning to trial a teaching strategy that presents complex information in 2 stages to enhance learning and understanding.

I would like to request access to three Year 10 classes in science for this study at your school. I am planning to use Year 11 physics material with Year 10 students to ensure that it is appropriately complex. The study will involve one period of teaching and I would like to request permission that I be responsible for teaching the three classes and that this is a compulsory activity for these classes. There will be one class that receives the pretraining strategy and a PowerPoint presentation, the other two classes will receive the same PowerPoint presentation, one class once through the presentation and the other twice through. If the new teaching strategy should prove to be successful then a summary of the research findings will be made available to your school so that your science department can benefit from the results of this study and I will develop guidelines for your teachers so they could use this strategy with their classes to enhance the learning of complex information. I will offer this as professional development for your science teachers regardless of the decision to participate or not made by you. All students in the allocated classes will take part in the instructional activities. I will be asking them for permission to use their data in my study. I will ask them to sign a Consent Form to confirm that they understand the research that they are being asked to be involved in. In addition, I will ask the students to give a parental information sheet to their parents/guardians in order to inform their parents/guardians of the nature and value of their participation in the targeted instructional activities.

I will visit the classes prior to the study to introduce myself, to explain the method, purpose and use of the data collected. In addition I will answer any questions the students may have and give them the opportunity to view the materials that will be used. This will take about 15 minutes. In a period before the teaching I will also ask the classroom science teacher to administer a pre-test, which will take about 10 minutes. During the teaching I would be asking the science teacher to be present in the room and to record what is happening as a running log in order to allow me to document the integrity of the treatment. At the end of the instructional activity, I will administer a post-test. In addition, I will be asking the classroom teachers to administer the same post-test about 3 weeks after the
study to assess delayed recall and understanding. This will take 10 – 15 minutes. Total time for this study will be about one and a half hours. I also ask for permission to access PAT test results for the individual students who agree to participate, in the classes involved in this study to assess the moderating influences of these abilities on performance on the post-test. The timing of the data collection will be in consultation with the classroom teachers so as not to disadvantage the students from their regular science learning. The benefit to the students will be that they will be introduced to material that they will likely be required to learn for NCEA level one science from AS 90940.

Your decision to participate is voluntary. The Principal/Board of Trustees has given assurance that participation or non participation by you or your Science Department will not affect your standing in the school. In addition I would be asking for your assurance that participation or non participation by science staff will not affect their standing in the school. If you consent to allow three classes to be involved in this study, you can choose to withdraw the data collected up until the 30th June 2014. The collated data will not be reported in ways that identify your school, students, teachers or yourself.

If you agree to allow me to conduct this research in your Science Department, please sign the Consent Form attached and I will collect it.

The data collected and Consent Forms will be kept separately for a period of 6 years in a secure storage space at The University of Auckland. Electronic data will be stored on a password-protected computer. After this period the data will be shredded and electronic copies erased.

My contact details should you wish to discuss this further are email c.haslam@auckland.ac.nz, telephone 6238899 extn. 83918, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150. My main supervisor, Richard Hamilton can be contacted on rj.hamilton@auckland.ac.nz, telephone 3737599 extn. 85619, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150. My Head of School, Judy Parr can be contacted on email jm.parr@auckland.ac.nz, telephone 6238899 extn. 88998, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150.

For any queries regarding ethical concerns you may contact The Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142. Telephone 093737599 extn. 83711

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE ON 7 Nov 2013 FOR (3) YEARS, Reference Number 2013/ 9814
PARTICIPANT INFORMATION SHEET
Science Teacher

Project Title: Can pretraining help science students learn complex information?  
Researcher: Carolyn Haslam

My name is Carolyn Haslam and I am both a staff member and PhD student at The University of Auckland. I am employed by The University of Auckland at the Faculty of Education and I am involved in Teacher Education in the curriculum area of Science. Your school has been chosen as it complements the other schools that have been invited to participate.

My aim for this study is to investigate whether a strategy called pretraining can be used in secondary science classrooms to help students learn complex information which could potentially cause cognitive overload and therefore interfere with learning and understanding in science. Much of the information presented in senior science is complex and often students are overwhelmed by this complexity and this can interfere with learning and understanding which is important for achievement in school and national assessments. I am planning to trial a teaching strategy that presents complex information in 2 stages to enhance learning and understanding.

I would like to request access to your Year 10 science class for this study at your school. I am planning to use Year 11 physics material with Year 10 students to ensure that it is appropriately complex. The study will involve one period of teaching and I would like to request permission that I be responsible for teaching your class and that this be a compulsory activity for your class. There will be one class that receives the pretraining strategy and a PowerPoint presentation, the other two classes will receive the same PowerPoint presentation, one class once through the presentation and the other twice through. If the new teaching strategy should prove to be successful then a summary of the research findings will be made available to your school so that your science department can benefit from the results of this study and I will develop guidelines for your teachers so they could use this strategy with your classes to enhance the learning of complex information. I will offer this as professional development to you and fellow teachers regardless of the decision to participate or not. All students in your class will take part in the instructional activities. I will be asking them for permission to use their data in my study. I will ask them to sign a Consent Form to confirm that they understand the research that they are being asked to be involved in. In addition, I will ask the students to give a parental information sheet to their parents/guardians in order to inform their parents/guardians of the nature and value of their participation in the targeted instructional activities.

