Evaluating Learning Theory-based Methods for Improving the Learning Outcomes of Introductory Statistics Courses

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

James Baglin

B.App.Sci (Psych)(Hons)

School of Mathematical and Geospatial Sciences

College of Science, Engineering and Health

RMIT University

June 2013

Declaration

I, James Baglin, declare that:

- except where due acknowledgement has been made, the work of this thesis is my own
- the work in this thesis has not been submitted previously, in whole or in part, to qualify for any other academic award
- the content of this thesis is the result of work which has been carried out since the official commencement date of the approved research program
- any editorial work, paid or unpaid, carried out by a third party has been acknowledged
- and any ethics procedures and guidelines have been followed.

James Baglin

June 2013

To my pop, the last true gentleman.

– Brian J. Walsh (Oct 25, 1934 - Aug 19, 2010)

Acknowledgements

I would like to acknowledge a number of very important people that contributed to and enabled the completion of this dissertation.

Firstly, I must acknowledge my primary supervisor, Associate Professor Cliff Da Costa. His support, wisdom and guidance throughout my candidature will never be forgotten. Cliff is a role model I will always admire. I am very fortunate and honoured to have been his student since the first year of my undergraduate degree. I must also thank my secondary supervisor, Associate Professor Anthony Bedford, for sharing his positivity when times were uncertain and keeping me busy in my "spare" time. Anthony's unwavering faith in my ability to complete this dissertation and the opportunities that he has provided to advance my early career will also never be forgotten.

Throughout my candidature I have been fortunate enough to work alongside and learn from some amazing people. I must thank Dr. Michael Bulmer and Associate Professor Matthew Linden for their collaboration. The opportunities and experience I have gained working with these brilliant people has been invaluable. I must also acknowledge the support and time of Matthew Ovens whom I spent countless hours with discussing statistics education research and learning from his extensive range of practical knowledge.

The tutors and administrative staff of the School of Mathematical and Geospatial Science have also been incredibly supportive over the years. Doing educational research is very challenging, but having their support made everything a little easier. I must also acknowledge the work of all the blind reviewers of the publications that arose from this dissertation. Each reviewer has helped contribute to and improve the work that is presented within.

To my family and friends, thank you for all your unconditional support. We have had some tough times throughout my journey, but never have I felt alone. Most importantly, I must acknowledge the support of my beautiful wife, Courtney. Words can't describe how much gratitude I feel. Courtney was there by my side every step of the way. Her support has been immeasurable.

Last, but certainly not least, I must thank all the selfless students who took the time to participate in my research. Without them, none of this dissertation would have been possible.

Abstract

Modern introductory statistics courses continue to evolve in order to reflect the progress of statistics education and the needs of modern students. Many of these developments relate to an increase in the use of technology and innovative teaching and assessment practices. However, while many of these changes have been informed by learning theories and extensive teacher experience, their efficacy has not been thoroughly evaluated. This thesis reports the findings of three major projects that have evaluated theory-based interventions aimed at improving the key learning outcomes of introductory statistics courses, namely statistical literacy, reasoning and thinking.

In Part I, the important topic of technological skills in statistics education is examined. As technology has become an inseparable part of modern statistical practice (Gould, 2010), so too has it become an integral part of modern notions of statistical literacy. From this perspective, understanding the development of technological skills in statistics education becomes a priority. Unfortunately, very little is known about the development of these skills (Chapter 3). Part I compares the effect of two different training methods, Error-management Training (EMT)and Guided Training (GT) on the development of students' ability to operate statistical packages. EMT is based on active-exploratory training principles where students develop skills through actively exploring a task domain (Dormann & Frese, 1994). Active-exploration is prompted by the use of minimal instruction. GT, on the other hand, is a passive form of training where students' proficiency is developed through comprehensive guided step-by-step instructions (Chapter 3, Keith & Frese, 2008). Previous studies in general software training (e.g. training to use word processors, spreadsheets and presentation software) suggested that EMT is superior to GT in promoting students' ability to adaptively transfer their skills outside of the training environment (Keith & Frese, 2008). A pilot

study was conducted to initially evaluate the feasibility of delivering statistical package training using minimal instructions required by EMT (Chapter 4). The pilot was conducted using a sample of 13 science and business university students who had previously completed an introductory statistics course.

Following the success of the pilot, Trial I compared EMT to GT using an explanatory mixed methods approach in a sample of 100 university psychology students enrolled in an introductory statistics course (Chapter 5). The quantitative phase of Trial I used a randomised experiment embedded in the course to compare measures of training transfer between students assigned to fortnightly EMT or GT for learning to operate the statistical package *SPSS*. The second qualitative phase used 15 in-depth interviews to help explain the quantitative results and explore the overall student experience of the statistical package training sessions (Chapter 6). While the quantitative results of Trial I were inconclusive, a thorough evaluation of Trial I laid the foundation for a second trial in the same course the following year. Trial II addressed the major limitations of Trial I using a quasi-experimental design in a sample of 115 psychology students (Chapter 7). EMT and GT were compared between two campuses of the same introductory statistics course. After controlling for important covariates, no difference in students' development of statistical package skills was found between the two training strategies. The outcomes of this series of studies suggested that other factors appeared to be playing a more important role than training strategies in the development of technology skills in statistics education.

In Part II of the dissertation, cognitive conflict strategies were evaluated for improving students' statistical reasoning by confronting students' misconceptions. Cognitive conflict strategies are designed to promote conceptual change by presenting contradictory information and replacing students' faulty conceptualisations with more scientifically valid understandings (Chapter 9, Limón, 2001). Cognitive conflict interventions had been identified by previous studies in statistics education as a promising method for reducing misconceptions related to a wide range of misunderstandings (e.g. Kalinowski, Fidler, & Cumming, 2008; Jazayeri, Lai, Fidler, & Cumming, 2010; Liu, Lin, & Kinshuk, 2010). Part II evaluated the use of brief conceptual change-based activities embedded in lectures for confronting a wide variety of misconceptions across the semester of an introductory statistics course for medical science students (Chapter 10). The study was conducted over two years on two separate student cohorts with a total sample size of 328. In the control cohort, baseline measures of statistical reasoning and misconceptions were included in an end of semester multiple choice exam. In the following year, the intervention cohort received eight brief cognitive conflict-based activities embedded in lectures and also completed the same select multiple-choice questions in the exam. The results of the study found two of the eight activities were associated with a statistically significant improvement in students' statistical reasoning. The results also suggested that the complexity of the misconception being targeted is likely to moderate the effect of a "brief" intervention format. More pervasive and difficult to change misconceptions related to statistical inference require longer and more intensive interventions.

Part III of the dissertation evaluated the impact of project-based learning on the development of statistical thinking. Project-based learning (PBL) is a form of experiential learning which is based on the concept of learning by doing (Blumenfeld et al., 1991). PBL has been used to help develop statistical thinking by engaging students in the entire data investigative cycle of statistical enquiry (MacGillivray, 2010; MacGillivray & Pereira-Mendoza, 2011; Snee, 1993). As a consequence of the difficulty of defining and assessing statistical thinking, empirical evidence of this proposed link is lacking (Chapter 12). In Study I an online virtual environment called the *Island* was first validated as a tool for delivering PBL in an online masters level introductory biostatistics course (Chapter 13). The quantitative and qualitative results of 42 student surveys and 5 in-depth interviews confirmed the validity of using the *Island* for PBL and provided qualitative evidence of the theoretical link between PBL and statistical thinking. In Study II this proposed link was initially tested using an experimental design. Participants from a large introductory statistics course for science students were randomly allocated as individuals or in small groups to complete two different types of research designs, observational or experimental, for an *Island*-based course project (Chapter 14). Study II hypothesised that a student's ability to think statistically about different research designs would depend on the project type they were allocated. Towards the end of the semester, 356 students completed a test of statistical thinking about experimental and observational studies. The results of Study II found that performance on the test of statistical thinking did not depend on students' allocated project type. While this study found inconclusive evidence of the proposed link between PBL and the development of statistical thinking, the outcomes of this study highlighted a number of major challenges facing this area of research.

The outcomes of these major parts provide valuable insight into the importance of evaluation research in statistics education and the challenges it presents to researchers. The findings discussed build upon statistics education research and suggest promising directions for future research.

List of Figures

3.1 Kanfer and Ackerman's Concepts of Task Difficulty and the Effect of Practice 21

4.1 Pilot Study Dot plots showing the distribution of each outcome variable between conditions. 36

5.1 Trial I flow chart. Note. RA = Randomly allocated, NRA = Not randomly allocated, DNC = Did not consent, EMT = Error-management training, GT = Guided Training, SA = Completed self-assessment 1 & 2. Semester 2 follow-up has been shaded. 43

5.2 Example of GT and EMT exercise instructions in *WebLearn* - Trial I . . 57

5.3 Trial I Covariate adjusted training transfer means with 95% CI between training approaches and across semesters. 62

5.4 Trial I mean ratings with 95% CI for training manipulation checks between trainng approaches. 65

6.1 A thematic map of the qualitative analysis of Trial I 74

7.1 Trial II Mean ratings with 95% CI for training manipulation checks between training strategies. Trial I results are included for comparison. 101

10.1 An example of a cognitive conflict-based activity for overcoming misconceptions of probability caused by equiprobability bias 124

10.2 Proportion of correct responses between cohorts for each of the conceptual change questions. Error bars show 95% CI for proportions. 130

10.3 Confidence Intervals I (Q7) Response Distributions 133

10.4 Probability (Q17) Response Distributions 134
10.5 Regression (Q37) Response Distributions 135

12.1 The *Island* (Bulmer & Haladyn, 2011) 151
12.2 Town and Islanders 152
12.3 The *Island* Interface 154

14.1 Study flow-chart. RA = Randomly allocated, TST = Test of Statistical Thinking 169
14.2 Adjusted Means with 95% *CI* for TST Experimental and Observational Subscale Scores Between Project Condition Allocation 174

List of Tables

1.1 A Comparison of the Revised Version of Bloom's Taxonomy to the Statistics Education Learning Outcomes 7

3.1 An Example of a Student's Metacognition for an Adaptive Transfer Task 27

4.1 Pilot Sample Group Characteristics 32
4.2 Descriptive Statistics Between Conditions on Outcome Variables 35
4.3 Pilot Study Student Question Themes and their Frequency Between Training Approaches 38

5.1 Statistical Package Training Trial I Sample Characteristics Between approaches 43
5.2 Trial I Analogical Transfer Self-Assessment 1 Exercises 48
5.3 Trial I Adaptive Transfer Self-Assessment 2 Exercises 50
5.4 Trial I Training Laboratory Schedule Across Campus and Condition . . 52
5.5 Trial I Statistical Package Training Laboratory Content 53
5.6 Trial I Intercorrelations and Descriptive Statistics of Study Variables Between Training Approaches 60
5.7 Trial I ANCOVA Models Predicting Training Transfer Measures 61
5.8 Descriptive Statistics and Two-sample t-tests Comparing Mean Ratings of Manipulation Checks between Training Approaches - Trial I 64
5.9 Descriptive Statistics and Two-sample t-tests Comparing Training approaches on Other Training Outcomes - Trial I 66

6.1 Interviewee Characteristics 74

7.1 Trial II Descriptive Statistics and Independent Samples t-tests Comparing Mean Ratings of Manipulation Checks between Training Approaches 100
7.2 Trial II Intercorrelations and Descriptive Statistics of Study Variables Between Training Approaches 102
7.3 ANCOVA Model Parameters Predicting Adaptive Transfer 104
7.4 Trial II Descriptive Statistics and Independent-samples t-tests Comparing Training approaches on Other Training Outcomes 106

10.1 Cohort Demographics 122
10.2 Course Topics, Conceptual Change-based Activity Schedule and Corresponding Conceptual Change Multiple Choice Exam Questions 126
10.3 Descriptive Statistics Across Cohorts and Programs 128
10.4 ANCOVA Model Parameters 129
10.5 Univariate and Multivariate Logistic Regression Models Predicting Correct Responses on Conceptual Change Questions 131

13.1 Eight Examples of Student Project Topics 158
13.2 *Island* Questionnaire Descriptive Statistics (Both Semesters Combined) 160

14.1 Mapping Wild and Pfannkuch's 1999 Fundamental Components of Statistical Thinking to *Island*-based Project Outcomes 170
14.2 Principal Components Analysis of the Test of Statistical Thinking . . . 172
14.3 Descriptive Statistics for Project Types and Intercorrelations Between Variables 173
14.4 ANCOVA Models Predicting TST Scores for Observational and Experimental Subscales 174

Contents

Declaration **i**

Acknowledgements **iii**

Abstract **v**

List of Figures **ix**

List of Tables **xi**

1 Overview **1**

1.1 The Field of Statistics Education and Statistics Education Reform . . . 2

1.2 The Challenge of Evaluation Research in Statistics Education 4

1.3 The Learning Outcomes of Statistics Education 6

1.4 A Brief Overview of the Major Parts 9

I Active-exploratory Training for Technological Skills **13**

2 Part I - Abstract **14**

3 Part I - Introduction **18**

3.1 Technological Skills in Statistics Education 18

3.2 Kanfer and Ackerman's Integrative Model of Skill Acquisition 20

3.3 Active-exploratory Training Approaches 22

3.4 Statistical Package Skills 29

3.5 Rationale and aims 30

4 Part I - Pilot Study **31**

4.1 Aims of the Pilot . . . 31

4.2 Method . . . 31

4.3 Results . . . 34

4.4 Discussion . . . 37

4.5 Conclusions . . . 39

5 Part I - Trial I - Quantitative Phase **41**

5.1 Rationale and Aims . . . 41

5.2 Method . . . 42

5.3 Results . . . 58

5.4 Discussion . . . 66

5.5 Conclusion . . . 71

6 Part II - Trial I - Qualitative Phase **72**

6.1 Rationale and Aims . . . 72

6.2 Method . . . 73

6.3 Qualitative Results and Discussion . . . 73

6.4 Conclusion . . . 85

7 Part I - Trial II **88**

7.1 Aims of Trial II . . . 88

7.2 Method . . . 89

7.3 Results . . . 98

7.4 Discussion . . . 106

7.5 Conclusion . . . 111

II Cognitive Conflict for Correcting Misconceptions **112**

8 Part II - Abstract **113**

9 Part II - Introduction **115**

9.1 Misconceptions and Cognitive Conflict for Conceptual Change . . . 115

9.2 Studies on Cognitive Conflict in Statistics Education 116
9.3 Rationale and Aims 119

10 Part II - Trial 121

10.1 Aim of Trial 121
10.2 Method 121
10.3 Results 127
10.4 Discussion 134
10.5 Conclusion 139

III Experiential Learning for Developing Statistical Thinking 141

11 Part III - Abstract 142

12 Part III - Introduction 144

12.1 Experiential Learning and Project-based Learning 144
12.2 Statistical Thinking 145
12.3 Project-based Learning for Statistical Thinking 148
12.4 The *Island* 150
12.5 Rationale and Aims 153

13 Part III - Study I 156

13.1 Aim of Study I 156
13.2 The Course 156
13.3 Method 158
13.4 Results and Discussion 159
13.5 Conclusion 165

14 Part III - Study II 166

14.1 Aim of Study II 166
14.2 Method 167
14.3 Results 172
14.4 Discussion 175

14.5 Conclusion 178

Conclusion **180**

Publications Arising **183**

References **184**

Appendices **201**

A Part I **201**

A.1 Pilot Questionnaire 202

A.2 Pilot Plain Language Statement 204

A.3 Pilot Consent Form 206

A.4 Pilot Training Exercise Examples 207

A.5 Trial I - Consent Form 211

A.6 Trial I - PLS 212

A.7 Trial I - Questionnaire Psychometrics 214

A.8 Trial I - Qualitative Interview Schedule 217

A.9 Trial I - Qualitative Interview PLS 218

A.10 Trial II - PLS 220

A.11 Trial II - Consent Form 222

A.12 Trial II - Performance Utility 223

A.13 Trial II - *SPSS* Certification Tasks 224

A.14 Trial II - *SPSS* Certification Task Scoring 248

B Part II **251**

B.1 Cognitive Conflict Study PLS and Consent 252

B.2 Multiple Choice Exam Questions 254

B.3 Adapted Multiple Choice Items 269

B.4 Cognitive Conflict-based Activity Slides 272

B.5 Conceptual Change Question Response Patterns 291

C Part III **297**

C.1 Study I - Plain Language Statement and Consent 298

C.2 Study I - Questionnaire 299

C.3 Study I - Semi-structured Interview Schedule 303

C.4 Study II - Participant Information Sheet 304

C.5 Study II - Consent Form 306

C.6 Study II - Test of Statistical Thinking 307

C.7 Study II - TST Grading Scheme - Observational 308

Chapter 1

Overview

This dissertation is split into three major parts reflecting three separate but related bodies of work. While each major part can be read independently in any particular order, this overview will highlight the common themes and rationale that connect each part. This dissertation deals with the evaluation of learning theory-based methods for improving the learning outcomes of introductory statistic courses. The studies reported herein are posited in the emerging field of statistics education which can be broadly defined as any research, both quantitative and qualitative, concerned with the "learning, teaching and assessment of statistical methods or statistical thinking" (Jolliffe, 2003, p. 49). The main theme driving the rationale for this dissertation is the evaluation of modern teaching and learning practices used in statistics education that, while based on sound learning theory and extensive teaching experience, still, to this day, lack extensive empirical verification (delMas, 2002). The context of these studies, the introductory statistics course, is also common across the three major parts. Introductory statistics courses are on the increase (Garfield & Ben-Zvi, 2005) as the importance of quantitative skills continues to grow in a world beset by a digital "data deluge". However, there are significant challenges facing the delivery of these courses and achieving positive student outcomes. The final theme of this dissertation is the many learning outcomes of statistics education that must be developed in students. These outcomes can be can be broadly categorised into three major areas - *statistical literacy*, *reasoning* and *thinking*. Each major part of this dissertation addresses learning outcomes under one of these domains. However, before the beginning of each major part, a brief overview of the themes will be presented.

1.1 The Field of Statistics Education and Statistics Education Reform

Statistics education is a research field in its own right, and is of vital importance to education, statistics, and any other discipline that uses statistics (Jolliffe, 1998; Ben-Zvi & Garfield, 2008). Statistics education research is multidisciplinary with contributions coming mainly from the fields of mathematics, statistics, education and psychology (Jolliffe, 2003; Ottaviani, 2005). The field has expanded substantially over the last couple of decades. Since 2000, over 83 dissertations in statistics education have been listed on the *International Association for Statistical Education* website (IASE, 2012). Today, the field of statistics education is served by several key professional associations acting in the interest of the community to help improve statistics education. Peak bodies include the IASE of the *International Statistical Institute* (ISI), which was established in 1991 and preceded by the *Education Committee* (1948 - 1991) (Vere-Jones, 1995) and the *Royal Statistical Society Centre for Statistical Education* (RSSCSE).

The continued expansion of statistics education has resulted in the establishment of a number of dedicated statistics education research outlets. The first peer-reviewed journal in statistics education, *Teaching Statistics*, was first published in 1978 by the *Teaching Statistics Trust* of the RSSCSE. This journal focused on the teaching of statistics to school-aged children (Ages 9 - 19). In 1993, the *American Statistical Association* published the first volume of the online *Journal of Statistics Education.* In 2002, statistics education research was further solidified in the literature with the publishing of the first edition of the *Statistics Education Research Journal* (SERJ) by the IASE and ISI. More recently in 2007, the journal *Technology Innovations in Statistics Education* (TISE) was established. Other dedicated statistics journals, such as *Journal of the Royal Statistics Society*, *American Statistician* and the *International Statistics Review*, also regularly publish special sections and articles related to statistics education.

Statistics education research conferences abound both nationally and internationally. The largest of its kind, the IASE's *International Conference on Teaching Statistics* (ICOTS), convenes every four years and has been running since 1982. The IASE also organises regular satellite meetings and special interest sessions around the ISI's *World Statistics Congress*. Other meetings and sessions by IASE include Round Table meetings and statistics education sessions at the *International Congress on Mathematical Education*. Examples of national conferences include the *Australian Conference on Teaching Statistics* (OZCOTS) the *United States Conference on Teaching Statistics* (USCOTS), and the section on statistical education of the the *Joint Statistical Meetings* of the ASA. All these conferences have played a part in enabling the continual development and progression of statistics education as a legitimate area of inquiry.

While cross-cultural variation must be kept in mind (see Vere-Jones, 1998), on the whole, modern statistics courses have continued to experience significant change in terms of their content and teaching practices. Today's modern courses are a stark contrast to their mathematically driven predecessors (Garfield, 2003). The *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) *Project* report outlines a modern course as one which emphasises statistical literacy and thinking, utilises real data, develops conceptual understanding, uses active learning, takes advantage of technology, and uses assessment to enhance learning. The modern course has evolved over the last couple of decades from the significant work of the international statistics education community and the increased accessibility of technology. Key stages of this evolution are strongly reflected in reports from the U.S. on statistics education reform (Cobb, 1992; American Statistical Association, 2005). In more recent times, technology continues to play a major role in shaping the modern course. This trend is most evident in the work of Wild, Pfannkuch, Regan, and Horton (2011) who propose to use dynamic computer visualisations and simulation for developing more accessible conceptions of statistical inference and the works of Gould (2010) and Nolan and Temple Lang (2010a) who raise the importance of technological literacy for the modern student.

Surveys from the U.S. also support the integral role of technology shaping the delivery of modern introductory statistics courses. Garfield, Hogg, Schau, and Whittinghill (2002) surveyed 243 U.S. statistics instructors from psychology, sociology, business, and economics backgrounds on the teaching practices in their courses. Approximately 50%

of respondents reported making use of technology in their courses. The respondents also indicated that many if their planned changes for the future also related to the further use of technology. A decade later, a survey of 227 instructors also from the U.S. found that the proportion of instructors using technology in their courses had risen, as predicted, to 76% (Hassad, 2012). Overall, the results of these surveys indicated the significant role of technology in the introductory statistics course.

Regardless, statistics education still faces many challenges. Much more research is still needed to evaluate the impact of changes brought about by the evolution of the modern introductory course, particularly in evaluating effective learning methods (Chance & Garfield, 2001). As delMas (2002) explains, while many of the recent changes in statistics education teaching and learning are based on learning theory and the expert knowledge of highly experienced instructors, many of these practices lack careful empirical evaluation. The most likely reason being the significant challenges associated with statistics education evaluation research.

1.2 The Challenge of Evaluation Research in Statistics Education

Research in statistics education is confronted by numerous challenges. While many of these challenges are not unique to statistics education, they are important to raise as they impact directly on the nature of research in the field. Very little is known and published on the methodology of research in statistics education (Jolliffe, 1998). A number of common challenges related to practical, ethical and assessment issues exist within the field. In terms of ethics and feasibility, the "gold-standard" of scientific research, the randomised-controlled experiment, is particularly difficult to implement (Jolliffe, 1998). Even when randomisation is possible, it is not always achievable in the typical semester-long investigations of a teaching and learning intervention (Chance & Garfield, 2001). The process of randomisation is also seen by many ethical bodies as a risk that could result in students in certain conditions being disadvantaged (Chance & Garfield, 2001), and, as such, students must be able to opt out or self-select. Opting-out, drop-outs and self-selection all impose serious threats to the internal validity of an educational experiment.

Non-randomised, observational or “quasi-experimental” studies overcome the ethical issues associated with randomisation, but at the drawback of being unable to directly control for pre-existing group differences. Thus, statistics education researchers are often faced with only being able to compare different methods between course cohorts (Jolliffe, 2003). While extraneous variables which may exert influences on group differences can be controlled for, they must be known in advance of the study and carefully measured or controlled. Research suggests that non-randomised studies which take proper control of known covariates provide reliable estimates of causal effects (Shadish, Clark, & Steiner, 2008; K. Benson & Hartz, 2000; Concato, Shah, & Horwitz, 2000). However, on a practical note, controlling for covariates that may confound the results of a study, might be difficult given institutional restrictions, limited resources and ethical constraints. Instructor changes, access to learning resources, classrooms, and computers are often outside the direct control of the statistics education researcher. Incorporating these issues in addition to the usual challenges associated with applied human research, such as non-compliance, drop-outs, and blinding, it emerges that statistics education researchers have their work cut out.

Regardless of these challenges, statistics education research must continue. The field draws on a broad range of research methodologies including both quantitative and qualitative modes of inquiry. Examples of these methods include, case studies, longitudinal studies, observational studies, and interviews (Garfield & Ben-Zvi, 2005). When statistics education researchers are considering research methods, the single most important consideration is to match the methodology with the research question being posed (Chance & Garfield, 2001). Designing a perfect study that unequivocally answers a particular research question is unrealistic. Instead, statistics education researchers must often settle for what is practical, ethical and achievable. Evidence is unlikely to come from one well designed study, but instead come from multiple independent lines of inquiry (Chance & Garfield, 2001). The field has made substantial inroads into this goal with the establishment of common learning outcomes which guide course assessment and lay the foundations for evaluation of outcomes of teaching and learning interventions.

1.3 The Learning Outcomes of Statistics Education

While there is no single agreed upon distinction between the major learning outcomes of statistics education (Ben-Zvi & Garfield, 2005), the concepts of statistical *literacy*, *reasoning* and *thinking* have been put forth as a useful framework for organising the these outcomes. These three "levels" have been likened to Bloom's (1956) hierarchical taxonomy of education objectives (Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluate) and the revised taxonomy by Anderson et al. (2000, Remember, Understand, Apply, Analyse, Evaluate and Create). The Revised Bloom's Taxonomy can be used to help define each learning outcome (Table 1.1). From here on, the Revised Bloom's taxonomy will be referred to as simply Bloom's taxonomy. These outcome levels structure this dissertation as each major part targets a different learning outcome that can be placed under one of these levels. However, due to the complexity and interrelationships that no doubt exist between each level, it must be cautioned that these levels are only used in a semantic sense to help organise the learning outcomes of an introductory statistics course, as opposed to providing concrete representations of independent domains of knowledge. For example, Marriott, Davies, and Gibson (2009) show how all levels of Bloom's Taxonomy can be applied to the statistical problem solving approach believed to be at the center of statistical thinking (Wild & Pfannkuch, 1999).

According to Rumsey (2002), *statistical literacy* is composed of an understanding of data, statistical concepts, statistical terminology, methods of data collection, computation of descriptive statistics, basic interpretation skills, and basic statistics communication skills. It also includes an understanding of using probability as a measure of uncertainty (Ben-Zvi & Garfield, 2005). The ability to use statistical technology, such as statistical packages, is also shaping modern notions of statistical literacy (e.g. Gould, 2010). Rumsey argued that statistical literacy is the foundation of the more higher level abilities of statistical reasoning and statistical thinking. Statistical literacy would occupy the first level, *Remembering*, of Bloom's taxonomy (Garfield, delMas, & Zieffler, 2010). Common words associated with the assessment of statistical literacy include, "identify", "describe", "interpret" and "compute" (Garfield, delMas, & Zieffler,

Table 1.1: A Comparison of the Revised Version of Bloom's Taxonomy to the Statistics Education Learning Outcomes

Bloom's Taxonomy Revised[1]	Definition	Verbs	Statistics Learning Outcome
1. Remember	Retrieval of relevant knowledge	Recognise Recall	Statistical Literacy
2. Understand	Determine the meaning of instructional messages	Interpret Exemplify Classify Summarise Infer Compare Explain	Statistical Reasoning
3. Apply	Carry out or use a procedure	Execute Implement	Statistical Thinking
4. Analyse	Breaking material into its constituent parts and detecting how the parts relate to one another and the overall structure and purpose	Differentiate Organise Attribute	
5. Evaluate	Making judgements based on criteria and standards	Checking Critiquing	
6. Create	Putting elements together to form novel, coherent whole or make original product	Generating Planning Producing	N/A

[1] Anderson et al. (2000)

2010). Some of these words reflect verbs associated with the *Understand* outcome of Bloom's taxonomy which reflect a relationship to statistical reasoning.

Garfield and Gal (1999b), defined *statistical reasoning* as the thought processes people utilise to understand statistical inference. Garfield (2002) further defined this concept as being the product of the conceptual understanding of the important statistical concepts of distributions, central tendency, variation, association, uncertainty, randomness, and sampling. Statistical reasoning can be assessed through asking question relating to the "how" and "why" of statistics (delMas, 2002; Garfield, delMas, & Zieffler, 2010). Statistical reasoning moves beyond Bloom's *Remembering* domain and into *Understanding* (Garfield, delMas, & Zieffler, 2010). Statistical reasoning has been the subject of much research in the statistics education field (Garfield, 2002, 2003) as it provides valuable insight into the way people make decisions and judgements using statistics (Garfield, 2002).

Statistical thinking involves a more advanced way of thinking compared to statistical reasoning (Garfield, delMas, & Zieffler, 2010). Chance (2002) concluded from a review of the literature that statistical thinking is largely an understanding of what a statistician does. Chambers (1993), and later Cameron (2009), are more specific, listing five categories of work characteristic of being a statistician. These include the following:

1. Preparing data, including planning, collection, organisation and validation
2. Analysing data, by models or other summaries
3. Presenting data in written, graphical or other form
4. Formulating a problem so that it can be addressed through statistical means
5. Carrying out research to develop new statistical methods

These conceptions of statistical thinking attempt to reflect the real way a statistician problem solves with data. Similar models have been adapted in statistic education to capture the essential features of what is referred to as the data investigative process. Wild and Pfannkuch (1999) adopt the PPDAC (Problem, Plan, Data, Analysis, Conclusions) model proposed by MacKay and Oldford (1994, as cited in Wild & Pfannkuch,

1999) for explaining this problem-solving approach (PSA). Marriott et al. (2009) use a similar framework, but with only four stages: Specify the problem and plan, collect data, process and represent data, interpret and discuss (PCPD).