I will visit the classes prior to the study to introduce myself, to explain the method, purpose and use of the data collected. In addition I will answer any questions the students may have and give them the opportunity to view the materials that will be used. This will take about 15 minutes. In a period before the teaching I will also ask you to administer a pre-test, which will take about 10 minutes. During the teaching I would be asking you to be present in the room and to record what is happening as a running log in order to allow me to document the integrity of the treatment. I will administer a post-test at the end of the instructional activities. In addition I will be asking you as the classroom
teacher to administer a post-test about 3 weeks after the study to assess delayed recall and understanding. This will take 10 – 15 minutes. Total time for this study will be about one and a half hours. I also ask for permission to access PAT test results for the individual students who agree to participate, in the classes involved in this study to assess the moderating influences of these abilities on performance on the post-test. The timing of the data collection will be in consultation with you so as not to disadvantage the students from their regular science learning. The benefit to the students is that they will be introduced to material that they will likely be required to learn for NCEA level one science from AS 90940.

Your decision to participate is voluntary. The Principal/BOT and Head of Department: Science have given assurance that participation or non participation by you will not affect your employment or standing in the school. I would also ask for your assurance that participation or non participation in this study by individual students will in no way affect their school science results this year. If you consent to be involved in this study, you can choose to withdraw the data collected from your class up until the 30th June 2014. The collated data will be assigned a code and therefore individual results will be anonymous, any reporting will be in ways that do not identify your school, students, or yourself.

If you agree to allow me to conduct this research in your Science Department, please sign the Consent Form attached and I will collect it.

The data collected and Consent Forms will be kept separately for a period of 6 years in a secure storage space at The University of Auckland. Electronic data will be stored on a password-protected computer. After this period the data will be shredded and electronic copies erased.

My contact details should you wish to discuss this further are email c.haslam@auckland.ac.nz, telephone 6238899 extn. 83918, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150.

My main supervisor, Richard Hamilton can be contacted on rj.hamilton@auckland.ac.nz, telephone 3737599 extn. 85619, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150.

My Head of School, Judy Parr can be contacted on email jm.parr@auckland.ac.nz, telephone 6238899 extn. 88998, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150.

For any queries regarding ethical concerns you may contact The Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142. Telephone 093737599 extn. 83711

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE ON 7 Nov 2013 FOR (3) YEARS, Reference Number 2013/ 9814
My name is Carolyn Haslam and I work and am studying for my PhD at The University of Auckland. I work at the Faculty of Education and I teach secondary and primary Student Teachers how to teach science. Your school has been chosen as it complements the other schools that have been invited to participate.

My aim for this study is to look at whether a strategy called pretraining can be used to help you learn difficult science ideas, ones that really strain your brain and make it hard to learn and understand.

A lot of ideas in physics are quite difficult to understand so I am planning to trial this strategy which has not been used in a classroom yet, while teaching you some Year 11 physics ideas. I would like to invite you and your class to be part of this study at your school. The study will involve one period of science which will be part of your science programme this year. I will be teaching your class and two other classes at your school. There will be one class that I will use the pretraining strategy and a PowerPoint presentation with and the other two classes will receive the same PowerPoint presentation, one of them once and the other class twice through. Your class could be involved in any one of these. If the new teaching strategy is successful I will come back and share the results with your science teachers so they can help you learn difficult ideas in other topics in science.

This is what will happen. I will visit your class before the study to introduce myself and explain what the study is all about and why I am doing it. I will also answer any questions you may have. I will also show you the sheets that you will be filling out during the study. These are very short. This will take about 15 minutes. Your teacher will give you a short pre-test to fill in just before I come, which will take about 10 minutes. During the teaching, your science teacher will be in the class. At the end of the teaching I will ask you to complete a short post test which will take you about 10 – 15 minutes. The total time for this study will be about one and a half hours of your science time. Taking part in this study will not affect your school science results. To make sure that the classes who take part in the study are similar I am also asking for permission from you to look at your individual PAT test results. I am doing this study in term 1 at a time which your teacher suggests will not interfere with your class’ science learning for tests and exams.

You do not have to allow me to use your results from the pre and post-test, questionnaire and PAT results for my study if you do not want to, the choice is yours. If you do choose to let me use your results, even after the study is finished up until the 30th June 2014 you can choose to remove your results from the study. I will be the only one who sees your results, your results will be given a code and no one except me will know what these codes are for. I can assure you that when I am sharing the final results of all the analysed data with other people they will not know that the study was done in your school with your teachers or that you were involved.
I am asking you to sign a Consent Form to allow me to use your results from the pre and post-test, questionnaire and PAT results in my study. I have also made a sheet explaining my study for your parents/guardians. Please take this sheet home to share with your parents/guardians. If you agree to help me with this study by letting me use your results then you need to sign the Consent Form attached to this letter and give it to your science teacher.