In the most comprehensive treatment of this topic, Wild and Pfannkuch (1999) also propose five types of thinking fundamental to statistical thinking - recognising the need for data, transnumeration, consideration of variation, reasoning with statistical models, and integrating the statistical and contextual. *Recognising the need for data* refers to the understanding that data beats anecdote. *Transumeration* involves the process of finding appropriate data and then transforming the data into information that improves our understanding of a phenomena under investigation. *Consideration of variability* is the understanding of the omnipresence of variability in data and how this variability leads to uncertainty. *Reasoning with statistical models* is an understanding of the models that statisticians use to gain knowledge. Finally, *Integrating the statistical and contextual* refers to the ability to synthesise the context of a problem or study with the statistical analysis. Taking all these various explanations of statistical thinking into account, Bloom's domains of *Apply*, *Analyse*, and *Evaluate* all constitute what it means to think statistically (Jolliffe, 2010).

Statistical thinking seems to be as difficult to assess as it is to define. Chance (2002) claimed that "evidence of statistical thinking lies in what students do spontaneously, without prompting or cue from the instructor" (p. 130). This implies that statistical thinking is not amenable to traditional assessment items (e.g. multiple choice exams). Consequently, many researchers have proposed innovative models for developing and assessing statistical thinking that have mainly come in the form of problem-based learning (Bowman & Gilmour, 1998; Marriott et al., 2009) and project-based learning (MacGillivray, 2010; MacGillivray & Pereira-Mendoza, 2011; Snee, 1993).

1.4 A Brief Overview of the Major Parts

Each major part of this dissertation is connected by the themes of "statistics education", "introductory statistics courses", "evaluation of learning theory-based methods", and the "outcomes of statistics education". Each major part can be read as a stand-alone body of research, and, therefore, can be read in any particular order. However, the parts have been structured in terms of the previously discussed levels of statistics education learning outcomes - statistical literacy, reasoning and thinking. Each major part is outlined here to provide a brief overview of the theories and studies that comprise it.

Part I of the dissertation deals with the important topic of technological skills in statistics education. As technology has become an inseparable part of modern statistical practice (Gould, 2010), so too has it become an integral part of modern notions of statistical literacy. From this perspective, understanding the development of technological skills in statistics education becomes a priority. Unfortunately, very little is known about the development of these skills (Chapter 3). Part I compares the effect of two different training methods, Error-management Training (EMT)and Guided Training (GT) on the development of students' ability to operate statistical packages. EMT training is based on active-exploratory training principles where students develop skills through actively exploring a task domain (Dormann & Frese, 1994). Active-exploration is prompted by the use of minimal instruction. GT, on the other hand, is a passive form or training where students' proficiency is developed through comprehensive guided step-by-step instructions (Chapter 3, Keith & Frese, 2008). Previous studies in general software training (e.g. training to use word processors, spreadsheets and presentation software) suggested that EMT is superior to GT in promoting students' ability to adaptively transferring their skills outside of the training environment (Keith & Frese, 2008). A pilot study was conducted to initially evaluate the feasibility of delivering statistical package training using minimal instructions required by EMT (Chapter 4). The pilot was conducted using a sample of 13 science and business university students who had previously completed an introductory statistics course.

Following the success of the pilot, Trial I compared EMT to GT using an explanatory mixed methods approach in a sample of 100 university psychology students enrolled in an introductory statistics course (Chapter 5). The quantitative phase of Trial I used a randomised experiment embedded in the course to compare measures of training transfer between students assigned to fortnightly EMT or GT for learning to operate the statistical package *SPSS*. The second qualitative phase used 15 in-depth interviews to help explain the quantitative results and explore the overall student experience of the statistical package training sessions (Chapter 6). While the quantitative results of Trial I were inconclusive, a thorough evaluation of Trial I laid the foundation for a second trial in the same course the following year. Trial II addressed the major limitations of Trial I using a quasi-experimental design in a sample of 115 psychology students (Chapter 7). EMT and GT were compared between two campuses of the same introductory statistics course. After controlling for important covariates, no difference in students' development of statistical package skills was found between the two training strategies. The outcomes of this series of studies suggested that other factors appeared to be playing a more important role than training strategies in the development of technology skills in statistics education.

In Part II of the dissertation, cognitive conflict strategies were evaluated for improving students' statistical reasoning by confronting students' misconceptions. Cognitive conflict strategies are designed to promote conceptual change by presenting contradictory information and replacing students' faulty conceptualisations with more scientifically valid understandings (Chapter 9, Limón, 2001). Cognitive conflict interventions had been identified by previous studies in statistics education as a promising method for reducing misconceptions related to a wide range of misunderstandings (e.g. Kalinowski et al., 2008; Jazayeri et al., 2010; Liu et al., 2010). Part II evaluated the use of brief conceptual change-based activities embedded in lectures for confronting a wide variety of misconceptions across the semester of an introductory statistics course for medical science students (Chapter 10). The study was conducted over two years on two separate students cohorts with a total sample size of 328. In the control cohort, baseline measures of statistical reasoning and misconceptions were included in an end of semester multiple choice exam. In the following year, the intervention cohort received eight brief cognitive conflict-based activities embedded in lectures and also completed the same select multiple-choice questions in the exam. The results of the study found two of the eight activities were associated with a statistically significant effect on improving students statistical reasoning by reducing misconceptions. The results also suggested that the complexity of the misconception being targeted is likely to moderate the effect of a "brief" intervention format. More pervasive and difficult to change misconceptions related to statistical inference require longer and more intensive interventions.

Part III of the dissertation evaluated the impact of project-based learning on the development of statistical thinking. Project-based learning (PBL) is a form of experiential learning which is based on the concept of learning by doing (Blumenfeld et al., 1991). PBL has been used to help develop statistical thinking by engaging students in the entire data investigative cycle of statistical enquiry (MacGillivray, 2010; MacGillivray & Pereira-Mendoza, 2011; Snee, 1993). As a consequence of the difficulty of defining and assessing statistical thinking, empirical evidence of this proposed link is lacking (Chapter 12). In Study I an online virtual environment called the *Island* was first validated as a tool for delivering PBL in an online masters level introductory biostatistics course (Chapter 13). The quantitative and qualitative results of 42 student surveys and 5 in-depth interviews confirmed the validity of using the *Island* for PBL and provided qualitative evidence of the theoretical link between PBL and statistical thinking. In Study II this proposed link was initially tested using an experimental design that randomly allocated individuals or small groups of students in a large introductory statistics course for science students to complete two different types of research designs, observational or experimental, for an *Island*-based course project (Chapter 14). Study II hypothesised that a student's ability to think statistically about different research designs would depend on the project type they were allocated. Towards the end of the semester, 356 students completed a test of statistical thinking about experimental and observational studies. The results of Study II found that performance on the test of statistical thinking did not depend on students' allocated project type. While this study found inconclusive evidence of the proposed link between PBL and the development of statistical thinking, the outcomes of this study highlighted a number of major challenges facing this area of research.

Part I

Active-exploratory Training for Technological Skills

Chapter 2

Part I - Abstract

As technology has become an inseparable part of modern statistical practice, so too has it become an integral part of modern notions of statistical literacy (Gould, 2010). From this perspective, understanding the development of technological skills in statistics education becomes a key priority. The best example is perhaps the important ability to use a statistical package. Unfortunately, very little is known about the development of these skills. This first major part of the dissertation reports the results of a series of studies investigating the impact of training approaches on the development of training transfer. Training transfer is evident in students who are able to transfer their skills outside the training environment (Hesketh, 1997). There are two major types of transfer, *analogical* and *adaptive*. Analogical transfer is the ability to transfer the same skills covered in training outside of the training environment, whereas, adaptive transfer is the ability to adapt one's skills to confront novel situations (Keith, Richter, & Naumann, 2010). Adaptive transfer is considered the most desirable outcome given the brevity of most training programs. This first major part compares two types of training approaches prevalent in the training literature, Error-management training (EMT) and Guided Training (GT). EMT is based on active-exploratory training (AE-T) principles where students develop skills through active exploration prompted by the use of minimal instruction (Dormann & Frese, 1994). GT, on the other hand, is a passive form of training where students' proficiency is developed through comprehensive, guided, step-by-step instructions (Keith, Richter, & Naumann, 2010). Previous studies in general software training (e.g. training to use word processors, spreadsheets and presentation

software) suggest that EMT is superior to GT in promoting students' ability to adaptively transfer their skills (Keith & Frese, 2008). However, the extent to which this applies to the development of technological skills in statistics education is unknown. Therefore, the overall aim of Part I of the dissertation is to compare EMT to GT for the development of statistical package skills in introductory statistics courses.

A pilot study was initially needed to determine the feasibility of using active-exploratory training (AE-T) approaches and the use of minimal instruction for the development of statistical package skills (Chapter 4). There were initial concerns that AE-T may be too difficult for students which would lead to higher training anxiety and frustration, lower statistical package self-efficacy, and therefore, poorer statistical package training outcomes. The pilot study randomly allocated thirteen participants, who had previously completed an introductory statistics course for science or business, to either AE-T or GT for a one hour training session covering basic operation of the statistical package *SPSS*. Training was scheduled outside of regular classes and all participants had previously completed an introductory statistics course. During the training session participants rated the perceived difficulty of the training, their statistical package anxiety, and their statistical package self-efficacy. One week following the session, participants also completed an online quiz to measure analogical transfer. The outcomes of the Pilot suggested that AE-T approaches were feasible to implement in statistics education for training to use statistical packages. With this reassurance in mind, plans went ahead for Trial I.

Trial I aimed to evaluate the effectiveness of Error-management training (EMT) for learning to operate the statistical package *SPSS* (Chapter 5). An explanatory mixed methods approach was used. This involved gathering quantitative data from a sample of 100 psychology students enrolled in a first year introductory statistics course. These students were randomly allocated to either EMT or GT computer laboratory training sessions. In a secondary qualitative phase, fifteen students participated in semi-structured interview following training. During the semester, participants completed five fortnightly *SPSS* training sessions. Prior to the last training session, participants completed a post-training self-assessment task that assessed training transfer. The same self-assessment task was also completed as a follow-up in semester two. Quantitative results indicated that after controlling for covariates, the results of Trial I found no statistically significant difference between the training approaches on measures of post-training and follow-up training transfer. However, a number of key limitations were identified that suggested a need for a follow-up study. The in-depth interviews from the qualitative phase were analysed using thematic analysis to help explain the results of the quantitative phase as well as explore the overall student experience of the statistical package training. The qualitative evidence reinforced key limitations previously identified and provided unique insight in students' perceptions of technology training in statistics education. Based on the major findings of Trial I, a second follow-up study was planned.

Trial II re-evaluated the effect of GT and EMT approaches on statistical package training transfer by addressing key limitations identified in the quantitative and qualitative phases of Trial I (Chapter 7). Trial II employed a quasi-experimental design using a sample of 115 psychology students enrolled in an introductory statistics course which ran concurrently across two campuses. The EMT and GT approach was implemented in Campus A and B respectively. Students completed weekly, one-hour training sessions learning to use the statistical package *SPSS*. In the final week of the semester, students completed an *SPSS* certification task to measure adaptive skill transfer. Due to non-random allocation, the covariates of gender, personal access, statistical knowledge, and training progress were taken into account when modelling adaptive transfer between training approaches. After controlling for these covariates, no difference in adaptive transfer was found between training approaches. The potential moderating effect of prior statistical knowledge on EMT is raised as a possible explanation for the null finding of Trial I and II. The Trial II re-evaluation and qualitative results of Trial I suggested that improving access to technology may provide a more powerful way to improve the development of technological skills in statistics education than training approaches alone. These series of studies that compose Part I lay a solid foundation for future research looking into technological skills for statistical literacy.

Publications

Reference to works in Part I should cite the following peer-reviewed papers that arose throughout the course of the dissertation. A paper outlining the adaptation of a theoretical framework for the development of technological skills in statistics education was presented at the 2012 *International Association for Statistics Education (IASE) Roundtable Conference*, held in Cebu, Philippines (Baglin & Da Costa, 2012b). An amended version of this paper was submitted for peer-review in a special edition of *Technology Innovations in Statistics Education* (Baglin & Da Costa, n.d.). The outcomes of the Pilot were presented at the 2010 *7th Australian Conference on Teaching Statistics (OZCOTS)* in Fremantle, Western Australia. A peer-reviewed paper of this presentation was published in the proceedings (Baglin & Da Costa, 2010). Preliminary quantitative outcomes of Trial I were presented as a poster at the 2011 *Australian Conference on Science & Mathematics Education (17th Annual UniServe Science Conference)* held in Melbourne, Victoria. A peer-reviewed paper of this poster was published in the proceedings (Baglin, Da Costa, Ovens, & Bablas, 2011). Following this conference, a expanded version was invited and accepted for publication into a special edition of the *International Journal of Innovations in Science and Mathematics Education* (Baglin & Da Costa, 2012a). A brief report of the qualitative phase of Trial I was presented and published in the proceeding of the *8th Australian Conference on Teaching Statistics (OZCOTS)* in Adelaide, South Australia (Baglin & Da Costa, 2012c). A report on the outcomes of Trial II was published in *Technology Innovations in Statistics Education* (Baglin & Da Costa, 2013).

Chapter 3

Part I - Introduction

3.1 Technological Skills in Statistics Education

Technology use is an inseparable part of modern statistics courses (Gould, 2010). Its use has been on the rise for the last couple of decades best reflected in the recommendations laid out by the statistics education reform (Cobb, 1992; American Statistical Association, 2005). Statistics education has readily adopted the use of technology in introductory statistics courses as a way of fostering students' conceptual understanding and moving the focus of courses away from computation. Recent survey reports from the U.S. have surfaced suggesting that up to 76% of statistics courses regularly use technology (Hassad, 2012). This has risen from an estimate of 50% identified in a previous U.S. survey ten years earlier (Garfield, Hogg, et al., 2002). The types of technology that are utilised in introductory courses vary but the most common examples include statistical packages, educational software, spreadsheets, applets, graphing calculators, multimedia material and data repositories (Chance, Ben-Zvi, Garfield, & Medina, 2007). The use of statistical packages is probably the most ubiquitous because they provide the widest range of benefits to instructors. Statistical packages automate difficult statistical formulae so as to allow for a greater focus on interpretation (B. Smith, 2003), provide instructors with unique tools for demonstrating statistical concepts (B. Smith, 2003), and familiarises students with technology commonly used in statistical practice (Oswald, 1996).

Chance, Ben-Zvi, et al. (2007) claim that the use of technology in statistics education has been to focus on “the content, and not the tool” (p.4), but recently, some instructors have begun to challenge this view. These instructors cite that the changing nature of statistical practice and an unprecedented access to data will require modern concepts of statistical literacy to expand beyond just conceptual understanding (Gould, 2010; Nolan & Temple Lang, 2010b, 2010a). As Gould (2010) explains, the ability to use statistical technology is now a fundamental component of statistical literacy, not a mere “hurdle” (p. 309) suggested by the prevailing attitude. The best example to illustrate this point is the ability to operate statistical software packages, e.g. *SPSS*, *Minitab*, *SAS*, *Stata*, and *R*. This ability is a vital skill that students must develop if they are to become statistically literate. Without this technological skill, students cannot meaningfully and practically analyse complex real-world data. In many cases, implementing modern statistical methods is completely impractical without the aid of a statistical package (e.g. creating plots, running simulations, statistical modelling, and bootstrapping). While much literature exists on the use of technology in statistics education, little has focused on the development of the technological skills required to use it. This might be construed as suggesting that most instructors assume students will just “pick up” (Gould, 2010) these skills and carry them throughout their career. Sadly, the opposite is most likely true.

If technological skills, such as statistical package skills, are fundamental to modern notions of statistical literacy, these skills need to be fostered in introductory statistics courses. The statistics education literature has fallen behind on understanding how these skills can be developed and how they interact in introductory statistics courses. Many fundamental questions must be addressed. How do students learn to use technology? What are the barriers to developing technological skills? How can instructors better foster students’ technological skills? Many of these questions have been addressed in the general software training literature (e.g. organisational training for word processors, email, internet use, spreadsheets, and presentation software). While statistics education can draw from this knowledge base, the unique environment of the introductory statistics course is likely to present many challenges. For one, the ability to use statistics technology is likely to be highly dependent on statistical knowledge.

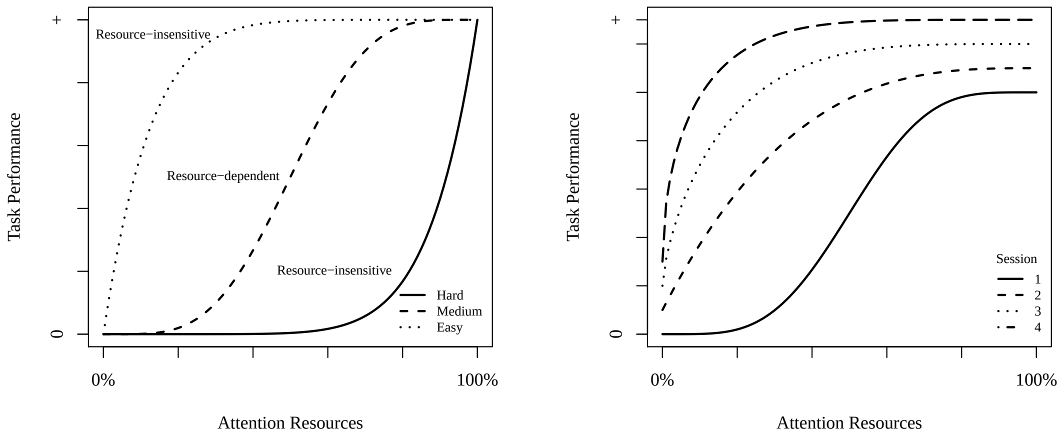
Separating technological skills from students' conceptual understanding of statistics will present a major challenge which will make understanding the development of these skills difficult. A theoretical framework will help focus research efforts.

3.2 Kanfer and Ackerman's Integrative Model of Skill Acquisition

A theoretical framework can help guide research efforts in the area of technological skills for statistics education. Kanfer and Ackerman's 1989 integrative model of skill acquisition is consistent with learning to use statistical packages and has been used to explain technological skill acquisition for general software (e.g. spreadsheets, presentations and word processing Keith, Richter, & Naumann, 2010). Kanfer and Ackerman's model explains technological skill acquisition by integrating students' cognitive ability and motivation within an information processing framework. According to Kanfer and Ackerman, skill acquisition is explained by four notions: *attentional resources*, *task demand*, *resource allocation* and *the effect of practice.*

All tasks (e.g. training exercises) require a certain level of attentional resources. Some tasks demand a high level of attention, while other tasks require less. Learners internally regulate attentional resources dedicated to a task and can choose to focus attention or divide attention between competing tasks. As a learner practices a task, the required level of attentional resources allocated to that task lowers, i.e. the effect of practice. The model assumes that there is a relationship between resource allocation and task performance, i.e. the more resources allocated to a task, the better the performance. However, this relationship is moderated by the nature of the task, motivation and cognitive ability.

Tasks can be divided into being either *resource-dependent* or *resource-insensitive.* Resource-dependent tasks are those tasks where an increase in attentional resources corresponds with a large performance gain. These tasks are generally those which are moderately difficult. On the other hand, resource-insensitive tasks are those where a change in attentional resources is associated with minimal changes in performance. Easy and difficult tasks are resource-insensitive as in both cases performance is relatively independent from attentional focus (see Figure 3.1a). Training should begin with resource-dependent tasks which will require the commitment of attentional resources. As the trainee practices, the resource-dependency of the task changes to become more resource-insensitive. It is this shift in attentional resources (see Figure 3.1b) that is referred to as the *effect of practice*.



(a) Task Difficulty (b) Effect of Practice

Figure 3.1: Kanfer and Ackerman's Concepts of Task Difficulty and the Effect of Practice

Kanfer and Ackerman (1989) proposed two major factors, distal motivation and cognitive ability, that regulate attentional resources allocated during training. Distal motivation determines the level of attentional resources allocated early on in training. Keith, Richter, and Naumann (2010) discuss the distal motivation of perceived performance utility. Perceived performance utility relates to level of belief that a task will be important to an individual. For example, a trainee with high perceived performance utility regarding statistical packages will view training as being beneficial to their career. Thus, they will be more likely to allocate a high level of attentional resources when tasks are resource-dependent. Those with low perceived performance utility will be less inclined to dedicate the required attentional resources to training. For example, a trainee who believes knowledge of statistical packages outside of a statistics course is of no use will be less inclined to commit attentional resources to training. A lack of attentional resources dedicated to resource-dependent tasks will retard the effect of practice leading to poor training transfer. *Training transfer* is defined as the ability to transfer skills outside of the training environment (Hesketh, 1997).

Cognitive ability determines the capacity of a learner to allocate attentional resources to any given task. High cognitive ability trainees have more attentional resources to offer, while those with low cognitive ability have less to offer. Because of this relationship between attention allocation and cognitive ability, task performance can largely become a function of cognitive ability. This relationship has been established in a large body of literature showing a strong relationship between job performance and cognitive ability (e.g. Hunter, 1986). Unfortunately, cognitive ability is not something that is amenable to change, but its effect tends to be less pronounced in tertiary populations where most students are expected to have average to high cognitive ability.

In summary, Kanfer and Ackerman's theory predicts that motivation and cognitive ability interact with early training performance when tasks are resource-dependent. As the trainee practices, tasks begin to become more resource-insensitive (Figure 3.1b). Therefore, the role of training is to transform resource-dependent tasks into resource-insensitive tasks. Kanfer and Ackerman's theory also predicts that trainees' performance early in training will be highly influenced by their motivation and cognitive ability. Poorly motivated and academically weaker students might struggle early in statistical package training which will later lead to poor transfer of skills. Fortunately, Keith, Richter, and Naumann (2010) demonstrate that AE-T can help moderate the effect of motivation and cognitive ability on training performance.

3.3 Active-exploratory Training Approaches

Kanfer and Ackerman's 1989 model can be used to guide the selection of effective statistical package training approaches. According to Kanfer and Ackerman's 1989 model, training transfer performance is likely to be influenced by a student's motivation and cognitive ability. This provides instructors with possible targets for fostering the development of technological skills. Improving students' cognitive ability may present challenges, but specialised approaches to training which compensate for lower cognitive ability might be possible. Improving students' motivation towards technology may provide the most practical target, but there will always be students difficult to motivate. Fortunately, previous studies looking at technological skill acquisition suggest that skill transfer can be improved by using different training approaches (Bell & Kozlowski, 2008).

A training approach is a theoretical framework that guides the design and delivery of technology training. This can be contrasted with training delivery methods, e.g. computer laboratory sessions, in-class demonstrations, and self-guided modules. A large body of research that has looked at general software training has found that active-exploratory training (A-ET) approaches appear to have superior outcomes for training transfer when compared to traditional guided training (GT) approaches (Bell & Kozlowski, 2008; Chillarege, Nordstrom, & Williams, 2003; Frese, Brodbeck, et al., 1991; Heimbeck, Frese, Sonnentag, & Keith, 2003; Keith & Frese, 2008; Keith, Richter, & Naumann, 2010; Nordstrom, Wendland, & Williams, 1998; Wood, Kakebeeke, Debowski, & Frese, 2000). GT is founded on the programmed learning method developed by the famous behaviourist Burrhus Frederic Skinner (1968). GT views the student as a passive participant during training. The student is presented with step-by-step, comprehensive and explicit instructions which guide them through learning to operate a statistical package. The GT approach is error-avoidant, i.e. errors are viewed as a non-productive waste of time. Students' skills are developed through repeated practice where operational errors are minimised. The GT approach is embodied in a large number of textbooks available for popular statistical packages such as *SPSS* (e.g. Allen & Bennett, 2008; Francis, 2007). The majority of these specialised software training books include step-by-step instructions supplemented by screenshots and output from the statistical package (Mills, 2003).

On the other-hand, A-ET presents the student with minimal instruction which engages them in actively-exploring the statistical package (Bell & Kozlowski, 2008). As comprehensive instructions are avoided, the student becomes an active participant in the development of their skills. A-ET approaches are in line with key recommendations made by both the Cobb Report (1992) and GAISE Report (2005) to promote active learning in the introductory statistics classroom. There are a number of different subtypes of AE-T including *pure active-exploratory learning* (e.g. Frese, Albrecht, et al., 1988; Kamouri, Kamouri, & Smith, 1986; McDaniel & Schlager, 1990), *error management training* (e.g. Frese, Brodbeck, et al., 1991; Gully, Payne, Koles, & Whiteman, 2002; Heimbeck et al., 2003; Keith & Frese, 2005; Keith, Richter, & Naumann, 2010), *guided exploration* (e.g. Bell & Kozlowski, 2002; Debowski, Wood, & Bandura, 2001; Wood et al., 2000) and *mastery training* (e.g. Chillarege et al., 2003; Kozlowski & Bell, 2006; Kozlowski, Gully, et al., 2001; Martocchio, 1994; Stevens & Gist, 1997; Tabernero & Wood, 1998). *Error-management training* has been the most successful.

Error-management training (EMT) goes one step further than pure AE-T by paying special attention to the function of errors. As students actively-explore the statistical package with minimal instruction, they will invariably commit errors. Frese, Brodbeck, et al. (1991) describes four reasons why errors are positive to training. First, errors draw attention to areas of learning that need further attention. The learner is prompted to start exploring these unknown areas leading to a better understanding of a system. Second, making errors help learners avoid making further errors in the future and trains the learner how to deal with errors once they have occurred. Third, errors promote exploratory learning which research suggests is superior to error avoidant GT. Thus, error management training by nature is an active-exploratory method of training. Fourth, the learner will invariably need to deal with errors in their work environment once training has finished. The learner will no longer have the assistance of an instructor, but instead will need to handle errors and other problems themselves. Learning effective strategies for avoiding and dealing with errors should therefore be an integral part of any training program. To help deal with the typical negative emotions experienced after making an error, EMT incorporates emotional control strategies. This involves normalising and positively framing errors. To achieve this, heuristics are presented to students during training, such as "Errors are a natural part of learning. They point out what you can still learn!" (Dormann & Frese, 1994, p. 368). These heuristics are delivered to students in training material and through encouragement by trainers (e.g. tutors).

Research suggests that A-ET approaches, such as EMT, are superior to GT approaches when considering *adaptive transfer*. Adaptive transfer is demonstrated in a student's ability to adapt limited training skills in order to confront novel situations outside of training (Keith, Richter, & Naumann, 2010). For example, a student may have covered conducting two-sample t-tests in a statistical package. Suppose they learn about one-way ANOVA in another course and want to use the statistical package to run a test. Adaptive transfer would be evident if the student could adapt their skills of conducting two-sample t-tests to figure out how to operate the statistical package to perform the one-way ANOVA. Another example of adaptive transfer would be a student transferring their knowledge of one statistical package to learn a different statistical package. Adaptive transfer is the most desirable outcome of training as it promotes sustainable learning beyond the brief experience afforded by most training. Training should provide students with a foundation that they can continue to adapt and build upon outside of the training environment. The other type of training transfer, *analogical transfer*, is simply the ability to transfer the same skills covered in training (Keith, Richter, & Naumann, 2010). For example, if a student covered correlation in training, analogical transfer is evident if the student can perform correlation outside of training.

A meta-analysis which combined the results of 24 studies looking at the effect of EMT found an overall significant and positive effect over GT (Keith & Frese, 2008). Keith and Frese combined the results of experiments looking at general software training including simulation, word processing, databases, presentations, spread sheets, e-mail, web browsers, and programming languages. The outcome of this analysis found that EMT was overall significantly superior to GT for promoting adaptive transfer, and, to a lesser extent, analogical transfer. The study also found that the two core components of EMT, active-exploration and error-encouragement, contributed unique training effects suggesting that EMT is more effective than A-ET alone. Keith and Frese concluded that EMT is the preferred method of training when adaptive transfer is the goal. The development of self-regulatory skills has been posited to explain the superiority of EMT.

According to Keith, Richter, and Naumann (2010), A-ET approaches, such as EMT, work by developing students' self-regulatory skills. Self-regulatory skills in a training context can be defined as a student's ability to guide their engagement in training activities by controlling cognition, mood, behaviour and focus (Karoly, 1993, p. 25). This involves both metacognition and emotional control. Ford, Smith, Weissbein, Gully, and Salas (1998) define metacognition as a student's ability to exert "control over his or her cognitions" (p. 220) by planning, monitoring and evaluating task performance (Brown, Bransford, Ferrara, & Campione, 1983). Emotional control can be defined as "the use of self-regulatory processes to keep performance anxiety and other negative emotional reactions (e.g. worry) at bay during task engagement" (Kanfer, Ackerman, & Heggestad, 1996, p.186). As Keith, Richter, and Naumann (2010) explains, minimal instruction promotes active-exploration which requires students to practice metacognitive skills. Students must plan, monitor and evaluate how they are progressing through the training activities. GT, on the other hand, creates a passive training environment where students progress by following instructions. They do not engage at the same level of metacognitive activity required by EMT. Students in EMT are also required to develop emotional control strategies to deal with negative emotions created by errors, e.g. anxiety. The EMT approach achieves this by creating an environment where students practice dealing with negative emotions, become habituated to inevitable commitment of errors and by helping students realize their positive functions. Emotional control may be particularly important for learning statistical packages as numerous studies have found a significant negative relationship between statistics anxiety and statistics course performance (J. Benson, 1989; Tremblay, Gardner, & Heipel, 2000; Onwuegbuzie & Seaman, 1995; Pretorius & Norman, 1992). Students in a GT approach avoid errors and become accustomed to the artificial use of guided instructions. They are not presented with the opportunity to develop emotional control strategies that are required when transferring skills in real-world situations outside of a "safe" error-free training environment.

Lending further to the promise to the EMT approach, Keith, Richter, and Naumann (2010) found that A-ET curbed the effect of low motivation and low cognitive ability on adaptive training transfer. Kanfer and Ackerman's model suggests that the efficacy of training can be reduced for students who lack motivation to develop statistical package skills and students who may have lower cognitive ability. Keith, Richter, et al. found that participants' trained using EMT for presentation and word processing software exhibited no relationship between adaptive transfer and participants' motivation or cognitive ability. On the other hand, participants' adaptive transfer for the GT condition was correlated with the participants' willingness to learn and their general cognitive ability. Keith et al. explains that the effects of motivation and cognitive ability depend on the degree of overlap between training tasks and transfer tasks. When training tasks overlap transfer tasks, i.e. analogical transfer tasks, the influence of cognitive ability and motivation on transfer performance is minimal. On the other hand, when there is little overlap between training and transfer tasks, i.e. adaptive transfer tasks, cognitive ability and motivation have a noticeable impact. When dealing with difficult or novel situations, trainees will activate their self-regulatory skills to get the job done, i.e. emotional control and metacognition – planning, monitoring, and evaluating. Table 3.1 provides an example of a student thinking through adapting their knowledge of creating histograms to create a side-by-side comparison in *SPSS*. However, the degree to which these findings extrapolate to the development of technology skills for statistical packages remains in question.

Table 3.1: An Example of a Student's Metacognition for an Adaptive Transfer Task

Activity	Example
Planning	I know how to obtain a histogram in SPSS, but how do I split the histogram by a grouping variable? I will need to try changing some options.
Monitoring	I will try putting the grouping variable in the panel by option and see what happens.
Evaluating	That seems to have done the trick. I will now be able to compare histograms between groups.

While the overall meta-analytic consensus might be that EMT is superior to GT, a number of well controlled studies have failed to support this finding raising a number of potential issues for EMT. One study by Debowski et al. (2001) investigated the effect of EMT versus GT for electronic search in bibliographic databases. In contrast with previous research, the authors hypothesised that GT would be more effective than error management training. The authors argued that because of the lack of effective error feedback from electronic search, trainees in an EMT condition would fail to receive clear feedback that is needed for EMT to work effectively. On the other hand, GT for electronic search would enable trainees to model search strategies that are most effective. In other words Debowski et al. predicted that EMT would be moderated by the quality of task feedback. The sample consisted of 48 university students randomly assigned to the training conditions. The participants took part in two sessions. The first session introduced the participants to electronic search, and then in the second session, participants were given five practice tasks. The training conditions were manipulated in the practice session. Directly following the practice session, trainees were assessed on a further two search tasks for performance measures. The results indicated that the GT group scored significantly higher on search performance across both performance tasks when compared to the error management training group. Debowski et al. concluded that the quality of task feedback moderates the effectiveness of EMT. This is important for statistical package training as the quality of feedback from a package or from training will impact the effectiveness of EMT. Training needs to be designed to ensure that students receive immediate feedback when errors are committed. The statistical package itself will do this to a large degree (e.g. warning messages), however, training feedback will need to address other errors that may go unchecked (e.g. compute means on nominal variables). This is where students' knowledge of statistics is also likely to aid students' during training.

Other studies have failed to support the use of an error management component of EMT. Lazar and Norcio (2003) assessed the effect of EMT, exploratory training, and traditional GT approaches for training novices to use the internet. These two conditions were compared to traditional training which was error-avoidant. The sample consisted of 263 participants recruited from the general population. Participants took part in a single three hour training session based on the groups that they had been randomly assigned. Following training, participants were then given 1 hour to complete 10 information gathering search tasks to measure training performance. The results of the experiment found no significant difference between traditional training and EMT. However, there was a statistically significant difference between exploratory training and traditional training. The authors concluded that exploratory training helps people to learn to deal with ambiguity presented by tasks such as internet navigation. The authors concluded that EMT was probably ineffective because of the ingrained tradition of viewing errors as being bad. One training session would be very unlikely to change this view. This suggests that properly implementing error-management into EMT for statistical packages may present challenges. Especially in a university environment where other courses are likely to send conflicting messages. Therefore, the efficacy of EMT over GT requires further investigation in the statistics education context.

3.4 Statistical Package Skills

One published study has looked at the effect of training approaches on the development of statistical package skills. Dormann and Frese (1994) randomly assigned 30 psychology students to either GT or EMT for learning to use the statistical package *SPSS*. Participants completed a single training session that lasted two hours. In the following hour training transfer was evaluated. The study did not specifically measure adaptive transfer, but instead, divided tasks between easy, moderate and difficult. The results indicated that participants in the EMT condition performed significantly better on measures of moderate and difficult training transfer tasks. The authors concluded that EMT was superior to error-avoidant GT approach for statistical package training. However, there were a number of limitations to this study.

Dormann and Frese's experiment suffered from a small sample size, the use of an now out-dated version of the statistical package, the immediate evaluation of training outcomes, no specific attempt to differentiate between analogical and adaptive transfer outcomes and the use of a one-off training session outside of a real statistics course. Studies using larger samples and up-to-date versions of statistical packages are required. Training outcomes need to differentiate between analogical and adaptive transfer and be assessed at more meaningful follow-up periods. For introductory statistics courses this would involve end of semester and between semester evaluations. Training should also be embedded within a real introductory statistics course to evaluate the ecological validity of EMT. Until EMT has been demonstrated to be efficacious in introductory statistics courses, it cannot be recommended over GT. Dormann and Frese also only considered training transfer outcomes.

Other training outcomes besides transfer are important to instructors. Students perceptions of training also requires consideration. For example, Debowski et al. (2001) found that trainees' perceptions of self-efficacy and overall training satisfaction were significantly higher for GT compared to pure active-exploratory training after an electronic search training session. Debowski et al. explains that the lack of task feedback provided by electronic searches restricts the effectiveness of AE-T approaches. AE-T approaches require immediate and direct feedback on task performance to enable trainees to identify effective strategies and correct their errors. Without knowing their errors or whether their strategy was effective, trainees will fail to develop self-efficacy and their overall perceptions of the quality of training will be diminished. Fortunately, statistical packages typically provide immediate and useful feedback regarding errors and online training environments which assess students' solutions can provide immediate feedback regarding correct solutions. However, future research is needed to evaluate students' perceptions of self-efficacy and satisfaction. There is a concern that AE-T approaches may be perceived as more difficult to GT which may lead to higher anxiety, lower self-efficacy and lower overall training satisfaction. These other outcomes are likely to have an important impact on students perceptions of training and therefore might impact students' attitudes towards technology and the course.

3.5 Rationale and aims

As technological skills in statistics education are becoming more and more important, understanding how these skills can be effectively developed in statistics courses is a key priority. As very little is known about this area of statistics education, the rationale for the following three studies stems from the requirement to gain insight into the effects of training strategies on the development of technological skills. The studies reported herein focused on the important and ubiquitous statistical package. These studies considered training transfer as well as other important outcomes, e.g training difficulty, self-efficacy, satisfaction and anxiety. A second aim was to begin exploring students' perceptions of training to gain a better insight into the student experience of learning to use technology in statistics courses. This was done in order to help identify important factors that may lead to further knowledge of the development of these skills.

Chapter 4

Part I - Pilot Study

4.1 Aims of the Pilot

Prior to Trial I, a pilot study was necessary to determine the feasibility of using active-exploratory training (AE-T) approaches for statistical package skill development. There were initial concerns that AE-T may be too difficult for students which would lead to higher training anxiety and frustration, lower statistical package self-efficacy, and therefore, poorer statistical package training outcomes. On a more practical note, there was a concern that AE-T may increase the time needed for students to complete training and also increase the number of questions directed at tutors due to the increased difficulty. A small pilot study was conducted to test these concerns. The outcomes of the Pilot would also help inform the design of Trial I.

4.2 Method

4.2.1 Participants

The pilot sample consisted of 15 participants who had previously completed a first year introductory statistics course. Participants were approached to participate in the pilot study during the start of a lecture at the end of semester one in 2010. Two students were dropped from the study leaving a final sample of 13 (one student failed to complete the pilot follow-up quiz and another was non-compliant during the training session). The students came from Business ($N = 4$) and Applied Science programs ($N = 9$).

There were 6 males and 7 females. All students were full-time first year students and two were international students. The sample mean age was 25.9 years (SD = 8.3). The mean time taken to complete the training was 64 minutes (SD = 18.8). The mean follow-up time for the pilot training quiz was 8 days (SD = 2). All participants had prior knowledge of statistical software packages (i.e. *Excel* and *MINITAB*), but none had experience with *SPSS*, the package used in the pilot study. *SPSS* was also selected as it would be the package to be used in future trials. Table 4.1 shows the breakdown of students in the two groups used in the study.

Table 4.1: Pilot Sample Group Characteristics

	GT	AE-T
N	6	7
Business	2	2
Applied	4	5
Males	4	2
International	1	1
Age: Mean (SD)	26.3 (10.8)	25.4 (6.4)
Tute Time Mins: Mean	63.3 (15.4)	65.0 (22.6)
Follow-up Days: Mean	8.5 (2.7)	7.7 (1.5)

4.2.2 Measures

Participants were given a tutorial booklet which contained the training approaches instructions, tutorial activities and outcome measures. This tutorial booklet gathered demographic information as well as measures of statistical package self-efficacy, statistical package anxiety, and perceived level of difficulty.

Statistical Package Self-efficacy. A measure of the change in statistical package Self-efficacy was adapted from three items of Finney and Schraw's 2003 *Current statistics self-efficacy* (CSSE) and *Self-efficacy to Learn Statistics* (SELS) scales. These scales have evidence of good psychometric properties (Finney & Schraw, 2003). The 3 items from the SELS and CSSE were modified to relate specifically to conducting statistical analysis using a statistical package (see Appendix A.1). The participants were asked to rate their self-efficacy on a 10-point scale ranging from 1 (no confidence at all) to 10 (complete confidence) before and after the training session. The self-efficacy change score was calculated by subtracting the self-efficacy rating taken before the session from the self-efficacy rating given upon session completion. Scores could range from -27 to 27. High scores are indicative of self-efficacy improvement. A score of 0 indicates no change in statistical package self-efficacy.

Statistical Package Anxiety. Statistical Package anxiety was measured using two items adapted from Cruise, Cash, and Bolton (1985) *Statistics Anxiety Rating Scale* (STARS). Once again, these two items were modified to relate specifically to statistical packages (see Appendix A.1). The STARS has well established psychometric properties (Baloglu, 2002). Participants responded to these items on a 10-point likert scale ranging from 1 (no anxiety) to 10 (very strong anxiety) with scores ranging from 2 to 20.

Perceived Training Difficulty. The final item in the tutorial booklet required participants to rate the perceived difficulty of the tutorial on a 10-point likert scale ranging from 1 (very easy) to 10 (very difficult)(see Appendix A.1).

Training Transfer. Analogical transfer was measured using a 10-question online quiz assessing the participant's recall of the previous week's training. The quiz had 14 possible marks. The quiz focused on assessing the participant's recall of operational knowledge of the statistical package. For those who are familiar with *SPSS*, this quiz covered differentiating variable view and data view, labelling variables, matching different descriptive outputs with the correct commands, using split file and select cases, running t-tests, finding p-value in *SPSS* output, and setting up basic graphical displays. Questions were a combination of multiple-choice and text/numerical responses. All questions contained an "I do not know" option to minimise guessing.

4.2.3 Procedure

After obtaining ethics approval, participants were recruited following the completion of a one semester introductory statistics course. Involvement in the study was strictly voluntary and was completed outside regular university attendance. A raffle for a major prize was used as an incentive for students to participate. All participants were randomly allocated to a training approach prior to attending. Participants were informed that they were involved in a study investigating how students learn to use statistical software packages. All participants were blinded to the exact nature of the study and the differences between training approaches (see Appendix A.3 and A.2 for a copy of the consent from and plain language statement used for this study).

The statistical package *SPSS for Windows Version 17* was used in all sessions. *SPSS* was chosen as this package was not taught in the participant's previous introductory statistics course. The two tutorial conditions were designed to take approximately one hour to complete. However, participants were given as much time as they needed. Both conditions covered the following topics in *SPSS*: Entering data, editing datasets, descriptive statistics, data file manipulation (select cases and split file), comparing means via t-tests (Paired t-tests and Two-sample t-tests) and basic graphical displays.

In the AE-T condition, participants were first given a question to answer and then given a few prompts to get started (e.g. Use the Analyse - Compare Means command). The idea was to give participants minimal information and a few pointers to get started. The participant would then attempt the exercise and in the process actively explore the statistical package. Students were encouraged to seek assistance only when they were really stuck. An example of an AE-T exercise is given in AppendixA.4.

In contrast, the GT approach had explicit explanations, screen shots and step-by-step instructions on how to do a specific analysis which the participant would work through. The participant would then be given another activity to practice the analysis that was previously explicitly explained. This condition deliberately avoided uncertainty and aimed to explain every aspect of the statistical package in the implementation of an instruction. An example of a GT exercise is given in A.4

The tutor present at the sessions recorded the time taken to complete the tutorial and the number of times the participant sought the tutor's assistance. This was to give an indication of the practical issues relating to the implementation of the training approaches.

4.3 Results

The results of this study were restricted to a descriptive analysis due to the small sample size. Descriptive statistics comparing the two conditions are show in Table 4.2. The median was the preferred measure of central tendency due to the susceptibility of the mean to outliers in small samples. Dot plots were also included to gain an insight into the variability of the outcomes between conditions and to also present the raw data which would otherwise be encompassed in the descriptive summaries (see Figure 4.1).

Table 4.2: Descriptive Statistics Between Conditions on Outcome Variables

Outcome	GT			AE-T Training		
	Median	*Mean*	*SD*	Median	*Mean*	*SD*
Time (Mins)	67.5	63.3	15.4	55	65	22.6
Questions	3	3	2.5	3	3.3	2.6
Difficulty	2.5	3.3	2.0	5	5.3	2.1
Anxiety	7	7.7	3.8	11	9.7	3.0
Self-efficacy	5	7.5	8.1	4	5.9	4.7
Quiz Score	7.5	8.2	3.3	8	7.9	1.7

For tutorial session times, the median was higher in the GT group (Median = 67.5 mins) compared to the AE-T group (Median = 55 mins). Inspection of the dot plot reveals less variability in the GT condition. With the exception of an outlier in the AE-T group, both training approaches appeared to finish within a similar time frame. This runs counter to intuition which dictates that the more difficult AE-T condition would take longer.

The median number of questions asked was of tutors the same between conditions (Median = 3). The AE-T group would be expected to ask many questions given the nature of the training approach, but the similar number of question in the GT group needs explaining. The researcher supervising the training sessions reported that the GT group would ask the tutor for validation in what they were doing. Participants in the GT group would also constantly run into problems because they had not followed the instructions properly. They would then persist in asking the tutor where they had gone wrong. The nature of these questions is reported in the next section.

As one would expect, the median perceived difficulty of the training session was higher in the AE-T condition (Median = 5) compared to the GT condition (Median = 2.5). This provides evidence of the validity of the difference between the nature of the two training approaches. This probably led to a higher median statistical package anxiety rating in the AE-T condition (Median = 11) versus the guided condition (Median = 7).

In terms of statistical package self-efficacy change, both groups were comparable

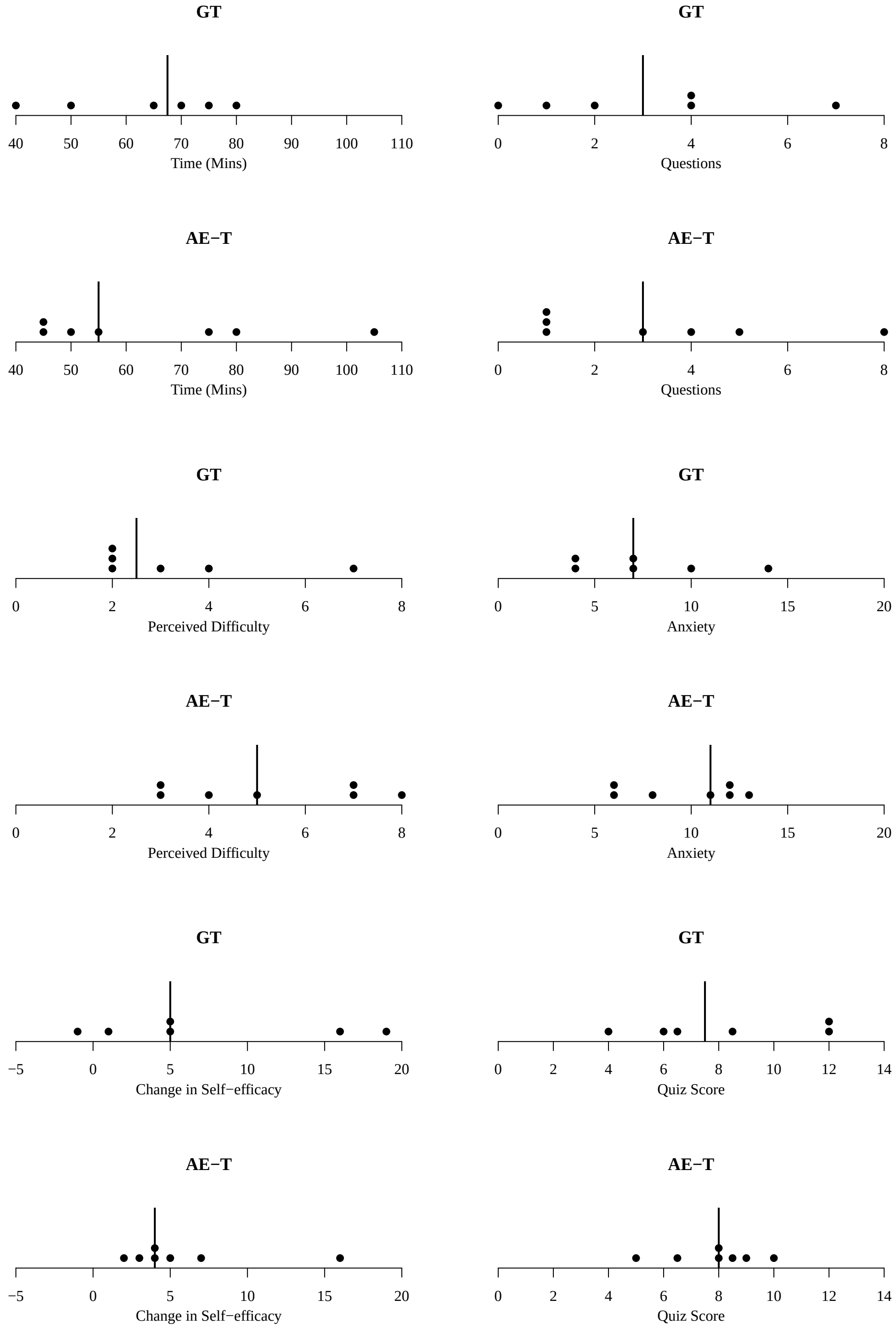


Figure 4.1: Pilot Study Dot plots showing the distribution of each outcome variable between conditions.

with a median of 5 and 4 for the GT and AE-T groups respectively. However, examination of the dot plot showed a large degree of variability in the GT condition, whereas the change score for the AE-T group, with the exception of one value, was clustered close to the median. One interpretation of this result is that the GT condition created very mixed perceptions of self-efficacy change ($SD = 8.1$), whereas the perceived change in the AE-T condition was more uniform ($SD = 4.7$).

The performance on the online quiz assessing analogical transfer one-week after the tutorial session between training approaches showed that the AE-T condition scored marginally higher (Median = 8) than the GT condition (Median = 7.5). Once again, the AE-T condition ($SD = 1.7$) was associated with a lower degree of variability as demonstrated in the dot plots (GT $SD = 3.3$). Overall, there appears to be little discernible difference between conditions on analogical scores taken from the quiz.

Questions

During the training session, the researcher acting as the tutor recorded the nature of questions or difficulties being raised during training. These questions were grouped into themes. These themes, along with their frequency, are presented in Table 4.3. Difficulties with *SPSS*'s split file feature was the most common theme across both training conditions. Taken together, there was a large degree of similarity between conditions on the themes of questions and difficulties raised during the training session.

4.4 Discussion

The aim of this pilot study was to compare AE-T and GT approaches for training students to use the statistical package *SPSS*. The outcomes measured included both contextual and performance indicators. As expected, the AE-T approach was associated with higher perceived difficulty and higher statistical package anxiety. This was expected as the idea of the AE-T approach was to engage students by giving them minimal information. Also, as anxiety was only measured after the session, the discrepancy might be explained by individual differences even though random allocation was used. This type of question could only be addressed with the use of a larger sample where the probability of pre-existing group difference in randomly allocated designs reduces as the

Table 4.3: Pilot Study Student Question Themes and their Frequency Between Training Approaches

Theme	GT ($N = 6$)	AE-T ($N = 7$)
Did not understand the split file feature	5	5
Had difficulties relabelling variables	3	2
Did not turn split file off	2	1
Had difficulties performing *t*-test	2	2
Did not set-up up box plot correctly	1	3
Did not turn the select cases filter off	1	2
Asked tutor to validate correct output	2	-
Did not understand difference between data and output	1	-
Did not understand select cases filter feature	1	-
Had trouble deleting variable	1	-
Did not understand the difference between data view and variable view	-	2
Asked how to save the data and output	-	2
Did not know how to load *SPSS*	-	1
Did not understand select cases filter feature	-	1

sample size increases. The trade-off of higher anxiety is partially compensated by no perceivable difference between training approaches on statistical package self-efficacy change. Both training approaches seemed to have a similar positive effect, but this effect seemed vastly more variable in the guided training approach.

The increased difficulty and anxiety in the AE-T condition is not necessarily negative. It may be reflecting increased trainee engagement which would be expected following the allocation of attentional resources to resource dependent tasks. This then activates students' self-regulatory skills ultimately leading to better training transfer outcomes (Keith, Richter, & Naumann, 2010).

There were also a few unexpected results. There were no noticeable differences between the training approaches on the number of questions asked during the tutorial sessions. From a practical perspective, this challenges the notion that using an AE-T approach would increase the demand on tutors supervising sessions. The results for session time also suggest that an AE-T approach would not necessarily increase the required time taken to complete a set tutorial.

Another outcome of interest in this study was the difference in performance between the two training approaches on an analogical transfer quiz. The AE-T group did score higher, but the large degree of variability in the GT group made any definitive conclusions difficult. Overall, the results from the quiz suggest no clear difference between the two training approaches. These results were in agreement with what would be excepted in a low powered study (Keith & Frese, 2008).

4.5 Conclusions

The results of the pilot study were used to inform the design Trial I. Practically, AE-T resulted in finishing times and a number of the questions posed by students which was comparable to the GT condition. This was an important consideration to address as it was crucial that imposing the AE-T in the full trial would not result in the need to allow students more time to complete training nor would it result in greater strain on tutors in the AE-T condition. This finding was surprising as one would expect a more difficult condition to be associated with longer completion times and more questions posed. These results were very reassuring for Trial I.

It came as no surprise that the AE-T resulted in higher perceived difficulty. In fact one participant even commented at the completion of the training session that more instruction was needed. This finding is not necessarily negative. In fact, it may be the product of increased student engagement during training. This finding validated the differences between the imposed conditions. The method of minimal instruction used in the pilot appeared to be valid and could be applied in the full trial.

The only negative finding in the pilot suggested that the AE-T resulted in increased anxiety during training. This may just be the product of increased engagement and higher perceived difficulty, but it must be noted that high levels of anxiety have been shown to be a negative predictor of performance in introductory statistics courses (J. Benson, 1989; Tremblay et al., 2000; Onwuegbuzie & Seaman, 1995; Pretorius & Norman, 1992). Therefore, Trial I should impose approaches that aim to moderate training anxiety. As discussed in Chapter 3, Error-management training (EMT)incorporates emotional control strategies that suit this role (Keith & Frese, 2008).

The results of the pilot study were inconclusive in terms of differences in training effectiveness. This is not surprising given that the primary aim of the pilot was not to evaluate training transfer. The small sample size and use of an online quiz to assess analogical transfer prevented a reliable comparison. The online quiz relied on students' recall of the training session may not be a valid and reliable measure of a student's actual ability to transfer their skills. A more valid approach would be to assess students by requiring them demonstrate transfer using the actual statistical package. Therefore, Trial I aimed to provide a more comprehensive and rigorous scientific comparison of GT and EMT approaches.

Chapter 5

Part I - Trial I - Quantitative Phase

5.1 Rationale and Aims

The aim of Trial I was to investigate the ecological effectiveness of the EMT approach for learning to use a statistical package in an introductory statistics course. EMT was selected as the increased errors that would inevitably result from AE-T might lead to training frustration and anxiety. To help trainees counter these negative emotions, EMT includes an emotional control component. As discussed previously in Chapter 3, few studies on the effectiveness of AE-T approaches, such as EMT, have focused on statistical package training and no studies to date have evaluated EMT within the context of a real introductory statistics course. It was hypothesised that EMT would be comparable to GT for analogical transfer tasks, but that EMT would be superior to GT for adaptive transfer tasks. Trial I also assessed other training outcomes including training anxiety, self-efficacy, difficulty and satisfaction in order to explore potential advantages and disadvantages of using either training approach. An explanatory mixed methods design was utilised to allow for follow-up qualitative data to be collected to help explain the quantitative experimental results. Data collected from in-depth semi-structured interviews also aimed to provide a more general exploration of the overall student experience of statistical package training. The results of the qualitative phase are reported separately in Chapter 6.

5.2 Method

An explanatory sequential mixed methods design was used for Trial I and involved collecting quantitative experimental data first and then explaining the quantitative results with in-depth qualitative data (Creswell & Plano Clark, 2011). The quantitative phase reported in this chapter involved a randomised experiment which compared EMT to GT for learning to use the statistical package *SPSS*. The secondary qualitative phase, reported in Chapter 6, used semi-structured in-depth interviews to explain the quantitative results as well as conducting an in-depth exploration of the overall student experience of statistical package training.

5.2.1 Participants

Participants consisted of first year psychology students enrolled in an introductory statistics course which ran concurrently across two campuses. Students were randomly assigned to odd and even week computer laboratory sessions as part of a regular course requirement. Of the 151 students enrolled, 117 consented to participate in the experiment. Three of these consenting students were not randomly allocated but instead placed automatically into available laboratories due to space limitations. There were 14 consenting participants who did not finish training. Seventy-six of these consenting students who finished training completed a post-training follow-up questionnaire. Seventy-nine of the consenting students that finished training in semester one were followed-up in a semester two statistics course. A flowchart summarising the study is shown in Figure 5.1. Table 5.1 displays the characteristics of the sample across the EMT and GT approaches. Note that there was vastly more females in the courses compared to males. This is common for psychology courses. Also note that the smaller proportion of students randomly allocated to GT on Campus A was due to laboratory size limitations for that group.

5.2.2 Measures

The measures included in Trial I are split into four major categories - covariates, manipulation checks, training transfer and other training outcomes. Covariates included

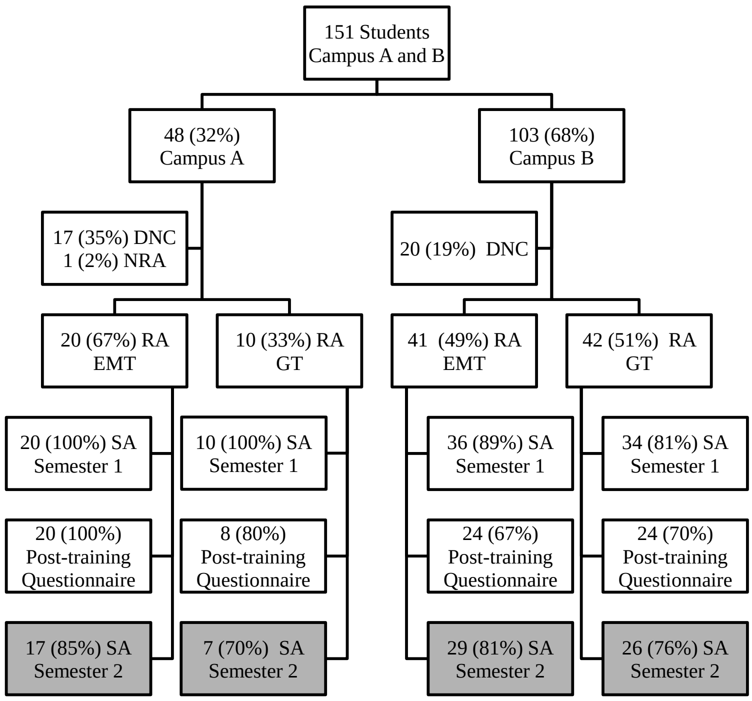


Figure 5.1: Trial I flow chart. Note. RA = Randomly allocated, NRA = Not randomly allocated, DNC = Did not consent, EMT = Error-management training, GT = Guided Training, SA = Completed self-assessment 1 & 2. Semester 2 follow-up has been shaded.

Table 5.1: Statistical Package Training Trial I Sample Characteristics Between approaches

		Strategy		
		GT	EMT	Total
Strategy	N (%)	44(44.0)	56(56.0)	100
Campus A	N (%)	10(33.3)	20(66.7)	30
Campus B	N (%)	34(48.6)	36(51.4)	70
Female	N (%)	31(44.3)	39(55.7)	70
Male	N (%)	13(43.3)	17(56.7)	30
Age	$M \pm SD$	19.84 ± 5.02	19.41 ± 5.07	19.60 ± 5.05

important variables to control for between training approaches and included measures of *statistical knowledge*, *training adherence* and *self-assessment compliance.* Manipulation checks consisted of self-reported measures to validate the correct implementation of training approaches and included *metacognitive activity*, *emotional control*, *error-orientation*, and *exploration* during training. Training outcomes included measures of *analogical*, *adaptive* and *total training transfer* across semester 1 and 2. Other training outcomes included *training anxiety*, *change in statistical package self-efficacy*, *training difficulty*, and *training satisfaction.*

Covariates

Statistical Knowledge. Statistical knowledge, which was defined as the proportion of marks obtained on the end of semester multiple-choice exam, was included as a covariate in the statistical analysis of the results. Statistical knowledge scores were used to control for the influence of statistical knowledge on operating the statistical package. Even though this study employed random allocation to training approaches to help reduce group bias, controlling this covariate would facilitate a more accurate comparison of the two training approaches.