The completed sheets from the study and your Consent Forms will be kept separately for a period of 6 years in a secure storage space at The University of Auckland. All data that I put on my computer is safe as it needs a password to access it and only I know this password. After 6 years the paper data will be shredded and the electronic data deleted.

Thank you for reading this.

Below are ways of getting hold of me if you have any more questions about this study. c.haslam@auckland.ac.nz, telephone 6238899 extn. 83918, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150.

If you need to talk to my PhD supervisor or my boss their contact information is below. My main supervisor, Richard Hamilton can be contacted on rj.hamilton@auckland.ac.nz, telephone 3737599 extn. 85619, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150.

My Head of School, Judy Parr can be contacted on email jm.parr@auckland.ac.nz, telephone 6238899 extn. 88998, The Faculty of Education, The University of Auckland, Private Bag 92601, Symonds St, Auckland 1150.

For any queries regarding ethical concerns you may contact The Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142. Telephone 093737599 extn. 83711

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE ON 7 Nov 2013 FOR (3) YEARS, Reference Number 2013/ 9814
CONSENT FORM

Principal/ Board of Trustees

THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS.

Project Title: Can pretraining help science students learn complex information?

Researcher: Carolyn Haslam

I have read the Participant Information Sheet, have understood the nature of the research, why I have been selected and that my involvement is voluntary. I have had the opportunity to ask questions and had them answered to my satisfaction.

• I agree to my school taking part in this research.
• I give consent for an approach to be made to the Head of Science to request participation in this research.
• I give consent to access PAT data on the individual students who agree to participate, in the classes involved in the study.
• I understand that it will take approximately one and a half hours to collect the data.
• I understand that this will be a compulsory activity for the three classes involved.
• I give my assurance that participation or non participation by the Head of Science or science teachers will not affect their employment or standing in the school.
• I give my assurance that the students’ grades and relationship with the school will not be affected by their participation or non participation in this study.
• I understand that participation in this research is voluntary and that I may withdraw my permission for you to approach the Head of Department: Science as well as withdraw any data collected at my school up until the 30th June 2014.
• I understand that data will be kept for 6 years after which they will be destroyed.
• I understand that the data will remain confidential, they will be collated with other school’s data and my school, teachers or students will not be identified in any publication or presentation.

Name ___________________________ School ___________________________

Signature ___________________________ Date ______________

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE ON 7 Nov 2013 FOR (3) YEARS, Reference Number 2013/ 9814
CONSENT FORM
Head of Department: Science

THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS.

Project Title: Can pretraining help science students learn complex information?

Researcher: Carolyn Haslam

I have read the Participant Information Sheet, have understood the nature of the research and why I have been selected and that my involvement is voluntary. I have had the opportunity to ask questions and had them answered to my satisfaction.

- I agree to take part in this research.
- I understand that it will take approximately one and a half hours to collect the data.
- I understand that this will be a compulsory activity for the three classes involved.
- I understand that the Principal has assured that my employment will not be affected regardless of my decision to participate or not participate in this study.
- I give my assurance that participation or non participation by the science teachers of the classes involved will not affect their employment or standing in the school.
- I understand that PAT data on individual students who agree to participate, in the classes involved in the study in my Department will be accessed.
- I understand that data may be withdrawn up until the 30th June 2014.
- I understand that data will be kept for 6 years after which they will be destroyed.
- I understand that the data will remain confidential, they will be collated with other school’s data and my school, teacher or students will not be identified in any publication or presentation.
- I understand that the research findings from this study will be made available to me and my Department.

Name ___________________________ School ________________________________
Signature _____________________________ Date __________________

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE ON 7 Nov 1023 FOR (3) YEARS, Reference Number 2013/ 9814
CONSENT FORM
Science Teacher

THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS.

Project Title: Can pretraining help science students learn complex information?
Researcher: Carolyn Haslam

I have read the Participant Information Sheet, have understood the nature of the research, why I have been selected and that my involvement is voluntary. I have had the opportunity to ask questions and had them answered to my satisfaction.

- I agree to take part in this research.
- I understand that it will take approximately one and a half hours to collect the data.
- I understand that this will be a compulsory activity for my whole class.
- I understand that I will be required to observe and record events in the classroom during the teaching.
- I understand that the Principal and HOD Science have assured that my employment will not be affected by my participation or non-participation in this study.
- I give my assurance that participation or non-participation by the students of my class will not affect their school science assessment results or their standing within the class.
- I understand that PAT data on individual students in my class will be accessed.
- I understand that data from my class may be withdrawn up until the 30th June 2014.
- I understand that data will be kept for 6 years after which they will be destroyed.
- I understand that the data will remain confidential, they will be collated with other school’s data and my school will not be identified in any publication or presentation.
- I understand that the research findings from this study will be made available to me.

Name ___________________________ School ___________________________
Signature _____________________________ Date _______________

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE ON 7 Nov 2013 FOR (3) YEARS, Reference Number 2013/ 9814
CONSENT FORM
Students
THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS.

Project Title: Can pretraining help science students learn complex information?

Researcher: Carolyn Haslam

I have read the Participant Information Sheet, and I know what I am going to be doing in this study. I know why I have been chosen and I have a choice about whether my results will be part of this study or not. I have been able to ask questions and had them answered, so I am happy about what I need to do.

• I understand that this activity will be compulsory for my class but that letting the researcher use my results is voluntary.
• I give permission for my results from the pre and post-test, questionnaires and PAT tests to be used in this study.
• I understand that it will take approximately one and a half hours.
• I understand that if I take part it will not affect my science marks.
• I understand that the Principal has given his assurances that my grades or relationship with the school will not be affected by my decision to let you use my data or not use my data in the study.
• I understand that I can decide to take back the information on any sheets I completed during the study up until the 30th June 2014.
• I understand that my PAT test results will be used in this study.
• I understand that this information will be kept for 6 years after which it will be destroyed.
• I understand that when the results from this study are shared with others that they will not know that my school, teachers or myself were involved.
• I understand that the results will be given to my science teacher so they can use this strategy themselves in their classes.

Name ___________________________ School__________________________________
Signature ___________________________ Date _______________

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN ETHICS COMMITTEE ON 7 Nov 2013 FOR (3) YEARS, Reference Number
Appendix E  Initial phase PowerPoint presentation

My PhD study

Carolyn Haslam

What is happening today

• Overview of the study
• Practice some questions I need you to answer
• Answer your questions
Overview 1

• Time: 1 period
• When: Monday 7\textsuperscript{th} April
• The teaching will be a physics topic
• Three classes from your school will be involved

Overview 2

In the period when I come back
1. I will teach you and during this time I will ask you to answer one question a few times
2. At the end of the teaching I will ask you to answer a few different questions about how hard it was
3. I will also ask you to answer a few questions about what I was teaching you
**THE TASK**

- I am going to ask you to rate things on a scale of 1-100.

Write your answer as a number between 1 and 100. eg 32, 45, 76, 92
THINK ABOUT THE LAST MOVIE OR DVD THAT YOU WATCHED

Question: How much did you enjoy the movie or DVD?

Not at all | 50 | very very much
0 | 50 | 100

Your answer
Fantastic = 91
Awful or rubbish = 12

Write your answer on the sheet as a number between 1 and 100 in the box on the right.
THINK ABOUT THE LAST SCIENCE LESSON YOU HAD WHERE YOU LEARNED SOMETHING NEW

• Question: How hard was it to understand the new science ideas in the whole lesson?

0                                  50                                  100
very very easy                        very very hard

• Write your answer on the sheet as a number between 1 and 100 in the box.

THINKING ABOUT THE LAST SCIENCE LESSON WHERE YOU LEARNED SOMETHING NEW

• Question: How hard was it for you to learn and remember the new ideas? Would you have to use your book to help you?

0                                  50                                  100
very very easy                        very very hard

• Write your answer on the sheet as a number between 1 and 100 in the box.
THINKING ABOUT THE LAST SCIENCE LESSON WHERE YOU LEARNED SOMETHING NEW

• Question: How successful do you think you were in learning the ideas? What percentage would you get if you had a test on these ideas?

Not at all successful

very successful

0 50 100

• Write your answer on the sheet as a number between 1 and 100 in the box.

THINKING ABOUT THE LAST SCIENCE LESSON WHERE YOU LEARNED SOMETHING NEW

• Question: How did you feel about this lesson? Were you angry, upset, discouraged, irritated or was it all good?

not at all frustrated

very frustrated

0 50 100

• Write your answer on the sheet as a number between 1 and 100 in the box.
• Question: How rushed was the lesson? Was the pace too slow or were you too rushed.

far too slow          about right          far too rushed
0                      50                       100

• Write your answer on the sheet as a number between 1 and 100 in the box.

Do you have any questions about any part of what we will do when I come back to teach you?
One other thing

- This will **not** be a BYOD lesson, *sorry*

Thank you
Appendix F  Practice cognitive load questions

**Question 1: Enjoyment of the video or DVD**

How much did you enjoy the movie or DVD?
Think: Was it fantastic or awful?

<table>
<thead>
<tr>
<th>Not at all</th>
<th>a lot</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Write your answer here

**Question 2: Difficulty of the ideas**

How hard was it to understand the new science ideas?
Think: How much effort did you have to put in to understand the science ideas?

<table>
<thead>
<tr>
<th>very very easy</th>
<th>very very hard</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Write your answer here

**Question 3: Effort to learn**

How hard was it for you to learn and remember the science ideas?
Think: Would I have to use my book to remember the science ideas?