Training Adherence. Training adherence was monitored throughout the semester in order to take into account the extent to which a participant engaged in training. Adherence was measured by two indicators - *laboratory completion* and *laboratory compliance.* Completion was defined as finishing a laboratory training session, whereas compliance was defined as attending an allocated laboratory training session. To construct this score, the number of completed training laboratory sessions was added to the number of times a participant completed training laboratory sessions during their designated laboratory times. If they completed any laboratory session in a different week or during their own time, compliance was scored as zero for that laboratory session. Due to a system error with logging laboratory session 1 grades, only laboratory sessions 2 - 5 were included for the calculation of this score. Therefore, the training adherence scores could range from no adherence (0) to perfect adherence (8)

Self-assessment Compliance. Self-assessment compliance was also taken into consideration. Compliance was defined as whether the student completed both self-assessment training transfer tasks in the allocated self-assessment laboratory session. If the students completed one or two of the self-assessment tasks outside of the allocated self-assessment laboratory session, they were classified as non-compliant. Compliance was important to take into account as students who did not attend the scheduled self-assessment laboratory sessions were not under supervision. These non-compliant students could have gone over the allocated time limit or received assistance from peers who had already completed the self-assessment tasks. Therefore, non-compliance was hypothesised to be associated with inflated self-assessment scores and would need to be controlled for when comparing training approaches on training transfer.

Manipulation Checks

Manipulation checks were measured across both training approaches using items contained in a self-reported post-training questionnaire. All items were responded to on a seven-point likert-type scale ranging from "strongly disagree" (1) to "strongly agree" (7). All items were borrowed or adapted from previous research. Scales composed of multiple item scores were averaged to get a final scale score. The manipulation checks were used to validate the correct implementation of the training approaches. It was hypothesised that the EMT approach would be associated with higher self-reported metacognitive activity, emotional control, error-orientation, and exploration.

Metacognition. The degree to with students engaged in metacognitive activity during training was measured using a self-report scale heavily adapted from Ford et al. (1998). Twelve items rated on a 7-point Likert-type scale ranging from "strongly disagree" (1) to "strongly agree" (7) asked questions relating to the extent to which a participant engaged to metacognitive activities during training (e.g. monitoring, planning and revising, see Appendix A.7). A sample item is "When my methods were not successful for completing statistical procedures in *SPSS*, I experimented with different approaches for completing the procedure". Item scores were averaged to get an overall metacognitive score. Higher scores indicate a higher self-reported level of metacognitive activity during training. Due to the substantial adaptation of the original Ford et al. items, the psychometric properties of the scale items were re-checked. A PCA extracted a single component using the eigenvalue greater than 1 approach which explained 50.54% of the variability in responses to metacognitive activity items. Cronbach's α for the adapted scale was .91(see Appendix A.7). These psychometric properties were regarded as being acceptable.

Emotional Control. The degree to which students developed self-regulatory skills related to emotional control was checked using 8 items adapted from Keith and Frese (2005). These eight items related to the degree to which participants regulated their emotions during training. An example of an item is "When difficulties arose during computer laboratories I did not allow myself to lose my composure". Items were responded to on a 7-point Likert-type scale ranging from "strongly disagree" (1) to "strongly agree" (7). According to a PCA of the adapted items, a unidimensional component explained 55.3% of the variation in responses to the emotional control items. The emotional control scale had high internal consistency with Cronbach's $\alpha = 0.89$ (Appendix A.7).

Error-orientation. As a manipulation check, error-orientation, or a participant's attitude towards errors made during training, was measured using two subscales adapted for statistical package training from the *Error Orientation Questionnaire* (EOQ, Rybowiak, Garst, Frese, & Batinic, 1999). The original EOQ was developed to measure how employees cope with errors committed in the workplace. The two subscales of EOQ, *Error Strain* (5 items, e.g. "When I made a mistake in *SPSS*, I lost my temper and got angry about it") and *Learning from Errors* (4 items, e.g. "From my errors, I have learned a lot about how to work with *SPSS*") had high internal consistency with Cronbach's $\alpha = .79$ and .89 respectively (Rybowiak et al., 1999). These original items were adapted to refer specifically to using the statistical package *SPSS* (see Appendix A.7). The nine items were rated on a 7-point Likert-type scale ranging from "strongly disagree" (1) to "strongly agree" (7). Scores on each item for each subscales were averaged to form an overall *Error Strain* and *Learning from Errors* subscale score. High scores for *Learning from errors* indicate a positive attitude towards errors and high score on *Error Strain* indicate an emotional intolerance for errors. A PCA confirmed the structure of the EOQ with *Learning from errors* accounting for 35.76% and *Error strain* accounting for 28.26% of the variability in responses. Cronbach's α for learning and error strain was .86 and .80 respectively (see Appendix A.7).

Exploration. The extent to which participants engaged in active exploration versus guided instruction during training was measured using six items relating to the following of step-by-step instructions (e.g. “I used step-by-step instructions when learning to use *SPSS*”), copying other students (e.g. “I copied how other students completed tasks in *SPSS*.”), seeking tutor assistance (“When I was unsure about how to complete a task in *SPSS*, I would immediately ask the tutor/or a friend for help”), and actively exploring *SPSS* (e.g. “I explored the features of *SPSS* without much instruction by changing options or trying different analyses in order to complete each laboratory exercise”). These items were loosely based on items adapted from Bell and Kozlowski (2008). A PCA revealed two components (eigenvalues greater than 1). The first component, labelled “Active” explained 35.85% of the variation in responses, whereas the second component, labelled “Guided” explained 20.76%. Cronbach’s α was .69 and .41 for Active and Guided components respectively (Appendix A.7). Due to the unimpressive α coefficients and the fact that these items appeared to assess somewhat unrelated aspects of guided and active-exploratory training, it was decided to individually assess each item when checking the validity of manipulations between training approaches.

Training Transfer

Self-assessment tasks which aimed to measure analogical, adaptive and total training transfer were completed in the final weeks of training between laboratory session 4 and 5 (see Table 5.4). Due to the lack of research in relation to the evaluation of statistical package training transfer and the effectiveness of EMT for statistical packages, designing tasks that aimed to measure training transfer proved quite challenging. When designing the self-assessment tasks, it was important that each task measured a student’s ability to successfully operate the statistical package and not be confounded by the student’s knowledge of statistics. For example, completing an exercise task that gets a student to find the median IQ of the sample may be confounded by the student’s knowledge of the median. While it was virtually impossible to eliminate this statistical knowledge dependency, each exercise task was designed to minimise its effect. For example, exercise questions which were used to score someone on their ability to operate *SPSS* asked questions relating to the acquired output from *SPSS* that proved they had completed the analysis correctly. The questions avoided interpretation of statistics or graphs which would be dependent on a student's statistical knowledge.

The first self-assessment task consisted of eight exercises that measured a student's analogical transfer. These exercises required students to complete similar tasks that had been covered during training. The exercises were based on analysis of a data file provided for the purpose of self-assessment. *SPSS* procedures covered in the analogical transfer self-assessment exercises are shown in Table 5.2. Tests were auto-marked by the online *WebLearn* assessment system described in Section 5.2.2. Exercise questions consisted of a combination of numeric responses and multiple-choice formats. Numeric response questions were given a decimal point tolerance to take into account different rounding precision given by default in *SPSS*. An analogical transfer score was calculated as the total number of questions correct with 8 being the highest possible score. In semester 2, the same analogical transfer items were used along with the conversion of adaptive transfer items 5 and 6 (see Table 5.3) into additional analogical items. These items became analogical because after the self-assessment tasks in semester 1, the participants trained how to complete these tasks in Laboratory 5 (see Table 5.4). Therefore, semester 2 analogical transfer scores were out of 10.

Table 5.2: Trial I Analogical Transfer Self-Assessment 1 Exercises

Task	Description	*SPSS* Procedure
1	Generate basic descriptive statistics	Analyse ⇒ Descriptive Statistics ⇒ Descriptives
2	Compare means between groups	Analyse ⇒ Compare Means ⇒ Means
3	Explore data between groups	Analyse ⇒ Descriptive Statistics ⇒ Explore
4	Generate box-plot of variable between groups	Graphs ⇒ Boxplot
5	Compare distributions between groups	Graphs ⇒ Histogram and Data ⇒ Split File
6	Frequency distributions	Analyse ⇒ Descriptive Statistics ⇒ Frequencies
7	Generate clustered bar chart	Graphs ⇒ Bar
8	Cross-tabulation and χ^2 test	Analyse ⇒ Descriptive Statistics ⇒ Crosstabs

The second self-assessment task originally consisted of 8 exercises that aimed to measure adaptive transfer, but 4 of these exercises were eliminated due to an incorrect data file being linked to the self-assessment in first semester. This error resulted in EMT groups not being able to attempt the last four exercises. Therefore, Adaptive transfer was scored out of 4 in the first semester. Adaptive transfer tasks were designed to be structurally distinct from training and required students to complete tasks and analyses in *SPSS* that were not strictly covered during training. The idea behind these tasks was to get students adapting their knowledge gained from training and applying it in novel situations. To achieve this the adaptive transfer tasks involved procedures in *SPSS* that were not covered in training or the adaptation of previously covered *SPSS* procedure through use of advanced options or chaining (i.e. combining multiple procedures together in unique ways, see Table 5.3). Scores on each adaptive training transfer exercise were summed to form a total adaptive transfer score out of 4. A total transfer score was also computed by summing together analogical and adaptive transfer scores. For semester 2 follow-up, adaptive training transfer was measured using 6 of the original items in Table 5.3. Items 5 and 6 were converted to analogical as how to complete these tasks were eventually covered in Laboratory 5 at the end of first semester.

Other Training Outcomes

Anxiety. Anxiety towards statistical package training was measured using four items adapted from the Tension-pressure dimension scale of the *Intrinsic Motivation Inventory* by Deci and Ryan reported in McAuley, Duncan, and Tammen (1989). These items have been used in previous experimental research (Deci, Eghrari, Patrick, & Leone, 1994; Ryan, 1982) to measure the degree to which participants feel anxiety while completing certain tasks or behaviors. A sample item that was adapted is "I felt pressured when training to use *SPSS*" (see Appendix A.7). These items were responded to on a 7-point Likert-type scale ranging from "strongly disagree" (1) to "strongly agree" (7). Ratings on each of these four item scores were averaged to get an overall statistical package anxiety rating score where higher scores are indicative of higher anxiety. While the original items had good evidence of reliability and validity in sport competition settings (McAuley et al., 1989), the adaptation and application of the scale items to statistical package training required the psychometric properties of the scale to be further validated. The results of a principal components analysis (PCA),

Table 5.3: Trial I Adaptive Transfer Self-Assessment 2 Exercises

Task	Description	*SPSS* Procedure	Adaptive Rationale
1	Multiple line graph	Graphs ⇒ Line	Required adapting knowledge of constructing clustered bar charts (Laboratory 4)
2	Clustered boxplot	Graphs ⇒ Boxplots	Required adapting knowledge of single factor box plots (Laboratory 2) and clustered bar charts (Laboratory 4)
3	Compare means across two factors	Analyse ⇒ Compare Means ⇒ Means	Required adapting knowledge of comparing means across a single factor (Laboratory 1) or by combining Split File (Laboratory 2) and Compare Means (Laboratory 1)
4	Normality test across two factors	Analyse ⇒ Explore	Required combining knowledge of normality tests (Laboratory 2) and split file (Laboratory 2)
5[a],[b]	Create scatterplot	Graphs ⇒ Scatter/Dot	Required adapting knowledge of creating plots from all modules. Bivariate features of *SPSS* were not covered until Laboratory 5
6[a],[b]	Compute correlation	Analyse ⇒ Correlate	Required adapting knowledge of the analyse menu. Bivariate features of *SPSS* were not covered until Laboratory 5
7[a]	Select cases based on a condition	Data ⇒ Select Cases	A highly difficult task by itself (Laboratory 2)
8[a]	Select cases based on two conditions	Data ⇒ Select Cases	Required adapting knowledge of Select cases (Laboratory 2)

[a] Removed from first semester adaptive transfer score due to IT issue.
[b] Converted to analogical tasks in 2nd semester.

using an eigenvalue greater than one criteria for component selection, resulted in a single component which explain 56.01% of the variation in statistical package anxiety scores. The reversed item 4 had the lowest component loading. Internal consistency of the scale revealed that Cronbach's $\alpha = .74$. (see Appendix A.7)

Statistical package self-efficacy Self-efficacy, defined as a participant's confidence in their ability to operate a statistical package after training, was measured using three items from Finney and Schraw's 2003 *Current Statistics Self-efficacy* (CSSE) scale. Participants were required to rate their level of confidence in their current ability to use *SPSS* for generating descriptive statistics, graphical displays and statistical inference. An example of an item is "To use the statistical package to conduct statistical inference (e.g. generate p-values)". A similar seven-point likert scale ranging from (1) no confidence at all to (7) complete confidence was used. Scores for the three item scores were averaged to form a single self-efficacy score (Cronbach's $\alpha = .83$). A PCA extracted a single construct which explained 74.23% of the variation in responses (see Appendix A.7).

Perceived Difficulty. The perceived difficulty of the training conditions was measured by asking participants to rate the overall difficulty of training to use *SPSS* on a 7-point Likert-type scale ranging from (1) "very easy" to (7) "very difficult".

Training Satisfaction. Following the five training laboratories, students rated their perceived level of training satisfaction on a 7-point Likert-type scale ranging from (1) "not at all Satisfied" to (7) "very Satisfied". This item was used to assess the student's attitudes towards training.

Procedure

Following ethics approval by the RMIT College Human Ethics Advisory Network on the 22^{th} November 2010 (Project No: BSEHAPP 48-10) and random allocation to odd and even week computer laboratories, students were approached before their lecture to participate in the study (see Appendix A.5 and A.6 for the consent forms and plain language statement used in this study respectively). Non-consenting students were still required to complete training, but their data was not recorded. The allocation to different laboratories was due to limitations with size and availability of large computer rooms. This odd and even week group allocation allowed for the manipulation of training approaches. The ordering of EMT and GT to odd and even weeks was counterbalanced between the Campuses which was considered a confounding variable (see Table 5.4). Campus A had GT on odd weeks and EMT on even weeks. On campus B the order was reversed. Counterbalancing the order controlled for possible time effects introduced by using odd and even weeks. For Campus A, there were a few issues with the training schedule. The day of Week 3 laboratories fell on a public holiday which meant that both the GT and EMT had to be accommodated into laboratories on the same day in Week 4. The same was completed in Week 7 after an IT issue with the university network prevented students from accessing their online training material during Week 6.

Table 5.4: Trial I Training Laboratory Schedule Across Campus and Condition

	Campus A		Campus B	
Week	GT	EMT	GT	EMT
Week 1	Laboratory 1			Laboratory 1
Week 2		Laboratory 1	Laboratory 1	
Week 3	Public Holiday			Laboratory 2
Week 4	Laboratory 2	Laboratory 2	Laboratory 2	
Week 5	Laboratory 3			Laboratory 3
Week 6		IT Error	Laboratory 3	
Week 7	Laboratory 4	Laboratory 3		Laboratory 4
Week 8		Laboratory 4	Laboratory 4	
Break				
Week 9	SA 1 + 2			SA 1 + 2
Week 10		SA 1 + 2	SA 1 + 2	
Week 11	Laboratory 5			Laboratory 5
Week 12		Laboratory 5	Laboratory 5	

Note. SA = Self-assessment.

Training consisted of five laboratories which corresponded procedures in *SPSS* with course content (see Table 5.5). Self-assessment tasks were completed towards the end of the semester between Laboratory 4 and 5 (see Table 5.4). Laboratories were scheduled for one hour per week, however students were permitted to stay longer to finish or catch-up. Students who missed their designated laboratory needed to ask permission to attend a non-designated laboratory. This was done so as to not disadvantage students and was a condition for ethics approval. This meant that some students were mixing approaches. This issue and the fact that students talk to each other, made blinding participants to the approaches impossible. However, the exact nature of the Trial was never explained to participants during the trial.

Table 5.5: Trial I Statistical Package Training Laboratory Content

Laboratory Title	Topics
1. *SPSS* Introduction	An overview of *SPSS*
	Entering data
	Editing variable properties
	Saving your work
	Descriptive statistics
	Editing graphs
	Exporting Analysis
2. *SPSS* Basics	Revision from Laboratory 1
	Normality Tests
	Box plots
	Histograms
	Split File
	Select Cases
3. Frequencies and Bar Charts	Revision from Laboratory 1 and 2
	Frequencies
	Recoding variables
	Bar charts
4. Cross-tabs and χ^2 tests	Revision from Laboratory 1, 2, and 3
	Cross-tabs
	Custom Tables
	χ^2 tests of association
	Clustered Bar Charts
5. Correlation and Regression	Revision from Laboratory 1, 2, 3, and 4
	Scatter plots
	Correlation
	Regression
	Testing assumptions of regression

Training was delivered using a proprietary, online assessment system called *WebLearn*. *WebLearn* is similar to a streamlined version of Blackboard's quiz, test and assignment features. Each laboratory consisted of objectives, instructions and exercises embedded with the approaches' instructions. Students would sequentially work through each exercise which were designed to introduce and get them practising the procedures of *SPSS*. To show that the student had successfully completed the procedure in *SPSS*, each exercise required students to answer a question that could only be answered if they had correctly operated *SPSS*. Students were required to score 70% or above for each laboratory to obtain a pass grade. If they passed the laboratory, they would get their participation mark for the course. Student's were allowed to reattempt a laboratory if they failed their first attempt. To find out if they had passed the laboratory, the student would submit all their answers once they had completed the laboratory exercises to the *WebLearn* system for marking. Marking was done automatically by the *WebLearn* server where the correct answers were stored.

Training Approaches

The GT group received step-by-step comprehensive instructions and screen shots summarising each exercise in *SPSS* (Figure 5.2a). The students were instructed to follow these steps and answer questions that confirmed they had completed the exercise correctly. GT was designed to minimise errors during training by showing students exactly how every exercise needed to be completed. Prior to each laboratory the following GT instructions were given to students.

GT Instructions

Follow the instructions carefully to avoid making mistakes. The instructions have been designed to keep you on track and learning *SPSS* in an efficient manner. If you get stuck, politely raise your hand for the tutor's assistance. Please remember to be patient as there are many other students who may also need the tutor's attention. Remember, for training to be the most effective, you should try not to make errors. The instructions will help you avoid them.

A laboratory tutor was also present during each scheduled session. In the GT approach, the tutor was instructed to help the participants as much as they needed in line with the theory of GT. GT conditions were also provided with feedback to their errors based on the theory that errors should be avoided. These heuristics reinforced Skinner's (1968) theory of programmed learning. Some examples of GT heuristics included the following:

- "Try again. Follow the instructions carefully"
- "Concentrate and try again"

- “If you are stuck, ask the tutor for assistance”
- “Read the instructions again”
- “Make sure you have followed the steps”

The EMT approach was given the exact same exercises but with modified instructions and no screen shots. The EMT approach used minimal instruction to get the participant actively exploring *SPSS* (Figure 5.2b). Instructions were designed to point the students in the right direction (minimal instruction), but left them to work out the specifics. Sometimes for difficult procedures or analyses, hints were given to help students get back on track if they got stuck. Each EMT laboratory began with the following instructions to trainees:

EMT Instructions

During training, you should expect to make errors as you learn to use *SPSS*. If you make an error, that’s great! Errors are a positive part of any learning experience. As a result of making errors, you can learn from your mistakes. If you do make an error, you are encouraged to find the solution yourself. Relax, think about the problem you are having and attempt to overcome it by trying something new. Don’t be afraid to make a few more mistakes attempting to solve the issue. Eventually you will figure it out. If you cannot find the solution within a few minutes, raise your hand for the tutor’s assistance. Please remember to be patient as there are many other students who may also need the tutor’s attention.

Students were also presented with error framing heuristics listed at the top of each exercise (Figure 5.2b). These were presented to students to assist them in framing errors in a positive light and enable them to deal with negative emotions associated with making errors. The heuristics were sourced from the literature as well as two others being created from the purpose of this study. These heuristics included the following:

- “If you have a problem, regard it as a learning opportunity” (Wood et al., 2000)

- "Errors are a natural part of learning. They point out what you can still learn!" (Dormann & Frese, 1994)
- "The more errors you make, the more you learn!" (Heimbeck et al., 2003)
- "The only bad errors are the ones you don't learn from"
- "Don't discount your errors. Acknowledge and learn from them"

In the EMT approach, the same tutor used in GT was instructed to encourage the students to find the solution themselves. If the participant was struggling after multiple attempts, the tutor was allowed to give them a hint to get them back on track. The tutor was also trained to reinforce the positive error framing heuristics by encouraging students to learn from their mistakes.

Self-assessment Tasks

Both approaches completed the same self-assessment tasks in weeks 9 and 10 after Laboratory 4. Self-assessment tasks were used for ethical reasons due to the potential of graded exams to be influenced by difference between approaches. The self-assessment tasks were administered online using the *WebLearn* system. Students answered the exercises by submitting answers generated from analysing the provided data file using *SPSS*. Each of the eight exercises included in self-assessment 1 was randomly drawn from a pool of similar questions. This was done to prevent students from copying other students' answers. The same randomisation procedure was used in Self-assessment 2.

Students were given 25 minutes to complete each self-assessment task. However, as students were able to complete laboratories outside of allocated laboratory times, those students may have gone overtime or got assistance from peers. Therefore, it was important to control for self-assessment compliance when analysing the results of the trial. Students were instructed that while they should aim to get all questions correct, to get the grade for the self-assessment, they would need to get 4/8 on the first assessment and 2/4 on self-assessment 2. Students were instructed to work on the self-assessment task themselves and were not permitted to talk or seek assistance from other students. Participants were allowed to attempt each self-assessment up to five times as the laboratories and self-assessment were graded on completion (formative

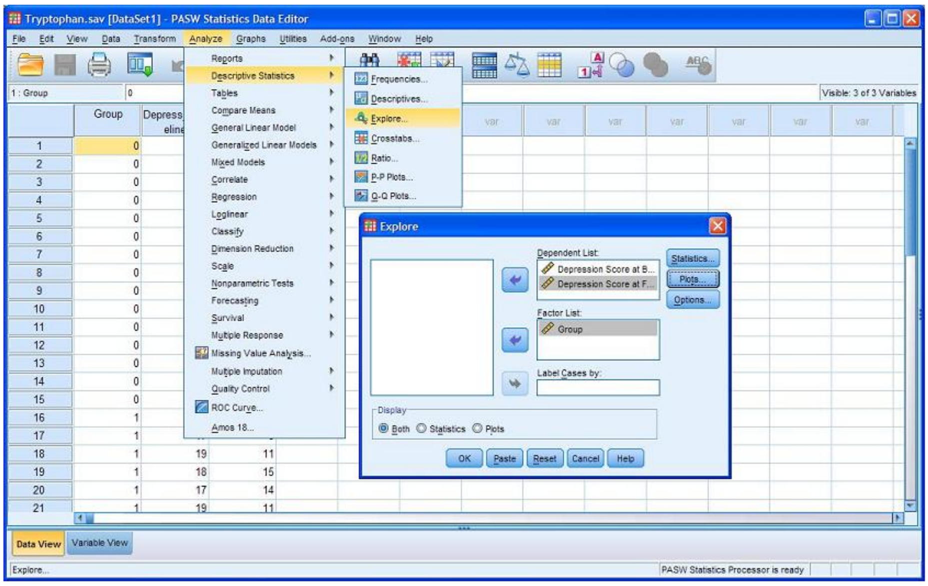
Revision - Explore

Let's run **Explore** on "Depression Score at Baseline" and "Depression Score at Follow-up" between the treatment and placebo group.

Follow these steps:

1. Click **Analyse⇒Descriptive Statistics⇒Explore**
2. Move "Depression Score at Baseline" and "Depression Score at Follow-up" into the **Dependent List** box
3. Move "Group" into the **Factor List** box
4. Click **OK** to explore your data.

These steps are summarised in the screen shot below.



What is the **variance** of Depression Score at **Baseline** for the **placebo** group?

Enter your response below:

(a) GT Exercise Example

"Don't discount your errors. Acknowledge and learn from them."

Revision - Explore

Let's run **Explore** on "Depression Score at Baseline" and "Depression Score at Follow-up" between the treatment and placebo group.

Location: Analyse⇒Descriptive Statistics⇒Explore

What is the **variance** of Depression Score at **Baseline** for the **placebo** group?

Enter your response below:

(b) EMT Exercise Example

Figure 5.2: Example of GT and EMT exercise instructions in *WebLearn* - Trial I

assessment). Obviously, it was in the student's best interest to pass on the first attempt or else they would need to retry in their own time. Only a participant's score on their first attempt was recorded for measuring training transfer.

Self-assessment tasks 1 and 2 were given to students again in the first laboratories of semester 2 as a follow-up. As two of the original adaptive exercises given in Self-assessment 2 were tasks covered in Laboratory 5 (see Table 5.3), they could no longer be regarded as adaptive. These exercises were maintained, but their scores were transferred to analogical scale. Therefore, for semester 2 self-assessment tasks, analogical transfer was marked out of 10 and adaptive transfer was marked out of 6. Students were instructed to complete these task as a revision exercise. There was no requirement to get a certain score to pass. Students were permitted to attempt the tasks as many times as they liked, however, only the students' first attempts were recorded for follow-up.

In the final lecture following semester one's training, students were approached to fill out the self-reported post-training questionnaire which contained the manipulation check and other training outcome items (difficulty, satisfaction, self-efficacy and anxiety). An online version of this post-training questionnaire was also used to follow-up students who did not attend the final lecture.

5.3 Results

5.3.1 Data Analysis

Results are presented in four sections. In the first section, descriptive statistics and intercorrelations between the study variables are reported. In the next section, analysis of covariance (ANCOVA) models are used to model the difference between training approaches on training transfer outcomes after controlling for the effects of training covariates. To validate the correct manipulation of training approaches, average student scores on manipulation check are compared between approaches using two-sample t-tests in the third section. Other training outcomes, i.e. training difficulty, satisfaction, self-efficacy, and anxiety (see Section 5.2.2), are finally compared between training approaches in the fourth section also using two-sample t-tests. As these t-tests were not independent between each other (i.e. multiple-comparisons), the p-values were used as indicators only.

5.3.2 Descriptive Statistics and Intercorrelations

Descriptive statistics and intercorrelations for training transfer outcomes and covariates are shown in Table 5.6. For the covariates, the EMT approach had higher mean training adherence and post-training compliance, but lower statistical knowledge and follow-up training compliance when compared to the GT approach. Descriptively at post-training, the EMT approach outscored the GT approach on analogical and total training transfer scores, but not on adaptive transfer. At follow-up, the EMT group out-scored the GT group on adaptive transfer, but the GT group appeared to do better on analogical and total transfer scores.

5.3.3 Modelling Training Transfer

One-way analysis of covariance (ANCOVA) was performed to assess for significant differences between the GT and EMT approaches on mean post-training and follow-up transfer outcomes (see Table 5.7). The ANCOVA models used training adherence, self-assessment compliance and statistical knowledge as covariates. Table 5.7 contains the ANCOVA model parameters and covariate adjusted means with 95% CI for all three training transfer outcomes across post-training and follow-up. Each model's assumptions were checked for evidence of any strong violations to the assumption of homogeneity of variance between cohorts, homogeneity of regression slopes, and approximate normality of residual error. No strong evidence of any violated assumptions emerged. The partial η^2 statistic has been included as an estimate of effect size. The η^2 statistic reflects the proportion of variability in an outcome variable that can be explained by its relationship with a particular variable after controlling for the effects of other variables in a model. All covariance adjusted outcome means between groups and across semesters are summarised in Figure 5.3.

The primary focus of the ANCOVA models was to compare the training approaches on training transfer outcomes after controlling for statistical knowledge, training adherence, and self-assessment compliance (Table 5.7). According to the first semester post-training outcomes there were no statistically significant differences between approaches

Table 5.6: Trial I Intercorrelations and Descriptive Statistics of Study Variables Between Training Approaches

Variable		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Training Adherence		-	.270**	.137	.146	.174	.122	.061	.104	.379**	.168
2. Statistical Knowledge			-	.299**	.219*	.325**	.451**	.431**	.507**	.124	.043
3. Analogical Transfer 1				-	.321**	.917**	.433**	.369**	.460**	-.176	-.177
4. Adaptive Transfer 1					-	.663**	.214	.412**	.364**	-.181	-.119
5. Total Transfer 1						-	.426**	.459**	.509**	-.205*	-.195
6. Analogical Transfer 2							-	.513**	.861**	-.097	-.024
7. Adaptive Transfer 2								-	.878**	.006	-.059
8. Total Transfer 2									-	-.050	-.049
9. SA Compliance 1										-	-.042
10. SA Compliance 2											-
EMT	*M*	7.05	0.69	5.41	1.66	7.07	7.54	2.89	10.43	71.4%	84.8%
	SD	1.38	0.15	1.66	0.94	2.16	1.53	1.90	2.93		
	N	56	56	56	56	56	46	46	46	56	46
Guided	*M*	6.84	0.74	5.23	1.72	6.91	7.75	2.69	10.44	47.7%	96.9%
	SD	1.40	0.15	1.92	0.88	2.31	1.97	1.73	3.33		
	N	44	41	44	43	44	32	32	32	44	32
Total	*M*	6.96	0.71	5.33	1.69	7.00	7.63	2.81	10.44	61.0%	89.7%
	SD	1.38	0.15	1.77	0.91	2.22	1.71	1.82	3.08		
	N	100	97	100	99	100	78	78	78	100	78

Note. 1 refers to Post-training (1^{st} Semester) and 2 refers to Follow-up (2^{nd} Semester).
* $p < 0.05$. ** $p < 0.01$.