<table>
<thead>
<tr>
<th>very very easy</th>
<th>very very hard</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Write your answer here

**Question 4: Performance**

How successful do you think you were in learning the science ideas?
Think: What percentage do I think I will get in a test on these ideas right now?

<table>
<thead>
<tr>
<th>Not at all successful</th>
<th>very successful</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

Write your answer here

**Question 5: Frustration**

How did you feel about this lesson?
Think: How discouraged, upset, angry or irritated did I feel?

<table>
<thead>
<tr>
<th>Not at all frustrated</th>
<th>very frustrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

Write your answer here
Question 6: Timing

How rushed did you feel the whole lesson was?
Think: How was the pace of the teaching and the time given to do the tasks?

far too slow                                  about right                               far too rushed

0                                              50                                       100

Write your answer here
Appendix G Pretest

Name ___________________________  School ___________________________  Class ____

This test is to see how much you already know about the topic of speed. Write your answers on this sheet. If you are not sure of the answer to any question please leave it blank.

1. Give an example of a unit that is used to measure speed. ______________

2. What is constant speed?

3. A boy runs along a track as shown below

   section X     section Y

   a. During section X he runs at a constant speed of 2 m/s for 15 seconds. What distance does he cover?

   b. During section Y he runs at 2.5 m/s and covers 12.5m. How long does this take him?

4. A child plays with a remote control car. The car starts from rest and travels a distance of 6m in 3 seconds. The car then travels a further 28m at a constant speed of 4m/s for 7 seconds. On the axes below complete a graph of the car’s journey.

Write the values of A and B on the lines below.

A=_____  
B=_____

Please turn over
5. A parachutist jumps from a plane and falls through a distance of 240m in 60 seconds.
   a. Calculate her average speed.
   b. On the axes below complete a sketch graph of her 60 s fall. (axes are not drawn to scale)
   c. Compare the sketch graph you drew and the actual graph of her fall. Give one reason why the actual graph of her fall may not look exactly like the one you drew above in part 5b.

6. The graph below shows the speed of 3 different cars. Look at the graph and answer the questions below.
   a. Which car is travelling the fastest? Car ______
   b. Why?
   c. Which car is travelling at 10m/s? Car ______
   d. Why?
What is pretraining?

+ An Introduction to the words that will be used in teaching
+ “The need for Speed”
+ Looking at the words and their meanings
+ Looking at the units (what we measure things in) that will be used
Distance

- How far something goes
- It is measured in metres – m
- Kilometres – km
- Centimetres – cm
- Millimetres – mm

Auckland - Wellington

Distance = 500km

Time

- How long it takes
- Measured in seconds – s
- Minutes – min
- Hours – h

Time flying Auckland to Wellington

Time = 1 hour
Speed

- Speed is how fast something is travelling (how quickly something moves)
- Speed is how far (distance) something can go in a certain time
- distance X

Speed example using a set distance

If 2 cars are travelling 100km
- Car A gets there in 1 hour
- Car B gets there in 1 ¼ hours
- Which car was travelling the fastest?
Question: How hard was it to understand what I was just teaching you?

very very easy  |  very very hard

0           | 50           | 100

Write your answer in the box on the sheet as a number between 1 and 100.

Speed example using a set time

If 2 cars travel for **10 minutes**

- Car A goes **10km**
- Car B goes **12 km**

Which car was travelling the fastest?

**10 mins**
Average speed

Car journey showing speed and average speed

Distance (m)

Time (s)

Average speed

Car journey showing actual speed and average speed

Distance (m)

Time (s)
**Constant speed**

- The same speed the whole time, no change in speed
- On the motorway you travel at 100km/h
- This is a constant speed of 100km/h

**Graphs**

- Axes
- y – up
- x - across

![Graph of x and y axes](image)
2. What is speed?

- Speed is how fast an object is travelling (how quickly something moves)
- What speed are you allowed to go on the motorway?
- 100 km/h (measure of speed or the unit of speed)

\[ \text{kilometres per hour} = \frac{\text{km}}{\text{h}} \]

- Speed includes a distance and a time measure
3. Usain Bolt’s Speed

- **Distance**: How far did Usain Bolt run?
- **Time**: How long did it take him? (We will use 10 seconds)
- **What speed did he run the race?**
- **Speed** is how far (distance) something can go in a certain time

\[
\text{Speed } (v) = \frac{\text{distance } (d)}{\text{time } (t)}
\]

Usain Bolt’s speed = \(100 \text{ metres (m)}\) \(=\) \(10 \text{ m/s}\)

If \(d = \text{(kilometre)}\) and \(t = \text{(hour)}\) \(s = \text{kilometres per hour}\)

\[
\begin{align*}
\text{km} & \quad \text{h} \quad \text{km/h}
\end{align*}
\]
4. Calculating speed

Average speed = \frac{\text{total distance}}{\text{total time}}

Sam ran 400 metres in 80 seconds, what is his average speed?