Table 5.7: Trial I ANCOVA Models Predicting Training Transfer Measures

	Analogical			Adaptive			Total Transfer		
Parameter	*B*	95% *CI*	η^2	*B*	95% *CI*	η^2	*B*	95% *CI*	η^2
Post Training (Semester 1)									
Statistical Knowledge	2.42*	(1.47, 6.13)	0.10	1.26	(-0.01, 2.52)	0.04	4.97**	(2.12, 7.82)	0.12
Training Adherence	0.16	(-0.1, 0.42)	0.02	0.13	(-0.01, 0.27)	0.03	0.30	(-0.02, 0.62)	0.04
SA Compliance 1[a]	-1.08**	(-1.82, -0.34)	0.08	-0.50*	(-0.91, -0.10)	0.06	-1.54**	(-2.44, -0.64)	0.11
Strategy[b]	-0.52	(-1.21, 0.17)	0.02	-0.06	(-0.43, 0.32)	0.00	-0.61	(-1.45, 0.23)	0.02
GT Adjusted Mean[c]	5.06	(4.55, 5.57)		1.66	(1.38, 1.94)		6.69	(6.06, 7.32)	
EMT Adjusted Mean[c]	5.58	(5.15, 6.02)		1.72	(1.49, 1.96)		7.30	(6.77, 7.83)	
Follow-up (Semester 2)									
Statistical Knowledge	5.83**	(3.01, 8.66)	0.19	6.45**	(3.44, 9.46)	0.20	12.28**	(7.41, 17.16)	0.26
Training Adherence	0.09	(-0.23, 0.41)	0.00	-0.04	(-0.38, 0.30)	0.00	0.05	(-0.50, 0.60)	0.00
SA Compliance 2[a]	-0.30	(-1.51, 0.92)	0.00	-0.29	(-1.58, 1.00)	0.00	-0.59	(-2.68, 1.51)	0.00
Strategy[b]	-0.01	(-0.77, 0.75)	0.00	-0.50	(-1.31, 0.32)	0.02	-0.51	(-1.82, 0.81)	0.01
GT Adjusted Mean[c]	7.62	(7.05, 8.19)		2.52	(1.91, 3.13)		10.14	(9.15, 11.13)	
EMT Adjusted Mean[c]	7.63	(7.16, 8.11)		3.01	(2.51, 3.51)		10.64	(9.83, 11.46)	

[a] Compliant students = 1. [b] GT = 1.

[c] Means after adjusting for the covariates of training adherence, SA compliance, and statistical knowledge.

1 refers to Post-training (1^{st} Semester) and 2 refers to Follow-up (2^{nd} Semester)

* $p < .05$, ** $p < .01$

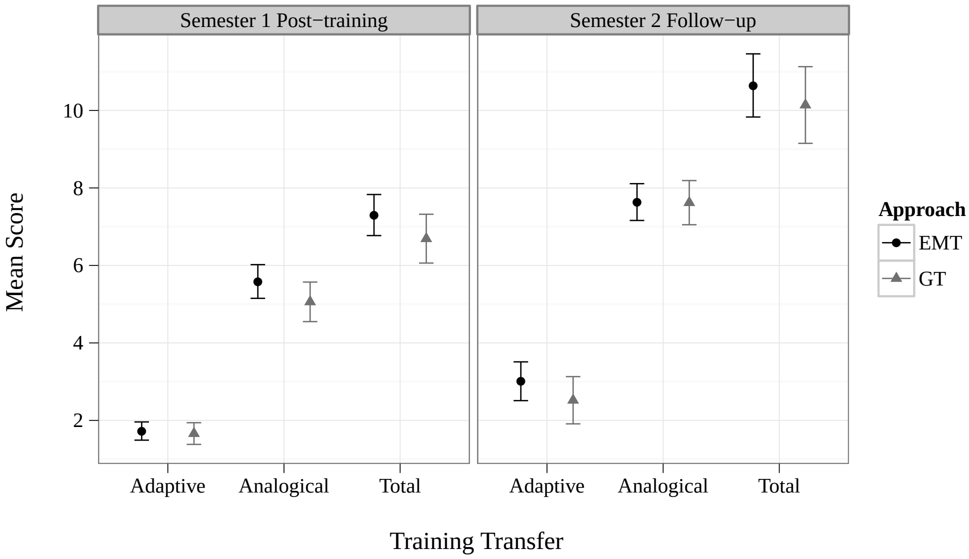


Figure 5.3: Trial I Covariate adjusted training transfer means with 95% CI between training approaches and across semesters.

on mean analogical, $F(1, 92) = 2.25, p = 0.137, \eta^2 = .02$, adaptive, $F(1, 91) = 0.10, p = .754, \eta^2 = .00$ and total training transfer scores, $F(1, 92) = 2.08, p = 0.153, \eta^2 = .02$, after controlling for covariates (Figure 5.3). The same non-significant trend was found at second semester follow-up, analogical, $F(1, 73) = 0.001, p = .978, \eta^2 = 0$, adaptive, $F(1, 73) = 1.47, p = .23, \eta^2 = .02$ and total training transfer scores, $F(1, 73) = 0.59, p = 0.447, \eta^2 = .008$ (Figure 5.3).

In all models, except for adaptive transfer at post-training, statistical knowledge was a statistically significant positive covariate (Table 5.7). This indicated that there was a positive relationship between training transfer outcomes and statistical knowledge. In addition to this finding, at follow-up in semester two the effect of statistical knowledge increased (see η^2 in Table 5.7). This suggests that as the gap between training completion and follow-up increases, the ability to operate a statistical package becomes more dependent on a student's knowledge of statistics. Compliance was also a statistically significant covariate for all outcomes at post-training, but not for follow-up. According to the ANCOVA models in Table 5.7, compliance was associated with lower transfer scores. This suggested that non-compliers would be at a significant advantage on self-assessment tasks when compared to participants that completed self-assessment tasks under controlled conditions. The effect of self-assessment compliance at follow-up was probably less pronounced as overall compliance at follow-up was much higher (see Table 5.6).

5.3.4 Manipulation Checks

Self-reported measures of metacognition, emotional control, active exploration, error strain, learning from errors and guided instruction were analysed to determine the validity of the differences between training approaches (see Appendix A.7 for full item descriptions). Assuming the approaches were imposed correctly, the EMT group would be expected to have higher mean ratings on metacognition, emotional control, learning from errors, exploring without instruction, operating without instruction and actively exploring *SPSS*. The EMT would also be expected to have lower mean ratings on Error Strain, the use of step-by-step instructions, copying from other students and immediately seeking assistance.

A series of two-sample *t*-tests found only one statistically significant difference in mean responses to the "Used step-by-step instructions" item (Table 5.8 and Figure 5.4). No significant differences were found between approaches on mean ratings of emotional control, error strain, learning from errors, copied other students, immediately sought assistance, explored without instruction, operate with instruction, and actively explored SPSS.

5.3.5 Other Training Outcomes

Differences between training approaches on measures of student perceptions towards training difficulty, satisfaction, self-efficacy and anxiety (see Section 5.2.2) were also assessed using items from the follow-up questionnaire validly completed by 78/100 (78%) of the original consenting sample. It was important to look at these subjective outcomes to explore potential advantages and disadvantages of using either training approach. Given the expected increased uncertainty presented by the EMT condition, there was a concern that students in that approach may experience greater levels of perceived training difficulty which may lead to higher levels of training anxiety, lower statistical package self-efficacy and lower overall training satisfaction. The results of the two-sample *t*-tests comparing the conditions indicated otherwise (see Table 5.9). On

Table 5.8: Descriptive Statistics and Two-sample t-tests Comparing Mean Ratings of Manipulation Checks between Training Approaches - Trial I

								95% *CI* of Difference	
Manipulation Variable		M	SD	N	SEM	t	p	Lower	Upper
Metacognition	GT	4.06	1.03	33	0.18	-0.98	.33	-0.69	0.24
	EMT	4.29	1.01	45	0.15				
Emotional Control	GT	3.93	0.60	33	0.10	-0.70	.49	-0.34	0.17
	EMT	4.02	0.52	45	0.08				
Learning from Errors	GT	4.01	1.26	33	0.22	-1.65	.10	-1.05	0.10
	EMT	4.48	1.25	45	0.19				
Error Strain	GT	3.47	1.17	33	0.20	-0.47	.64	-0.76	0.47
	EMT	3.62	1.47	45	0.22				
Used step-by-step instructions	GT	6.58	.66	33	0.12	3.23[a]	<.001**	0.33	1.40
	EMT	5.71	1.62	45	0.24				
Copied other students	GT	3.52	2.06	33	0.36	1.22	.23	-0.35	1.48
	EMT	2.95	1.94	44	0.29				
Immediately sought assistance	GT	5.15	1.62	33	0.28	1.64	.11	-0.14	1.47
	EMT	4.49	1.87	45	0.28				
Explored without instruction	GT	3.39	2.06	33	0.36	-1.34	.18	-1.40	0.27
	EMT	3.96	1.64	45	0.24				
Operate without instruction	GT	3.61	1.98	33	0.35	-1.10	.28	-1.36	0.39
	EMT	4.09	1.87	45	0.28				
Actively explored SPSS	GT	3.91	1.79	33	0.31	-0.12	.90	-0.80	0.70
	EMT	3.95	1.51	44	0.23				

[a] Equal variances not assumed. * $p < .05$, ** $p < .01$

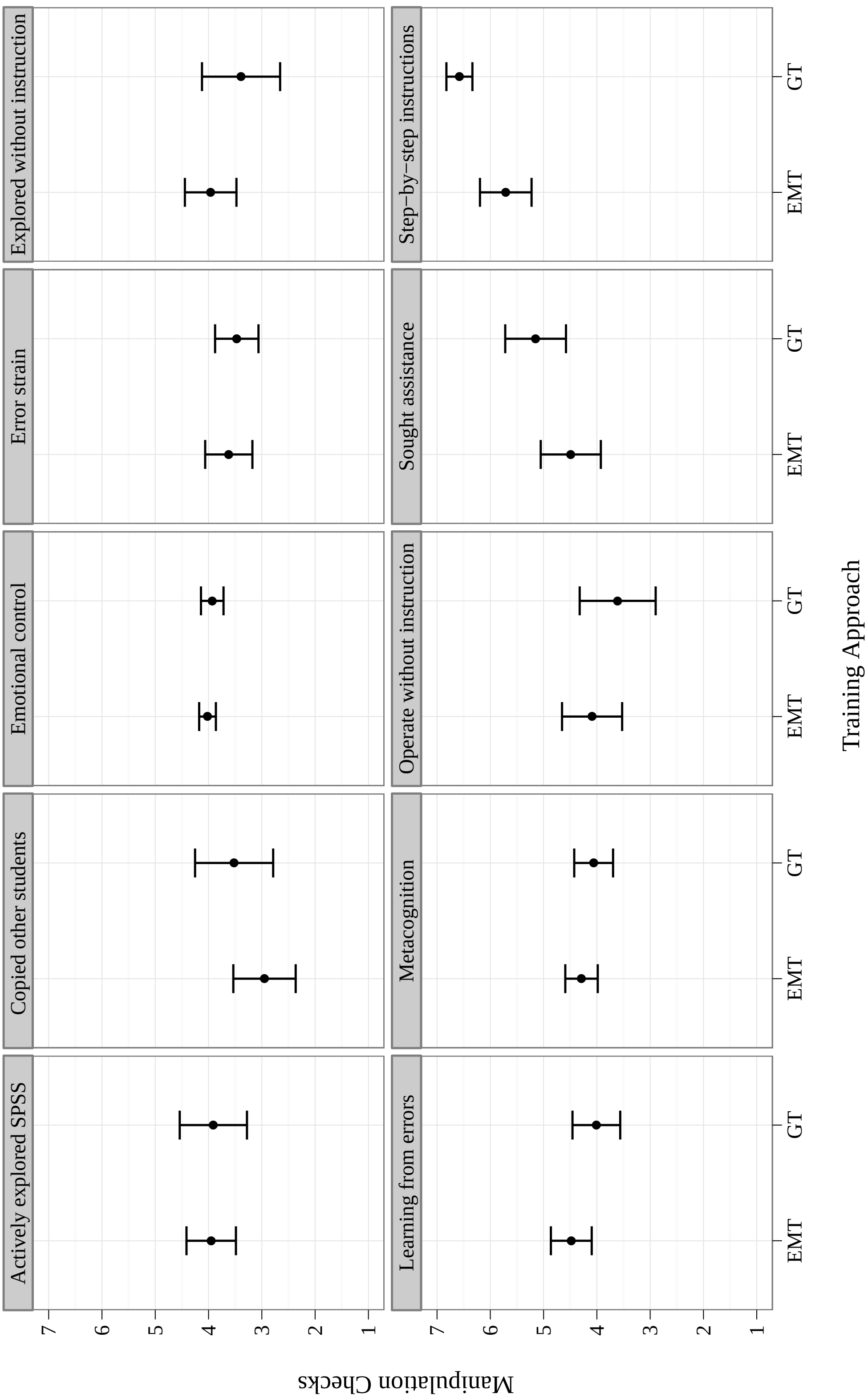


Figure 5.4: Trial I mean ratings with 95% CI for training manipulation checks between trainng approaches.

average, the EMT group rated training difficulty and training anxiety higher compared to GT, but the difference in means was not statistically significant (see Table 5.9). The EMT group rated their statistical package self-efficacy lower than the GT approach on average, but once again, this difference was not statistically significant. Surprisingly, it was the EMT group that scored a higher mean training satisfaction score when compared to GT, but the results of the means comparison was not statistically significant. Overall, across all measures of student perceptions of training, there was no evidence of any statistically significant differences.

Table 5.9: Descriptive Statistics and Two-sample *t*-tests Comparing Training approaches on Other Training Outcomes - Trial I

								95% *CI* of Difference	
Manipulation Variable		*M*	*SD*	*N*	*SEM*	*t*	*p*	Lower	Upper
Training Difficulty	GT	3.85	1.58	33	.28	-1.63	.11	-1.18	0.12
	EMT	4.38	1.28	45	.19				
Training Satisfaction	GT	4.24	1.66	33	.29	-0.51	.61	-0.88	0.52
	EMT	4.42	1.44	45	.21				
Self-efficacy	GT	4.64	1.13	33	.20	0.16	.88	-0.51	0.60
	EMT	4.60	1.26	45	.19				
Anxiety	GT	4.03	1.23	33	.22	-1.41	.16	-0.96	0.17
	EMT	4.43	1.22	45	.18				

* $p < .05$, ** $p < .01$, CT = Certification Task

5.4 Discussion

The results of Trial I found no statistically significant difference between EMT and GT approaches on measures of analogical, adaptive, and total training transfer at both post-training and follow-up after controlling for statistical knowledge, training adherence and self-assessment compliance. These findings failed to support the research hypothesis of this study and failed to support the findings of previous research (Keith & Frese, 2008; Keith, Richter, & Naumann, 2010; Dormann & Frese, 1994).

Statistical knowledge was the only reliable and significant predictor of training transfer performance. This study also showed that this dependency became stronger with time between post-training and follow-up in second semester. There are two likely interpretations for this finding. The first suggests that a student's ongoing ability to operate a statistical package is largely dependent on their knowledge of statistics. However, an alternate interpretation is that the self-assessment tasks were largely measuring statistical knowledge instead of the ability to operate a statistical package. This study assumed that after controlling for statistical knowledge, the remaining variability in transfer scores could be attributed to variability in statistical package skills. However, there is no direct way to test this assertion. Further research is needed to better understand this relationship and its implications on training design and outcomes. Future research also needs to examine how statistical package skills can be properly assessed incorporating this very likely dependency. Regardless, this study was the first to provide evidence of a relationship between statistical package skills and knowledge of statistics. This relationship will be important to control for in future studies that compare the effectiveness of different training approaches.

The second aim of this study was to investigate important advantages and disadvantages to implementing either of the training approaches into an introductory statistics course. This study looked at students' self-reported perceptions of training difficulty, training satisfaction, training anxiety and statistical package self-efficacy. Some instructors might be concerned that the EMT approach might be more difficult for students leading to increased anxiety and lower self-efficacy. This may then lead to lower overall student satisfaction towards training. However, the results of this study failed to find any statistically significant evidence to support this concern. There were no significant differences between students' mean self-reported ratings of these outcomes.

A number of limitations to the study and training design must be considered before drawing conclusions. This study used a sample of psychology students, which are unlikely to reflect the diverse characteristics of all students who take introductory statistics courses. Therefore, the results must be cautiously generalised to other backgrounds. EMT was hypothesised to have the greatest effect on adaptive transfer, but with four out of the eight adaptive transfer tasks being removed due to online technical difficulties for post-training self-assessment, the exact effect of EMT on adaptive transfer at post-training remains to be seen. It is difficult to determine what would have happened if the IT issue did not occur, but it would be safe to assume that the inclusion of four more adaptive transfer tasks would have introduced more variability in adaptive transfer scores and made it easier to detect differences between approaches if those differences existed.

In terms of the study design, this experiment was un-blinded. While students were never explicitly made aware of the nature of this study, it is highly probable that students became aware of the difference between approaches as the semester progressed. The tutor was also un-blinded to the nature of the approaches. While it is difficult to speculate the exact influence this might have had on the results, the potential for bias cannot be ruled out. However, this type of experimental control is always going to be difficult to achieve in real-world educational research.

The major strength of this study, ecological validity, i.e. embedding the evaluation of EMT into a real introductory statistics course, was also its greatest limitation. Due to limited laboratory availability, training was scheduled on a fortnightly basis for each group. This meant that students had only a minimum estimated training time of four hours with *SPSS* before taking the self-assessment tasks. Given the large time intervals between training and the relative shortness of training, it is possible that the effects of training were interrupted and poorly consolidated. Future studies need to provide more frequent and consistent training throughout a course.

The training laboratory sessions were compulsory, but a large number of students missed laboratory sessions on a regular basis. This raised issues with training compliance. Due to ethical reasons, these students were permitted to attend laboratory sessions of the opposite training approach or complete the laboratory sessions in their own time. However, these students still received their respective approaches' instructions as the laboratory sessions were delivered through an online learning system which based laboratory session instructions (GT vs. EMT) on their allocated training approach. The results of the statistical models predicting training transfer performance at post-training found that non-compliance with the self-assessment, i.e. doing the self-assessment outside of the designated laboratory session, was associated with higher self-assessment scores. Non-compliant students probably did not adhere to the self-assessment time limit or received help from peers who had already completed the self-assessment tasks. As attendance was recorded at all laboratory sessions, controlling for measures of training adherence and self-assessment compliance in the statistical models have at least partially taken these limitations into account. However, future research could benefit by ensuring students remain blinded and are given extra incentive to attend allocated laboratory sessions.

The laboratory sessions were scheduled for one hour. While the training was designed to fit within this time period, anecdotally many students reported feeling under time pressure which resulted in them rushing through laboratory sessions and using guesswork to get the laboratory sessions done in the designated time. It is possible that time constraints negatively impacted the EMT approach and violated the error framing instructions. Under time constraints, it would be very difficult for a student to view errors as anything else but a waste of time. While the availability of computer laboratories was outside the control of the researchers, a possible solution to this problem would be to provide further training opportunities so that students had adequate time to work through training material.

All training was graded in terms of satisfactory completion and students were allowed multiple attempts at the training laboratory sessions and self-assessment tasks. This feature of training may have resulted in unmotivated students not expending their greatest effort on self-assessment tasks. Instead, they may have done just enough to attain a level of satisfactory completion. The issues of low incentive may have masked a participant's true ability on the self-assessment tasks. While randomisation provided some level of protection against this issue biasing a particular training approach, in the future, assessment that better engages students in demonstrating their ability to operate a statistical package should be used.

There were also a number of important limitations related to the delivery of training approaches and the assessment of statistical package training transfer. While the researchers of this study were familiar with active learning approaches, this was the first time EMT was implemented for statistical package training at the study's institution. It was also the first time, to the authors' knowledge, that statistical package adaptive training transfer outcomes were formally assessed and reported in the literature. As such, many aspects of this study required the adaptation of methods and measures used in previous research. Only one study by Dormann and Frese (1994) related specifically to statistical package training. However, due to the age of this study, the absence of a specific mention of adaptive transfer and implementation of a one off training session outside of a statistics course, the Dormann and Frese experiment provided only limited insight into the delivery of EMT and assessment of training transfer outcomes. Therefore, the delivery and assessment of training transfer required careful evaluation and reflection.

The results of the manipulation checks brought the validity of the EMT approach into question. If this study implemented EMT successfully then, when compared to participants in GT, participants in the EMT approach would be hypothesised to self-report more metacognitive activity, evidence of exploratory behaviour, positive attitudes towards making errors and better emotional control. The only difference observed between approaches on the manipulation checks was for the use of step-by-step instructions. While the EMT group scored significantly lower, they still had a highly positive average level of agreement. This rating seemed too high assuming minimal instruction had been used correctly in the EMT approach. It is likely that participants in the EMT approach perceived the sequential delivery of exercises during training and the provision of training hints as providing guidance similar to step-by-step instructions. The results of the manipulation checks indicate that there may be a problem with the validity of the EMT approach.

The self-assessment tasks used as measures of training transfer outcomes were also limited. As there was no literature to base the design of these tasks on, their validity as measures of analogical and adaptive transfer for statistical package training only extends to face validity. The strong relationship between statistical knowledge and training transfer suggests that less dependent methods need to be explored in order to get a more valid measure of a student's ability to operate a statistical package. The degree to which the self-assessment tasks captured analogical versus adaptive transfer was also an issue. Adaptive transfer is likely to be demonstrated by what students do spontaneously when working on their own statistical analysis problems outside of training. The degree to which this ability was captured using the self-assessment tasks used in this study was questionable. Future research on the assessment of statistical package training transfer is needed so that these outcomes can be reliably and validly measured in the future.

5.5 Conclusion

After a critical analysis of the results, manipulation checks and methods, it is clear that further research is needed before a clear conclusion is reached about the relative merit of EMT over GT for statistical package training. While Trial I may have been unsuccessful in detecting the true effect of EMT, it did provide a valuable foundation to support future studies in this fertile area of statistics education research.

Therefore, Trial II, was conducted to build upon the results of Trial I and address the following major limitations:

- Improve the validity of imposing EMT for statistical packages to ensure it abides by the principles of active-exploration, minimal instruction and positive error-framing.
- Increase training time and practice opportunities.
- Ensure students are properly blinded to the differences between training conditions
- Design training sessions to reinforce statistical knowledge as it may help enhance the effectiveness of training.
- Design and utilise an improved measure of statistical package training transfer that aims to better engage students and provide a more valid measure of students' technological skills.

Trial II continued to address ecological validity as the literature is already flooded with studies demonstrating the external validity of the superiority of EMT over GT in highly controlled studies (Keith & Frese, 2008). However, until the superiority of EMT can be demonstrated in real-world introductory statistics courses, EMT cannot be recommended over GT. At the conclusion of Trial I, it still remained to be seen whether "less guidance is more" when it comes to training students how to use statistical packages in introductory statistics courses.

Chapter 6

Part II - Trial I - Qualitative Phase

6.1 Rationale and Aims

Limitations with Trial I that were identified included unblinded participants, IT issues, limited computer laboratory resources, noncompliant students, and poor student engagement with the self-assessment exercises. These limitations highlighted the challenges of embedding randomised experiments in real education settings. In Trial I possible problems with the manipulation of training approaches and the effect of time pressure were also reported. The validity of measures of training transfer were also called into question. The inclusion of the qualitative phase to Trial I allowed further critical evaluation of these quantitative results as well as the opportunity to explore the student experience of technology training in a more general sense.

This chapter reports the results of the qualitative phase of Trial I. The primary aims were as follows:

1. to document an in-depth exploration of the overall student experience of statistical package training
2. to further evaluate the possible impact of training approaches used in computer laboratory sessions on students' experiences and skill development

6.2 Method

In the Trial I follow-up questionnaire given in the final lecture of semester one, students were invited to participate in semi-structured in-depth interviews. The inclusion of interviews was approved by the RMIT College Human Ethics Advisory Network (CHEAN) on 8^{th} April, 2011. All interviewees received a free movie ticket as a token of appreciation for their time. Interviewees were provided with a plain language statement summarising the qualitative phase prior to the interview commencing (See Appendix A.9). Verbal consent to record the interview was obtained from each interviewee.

6.2.1 Interviews and Data Analysis

Fifteen interviewees (GT $N = 9$ and EMT $N = 6$) volunteered to participate in semi-structured interviews following training. Table 6.1 shows a break down of the characteristics of the sample. Interviews were conducted face-to-face and over the telephone during the exam period. Interview questions covered a range of topics including attitudes towards training, confidence in operating *SPSS*, emotions experienced during training, training difficulties, assistance required, problem solving and suggested improvements (see Appendix A.8 for the complete interview schedule). All interviews were audio-recorded and transcribed verbatim. Qualitative data was analysed using a six-step inductive thematic analysis method described by Braun and Clarke (2006). The six steps included: 1) data familiarisation, 2) initial code generation, 3) theme searching, 4) theme revision, 5) theme definition and naming, and 6) reporting. Once the overall analysis had been completed, coded extracts for each main theme were compared across the different training approaches to consider possible moderating effects on the themes. Any differences in the theme trends between the training approaches were recorded.

6.3 Qualitative Results and Discussion

Eight major themes emerged from summarising the qualitative data via the thematic analysis. A thematic map is provided in Figure 6.1. Each theme will now be defined and discussed along with any major trend differences observed between training approaches.

Table 6.1: Interviewee Characteristics

ID	Age	Condition	Campus	Gender
GT1	18	GT	Campus B	Female
GT2	54	GT	Campus A	Female
GT3	21	GT	Campus B	Female
GT4	23	GT	Campus A	Male
GT5	18	GT	Campus B	Male
GT6	18	GT	Campus A	Male
GT7	24	GT	Campus A	Female
GT8	46	GT	Campus B	Female
GT9	18	GT	Campus B	Female
EMT10	21	EMT	Campus A	Female
EMT11	18	EMT	Campus A	Female
EMT12	18	EMT	Campus B	Female
EMT13	33	EMT	Campus A	Female
EMT14	28	EMT	Campus A	Female
EMT15	18	EMT	Campus B	Male

Where appropriate, these themes will be related back to the quantitative outcomes of Trial I. Quotes are labelled using identification codes (e.g. EMT − 14 refers to interviewee 14 from the EMT approach).

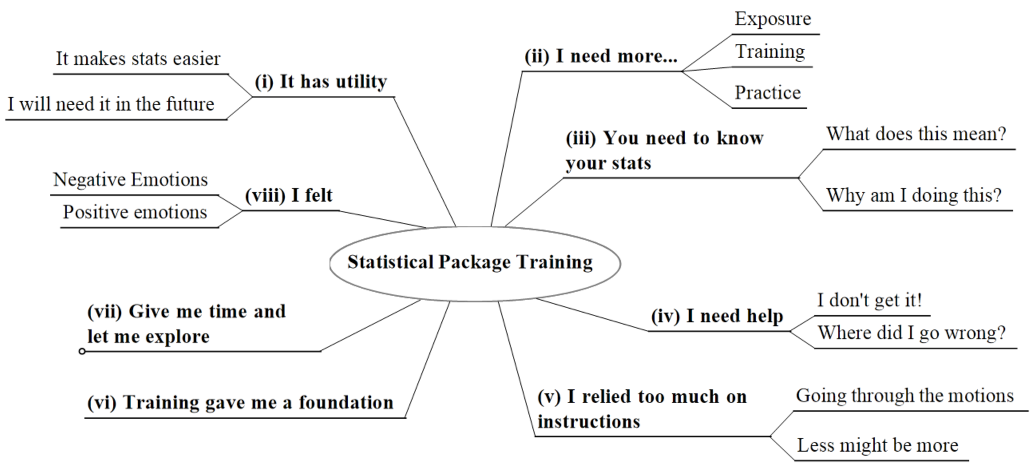


Figure 6.1: A thematic map of the qualitative analysis of Trial I

(i) It has utility

This theme referred to the students' perceptions of the utility of training. Almost unanimously all interviewees, regardless of training approach, agreed that learning to use *SPSS* was important for their future academic careers and it would make doing statistical analysis easier:

> Because it [*SPSS*] makes it easier in the future if we have lab reports and stuff like that without having to manually input the data and make up our own graphs; the system will do it more accurately than I guess we would. [GT - 2]

This was good news as instructors typically spend a lot of time justifying statistical package utility, not to mention the need to learn statistical concepts itself. The participants appeared to have recognised the importance of learning *SPSS*. However, a closer inspection of this theme revealed an interesting trend.

Students continually referred to the "future" applicability of this skill. They did not appear to see its current relevance. One student questioned whether developing these skills could wait until later in their degree course.

> [E]ventually we'll start to need to know how to use all this researchy [sic] stuff. And when we do [. . .] experiments we need this, so I think it'll be good for the future, but I think it would better if we had this later on instead of now, I think. [EMT - 15]

This was an important insight into the motivation of students and the likely level of engagement that they will exhibit during training. Perhaps more effort is needed to enhance the immediate perceived relevance of statistical package skills. For example, if these skills were necessary to complete data analysis projects early in the course. Regardless of students' ability to see the long-term benefit of a skill, there is little doubt that student engagement could be enhanced earlier by a perception of immediate utility.

(ii) I need more. . .

Almost all participants expressed the need for further exposure, training and practice using *SPSS*:

> I think if we had more labs, that would be very helpful, because that would give us more exposure to the actual product. Because, outside of the labs, we don't really use *SPSS* or we can't really see it at home, but if we had more exposure to it, I think we would learn it a lot better [GT - 5].