Average Speed = \frac{400\text{ m}}{80\text{ s}} = 5 \text{ m/s}

Show your working (thinking) – don’t just write the answer

5. Calculating distance and time

speed (v) = \frac{\text{distance (d)}}{\text{time (t)}}

distance = v \times t

time = \frac{d}{v}

Try questions a. and b. at the bottom of page 1 of the worksheet, remember to show your working
6. Questions

a. Keri ran at 8 m/s for 5 seconds, how far did she run?

\[ v = 8 \text{ m/s} \quad t = 5 \text{s} \quad d=? \]

\[ d = v \times t \]

\[ d = 8 \text{ m/s} \times 5 \text{s} = 40 \text{m} \]

b. Jono drove 630 m to the shops at a speed of 7 m/s, how long did it take him?

\[ d = 630 \text{m} \quad v = 7 \text{ m/s} \quad t=? \]

\[ t = \frac{d}{v} = \frac{630 \text{m}}{7 \text{ m/s}} = 90 \text{s} \]

\[ \text{Question): How hard was it to understand what I was just teaching you?} \]

\[ \text{very very easy} \quad 0 \quad 50 \quad 100 \quad \text{very very hard} \]

\[ \text{Write your answer in the box on the sheet as a number between 1 and 100.} \]
7. Graph of Usain Bolt’s speed over 100m

Usain Bolt running 100m

8. Sketching graphs

+ Question

+ A toy car travels 40m in 4 seconds. Draw a sketch of this graph
+ Don’t use a ruler
+ Sketch the axes like example in the title of this slide
+ Add the correct numbers to each axis and draw in the line starting from (0,0)
9. Answer

Distance (m) vs. time(s)

Distance = 40 m, time = 4 s

Toy Car

Speed = slope of the line

Slope = speed
10. 3 Different average speeds

Different car’s journeys

Distance (m) vs. Time (s)

Car 1
Car 2
Car 3

0 10 20 30 40 50 60 70 80 90
0 1 2 3 4

11. What happens if we look at a real or actual journey and not average speed?

A journey of a car travelling at 50km/h then slowing down for a red light and stopping...
11. What happens if we look at a real or actual journey and not average speed?

A journey of a car travelling at 50km/h then slowing down for a red light and stopping.

12. Another real or actual car journey

A journey for a car starting from stopped at the lights getting onto the motorway.
Question: How hard was it to understand what I was just teaching you?

very very easy  |  very very hard

0  |  50  |  100

Write your answer in the box on the sheet as a number between 1 and 100.

End of teaching questions

Please fill in the A4 sheet which asks questions about the whole teaching time

Don’t forget to put your name at the top of the sheet
Post test

speed (v) = distance (d) / time (t)

distance = v × t

time = d / v

You have 10 minutes

All done
Appendix J  Student Worksheet

The Need for Speed

Slide 2. What is speed?  Example : 100km / h

= 100 kilometres per hour ( km ÷ hour or km/h)

Speed is how fast an object is travelling (how quickly something moves)

Slide 3. Usain Bolt

Speed is how far (distance) something can go in a certain time

distance = 100m  time = 10 s

Speed = distance (m) ÷ time (s)  Usain Bolt’s speed = 100m ÷ 10s = 10m/s

Slide 4. Average Speed = total distance ÷ total time

eg. 400m ÷ 80s  average speed = 5m/s

Slide 5. Calculating distance and time

speed (v) = distance (d) ÷ time (t)

distance = v x t

time = d ÷ v

Slide 6. Questions

a. Keri ran at 8m/s for 5 seconds, how far did she run?

b. Jono drove 630m to the shops at a speed of 7 m/s, how long did it take him?

Slide 8. Sketch of a toy car’s journey
7. Graph of Usain Bolt's speed over 100m

8. Sketching graphs

- Question
  - A toy car travels 40m in 4 seconds. Draw a sketch of this graph.
  - Don't use a ruler.
  - Sketch the axes like example in the title of this slide.
  - Add the correct numbers to each axis and draw in the line starting from (0,0).

9. Speed = slope of the line

10. Different average speeds

11. What happens if we look at a real journey not average speed?

12. A real car journey
Appendix K Online cognitive load question

Name__________________________  School__________
Class________________

At several different times during the teaching today you will be asked to think about how difficult it is for you to understand what I am teaching at that time. You need to choose a number from between 0 and 100 and write it in the box below

**QUESTION:** “How hard was it to understand what I was just teaching you?”

Very very easy  |  0  |  50  |  100  |  very very hard

Time 1

Time 2

Time 3

Time 4
Appendix L  Summative cognitive load questionnaire

Name _________________________  School _______________  Class____

For all the questions below think about the whole teaching time today. Give your answer as a number between 0 and 100 and write it in the grey box at the end of the scale.