This theme was also consistent between training approaches, with neither approach being more or less likely to express this need. Students suggested that more training lab sessions would have been beneficial, and some participants proposed embedding *SPSS* demonstrations into lectures, laboratory sessions or tutorials to increase exposure and familiarity of the package:

> I suppose maybe with the tutor showing us how to do it first, rather than just using the instructions and getting their help if needed.[GT - 9]

Doing so might also help students to see a stronger link between statistical concepts covered in lectures and the exercises covered during training. Identification of this theme reinforced it as a significant limitation raised in the discussion of the quantitative phase of Trial I.

When discussing the need for further practice, a few participants raised the inaccessibility of *SPSS* from home as being a major limitation, or as one student explained:

> I don't have access to it [*SPSS*] anywhere else so that's kind of the only practice I got with it and I don't think that's enough. [EMT - 10]

The price of a personal license for many industry-based packages is a barrier to students. There is no doubt in our minds that the convenience of home access to a statistical package would present valuable practice opportunities outside of regular training.

The eventual goal of statistical package training or any technology training in statistics education should be to provide students with the necessary skills and dispositions required to master the technology and make it a part of the students' regular repertoire of technology skills. For a statistically literate person, the ability to operate a statistical package should be as common as the ability to operate a word processor. Access can present a major barrier to this eventual goal. Perhaps instructors have grossly underestimated the importance of access and its impact on students' skill development.

(iii) You need to know your stats

This theme consistently appeared when discussing students' experiences during training. Participants from both EMT and GT acknowledged a strong dependency between understanding statistical concepts and understanding the *SPSS* training:

> I think maybe because my confidence for maths and statistics anyway is pretty low, I was just like 'I don't understand this, so how am I going to understand the program?' [EMT - 10]

Another participant explained how a strong understanding of the content made training easier:

> I think I have a [sic] better confidence than my friends, but I think that's mainly because I have a grasp on the actual theory behind it, rather than just the steps. [GT - 3]

Participants also talked about their difficulty linking the training with their lecture content. They sometimes failed to understand not only what they were doing, but more importantly putting it all together to understand why:

> I especially realised when I completed the quizzes, the self-review quizzes, how much I didn't actually understand it. I sort of just basically learned how to follow the steps but I didn't have a good foundation of understanding as to why I was doing it. So when I was given the task of doing it without the steps I realised how much I didn't get it. [EMT - 12]

This finding was not surprising having been observed in the quantitative phase, but it does have a very important implication. It implies a strong dependency between knowledge of statistics and the ability to be able to operate a statistical package. Chance, Ben-Zvi, et al. (2007) claims that introducing technology too early can overwhelm students who are still developing their understanding of statistical concepts. Students can become lost in the technology and lose sight of the bigger statistical picture:

> I sort of really didn't get the lectures and then I'd go to the lab and I felt like it was completely different and I didn't really get that. [EMT - 12]

This dependency needs further investigation and effective approaches for moderating its proposed effect should be explored. Delaying the introduction of technology is probably not a feasible option given the importance of fostering these skills. Moore's (1997) conjecture that effective learning emerges from the right balance and alignment of content, pedagogy and technology reiterates this continual challenge.

(iv) I need help

The majority of participants reported seeking assistance during the lab training sessions mostly from the tutor and sometimes from their peers. The degree of reliance on assistance varied between students:

> I asked for help straight away, which is probably not a good thing because I could have worked it out for myself but I'd just sort of look at it and think "OK that doesn't marry up" and then I'd look again quickly and freak out a little bit and then put up my hand and the teacher would come over [GT - 2].

> I just asked, I didn't even bother trying to figure it out myself because the one experience I did have of trying to fix it myself I made it worse. So I learnt and put my hand up and [the tutor] would be like "I'll be with you in a minute" [GT - 1].

> I asked the tutor sometimes, when I was really stuck. If I was just kind of stuck I'd still try to do it myself. But only if it was a really difficult situation, I'd ask the tutor [EMT - 15].

The first and second quote reflects a disposition in trainees that should not be reinforced. These students were clearly not engaged with training and perhaps the easy access to assistance from the tutor enabled or exacerbated this poor engagement. The third quote is more aligned with a desirable work-ready disposition. This student persisted in the face of difficulties, but knew when it was time to seek help. The most common reason stated for seeking assistance was help to identify where participants had gone wrong and when participants didn't understand the exercise or didn't understand the output:

> And a lot of the time, when you have to look at tables and stuff, I wouldn't know which number I was meant to be looking at. So I guess it was more to do with the theory. [GT - 3]

An apparent difference between EMT and GT emerged in this theme. The GT respondents were more likely to seek immediate assistance for the problems they faced. Respondents from the EMT condition were more inclined to identify themselves as "problem solvers" who would only seek help after first giving it a try.

> I asked a couple of questions, of tutors just when I had no idea what I was doing and had tried about a million times [EMT - 10].

> No, look I'm a problem solver, so I really wanted to try and do it myself, so I only asked for help from the tutor as a last resort [EMT - 11].

> Not really. The first few I sat by myself and did it all by myself. I asked the tutor one or two questions but mainly just worked it out myself [EMT 13].

This demarcation between approaches in reports of seeking assistance suggested the volunteer interviewees were adopting the behaviours largely consistent with their allocated training approach. This difference was not consistent with manipulation checks reported in the post-training questionnaire for the quantitative phase (see Table 5.8) and is therefore mostly likely an artefact of the use of volunteers. The need for help is also likely to vary between students from different academic backgrounds who may be more or less familiar with technology than psychology students.

(v) I relied too much on instructions

When asked about how they managed the self-assessment tasks at the end of the semester, many interviewees from GT talked about how difficult some of the tasks were after their instructions were taken away:

> Yes. When we didn't have the exact instructions I was a bit lost, so that just said to me I relied too much on the instructions before. [GT - 2]

> I've been going by the instructions, and when there was no instruction I found it really difficult, like I realised I hadn't really remembered how to do it on my own, and sometimes I could figure it out and obviously, that was fine, but then there were times when I just had no idea what I was doing. [GT - 8]

> I don't think I actually learned how to use it. I think I learned how to follow steps but if I were to sit down in front of the package now I could probably do one thing that we did in the first lab and then continued it throughout, comparing the means or something, but I could not do anything else because basically, what I found, all you were doing was looking at the steps and then just following it one step at a time, not as a whole. [GT - 3]

A few students explained that they had developed an overreliance on the instructions which resulted in them just going through the motions during training:

> I was just learning how to follow the steps and just try to get a sufficient amount of right answers to pass each time. [EMT - 12]

> With the training there was a little bit of step by step and, personally, I didn't THINK a lot about it. [GT - 6]

As a result, one participant from the GT condition suggested using less instructions as a way for improving training [GT - 2]. Another GT student proposed to use instructions initially and then stop them later in the semester:

> I think the best way was what we did this semester - give instructions, follow certain instructions to do certain things and, after a while, just stop the instructions and see how the students go [GT - 4].

However, it's not hard to imagine how unpopular this would be. Building on the student's suggestion, a better approach might be slowly easing the instructions off as the semester progresses.

Statistical package training, given time, should allow students to transfer their skills outside of training without the need for comprehensive instructions used in training. Interviewees from the EMT condition didn't exhibit a strong overreliance on instructions. When asked about the difficulty of training one respondent stated:

> Not really, they sort of, I think they [the computer laboratories] were actually quite good. They were a good level, they weren't sort of like giving you exactly step by step, there was enough room to actually have a play around yourself I think. Yeah, I think they were actually at quite a good level, sort of that middle point where it wasn't too hard but it wasn't just take the steps [EMT - 10].

However, one interviewee from EMT felt that even the removal of minimal instruction made the self-assessment tasks more challenging:

> [Training] was positive when the instructions were there, like they were telling me what to do but, without it I don't think I can cope unless I have more training and get used to it more [EMT - 14].

While it was clear that the students needed more practice in this study, too much instruction may be inducing dependency and disengagement with technological training. These qualitative results suggest that the EMT interviewees were more engaged in training and less likely to develop a strong sense of reliance on instructions, but regardless, the results of the quantitative phase suggest that it had no impact on the average training transfer performance of students in either approach.

(vi) Training gave me a foundation

Participants discussing their level of preparation for the self-assessment task and use of *SPSS* outside of training had mixed perceptions about their ability to transfer their skills. However, a general perception that training had provided them with a basic foundation emerged. When asked if they felt they were ready to use *SPSS* beyond training, one student commented:

> A little bit, at least I'm a little bit more familiar, but I wouldn't say that I would be confident in going into an assignment where I would be expected to use *SPSS* for a lab report. I think I'd struggle a little bit. [EMT - 12]

This perception was mostly explained by the brevity of the training delivered in this course. Given more experience, it is likely that this self-efficacy towards operating statistical packages outside training would improve. At the very least, the training did manage to familiarise the students with the basic operations of the package and provided a foundation for future development. When comparing the responses between approaches, interviewees from the GT approach were more likely to initially present as confident users, but then be quick to point out that they were confident only in the basics:

> After the training I feel that I'm very confident in it, in the subjects we actually did I think I'm pretty good, but if I was asked to do something off that, maybe I would have a little bit of trouble - I'd have to find my way around but the basics I think I've got down pat [GT - 5].

Respondents from the EMT condition were not as certain:

> Well, I'm more confident than I was at the beginning, but I'm not very confident. [EMT - 16]

This was an interesting outcome as the quantitative phase found no difference on mean training transfer between conditions. It appeared that the volunteer interviewees from EMT were underestimating their ability. Once again, this difference observed in this theme conflicted with self-reported post-training questionnaire items measuring statistical package self-efficacy in the quantitative phase. No statistically significant difference in mean ratings were found between conditions (see Table 5.8).

(vii) Give me time and let me explore

Interviewees were asked how they went about solving problems that arose during training and a hypothetical question about how they would approach a novel statistical analysis not covered in training. Many participants reflected on an innate propensity to explore *SPSS* to solve their future problems:

> I guess I just tried to do it a different way, just kind of cover every possible option of doing something. [EMT - 12]

In terms of hypothetically figuring out how to do an analysis not covered during training, many participants were quietly confident they could figure it out for themselves if given enough time to explore:

> Given a reasonable amount of time, yes. If I had time to sort of play with it and make mistakes, cause that's how I've taught myself with everything else on a computer is I've had time to sit there and put things in and try different things, yeah I think I could. It would take me time, but I would get there [GT - 1].

Regardless of the underlying nature of the training approaches, many interviewees from GT reported using exploratory behaviours to solve problems that they faced:

> I started playing around with certain things - example, if you gave me a certain question and I had no idea and I just started playing around and I actually got it, that actually helped me to learn how to get that [GT - 4].

> Yeah, so you'd just try and apply a bit of logic; try and make an educated guess of what it would be and just go from there [GT - 6].

> It's just me - to learn on computers I just click every button to see what it does; that's how I learn, whereas when I'm doing a lab I'm not sure if I should do that because I may stuff up the test, so to me, if I'm doing it by myself . . . I don't know - I get distracted [GT - 7]

Students from both approaches reported using exploratory behaviour. Even in the presence of comprehensive instructions, many students appeared to be at ease with playing around with technology and exploring the technology on their own terms. In the previous "I need help" theme many students from GT reported a tendency to seek immediate assistance, while in this theme many GT interviews also reported using exploratory behaviour. It would be valuable to know what factors explain why some students choose to seek help while others appear happy to figure things out for themselves.

Given the open-ended nature of this question, it was surprising to find most students would first explore to see if they could figure out how to conduct a new analysis procedure. In retrospect, it was possible that this trend reflected students' inexperience with statistics and an attempt to find any solution that seems correct (Chance, Ben-Zvi, et al., 2007). However, the impression from the data was that the students were expressing a general approach to the use of technology. While some instructors might be concerned about the thought of their students stumbling around a little trying to find the correct procedure, this type of behaviour might be more conducive to training transfer as it moves away from the unsustainable use of step-by-step instructions used in conventional training (Dormann & Frese, 1994). It's clear that further research is needed to settle the issue of whether less instruction is more.

(viii) I felt...

Participants reported experiencing a wide range of positive and negative emotions during training. The similarity in experiences between the approaches was strong. Regardless of conditions, training was a very emotionally rich environment. Negative emotions were mostly related to anxiety or a fear of failure:

> Emotions? A bit of nervousness. A bit of an attitude of "what happens when I fail?" Fear that I won't actually understand the instructions and I'll have to constantly put up my hand for help. Just fear of really not understanding the questions, basically, and the instructions. [GT - 4]

As training progressed, anxiety shifted towards feelings of frustration, stress and annoyance. In contrast, many participants expressed positive attitudes and emotions towards training:

> I kind of did enjoy it actually. It was fun trying to solve the damn things, even though it was difficult, but still, I liked it I guess. [EMT - 15]

Some participants explained that their emotions helped them engage. When asked if their frustration was distracting, one participant explained:

> No I wouldn't say distracting, I think it just gives me motivation to knuckle down and do it again. [GT - 3]

Another interviewee answered when reflecting on the effect of their anxiety:

> Yes, probably beneficial because it was more motivating and it sort of encouraged me to take my time and read it slowly and work out what I'm doing without just rushing ahead which caused the anxiety to begin with [GT - 2].

Other participants explained how they used emotional control skills to deal with negative emotions as they arose during training:

> It was just . . . "I hate it [*SPSS*]" and then I'd go "this is ridiculous, I don't want to do this" and then sort of I'd have to talk myself into "well, you have to do it. Slow down, let's go back, let's have a look at why you've picked the wrong one" or, you know. And sometimes it was just that I had misread a step or skipped over a step, so I was thinking I was doing everything but I'd missed something in the instructions [EMT - 11].

There didn't appear to be any perceivable association between EMT and the qualitative data extracts related to emotional control. This finding was supportive of results obtained from the quantitative phase. No statistically significant difference was found in mean ratings of self-reported emotional control during training between EMT and GT approaches (see Table 5.8).

The validity of any training that doesn't make students feel a little bit uncomfortable (i.e. training that is too easy) should be questioned. Finding the right level of difficulty should be the instructor's goal. Adaptive emotional control strategies should be encouraged and developed in all students. Students need to become comfortable dealing with problems and being patient with themselves when they make mistakes. Future research is needed to determine if EMT can help develop these during statistical package training.

6.4 Conclusion

The results of this qualitative phase of Trial I have been insightful. Being the first qualitative study looking at the development of technology skills in statistics education, many interesting findings have emerged. The first point relates to the merit of employing mixed-method research in statistics education. Had only one method of research been employed an opportunity to gain further valuable insight into Trial I would have been missed. Sometimes the evidence obtained using both qualitative and quantitative methods did not converge, probably due to the small volunteer sample, but in most cases a large degree of agreement was observed. This was mostly evident in the overall high degree of similarity in students' experiences of statistical package training between approaches. This overall shared experience provided further support for the quantitative phase's major finding of no difference between training approaches.

When comparing the trends in themes between approaches, the overall experience reported by volunteer interviewees was largely the same for perceived utility, the need for more training, the importance of statistical knowledge, exploratory behaviour and emotional range. Differences between approaches for the themes of instructional reliance, the need for assistance and confidence in foundational skills emerged. While interviewees from EMT were more inclined to attempt to work through their problems and felt less reliance on instructions, they did show a trend in underestimating their ability compared to GT interviewees. This might suggest that minimising guidance may result in lower student self-efficacy even though in reality these students were no worse off in terms of skill transfer. However, as the interviewees were only a handful of volunteers, these findings must be interpreted with great caution.

The overall themes that emerged from the in-depth qualitative analysis provided thought-provoking insight into how statistical package training is perceived by students. These overall themes are summarised as follows. 1) Students understand the future utility of statistical package training but an effort should be made to make this utility felt sooner. 2) Instructors should not underestimate the time required for students to develop a sense of proficiency with a statistical package. Providing access to the statistical package and increasing training opportunities is important. 3) A single course is unlikely to develop a sense of proficiency, but instead will lay a foundation to be built upon. 4) Students need to understand statistical concepts to get the most from statistical package training. 5) Reducing instructions and access to immediate assistance might help students develop better persistence and problem solving skills which may lead to better training outcomes. 6) Students should be allowed time to explore, problem solve and learn to recognise when they are out of their depth. 7) Training is an emotionally rich environment. Effective training design and successful trainees will control and harness these emotions to maximise engagement.

The results of the quantitative phase reinforced the following recommendation for the design of Trial II:

1. Provide students with more training and practice opportunities

2. Build students' statistical knowledge alongside training to enhance its effect.

Chapter 7

Part I - Trial II

7.1 Aims of Trial II

The main aim of Trial II was to re-evaluate the effect of GT and EMT approaches on statistical package training transfer by addressing key limitations identified in the quantitative and qualitative phases of Trial I. Specifically, this study aimed to improve the validity of the implementation of the EMT approach, increased overall training time across the semester, blinded participants to the nature of the study, developed students' statistical literacy by embedding formative assessment questions throughout training sessions and developed an improved measure of adaptive transfer. This study opted for a quasi-experimental design due to practical and ethical issues imposed by implementing randomized studies in educational settings. While randomised studies are considered the gold standard for evaluating educational interventions, research suggests that quasi-experimental designs can provide reliable estimates of causal effects provided adjustment for known covariates has taken place (Shadish et al., 2008). Important and known covariates were measured and controlled for to improve the comparisons between training approaches. This study chose to focus only on adaptive transfer outcomes as these were considered the most pertinent outcomes of statistical package training. It was hypothesised that EMT would lead to significantly better statistical package adaptive transfer skills. To explore the possible implications of using EMT over GT, measures of student self-efficacy, training satisfaction, training anxiety, and training difficulty were also compared.

7.2 Method

7.2.1 Participants

This study received ethics approval from the RMIT College Human Ethics Advisory Network on the 8^{th} April, 2011 (Project No. BSEHAPP 48-10). Participants were recruited from an introductory statistics course for psychology students which ran concurrently across two campuses, A and B. The course covered exploratory data analysis, statistical inference for categorical variables and correlation. While not included in duration of Trial II, the course continues in second semester and covers inference of means and regression. Campus A had 41 students enrolled of which 35 (85%) consented to participate in the study. Campus B had 127 students enrolled of which 93 (73%) consented to participate. By the end of the study, 34 (97%) and 81 (87%) participants completed the requirements of the study from Campus A and B respectively ($N = 115$). Campus A had a mean age of 22.3 years ($SD = 7$) with 24 (74%) females. Campus B had a slightly lower mean age of 20.2 years ($SD = 3.2$) with 68% (55) being female. During the first lecture students were invited to participate in the Trial by providing them with a plain language statement and consent form (see Appendix A.10 for the PLS and Appendix A.11 for the consent form used). Those who chose to participate filled out a short pre-training questionnaire which asked them if they had been previously trained to use the statistical package *SPSS*. There were two (6%) participants from Campus A and nine (11%) participants from Campus B who reported being previously trained.

Campus A was arbitrarily designated the EMT approach and Campus B the GT approach. This non-random allocation meant that campus was a confounding variable. Major differences between campuses were present both between students and course delivery. Campus A (EMT) and Campus B (GT) tertiary program entrance requirement scores where 68 and 77 respectively for the year of the study. This difference reflects a greater preference for Campus B, meaning that it tends to attract students who performed better in their final year of secondary education. While the course was delivered by the same instructor, course contact hours were during the afternoons for Campus A and mornings for Campus B. This is important as, anecdotally, students prefer morning statistics sessions at the trial's institution. As will be discovered later in the chapter, the Campuses also differed on personal access to *SPSS* and the number of lab sessions completed outside of training. There were 13/32 (40.6%) students surveyed from Campus A that reported having personal access to *SPSS* versus only 9/81 for Campus B (11.1%, see Table 7.2). While it is difficult to speculate exactly why this difference existed, perhaps the afternoon scheduling of computer laboratory sessions for Campus A meant that many more students than Campus B sought to do these sessions at a more convenient time. Personal access to *SPSS* would be required to do so. This speculation is partially supported by the fact that students surveyed from Campus A reported completing an average of 4.41/10 (SD = 2.80) laboratory session outside of training when compared to an average of 3.09/10 (SD = 3.4) for Campus B (see Table 7.2).

7.2.2 Measures

Measures used in Trial II are categorised into covariates, manipulation checks, training transfer and other training outcomes. A pre-training questionnaire given in the first week of the semester along with the PLS and consent forms obtained participants' demographic information and measured the covariate of perceived performance utility. A post-training questionnaire given in the final week of the semester measured manipulation checks and other training outcomes.

Covariates

Due to the quasi-experimental design of this study, it was important to control for pre-existing differences between the training approaches which may explain variability in training transfer measures. Statistically controlling for these variables would enable a better estimation of the association between training approaches and training transfer. Based on Kanfer and Ackerman's model, a student's cognitive ability will explain a large degree of the variability in training transfer outcome measures. Cognitive ability is a broad general construct that requires specialized testing (e.g. IQ testing) which was beyond the scope of this study. Therefore, a substitute variable for controlling for this effect was needed. A student's knowledge of statistics, as measured by average test and exam performance across the semester was chosen for this purpose. This was calculated by averaging the student's grade percentage across test 1, test 2 and the final exam. If a student missed any assessment, they received the average of the assessment they had completed. While statistics exams scores have been found to be very weakly correlated with intelligence (e.g. Furnham & Chamorro-Premuzic, 2004), they do provide a more relevant way of controlling for the effect of student ability on training transfer. As discovered in Trial I, statistical knowledge was related to statistical package training transfer, suggesting that a student's knowledge of statistics will impact their development of statistical package skills. Therefore, to disentangle the effect of training approaches on adaptive training transfer, statistical knowledge was controlled for between training approaches.

Students' motivation to learn statistical packages was also taken into account as suggested by Kanfer and Ackerman's model. While there are many models of motivation which could be considered, this study took a direct approach similar to Keith, Richter, and Naumann (2010). This involved measuring students' self-reported perceived performance utility. Statistical package performance utility was defined as the extent to which a student viewed *SPSS* as being useful technology for doing statistics. This trial adapted items from the Questionnaire for the Content-Differentiated Assessment of Attitudes toward the Computer (Richter, Naumann, & Groeben, 2000, see Appendix A.12). An example of an item is "*SPSS* will be a useful tool for doing my statistical analysis". The seven items that made up this scale were rated on a 7-point likert-type scale ranging from "strongly disagree" (1) to "strongly agree" (7). Scores were averaged to get an overall performance utility score. High scores indicate a high perceived level of perceived performance utility. The original items from Richter et al. (2000) had evidence of good psychometric properties. However, these metrics were re-analysed following adaptation for the purpose of Trial II. A Principal Components Analysis (PCA) extracted a unidimensional construct using the eigenvalue greater than 1 approach which explained 62.6% of the variability in responses to performance utility items. The scale had a high internal consistency rating of Cronbach's $\alpha = 0.88$

Students' progress through the training was recorded by counting the number of training sessions each student had completed up to one week prior to assessment of adaptive training transfer. As there were a total of ten training sessions, scores on this covariate could range from 0 to 10. The post-training questionnaire also asked participants to self-report the number of training sessions that they completed outside of their designated training session times. This variable was included to take into account possible differences between the campuses that related to how the students completed the training. This was important to include as training was available online outside of scheduled training times. As this measure was self-reported on the post-training questionnaire, 32/93 (34.8%) participants in GT and 3/32 (9.4%) participants in the EMT approach were missing data. In the post-training questionnaire, participants were also asked if they had personal access to the statistical package. This was important to take into account as students with personal access may systematically differ from students who could only access the package on campus. Gender and age were also recorded.

Manipulation Checks

In line with Trial I, it was important to evaluate the validity of the imposed training approaches. Trial I reported limitations with the manipulation of training approaches as a possible explanation for the null findings. Therefore, it was important to include the same manipulation checks as a measure of internal validity. The same self-reported measures of metacognitive activity, emotional control, exploratory behaviour, the use of instructions and error orientation during training used in Trial I were included in the post-training questionnaire for Trial II (see Appendix A.7). All measures were rated on a seven-point likert-type scale ranging from (1) “strongly disagree” to (7) “strongly agree”. Scale scores were calculated by averaging participants’ responses across items. Items that needed to be reverse-coded were reversed prior to averaging.

In summary, metacognition was measured using 12 items adapted from Ford et al. (1998). An example of an item is “I tried to monitor closely the statistical procedures in *SPSS* where I needed the most practice”(Cronbach’s α = .89). The degree to which students exercised emotional control during training was measured using eight items originally adapted from Keith and Frese (2005) for Trial I. An example of an item is “When difficulties arose during computer labs I was able to focus all my attention” (Cronbach’s α = .82). Students’ attitudes towards errors made during training were measured using the *Error Strain* and *Learning from Errors* subscales of the *Error Orientation Questionnaire* (Rybowiak et al., 1999). These items were adapted in Trial I to refer to errors made during statistical package training. The Error Strain subscale measured the degree to which students felt negative emotions when making errors (e.g. "I was afraid of making errors when learning to use *SPSS*") using five items and the Learning from Errors subscale measured the degree to which participants viewed errors as being a valuable learning experience (e.g. "From my errors, I have learned a lot about how to work with *SPSS*"). The sample's internal consistency was Cronbach's α = .79 and .82 for Error Strain and Learning from Errors subscales respectively.

The degree to which students participated in exploratory or guided behaviour during training was measured using six self-reported items borrowed and adapted from Bell and Kozlowski (2008). Three of these items related to exploratory behaviour consistent with EMT, e.g. "I tried to discover how to operate *SPSS* without any instruction". The other three items measured students' behaviour consistent with GT, e.g. using instructions, modelling others and seeking assistance from tutors. An example of an item is "When I was unsure about how to complete a task in *SPSS*, I would immediately ask the tutor/or a friend for help". To aid the comparison with Trial I, the mean rating of individual items were considered when checking the validity of the training approaches.

Adaptive Training Transfer

An *SPSS* certification task was used to measure adaptive transfer (see Appendix A.13). Analogical transfer was not considered because adaptive transfer was the goal of training and the most important training outcome. The certification task was scheduled for the final week of the semester and participation in the task contributed to a 5% course grade. The certification task was included to increase students' engagement in training during the semester. A limitation of Trial I was the possibility of poor student engagement as an issue for measuring training transfer. The certification task was designed to increase student engagement by making students aware of the activity early in the semester, by making the task sound official, and attributing a higher grade to its completion than regular training. The task lasted one hour and was completed under exam conditions (no talking, no assistance). However, students were allowed to bring a copy of the course's *SPSS* quick guide which is described below. The certification task presented students with six exercises. For each exercise, *SPSS* output was presented on a printed handout. Using a data file provided to them, the students had to replicate the output using *SPSS* for each exercise as closely as possible. The closer the student replicated the output, the higher their training transfer. The first two tasks were designed to be very simple and were not included in adaptive transfer scores. The remaining four tasks were designed to measure adaptive transfer and were scored out of 32. The exercises were adaptive because students had to replicate output that required them to adapt their training knowledge. This involved being able to link multiple procedures together that were treated separately during training (e.g. segregate data file, filter out specific cases and create a plot) as well as manipulate and edit output (e.g. adding labels, reference lines and markers) in ways in which training did not cover.

Students were instructed to export their single closest replication of each exercise to a word processing document and upload it to an online submission site before leaving the certification session. There were three versions of the certification task worksheets (A, B, C, see Appendix A.13). Each version was slightly different to prevent students' collaborating with their neighbours. A grading code was developed to identify key elements of each exercise which indicated the student had successfully adapted their skills (see Appendix A.14). These key elements were scored higher than other elements of the output that did not require students to adapt their skills. The lead researcher completed all grading. All student attempts were labelled using student numbers. Attempts from each training approach/campus were mixed together. This was done to blind the lead researcher as to which training approach/campus each attempt belonged to. For student feedback purposes, participants were given a level, 0, 1, 2 or 3, which reflected their performance on the certification task. Students who scored $0 - 1$ were given the opportunity to complete further training between semesters to brush up on their *SPSS* skills before second semester.

Other Training Outcomes

Besides training transfer, it was important to consider other training outcomes that may impact on students and instructors. As in Trial I, this trial considered the association between training approaches and students' perceptions of statistical package self-efficacy, training anxiety, overall difficulty and satisfaction. Students' perceptions of the difficulty, anxiety experienced and level of training preparedness for the certification task were also evaluated. When giving their responses to the end of semester post-training questionnaire participants were asked to rate the overall difficulty and satisfaction of training on a scale ranging from (1) "very easy/not at all satisfied" to (7) "very difficult/very satisfied" respectively. On the same questionnaire, participants were also asked to rate their level of statistical package self-efficacy. Statistical package self-efficacy was defined as a participant's confidence in their ability to operate a statistical package after training. Three items from Finney and Schraw's (2003) *Current Statistics Self-efficacy* (CSSE) scale were adapted for this purpose. Participants were required to rate their level of confidence in their current ability to use *SPSS* for generating descriptive statistics, graphical displays and statistical inference. An example of an item is "To use the statistical package to conduct statistical inference (e.g. generate p-values)". A similar seven-point likert scale ranging from (1) "no confidence at all" to (7) "complete confidence" was used. Scores for the three items were averaged to form a single self-efficacy score (Cronbach's $\alpha = .78$).

Participants rated their level of anxiety that they experienced during training using the same four items used in Trial I from the Tension-pressure dimension scale of the *Intrinsic Motivation Inventory* created by Deci and Ryan and reported in (McAuley et al., 1989). A sample item adapted in Trial I is "I felt tense when training to use *SPSS*". Items were rated on a seven-point likert-type scale ranging from (1) strongly disagree to (7) strongly agree (7). Item ratings were averaged to obtain a scale score where higher scores equated to higher training anxiety (Cronbach's $\alpha = .73$).

Before leaving the certification task session, participants were asked to rate the perceived difficulty of the certification task along with the level of anxiety they experienced and the degree to which they felt training had prepared them for the certification exercises. All questions were rated on a similar seven-point scale used in the end of semester post-training questionnaire.