**Question 1 : Understanding the ideas**

**How hard was it to understand what I was teaching you today?**
Think:  How much effort did you have to put in to understand what I was explaining?

very very easy                                very very hard

0                                              50                                              100

Write your answer here

**Question 2 : Effort to learn**

**How hard was it for you to learn and remember the science ideas?**
Think: Would I have to use the worksheet to remember the science ideas?

very very easy                                very very hard

0                                              50                                              100

Write your answer here

**Question 3 : Performance**

**How successful do you think you were in learning the ideas?**
Think: What percentage do I think I will get in the post test?

Not at all successful                                very successful

0                                              50                                              100

Write your answer here

**Question 4 : Frustration**

**How did you feel about the teaching session?**
Think: How discouraged, upset, angry or irritated did I feel?

Not at all frustrated                                very frustrated

0                                              50                                              100

Write your answer here

**Question 5 : Timing**

**How rushed did you feel the whole lesson was today?**
Think: How was the pace of the teaching and the time given to do the tasks?

far too slow                                about right                                far too rushed

0                                              50                                              100

Write your answer here
Appendix M Monitoring of participant behaviour sheet

Teacher ____________________ School______________________ Class____

In my experience a teacher new to teaching this class would find the management of this class

- [ ] very difficult
- [ ] difficult
- [ ] reasonable
- [ ] straightforward
- [ ] extremely straightforward
Appendix N  Post test

Name ___________________________  School ______________________________  Class _____

This test is to see how much you already know about the topic of speed. Write your answers on this sheet. If you are not sure of the answer to any question please leave it blank.

1. Give an example of a unit that is used to measure speed. __________________

2. What is constant speed?

3. A boy runs along a track as shown below

| section X | section Y |

c. During section X he runs at a constant speed of 2 m/s for 15 seconds. What distance does he cover?

d. During section Y he runs at 2.5 m/s and covers 12.5m. How long does this take him?

4. A child plays with a remote control car. The car starts from rest and travels a distance of 6m in 3 seconds. The car then travels a further 28m at a constant speed of 4m/s for 7 seconds. On the axes below complete a graph of the car’s journey.

Write the values of A and B on the lines below.

A=_____  
B=_____

A distance 
(m)  

Please turn over
5. A parachutist jumps from a plane and falls through a distance of 240m in 60 seconds.
   
d. Calculate her average speed.

   e. On the axes below complete a sketch graph of her 60 s fall. (axes are not drawn to scale)

   ![Graph](image)

   f. Compare the sketch graph you drew and the actual graph of her fall. Give one reason why the actual graph of her fall may not look exactly like the one you drew above in part 5b.

6. The graph below shows the speed of 3 different cars. Look at the graph and answer the questions below.

   ![Graph](image)

   a. Which car is travelling the fastest? Car ____

   b. Why?

   c. Which car is travelling at 10m/s? Car ____

   d. Why?
Appendix O  Spatial ability Cube Comparison Test S2  
(SO1 and SO2)

Part 1 (5 minutes)

DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.  
STOP.

Copyright © 1962, 1976 by Educational Testing Service. All rights reserved.
Part 2 (5 minutes)

22. B O C
    S D D

23. X H F
    S D D

24. + +
    S D D

25. J K L
    S D D

26. A M 1
    S D D

27. B C
    S D D

28. 3 V 4
    S D D

29. B W C
    S D D

30. E O I
    S D D

31. G N X
    S D D

32. Y K L
    S D D

33. R B O
    S D D

34. + 2
    S D D

35. 3 +
    S D D

36. O A
    S D D

37. H T G
    S D D

38. I J X
    S D D

39. 8 N 9
    S D D

40. L P Q R
    S D D

41. H L K E
    S D D

42. X
    S D D

DO NOT GO BACK TO PART 1 AND
DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO. STOP.

Copyright © 1962, 1976 by Educational Testing Service. All rights reserved.
Appendix P  Paper folding test VZ3 (SV1 and SV2)

Part 1 (6 minutes)

1

2

3

Go on to the next page
DO NOT GO BACK TO PART 1, AND
DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO. STOP.

Copyright © 1962 by Educational Testing Service. All rights reserved.
Appendix Q Spatial ability PowerPoint presentation

Spatial ability Tests
Instructions

Important information

These tests are NOT part of your school assessment
Do the best you can
Answer sheet

Write your
• Name
• School and
• class
On your Answer sheet

The Tests

• There are 2 tests
  Cube comparison test
  Paper folding test
• Both have 2 parts
• The tests have a time limit
Test 1: Cube comparison Test

- This test is about comparing 2 cubes or blocks to see if they are the **same** or **different**
- There are 2 parts to this test
- You have **3 minutes** to do each part
- Listen to your teacher for when to start and stop each part

Time to Practice

Look at Set A
These cubes are not the same
they are different
So you would Circle D

Look at Set B
These could be the same cube
So you would circle S
Your turn to practice

Set C

\[
\begin{array}{c}
\text{A} \\
\text{B} \\
\text{X}
\end{array}
\quad
\begin{array}{c}
\text{A} \\
\text{B} \\
\text{X}
\end{array}
\]