7.2.3 Training

Participants completed weekly one-hour statistical package training sessions in designated computer laboratories under the supervision of tutors. These sessions were designed to introduce students to the use of the statistical package *SPSS* v. 20 as well as reinforce statistical concepts covered in lectures. The training was delivered using an online proprietary web-based assessment system called *WebLearn*. Participants completed five training modules made up of a training and practice session (10 weekly sessions in total). Training sessions introduced new *SPSS* procedures and practice sessions were used to consolidate the training material. Students completed the certification task in the final week of the semester. Completion of each laboratory session and the certification task contributed to a 20% (10 laboratory sessions = 15%, certification task = 5%) participation grade. The module topics included the following: Introduction to *SPSS* (overview, entering data, editing variables, saving files, descriptive statistics, basic plots, editing plots, exporting output), The Basics of *SPSS* (revision from lab 1, boxplots, histograms, segregating and filtering data), Frequencies in *SPSS* (revision from lab 1 and 2, frequency tables, bar charts, recoding variables, and computing new variables), Crosstabs in *SPSS* (revision from lab 1, 2, and 3, cross-tabulations, Chi-square tests of association, clustered bar charts), and Correlation in *SPSS* (revision from lab 1, 2, 3 and 4, scatter plots, matrix scatter plots, and correlations). To help reinforce statistical concepts covered in the course, formative multiple-choice questions were embedded throughout laboratory sessions for both training approaches. These questions pre-empted statistical concepts to be covered in training to help facilitate the correct interpretation of *SPSS* output. For example, before students created cross-tabulations of two categorical variables, participants were presented with questions that required them to practice interpreting row and column percentages.This was done to satisfy recommendations from the quantitative and qualitative phases of Trial I.

All training sessions were delivered online using *WebLearn*. The training sessions presented students with exercises that required them to learn to operate *SPSS*. Students either entered data or downloaded data files to use during the training and practice sessions. To confirm that the student had successfully operated the package, each exercise contained a question about the *SPSS* output generated. Students would enter their answer to receive immediate feedback on whether they had successfully completed the exercise. Each exercise was presented one-at-a-time and could be attempted multiple times. Students were advised to move through the exercises sequentially. To get their participation grades, students were required to attain 75% or above. Feedback for incorrect answers was provided in a form consistent with the training approach (described below). Both training approaches were provided with a copy of an *SPSS* quick guide reference. This guide listed and briefly described the features and procedures of *SPSS* that were covered throughout the entire semester of training. The guide was provided in response to previous course feedback. Electronic copies were linked to all training sessions.

EMT

Students in the EMT approach (Campus A) were presented with instructions at the beginning of training that established the conditions of the EMT approach. The instructions promoted active exploration and a positive attitude towards making errors. Students were told to expect to make errors and that these errors were a natural part of the learning process. Students were encouraged to try to rectify any errors or solve problems they had before seeking assistance from the tutors. At the beginning of each EMT session, students were provided with notes providing a minimal instructional overview of the features and procedures of *SPSS* that they would be covering. These notes contained screenshots showing students how to access these procedures, but the screenshots were not linked with exercises, nor were there any step-by-step instructions provided. This aimed to improve the conditions of minimal instruction and enhance exploratory behaviour. The exercises no longer directly linked students to the procedures required to complete training tasks as in Trial I. Students had to make educated guesses using the notes and screenshots given at the beginning of the training. Students needed to explore these features and adapt them to complete their training exercises. Tutors were not permitted to guide students, but instead to encourage students to find solutions themselves. Throughout training, error-framing heuristics were presented to students above the exercises they were completing, e.g. “Errors are a natural part of learning, they point out what you can still learn.” These heuristics were provided to remind students of the positive function of errors. If a student got an exercise wrong, feedback was provided in the form of a positive error-framing heuristic as well as a hint designed to help them rectify their error, e.g. "Try playing around with the order of the variables entered into your plot".

GT

Students in the GT approach (Campus B) were instructed to carefully follow the step-by-step instructions given to them and to avoid making errors where possible. If students made a mistake, they were told to read back through the instructions. If they were uncertain, they could ask the tutor for guidance. In the GT approach, each exercise provided students with comprehensive step-by-step instructions and screenshots guiding the student through the entire exercise. Students were given automatic feedback from *WebLearn* telling them to re-try the steps when they made an error. Students would then be given another exercise to practice the procedure covered by the step-by-step instructions. The goal of GT was to have students practising the statistical package in an error-avoidant environment.

7.3 Results

Data analysis comprised of the following three phases: validating training approaches, modelling adaptive transfer scores, and comparing training approaches on other outcomes. In order to assess training validity, mean ratings on manipulation check items were compared between training approaches using a series of independent sample t-tests. This was important as the correct manipulation of training approaches related directly to the internal validity of the study. Adaptive transfer scores were modelled using one-way analysis of covariance (ANCOVA). ANCOVA allowed the mean adaptive transfer scores to be compared between training approaches after controlling for the effect of training covariates. It was important to control for covariates in these models due to non-random allocation of participants to training approaches. The assumptions for ANCOVA were checked prior to reporting and interpreting models. No strong evidence of any violations to the assumptions of ANCOVA were found. Due to some covariates containing a high proportion of missing values, multiple imputation techniques were used to estimate missing values. This aimed to reduce possible bias introduced by standard listwise deletion in *SPSS* and improve the statistical power of the models. Finally, a series of independent sample t-tests were used to compare mean self-reported ratings on other training outcomes in order to explore the possible implications of implementing either of the training approaches.

7.3.1 Validating Training Approaches

In order to evaluate whether the training approaches had been conducted correctly, mean student self-report ratings on metacognition, emotional control, learning from errors, error strain, guided training behaviour and exploratory training behaviour were compared using a series of independent sample t-tests (Table 7.1). The results of these tests revealed that participants' mean ratings of the EMT approach were significantly different to the mean ratings of participants in the GT approach on items of active exploration, exploration without instructions, metacognition, operation without instruction, seeking assistance and the use of step-by-step instructions. Participants in the EMT approach reported significantly higher mean self-reported ratings of exploratory behaviour, metacognition, and operation without instructions. However, there were no significant differences on ratings of error strain, collaborating with other students, emotional control or learning from errors (see Table 7.1). These manipulation checks are reported along with the means found in Trial I. There was evidence of a vast improvement to the validity of EMT for Trial II.

7.3.2 Modelling Adaptive Transfer Scores

Before modelling adaptive transfer scores, the first step was to identify important covariates. Descriptive statistics and intercorrelations for covariates and adaptive transfer scores between training approaches are shown in Table 7.2. Covariates that were statistically significantly correlated with adaptive transfer scores were selected as covariates. Gender, personal access, training progress, and statistical knowledge were all significantly and positively correlated with adaptive transfer scores. The personal access variable contained a high degree of missing values, 32/93 (34.8%) for GT and 3/32 (9.4%) for EMT.

Table 7.1: Trial II Descriptive Statistics and Independent Samples t-tests Comparing Mean Ratings of Manipulation Checks between Training Approaches

								95% CI of Difference	
Manipulation Variable		M	SD	N	SEM	t	p	Lower	Upper
Metacognition	GT	4.46	1.06	57	0.14	-2.14	.035*	-0.87	-0.03
	EMT	4.91	0.73	32	0.13				
Emotional Control	GT	5.44	0.97	57	0.13	2.08	.041*	0.02	0.84
	EMT	5.01	0.86	32	0.15				
Learning from Errors	GT	4.83	1.05	57	0.14	-1.47	.146	-0.83	0.12
	EMT	5.18	1.14	32	0.20				
Error Strain	GT	2.91	1.46	57	0.19	-.97	.333	-0.89	0.31
	EMT	3.21	1.17	32	0.21				
Used step-by-step instructions	GT	6.21	1.45	57	0.19	4.34	<.001**	0.72	1.95
	EMT	4.87	1.29	32	0.23				
Copied other students	GT	2.19	1.55	57	0.21	-1.18	.240	-1.07	0.27
	EMT	2.59	1.50	32	0.27				
Immediately sought assistance	GT	4.63	2.02	57	0.27	2.02	.046*	0.02	1.81
	EMT	3.72	2.08	32	0.37				
Actively explored *SPSS*	GT	3.70	1.79	56	0.24	-2.67	.009**	-1.73	-0.25
	EMT	4.69	1.45	32	0.26				
Operate without instruction	GT	3.61	1.87	57	0.25	-4.10	<.001**	-2.34	-0.81
	EMT	5.19	1.47	32	0.26				
Explored without instruction	GT	2.98	1.70	57	0.22	-5.39	<.001**	-2.51	-1.15
	EMT	4.81	1.20	32	0.21				

* $p < .05$, ** $p < .01$

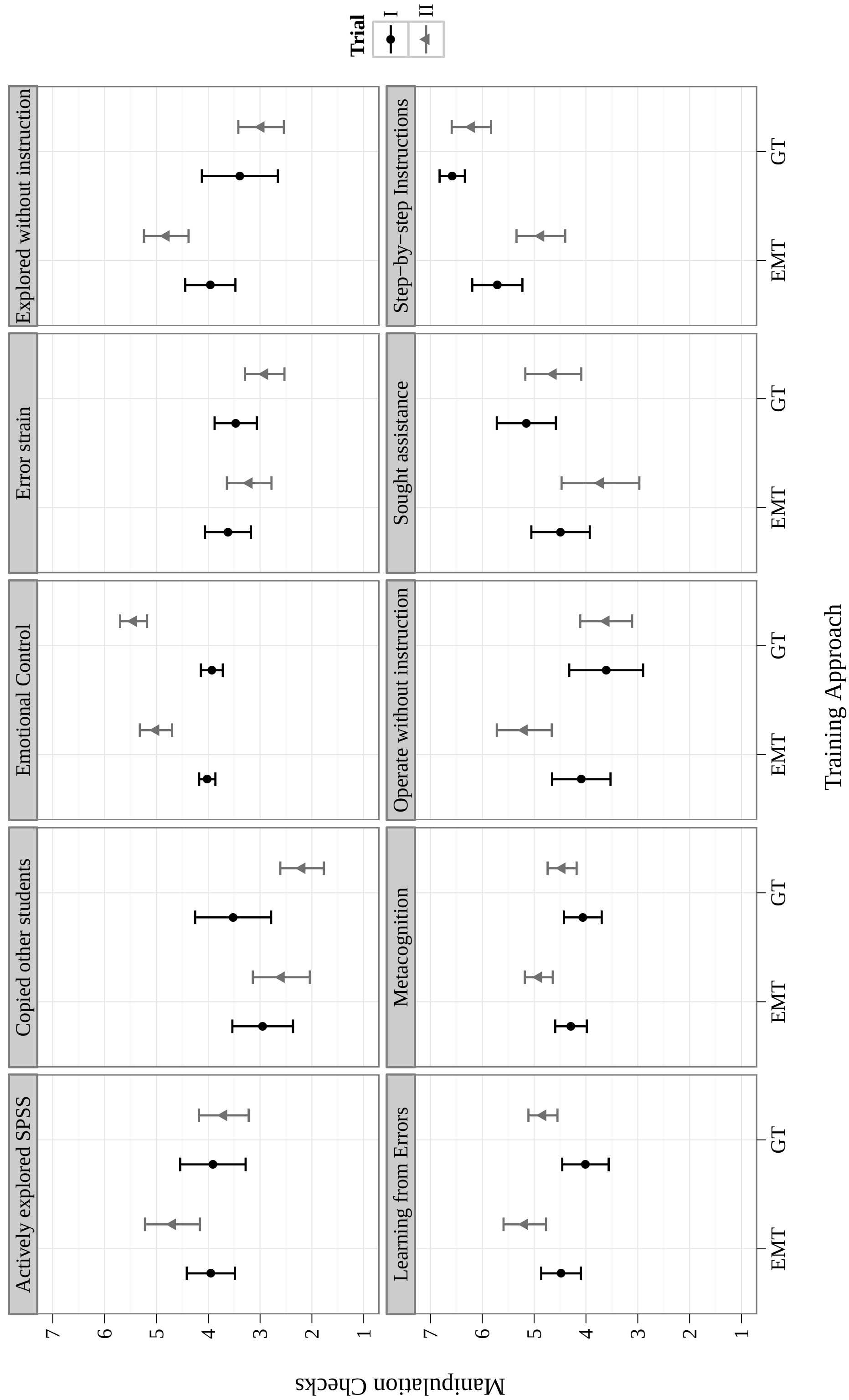


Figure 7.1: Trial II Mean ratings with 95% CI for training manipulation checks between training strategies. Trial I results are included for comparison.

Table 7.2: Trial II Intercorrelations and Descriptive Statistics of Study Variables Between Training Approaches

Variable		1.	2.	3.		4.	5.	6.	7.	8.	9.
1. Gender[a]		-	.109	.170		.093	.240*	.067	-.116	.085	.221*
2. Personal Access			-	.010		.355**	.215*	-.199	.241*	-.024	.296**
3. Previous Training Experience				-		.138	.128	-.160	-.003	.091	.068
4. Age						-	.165	-.013	.286**	.143	.055
5. Labs Completed Outside Training							-	-.322**	-.036	-.061	-.013
6. Training Progress								-	.059	.421**	.238*
7. Performance Utility									-	.079	.052
8. Statistical Knowledge										-	.485**
9. Adaptive Transfer											-
GT	N	26/81 Male	8/57 Yes	9/81 Yes	M	20.2	3.09	9.16	5.61	72.38	13.79
	%	32.1	14.0	11.1	SD	3.21	3.4	1.84	0.97	13.65	6.97
					N	81	57	81	79	81	81
EMT	N	9/34 Male	13/32 Yes	2/34 Yes	M	22.32	4.41	7.88	5.94	66.25	13.24
	%	26.5	40.6	5.9	SD	6.95	2.80	2.58	0.75	14.42	7.32
					N	34	32	34	34	34	34

* $p < .05$, ** $p < .01$, [a] Gender: Females = 1, Males = 2.

Adaptive transfer scores were modelled using one-way analysis of covariance (ANCOVA). ANCOVA allowed for the mean adaptive transfer scores to be compared between training approaches after controlling for the effects of gender, personal access, training progress and statistical knowledge. The first model employed traditional listwise deletion of cases with missing values present in the personal access covariate. While the overall model was statistically significant, $F(5, 83) = 8.93, p < .001, \eta^2 = .35$, $N_{GT} = 57$, $N_{EMT} = 32$, training approach was not a statistically significant predictor of adaptive training transfer scores, $F(1, 83) = 0.22, p = .64, \eta^2 = .003$. Personal access, $F(1, 83) = 9.34, p = .003, \eta^2 = .10$, and statistical knowledge, $F(1, 83) = 15.86, p < .001, \eta^2 = .16$ were both statistically significant covariates (see Table 3). Gender, $F(1, 83) = 3.80, p = .06, \eta^2 = .04$, and training progress, $F(1, 83) = 0.80, p = .37, \eta^2 = .01$, failed to reach statistical significance in the model suggesting that personal access and statistical knowledge better accounted for adaptive transfer scores (see Table 7.3).

A second model was refitted after removing the personal access covariate in order to study its influence in the initial model and employ a set of covariates with fewer missing values. The second model was also statistically significant, $F(4, 110) = 10.4, p < .001, \eta^2 = .27$, $N_{GT} = 81$, $N_{EMT} = 34$, but did exhibit a lower partial η^2 indicating a higher degree of unexplained variance (see Table 7.3). Once again, training approach was not statistically significant, $F(1, 110) = 0.91, p = .343, \eta^2 = .01$, but it did enter the model showing a slightly larger effect. With the removal of personal access, gender became statistically significant, $F(1, 110) = 5.02, p = .03, \eta^2 = .04$, and statistical knowledge remained in place as the strongest predictor, $F(1, 110) = 26.37, p < .001, \eta^2 = .19$. As per the initial model, training progress was not statistically significant, $F(1, 110) = 0.34, p = .56, \eta^2 = .00$.

A comparison of the two previous models suggested some important co-variation between adaptive transfer scores, personal access and gender. Given that personal access was highly correlated with adaptive transfer scores (see Table 7.2) and there was a large difference in the proportion of students with personal access between training approaches (40.6% EMT vs. 14% GT), both of the previous models suffered serious limitations. Model 1 was underpowered and possibly biased by the listwise removal of missing cases and Model 2 completely ignored the personal access covariate.

Table 7.3: ANCOVA Model Parameters Predicting Adaptive Transfer

	1. Listwise deletion					
Parameters	B	95% CI	SE	t	p	η^2
Gender[a]	2.74	(-0.06, 5.54)	1.41	1.95	0.055	0.04
Personal Access	4.82	(1.69, 7.96)	1.58	3.06	0.003**	0.10
Training Progress	0.30	(-0.37, 0.96)	0.33	0.90	0.373	0.01
Statistical Knowledge	0.21	(0.10, 0.31)	0.05	3.98	$< .001$**	0.16
Training approach[b]	-0.67	(-3.52, 2.17)	1.43	-0.47	0.640	0.00
GT Adjusted Mean	13.49	(11.88, 15.10)	$N = 57$			
EMT Adjusted Mean	14.16	(11.96, 16.37)	$N = 32$			
	2. Personal access removed					
Gender[a]	2.79	(0.32, 5.26)	0.03	2.24	0.027*	0.04
Training Progress	0.17	(-0.42, 0.77)	0.56	0.58	0.564	0.00
Statistical Knowledge	0.23	(0.14, 0.32)	0.00	5.14	$< .001$**	0.19
Training approach[b]	-1.24	(-3.83, 1.34)	0.34	-0.95	0.343	0.01
GT Adjusted Mean	13.26	(11.90, 14.62)	$N = 81$			
EMT Adjusted Mean	14.50	(12.36, 16.64)	$N = 34$			
	3. Multiple imputation of missing values					
Gender[a]	2.20	(-0.16, 4.56)	1.20	1.83	0.067	
Personal Access	5.32	(2.17, 8.47)	1.60	3.33	0.001**	
Training Progress	0.36	(-0.23, 0.94)	0.30	1.20	0.232	
Statistical Knowledge	0.21	(0.13, 0.30)	0.04	4.85	$< .001$**	
Training approach[b]	0.03	(-2.51, 2.57)	1.30	0.03	0.980	
GT Adjusted Mean	13.64	(12.34, 14.93)	$N = 81$			
EMT Adjusted Mean	13.60	(11.53, 15.68)	$N = 34$			

* $p < .05$, ** $p < .01$, [a] Females = 1, Males = 2, [b] GT = 1, EMT = 0

Consequently, a third model was fitted. The third model used a multiple imputation (MI) method to estimate missing values for the personal access covariate. While the assumption behind this procedure states that missing values are required to be missing at random (MAR) or missing completely at random (MCAR), studies suggest MI performs quite favourably in situations where data are not missing at random (non-MAR, Shrive, Stuart, Quan, & Ghali, 2006; Greenland & Finkle, 1995). As Schafer (1997) explains, multivariate data sets that exhibit robust associations between variables provide a useful basis for imputing missing values which aids in minimizing possible bias introduced by imputation of non-MAR values.

Multiple imputation was performed using the *IBM SPSS Missing Values 19* package. All covariates and outcome variables were specified in the model and ten imputations were obtained. Parameters estimates for the ten imputations were pooled together and used to construct the third ANCOVA model (see Table 7.3). The results of the ANCOVA using pooled parameter estimates from multiple imputations of missing values validated the results of Model 1. Personal access, $p < .001$, and statistical knowledge, $p < .001$, were the only statistically significant predictors of adaptive training transfer. There was no evidence of a statistically significant effect for training approach, $p = .98$.

7.3.3 Other Training Outcomes

Independent samples t-tests were used to compare mean self-reported ratings between training approaches on training difficulty, training satisfaction, training anxiety, and post-training self-efficacy (see Table 7.4). Mean self-reported ratings of participants' perceptions of certification task's difficulty, anxiety and degree of preparedness were also analysed (see Table 7.4). Evidence of a statistically significant difference in mean ratings were found for training difficulty ($p < .001$) and satisfaction ($p = .016$). There was no evidence of statistically significant differences in participants' ratings of training anxiety ($p = .79$) and statistical package self-efficacy ($p = .67$). In terms of participants' perceptions of the certification task, there was no statistically significant evidence of any differences existing between participants' mean ratings of difficulty ($p = .492$), anxiety ($p = .525$) and preparedness ($p = .655$).

Table 7.4: Trial II Descriptive Statistics and Independent-samples t-tests Comparing Training approaches on Other Training Outcomes

								95% CI of Difference	
Outcome		M	SD	N	SEM	t	p	Lower	Upper
Training Difficulty	GT	3.30	1.21	57	0.16	-3.47	0.001**	-1.46	-0.40
	EMT	4.23	1.18	31	0.21				
Training Satisfaction	GT	5.19	1.30	57	0.17	2.46	0.016*	0.14	1.31
	EMT	4.47	1.39	32	0.25				
Training Anxiety	GT	3.16	1.20	57	0.16	-0.27	0.788	-0.60	0.46
	EMT	3.23	1.22	32	0.22				
Self-efficacy	GT	4.98	1.13	57	0.15	-0.43	0.671	-0.55	0.35
	EMT	5.07	0.80	32	0.14				
CT Difficulty	GT	4.87	1.15	77	0.13	0.69	0.492	-0.30	0.62
	EMT	4.71	0.94	31	0.17				
CT Anxiety	GT	4.24	1.59	78	0.18	-0.64	0.525	-0.85	0.44
	EMT	4.45	1.39	31	0.25				
CT Preparedness	GT	4.48	1.37	77	0.16	0.45	0.655	-0.43	0.68
	EMT	4.35	1.17	31	0.21				

* $p < .05$, ** $p < .01$, CT = Certification Task

7.4 Discussion

The aim of this study was to evaluate the effect of training approaches for the development of technological skills in statistics education. This study specifically examined statistical package skills and how different training approaches might promote the development of sustainable outcomes, i.e. adaptive transfer. The EMT approach, a sub-type of active-exploratory training, was hypothesized to promote adaptive transfer above and beyond a conventional GT approach. The hypothesis of this study was based the positive outcomes of previous research which has looked at adaptive transfer for general software skills, e.g. computer simulations, word processors, database searches, and spreadsheets (Keith & Frese, 2008; Bell & Kozlowski, 2008; Chillarege et al., 2003; Frese, Brodbeck, et al., 1991; Heimbeck et al., 2003; Keith, Richter, & Naumann, 2010; Keith & Frese, 2005). However, after controlling for covariates, the results of this study found no statistical evidence of an association between the EMT approach and students' level of adaptive transfer. These results contradict an early experiment looking at statistical package skills by Dormann and Frese (1994), but confirm the results of Trial I.

The findings of the Dormann and Frese (1994) experiment suggested initial promise for EMT for statistical package skills. However, their experiment had many limitations which required further research. Short-term follow-up, a small sample, one-off training sessions, and no deliberate attempt to measure adaptive transfer seriously limited their conclusions. Trial I also had limitations. Due to significant constraints imposed on educational research, Trial I confronted issues with a short duration of training, un-blinded participants, questionable validity of training transfer measures, questionable student engagement during the evaluation of training transfer, and questionable validity of the imposed EMT approach. Hence, the aim of Trial II was to address these limitations.

The strengths of this study lie in its ecological validity (positioned within a real introductory statistics course), careful manipulation of training approaches, and improved validity of the evaluation of adaptive transfer for statistical package skills. Regardless, this study still had limitations. Once again, this study used a sample of psychology students, which are unlikely to reflect the diverse characteristics of all students who take introductory statistics courses. Therefore, any results must be cautiously generalised to other student backgrounds. While randomised experiments are highly regarded for this type of evaluation, randomized protocols are notoriously challenging to implement effectively in an educational setting. Quasi-experimental designs provide a feasible compromise. However, due to non-randomization, the potential for systematic bias between training approaches is high. Fortunately, research suggests that quasi-experimental designs can provide reliable approximations to randomized experiments providing proper adjustment to known covariates has taken place (Shadish et al., 2008). Trial II was designed prospectively to control for known covariates in the statistical analysis. Regardless, the degree to which this study has approximated a randomised study is difficult to ascertain.

There were a number of differences between the training approaches, or campuses, that were likely to impact on the development of adaptive transfer. As the Kanfer and Ackerman (1989) model suggests, the cognitive ability of trainees will have an effect on training performance and subsequent training transfer outcomes. While statistical knowledge is no substitute for a measure of general cognitive ability, it does provide insight into the academic and statistical ability of participants. The descriptive statistics show a difference of six percent on average statistical knowledge scores between training approaches/campuses. This highlights a key difference between the two approaches' participants' academic abilities. This difference is further supported by national tertiary entrance requirements for undergraduate university programs. Campus A (EMT) and Campus B (GT) entrance scores were respectively 68 and 77 out of a theoretical 100. This suggests that students who performed better in their final years of secondary school were more attracted to Campus B even though they are in the same psychology programs run across different campuses. Fortunately, the adjustment for statistical knowledge does reduce the possibility of bias attributed to differences in students' academic ability.

Differences between the campuses that could not be controlled for were the class and laboratory session times. Campus A lectures and computer laboratory sessions were scheduled from midday to mid-afternoon, and Campus B were scheduled during the mornings. Anecdotally, previous students from Campus A have raised concerns about the scheduling of the statistics course in the afternoon stating that they felt tired by the time they got into the computer laboratory sessions by late afternoon. Students had a clear preference for morning sessions. However, due to institutional constraints, the computer laboratory sessions could only be scheduled during the afternoon. This difference between the campuses could explain a number of the study's observations. It may explain why the overall perceived difficulty and satisfaction of training was lower for EMT/Campus A. There is no doubt that being tired would lower overall satisfaction and increase perceived difficulty. This may also explain why many of the Campus A participants reported completing training sessions outside of the scheduled times more frequently. Completing more training sessions outside of class would also explain why their average level of training progress was lower prior to the certification task. The structure and weekly progression of the scheduled laboratory sessions would be more likely to keep students up-to-date. Forcing students to attend the scheduled laboratory computer sessions would have been possible, but doing so would have violated the ecological nature of this study. It was important for these courses to allow students access to training sessions in their own time. This also addressed a key recommendation from Trial I to increase practice opportunities for students.

Comparison of mean ratings on the manipulation check items showed that participants in the EMT approach engaged in less guided instruction and more exploratory behaviour. This was a vast improvement on the manipulation checks reported in Trial I. Surprisingly, however, the error-framing aspect of EMT was not validated. The addition of an error-framing element to active-exploratory training has been found to provide a unique effect above and beyond active-exploratory training alone (Keith & Frese, 2008). The absence of an error-framing effect may have reduced the overall effectiveness of EMT. This study suggests that encouraging and promoting errors as a beneficial aspect of training for statistical package skills might present a unique challenge. Given that most students come from educational settings where errors are viewed as failure and something to be avoided, one semester of training may not have been enough to change students' perceptions and attitudes towards making errors.

The certification task, which aimed to measure statistical package adaptive transfer skills, was an improvement on the validity of the self-assessment exercises of Trial I. The certification tasks were designed to minimize the effect of statistical knowledge on operating the statistical package. While students still required a basic level of statistical knowledge to understand the output that was given, this dependency was reduced as students did not have to make statistical knowledge decisions about what statistical methods to use. The students could concentrate on demonstrating their ability to operate the statistical package. Anecdotally, student engagement during the certification task was reported to be high. The certification task was the only training session that was compulsory to attend in person. Tutors were present during these sessions to ensure exam conditions were imposed. Making the certification task worth 25% of the computer laboratory participation grade ensured that students took the task seriously.

Overall, this study failed to support the efficacy of EMT over GT. Therefore, it is important to consider possible explanations that may explain why an effect may not have been detected. One possible explanation that requires further investigation is the potential mediating effect of prior knowledge on EMT. The effectiveness of EMT is based on studies using technological skills that do not require specialised prior knowledge (e.g. word processors, presentation software, spreadsheets etc., Keith & Frese, 2008). Technological skills for statistics may present a special case as these skills are likely to be highly dependent on trainees' knowledge of statistics. This explanation concurs with the findings of Debowski et al. (2001) who found that low task feedback moderated the effect of active-exploratory training. Students with low statistical knowledge would need to rely on their limited understanding and the inbuilt error feedback of the statistical package. However, instructors know all too well the limitations of statistical package warnings. For example, if students do not understand the different types of variables, many statistical packages will happily calculate a "mean" gender where males and females have been coded numerically. Statistical knowledge enhances the task feedback of training and therefore may moderate the effect of EMT. This would explain the difference between this study and the findings of Dormann and Frese (1994). Dormann and Frese used participants who had already completed introductory statistics courses and may have already developed the necessary knowledge to enhance task feedback to benefit from EMT. On the other hand, Trial I and II trained students during the development of the required prior knowledge. These students may have missed out on the benefits of EMT as they were still coming to terms with understanding statistical concepts. Therefore, low prior statistical knowledge may moderate the effect of EMT in statistics education. Future research should test this hypothesis by evaluating EMT on students who already possessing prior statistical knowledge.

This study confirmed a moderate relationship between training transfer and statistical knowledge identified in Trial I. This relationship suggests that students who have a better understanding of statistical concepts tended to develop statistical package skills better than students with lower statistical knowledge. As discussed in the previous paragraph, this is likely due to the increased task feedback provided by having adequate contextual knowledge. However, there is still a large degree of unexplained variance suggesting that many other factors may come into play. This study asked participants if they had personal access to the statistical package. Students with personal access tended to perform better on measures of adaptive training transfer even after controlling for participants' statistical knowledge, gender, and training progress. Personal access may have provided students with greater opportunity to practice and the ability to better integrate the statistical package into their regular repertoire of software. This finding also emerged in the qualitative phase of Trial I and suggests an interesting avenue for future research. Future research should look at evaluating the importance of personal access to technology on the development of technological skills. The results of this study suggest that access will likely produce a greater effect than the use of different training approaches.