Set D

\[
\begin{array}{c}
\text{E} \\
\text{G}
\end{array}
\quad
\begin{array}{c}
\text{G}
\end{array}
\]

Answers
Set C = Different
Set D = Different

Your turn

Part 1 (1 page)
• You have 3 minutes

Start now
Your turn

Part 2 (1 page)
• You have 3 minutes
Start now

Test 2: Paper folding Test

• In this test you are trying to imagine how a piece of paper can be folded into a 3D shape
• The unfolded paper is on the left
• The 3D shape is on the right
• Your task is to match the numbers 1, 2, 3, 4 and 5 with the letters on the 3D shape
• Write your answers in the chart on your answer sheet
Time to Practice

Unfolded paper

3 D shape

You can use letters twice

<table>
<thead>
<tr>
<th>side</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>matches</td>
<td>H</td>
<td>B</td>
<td>G</td>
<td>C</td>
<td>H</td>
</tr>
</tbody>
</table>

Your turn

Part 1 (2 pages)
You have 6 minutes
Start now

Please use Capital letters eg. B
Your turn

Part 2 (2 pages)
You have 6 minutes
Start now

Please use Capital letters eg. B

All done
Appendix R Spatial ability Answer sheet

Spatial Ability Tests Answer Sheet
Name __________________________
School ___________________________ Class ____________

Cube comparison Test
Circle S if the cubes are the same
Circle D if the cubes are different

Part 1
1.   S    D
2.   S    D
3.   S    D
4.   S    D
5.   S    D
6.   S    D
7.   S    D
8.   S    D
9.   S    D
10.  S    D
11.  S    D
12.  S    D
13.  S    D
14.  S    D
15.  S    D
16.  S    D
17.  S    D
18.  S    D
19.  S    D
20.  S    D
21.  S    D
STOP

Part 2
22.  S    D
23.  S    D
24.  S    D
25.  S    D
26.  S    D
27.  S    D
28.  S    D
29.  S    D
30.  S    D
31.  S    D
32.  S    D
33.  S    D
34.  S    D
35.  S    D
36.  S    D
37.  S    D
38.  S    D
39.  S    D
40.  S    D
41.  S    D
42.  S    D
STOP

Paper Folding Test : Part 1

1. Side 1 2 3 4 5
   Matches

2. Side 1 2 3 4 5
   Matches

3. Side 1 2 3 4 5
   Matches

4. Side 1 2 3 4 5
   Matches

5. Side 1 2 3 4 5
   Matches

6. Side 1 2 3 4 5
   Matches

STOP

Part 2

7. Side 1 2 3 4 5
   Matches

8. Side 1 2 3 4 5
   Matches

9. Side 1 2 3 4 5
   Matches

10. Side 1 2 3 4 5
    Matches

11. Side 1 2 3 4 5
    Matches

12. Side 1 2 3 4 5
    Matches

STOP

299
Appendix S Overview of Thesis

Does Pre-training help secondary students learn complex science information?

Study One

Investigating graphing in textbooks, national exams, and science classrooms

- Textbook analysis
- National exam analysis
- Graphing questionnaire

To confirm if teaching graphing skills is a suitable context for an intervention study using pre-training

Study two

An intervention study using the strategy of pre-training

- Introductory (initial) phase
- Pilot study
  - Data collection and feedback phase
  - Changes to materials and methodology
- Main study
  - Initial phase
  - Data collection phase
  - Spatial ability testing
References


*Applied Cognitive Psychology, 26*(6), 827-832. doi:10.1002/acp.2882


doi:10.1126/science.1736359

*Trends in Cognitive Sciences, 4*(11), 417-423. doi:10.1016/S1364- 
6613(00)01538-2

doi:10.1007/BF00287729

(Eds.), *International Handbook of Science Education- Part Two* (pp. 869-893). 

Research in Science Teaching, 39*(8), 659-663. doi:10.1002/tea.10044

Baker, R., & Jones, A. (2005). How can international studies such as the international 
mathematics and science study and the programme for international student 
assessment be used to inform practice, policy and future research in science 
145-157. doi:10.1080/0950069042000276695

Research in Science Teaching, 32*(1), 3-27. doi:10.1002/tea.3660320104
Bannert, M. (2002). Managing cognitive load—recent trends in cognitive load theory. Learning and Instruction, 12(1), 139-146. doi:10.1016/S0959-4752(01)00021-4


Blayney, P., Kalyuga, S., & Sweller, J. (2010). Interactions between the isolated–interactive elements effect and levels of learner expertise: Experimental evidence


load. *Educational Psychology Review, 26*(2), 225-244. doi:10.1007/s10648-014-9262-6


Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? Learning and Instruction, 16(2), 104-121. doi:10.1016/j.learninstruc.2006.02.007


Simon, H. A. (1974). How big is a chunk? By combining data from several experiments, a basic human memory unit can be identified and measured. *Science, 183*(4124), 482-488. doi:10.1126/science.183.4124.482