7.5 Conclusion

Technological skills, such as the ability to operate statistical packages, are an important part of modern notions of statistical literacy. While the focus of statistics education is to teach the concepts, instructors can no longer ignore the importance of technological skills, especially, as students become more and more reliant on the technology. Statistics education research needs to play a key role in understanding how these types of skills interact in statistics courses and how these skills are best developed. This series of studies found no association between the development of statistical package skills and two different types of training approaches, error-management training and guided training. However, the findings identified important areas for future research. The potential moderating effect of prior knowledge on statistics technological skills require further investigation. Statistical knowledge was indeed the most important predictor of adaptive transfer. The importance of personal access technology may also prove to be an important determinant. Further research is needed to understand how these factors and many other undiscovered factors can be manipulated to foster students' development of technological skills in statistics education.

Part II

Cognitive Conflict for Correcting Misconceptions

Chapter 8

Part II - Abstract

Previous studies in science and statistics education suggest that cognitive conflict strategies may provide a quick and highly effective intervention for reducing common misconceptions related to students' statistical reasoning (Limón, 2001). Cognitive conflict interventions present conflicting or anomalous information to students which aim to promote conceptual change (Posner, Strike, Hewson, & Gertzog, 1982). However, previous research in statistics education has typically evaluated cognitive conflict interventions for only a few misconceptions using highly targeted, typically tutorial-based, sessions (e.g. Kalinowski et al., 2008; Jazayeri et al., 2010; Liu et al., 2010). Studies are needed to evaluate the effect of cognitive conflict-based activities for addressing a wider range of misconceptions and delivering them using different methods, e.g. via lectures. The aim of the trial for Part II was to evaluate the effect of brief lecture-based cognitive conflict activities aimed at addressing a wide range of misconceptions across an entire semester of an introductory statistics course.

The Part II trial (Chapter 10) compared two yearly cohorts of a large introductory statistics and epidemiology course for medical science students. The control cohort completed the course as normal and answered select multiple choice questions measuring statistical reasoning and misconceptions that would be compared to the following year's cohort. In the following year, an intervention cohort completed the same course but, in addition, received a series of eight brief lecture-based cognitive conflict activities throughout the semester targeting a wide range of misconceptions. The intervention cohort completed the same multiple choice exam questions that were linked to misconceptions targeted by each cognitive conflict-based activity.

The overall conceptual change scores for the intervention cohort were significantly higher than the control cohort. However, the effect was small. Individual question analysis revealed statistically significant associated effects for the cognitive conflict activities targeting probability and regression. Surprisingly, one question related to confidence intervals was significantly associated with poorer performance in the intervention cohort.

The trial found some promising evidence on the potential effect of brief lecture-based cognitive conflict activities for confronting students' commonly held misconceptions of statistics concepts. The activities that were associated with a statistically significant effect suggest that the complexity of the misconceptions being targeted may moderate the effect of the brief lecture-based interventions. Misconceptions related to more difficult concepts, i.e. statistical inference, may require more careful, intensive and targeted interventions. The associated poorer performance in the intervention group for a confidence interval question highlights the important role of evaluation research. Sometimes well meaning interventions may have unexpected effects which would otherwise be missed without careful evaluation.

Chapter 9

Part II - Introduction

9.1 Misconceptions and Cognitive Conflict for Conceptual Change

Learning statistics requires students to understand many difficult, complex and counter-intuitive concepts (Ben-Zvi & Garfield, 2005). Not surprisingly, the statistics education literature has documented a wide range of misconceptions that students may hold (e.g. Castro Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007; Fidler, 2006). A misconception can be defined as a "pattern of errors that reflects a misunderstanding of a statistical concept" (p. 35, Cohen, Smith, Chechile, Burns, & Tsai, 1996). Evidence suggests that misconceptions present in statistics education are highly pervasive, persistent and difficult to change (Garfield & Ben-Zvi, 2007). In a review of the literature, Castro Sotos et al. (2007) identified 17 studies documenting students' misconceptions relating to statistical inference. These included misconceptions about sampling distributions (e.g. the law of small numbers and sampling variability, Finch, 1998), hypothesis testing (e.g. misinterpretations of p-values, Haller & Krauss, 2002) and confidence intervals (e.g. the effect of sample size on confidence interval width, Fidler, 2006). Misconceptions are of great concern to statistics instructors because they reduce students' statistical reasoning skills or their ability to correctly understand and interpret statistical information (Garfield & Chance, 2000). Statistics instructors require interventions aimed at overcoming and reducing the occurrence of misconceptions.

Previous interventions have been based on conceptual change theory. Conceptual change theory aims to explain how students change their conceptions when presented with new information (Posner et al., 1982). Interventions have focused on creating cognitive conflict in students by presenting them with anomalous information (Limón, 2001). As Posner et al. (1982) explain, conceptual change is unlikely to take place unless students' previous conceptualisations become implausible and a new conceptualisation is presented. A typical cognitive conflict intervention for conceptual change is based on three major steps: (a) the students' current understanding of a concept is identified, (b) the students are presented with conflicting information that renders their prior conceptualisation implausible and a scientifically valid conceptualisation is introduced and (c) the extent of conceptual change is evaluated (Limón, 2001). The method of presenting conflicting information can vary from direct instruction, self-guided or group-based (Hirsch & O'Donnell, 2001). A number of researchers in statistics education have investigated the impact of cognitive conflict for conceptual change.

9.2 Studies on Cognitive Conflict in Statistics Education

Early studies by Watson (2002a, 2002b, 2007) employed interviews and video-based peer prompting to correct primary and secondary school children's misconceptions about sampling (Watson, 2002a), averages (Watson, 2007) and inferential comparison between two groups (Watson, 2002b). All studies employed a similar interview-based protocol that evaluated students' understanding of a statistical concept by presenting them with a series of questions. Students' responses to the question were graded on a hierarchy of conceptual understanding. Students who exhibited misconceptions in their initial responses to questions were shown video or textual prompts of other students explaining a concept in a more statistically valid way. These prompts aimed to create cognitive conflict in students. The interviewer would then ask the students what they thought about their initial answers after the prompts were presented. The interviewer recorded whether or not the students improved their conceptualisation after the prompts. Watson (2002a) found that 7/32 (22%) students improved their conceptualisation of sampling after prompting. Watson (2002b) found that 13/23 (57%) and 15/50 (30%) students improved their conceptualisations for moderate and difficult questions, respectively, for comparing two groups using a graphical format after conceptual change prompts. Watson (2007) found that 27/46 (59%) students improved their understanding of averages following cognitive conflict prompts. Overall, Watson's studies suggest that student and peer interactions may be an effective vehicle for creating cognitive conflict and conceptual change. However, the major limitation of Watson's work was the lack of a control group.

Hirsch and O'Donnell (2001) compared three different methods of cognitive conflict interventions for correcting misconceptions related to probability. The three interventions included direct instruction, individual activities and small group activities. These three interventions were compared to a control group where no cognitive conflict was created. All 103 students who participated in the study were identified as having prior misconceptions on a pre-test on probability. All students watched a one-hour video lecture on probability and completed a short 25 minute intervention based on their randomly allocated conditions. The cognitive conflict interventions required students to answer questions regarding probability and then a deliberate attempt was made to draw their attention to their misconceptions, thus creating conflict. In a following week, the participants completed another 45 minute intervention session and immediately completed a post-test of probability misconceptions. The results indicated no association between cognitive conflict interventions and the alleviation of misconceptions. However, in a subsequent follow-up of 27 of the original participants in the following weeks of the post-test, a statistically significant association was found between the absence of misconceptions and instructional intervention. Those in the cognitive conflict interventions were less likely to have misconceptions. Unfortunately, this result must be interpreted with caution due to the potential of follow-up bias.

Kalinowski et al. (2008) utilised two forms of cognitive conflict strategies for overcoming the inverse probability fallacy as it relates to hypothesis testing. The inverse probability fallacy occurs when a modus tollens argument is incorrectly applied to probabilistic reasoning, i.e. the illusion of probabilistic proof by contradiction (Castro Sotos et al., 2007). Kalinowski et al. randomly allocated six pre-existing tutorial groups in a third year undergraduate psychology program to two conditions. The first condition presented students with obviously false applications of the modus tollens argument and then explained how the argument creates the illusion of probabilistic proof by contradiction in hypothesis testing. This aimed to create conflict with students who had misconceptions about the nature of hypothesis testing. The second condition contrasted Bayesian posterior probabilities to p-values to create cognitive change through comparison to an alternate paradigm of statistical inference. This approach had been posited by a number of authors as a possible way of overcoming misconceptions about conventional statistical inference (Berry, 1997; Gigerenzer, Krauss, & Vitouch, 2004; Haller & Krauss, 2002; Lecoutre, 2006). Haller and Krauss dubbed this “insight by comparison” (p. 11). The researchers measured students’ misconceptions of hypothesis testing before intervention, post intervention and at five week follow-up. The intervention was administered in a single 45 minute tutorial. The researchers found both methods led to an equal and statistically significant reduction in misconceptions of hypothesis testing. The equal effectiveness of both methods suggests that the underlying mechanism behind the interventions, i.e. confronting misconception with conflicting information, was the cause. The anomalous information used to create the conflict did not appear to matter.

Jazayeri et al. (2010) studied the impact of cognitive conflict on students’ reasoning about sampling variability. The study used a sample of 185 psychology students enrolled in an introductory statistics course. These students participated in weekly tutorial classes of approximately 20 students per group. The study involved randomly allocating a cognitive conflict tutorial activity to these different tutorial groups early in the semester. The cognitive conflict groups were asked a question regarding the relationship between sample size and sampling variability. After the students answered the question, the lecturer directly confronted any misconceptions with conflicting information. The intervention was reported to take approximately 10 minutes of tutorial time. A standard instruction group, which acted as a control, received the same question, but the misconceptions were not directly confronted. Only the correct answer was shown to the students. Later in the semester, all students’ were followed up with a post-intervention sampling variability question. Students who were in the cognitive conflict-based tutorials earlier in the semester were significantly more likely when compared to the standard tutorial groups to reason correctly about the relationship between sample size and sampling variability. Jazayeri et al. (2010) concluded that cognitive conflict-based tutorial activities can have a significant and lasting effect on reducing misconceptions and improving reasoning about sampling variability.

Liu et al. (2010) and Liu (2010) created a computer-assisted learning program, named Simulation Assisted Learning Statistics (SALS), to address students' misconceptions about correlations by inducing cognitive conflict. A sample of 72 final year secondary school students were randomly allocated by Liu et al. (2010) to either a SALS-based learning program intervention or a lecture-based control group for correcting common misconceptions about correlation. The SALS condition completed ten learning activities that used computer-based learning and cognitive conflict to correct misconceptions. The lecture-based control group also received ten activities which involved reading, practicing, correcting and reviewing concepts of correlation. A pre-test measuring misconception about correlations was administered the day before the intervention. The same test of misconceptions was given as a post-test immediately after the intervention. The results found that the SALS-based intervention was statistically significantly more effective than the lecture-based learning group for correcting common misconceptions related to correlation.

9.3 Rationale and Aims

The literature reviewed in Section 9.2 evaluating the effectiveness of cognitive conflict as a method of conceptual change has provided overall supportive evidence. However for the few studies that do exist their scope has been limited to addressing a small range of misconceptions using highly targeted intervention sessions. Introductory statistics courses are likely to be filled with a variety of misconceptions and implementing these interventions in targeted sessions is often impractical. For example, accessing high quality tutors or training new tutors to implement these strategies may not be feasible. Limón (2001) also suggested future studies were needed to consider the time required to achieve conceptual change. Studies in statistics education have varied in length of cognitive conflict intervention from more than an hour to only ten to fifteen minutes. Given that Jazayeri et al. (2010) had success with brief (10 minutes) interventions for sampling variability, it was of great interest to this study to see if the brief format would be effective for different types of misconceptions. Therefore, the aim of this study was twofold. The first aim was to evaluate the effectiveness of a series of cognitive conflict activities addressing a wide range of misconceptions throughout an entire semester of an introductory statistics course. The second aim was to evaluate the effectiveness of brief cognitive conflict activities delivered during regular lectures by a statistics instructor instead of longer specialised sessions typically used by other studies. It was hypothesised that the use of cognitive conflict exercises across the semester of a large introductory statistics course would be associated with fewer misconceptions and better statistical reasoning when compared to a control course that did not receive the cognitive conflict activities intervention.

Chapter 10

Part II - Trial

10.1 Aim of Trial

The aim of this trial was to evaluate the effectiveness of brief cognitive conflict-based activities embedded in lectures for confronting and correcting a range of common statistical misconceptions in a large introductory statistics course.

10.2 Method

10.2.1 The Cohorts

This trial was conducted on two yearly cohorts of the same introduction to epidemiology and statistics course. The twelve week course covered an introduction to epidemiology, statistics, dose response, statistical inference via cross-tabulation, common statistics in epidemiology, one and two-sample statistical inference, regression, correlation and one-way ANOVA. The course was largely given to students from Laboratory Medicine (Lab Med), Pharmaceutical Science (Pharm Sci), Biomedical Science (Biomed Sci) and Pharmacy (Pharm) programs. Weekly course contact consisted of a two-hour lecture, one-hour computer laboratory session and one-hour tutorial session. Assessment included computer laboratory training task completion (25%), tutorial worksheets (20%) and a final exam (55%). The course topics and assessment structure were consistent across the two cohorts.

The control cohort comprised of 225 students of whom 161 (72%) students consented to have their data recorded for the purpose of this trial. In the following year the intervention cohort consisted of 241 students of whom 167 (69%) students consented to participate. Table 10.1 compares the demographics between the cohorts using an independent samples t-test and χ^2 tests of association. The sample cohorts were similar on mean age ($p = 0.595$) and the distribution of gender ($p = .344$) and residency ($p = .238$, see Table 10.1). However, the intervention cohort was associated with an under-representation of pharmacy students ($p = .001$, see Table 10.1). This difference in program distribution was important to control for when analysing the results of the trial as the pharmacy program has a much higher academic entrance requirement than the other three programs that enrolled in the course.

Table 10.1: Cohort Demographics

			Cohort			
			Control	Intervention	Total	p
Age	$M \pm SD$		19.98 ± 3.17	20.16 ± 3.18	20.07 ± 3.17	0.595[a]
Gender	Female	N	93	105	198	0.344[b]
		%	47.0	53.0		
	Male	N	68	62	130	
		%	52.3	47.7		
Residency	Domestic	N	152	152	304	0.238[b]
		%	50.0	50.0		
	International	N	9	15	24	
		%	37.5	62.5		
Program	Lab Med	N	36	20	56	0.001[b]
		%	64.3	35.7		
	Pharm Sci	N	42	65	107	
		%	39.3	60.7		
	Biomed Sci	N	37	54	91	
		%	40.7	59.3		
	Pharm	N	45	28	73	
		%	61.6	38.4		

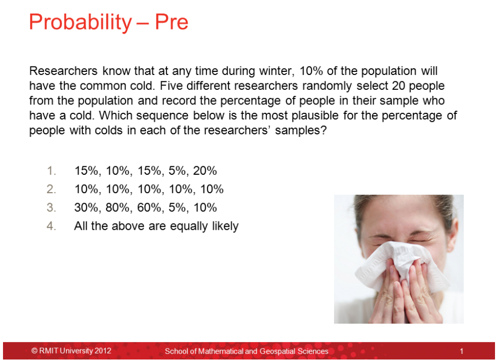
[a] means compared using independent samples t-tests assuming equal variance
[b] Pearson's χ^2 test of association

10.2.2 Outcomes Measures

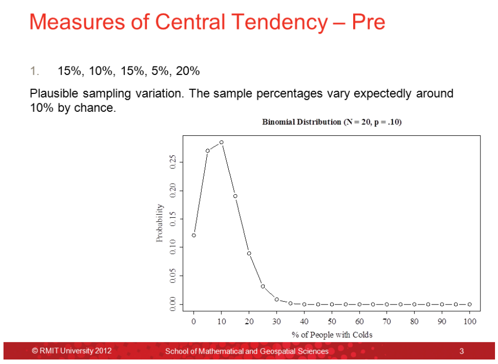
Statistical reasoning was measured using a conceptual change scale of 18 multiple-choice questions adapted from the Comprehensive Assessment of the Outcomes of a First Course in Statistics (CAOS) test (delMas, Garfield, Ooms, & Chance, 2006, 2007), Assessment Resource Tools for Improving Statistical Thinking (ARTIST) topic scales, and the ARTIST item database available from https://apps3.cehd.umn.edu/artist/index.html (see Garfield, delMas, & Zieffler, 2010, for a detailed discussion of these resources) . A further 22 items, also adapted from the CAOS test and ARTIST website, were included as a measure of baseline scores (see Appendix B.2 for all 40 multiple choice questions used in the exam). These items referred to other course concepts which were not targeted by the cognitive conflict activities. Therefore, the average scores on these questions should remain similar across the cohorts assuming there was no systematic difference between the ability of students and the delivery of the course. A baseline outcome aimed to help control for these potential types of confounding between the cohorts. The reasons for adapting the CAOS test and ARTIST questions included aligning the context of the questions to epidemiology, the addition of extra choices, removal of questions that did not relate to the learning outcomes of the course and the change of question wording to better align with definitions of confidence intervals covered in the course. Questions that were removed were replaced with new questions or adapted questions from ARTIST topic scales and the ARTIST item database (See Appendix B.3 for a detailed breakdown of the changes).

10.2.3 The Cognitive Conflict-based Activities

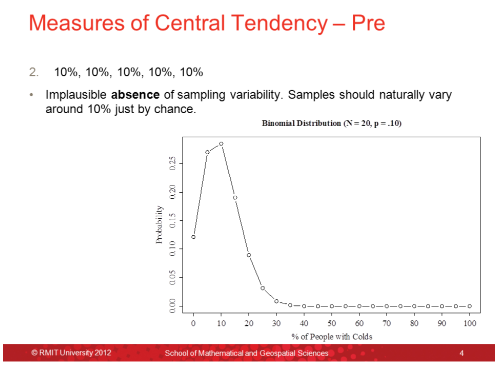
The cognitive conflict-based intervention activities aimed to improve students' statistical reasoning by confronting and correcting common statistical misconceptions. The activities were embedded in lectures throughout the semester. They were designed to be brief (approximately 10 mins) and took advantage of clicker technology built into the lecture venue for tallying responses. An example of the cognitive conflict activity used to correct misconceptions related to the equiprobability bias is shown in Figure 10.1 (see Appendix B.4 for the entire collection of activity slides). There were three stages to each activity. In the first stage the students' prior understanding was evaluated. This would help students determine if they held any misconceptions. In the second stage, anomalous, contradictory and conflicting information was presented which aimed to prompt students assimilate the correct conceptualisation being presented. This involved the presentation of prepared slides and lecturer-led discussion. As a follow-up, the third stage asked students a similar follow-up question to evaluate conceptual change. Each activity was uploaded online following the lecture for students to go through in their own time and for students who missed the lectures.



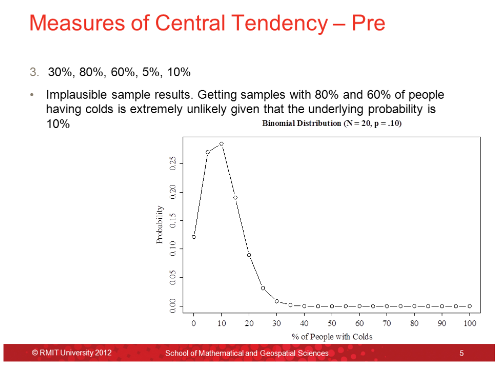
(a) Identify prior understanding



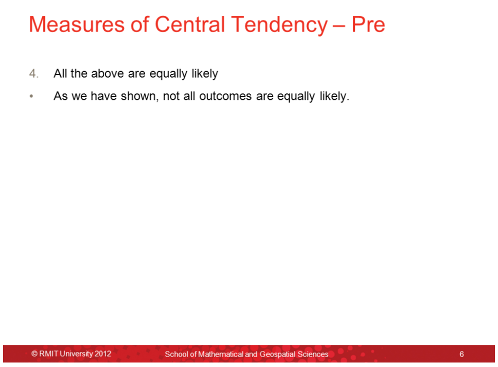
(b) Conflicting information



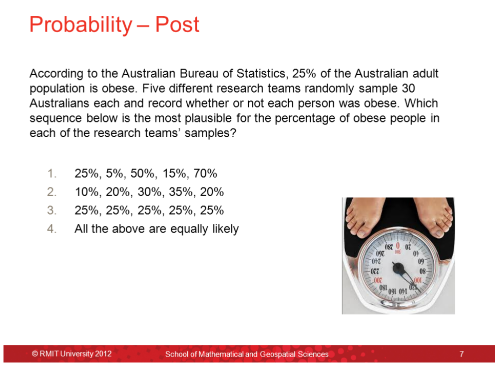
(c) Conflicting information cont.



(d) Conflicting information cont.



(e) Conflicting information cont.



(f) Conceptual change evaluated

Figure 10.1: An example of a cognitive conflict-based activity for overcoming misconceptions of probability caused by equiprobability bias

Each conceptual change-based activity was embedded in lectures close to topics where the misconceptions were more likely to arise (Table 10.2). As it would be impossible to address all possible misconceptions that may be exhibited in an introductory statistics course, this trial focused on the eight misconceptions listed in Table

10.2. These were selected because they included a diverse range of misconceptions and, based on previous experience and exam results, were known to be prevalent in the course. Each activity was linked to at least one conceptual change multiple-choice exam question that aimed to evaluate students' statistical reasoning and the presence of misconceptions (see Table 10.2). Note that a ninth pilot activity related to central tendency was included to initially practice the delivery and steps of the cognitive conflict activities. The pilot was not included in the evaluation or linked to a conceptual change exam question. It was assumed that confronting misconceptions would lead to better statistical reasoning as evidenced by a student being more likely to select the correct answer on the conceptual change multiple choice questions.

10.2.4 Procedure

Ethics approval for this project was provided by the RMIT College Human Ethics Advisory Network on the 27^{th} November 2009 (Project No: BSETAPP 64-09). The control cohort did not receive the cognitive conflict-based activities. The students received clicker-based questions just as in the intervention cohort, but the activities did not focus on the misconceptions covered by the cognitive conflict-based activities. The course topics, lecture schedule and other learning activities of the courses were largely the same between the cohorts. The multiple choice questions used to measure conceptual change and baseline scores were embedded at the end of the semester exam for both cohorts. The multiple choice questions comprised 50% of the marks for the exam. The other 50% came from a short-answer component.

Students were approached at the end of each semester and asked to consent to have their data recorded for the purposes of evaluating new course learning content delivered in introductory statistics courses (see Appendix B.1 for the plain language statement and consent form.). The students were not informed about the exact nature of the new learning content. Only the data from students who consented to participate in both cohorts are analysed in this trial.

Table 10.2: Course Topics, Conceptual Change-based Activity Schedule and Corresponding Conceptual Change Multiple Choice Exam Questions

Topic	Conceptual Change Activity	CC Outcome Questions[a]
Introduction to Epidemiology	-	-
Introducing Statistics	-	-
Dose Response	Central Tendency: Included as a pilot activity	N/A
	Distributions: Targeted students' understanding of how the shape of distributions relate to the nature of data	Distributions I - III
Statistical Inference using Crosstabs	Probability: Targeted the misconception that random events with different probabilities are all equally probable (equiprobability bias)	Probability
	p-values: Targeted the common misconception that the p-value is the probability of the null hypothesis being true.	p-values I - III
	Hypothesis testing – Targeted the misconception that rejecting the null hypothesis means the null hypothesis is false and the alternate is true (The illusion of probabilistic proof by contradiction).	Hypothesis Testing I - III
Common Statistics in Epidemiology	Confidence Intervals: Confusing interpretations of confidence intervals with Bayesian credible intervals	Confidence Intervals I – III
One and Two-sample Sample Inference	Sampling Distributions: Targeted the misconception that sampling distributions resemble sample distribution, i.e. belief in the law of small numbers and that a sampling distribution is the same thing as a sample distribution	Sampling Distributions I – III
Regression and Correlation	Correlation: Targeted the misconception that a correlation equals causation	Correlation
	Regression: The misconception that regression models can be used to predict beyond the range of data	Regression
One-way ANOVA	-	-

[a] See Appendix B.2-B.3

10.3 Results

10.3.1 Descriptive Statistics and Intercorrelations

The descriptive statistics for conceptual change scores and baseline multiple choice scores between cohorts and programs are shown in Table 10.3. Descriptively, the Pharmacy program is consistently associated with higher mean scores when compared to all other programs. Thus, underrepresentation of pharmacy students in the intervention cohort was important to control for when making comparisons.

10.3.2 Modelling Conceptual Change Scores

The first stage of analysis involved comparing the mean total conceptual change scores between the cohorts after controlling for program and baseline multiple choice scores. A one-way analysis of covariance (ANCOVA) was used for this purpose. There was no strong evidence to violate the assumption homogeneity of variance between cohorts or homogeneity of regression slopes. Residual errors of the model appeared approximately normal. The overall ANCOVA model was statistically significant, $F(5, 321) = 16.63, p < .001$, partial $\eta^2 = .21$. Table 10.4 reports the model parameter estimates. The program, $F(3, 321) = 6.64, p < .001$, partial $\eta^2 = .06$, and baseline multiple choice, $F(1, 321) = 47.29, p < .001$, partial $\eta^2 = .13$, covariates were both statistically significant. After controlling for these effects, a statistically significant mean difference between the control and intervention cohorts was found, $F(1, 321) = 6.017, p = .015$, partial $\eta^2 = .02$. However, the small partial η^2 indicated that the magnitude of the difference was small.

10.3.3 Individual Conceptual Change Question Analysis

The next stage was to drill down into the individual conceptual change questions to explore exactly which conceptual change questions were associated with higher proportion of correct response in the intervention cohort. Figure 10.2 shows that for 14/18 conceptual change questions the intervention cohort was associated with a higher proportion of correct responses. Univariate and multivariate logistic regression models were used to determine if any of these associations were statistically significant. Multivariate

Table 10.3: Descriptive Statistics Across Cohorts and Programs

		Conceptual Change Score				Baseline Multiple Choice Score			
Cohort	Program	*M*	*SD*	*N*	95% *CI*	*M*	*SD*	*N*	95% *CI*
Control	Lab Med	8.53	2.79	36	(7.58, 9.47)	13.58	2.39	36	(12.77, 14.39)
	Pharm Sci	8.24	2.95	42	(7.32, 9.16)	14.14	2.51	42	(13.36, 14.93)
	Biomed Sci	9.41	2.40	37	(8.61, 10.2)	14.14	3.04	37	(13.12, 15.15)
	Pharm	10.80	2.14	45	(10.16, 11.44)	15.18	2.25	45	(14.5, 15.85)
	Total	9.29	2.76	160	(8.86, 9.72)	14.31	2.59	160	(13.9, 14.71)
Intervention	Lab Med	9.85	3.48	20	(8.22, 11.48)	13.80	2.57	20	(12.6, 15)
	Pharm Sci	9.20	2.57	65	(8.56, 9.84)	13.94	2.49	65	(13.32, 14.56)
	Biomed Sci	9.94	3.13	54	(9.09, 10.8)	14.33	2.87	54	(13.55, 15.12)
	Pharm	10.89	2.71	28	(9.84, 11.94)	15.00	2.16	28	(14.16, 15.84)
	Total	9.80	2.94	167	(9.35, 10.25)	14.23	2.59	167	(13.83, 14.62)
Total	Lab Med	9.00	3.09	56	(8.17, 9.83)	13.66	2.44	56	(13.01, 14.31)
	Pharm Sci	8.82	2.75	107	(8.3, 9.35)	14.02	2.49	107	(13.54, 14.5)
	Biomed Sci	9.73	2.86	91	(9.13, 10.32)	14.25	2.92	91	(13.64, 14.86)
	Pharm	10.84	2.36	73	(10.29, 11.39)	15.11	2.20	73	(14.6, 15.62)
	Total	9.55	2.86	327	(9.24, 9.86)	14.27	2.59	327	(13.98, 14.55)

Table 10.4: ANCOVA Model Parameters

Parameters	B	95% CI	SE	t	p	η^2
Control	-0.71	(-1.29, -0.14)	0.29	-2.45	0.015	0.018
Intervention	0^a	-	-	-	-	-
Lab Med	-1.26	(-2.17, -0.35)	0.46	-2.72	0.007	0.023
Pharm Sci	-1.75	(-2.54, -0.97)	0.40	-4.40	< 0.001	0.057
Biomed Sci	-0.93	(-1.74, -0.12)	0.41	-2.27	0.024	0.016
Pharmacy	0^a	-	-	-	-	-
Baseline MC	0.39	(0.28, 0.50)	0.06	6.88	< 0.001	0.128
Control Mean[b]	9.25	(8.85, 9.65)	$N = 160$			
Intervention Mean[b]	9.97	(9.55, 10.38)	$N = 167$			

[a] Dummy coded, [b] adjusted for program and baseline multiple choice scores

models controlled for program and baseline multiple choice scores. Univariate models were included to highlight the possible confounding effects of differences in baseline multiple choice scores and program representation across cohorts.

Table 10.5 reports the parameter estimates for univariate and multivariate regression models predicting correct responses for each of the conceptual change questions. Three out of the eighteen questions were associated with a statistically significant difference in correct responses between cohorts, Confidence Intervals I, Probability and Regression. These effects were consistent across the univariate and multivariate models. Surprisingly, the Confidence interval I question was associated with significantly poorer performance in the intervention cohort. The overall small effect sizes observed in the logistic regression models findings were consistent with the overall small effect identified in the ANCOVA. Analysis of the odds ratio estimates for the multivariate models suggested that corrections for baseline multiple choice scores and program effects slightly inflated the associated effect for the intervention cohort when compared to the univariate estimates.

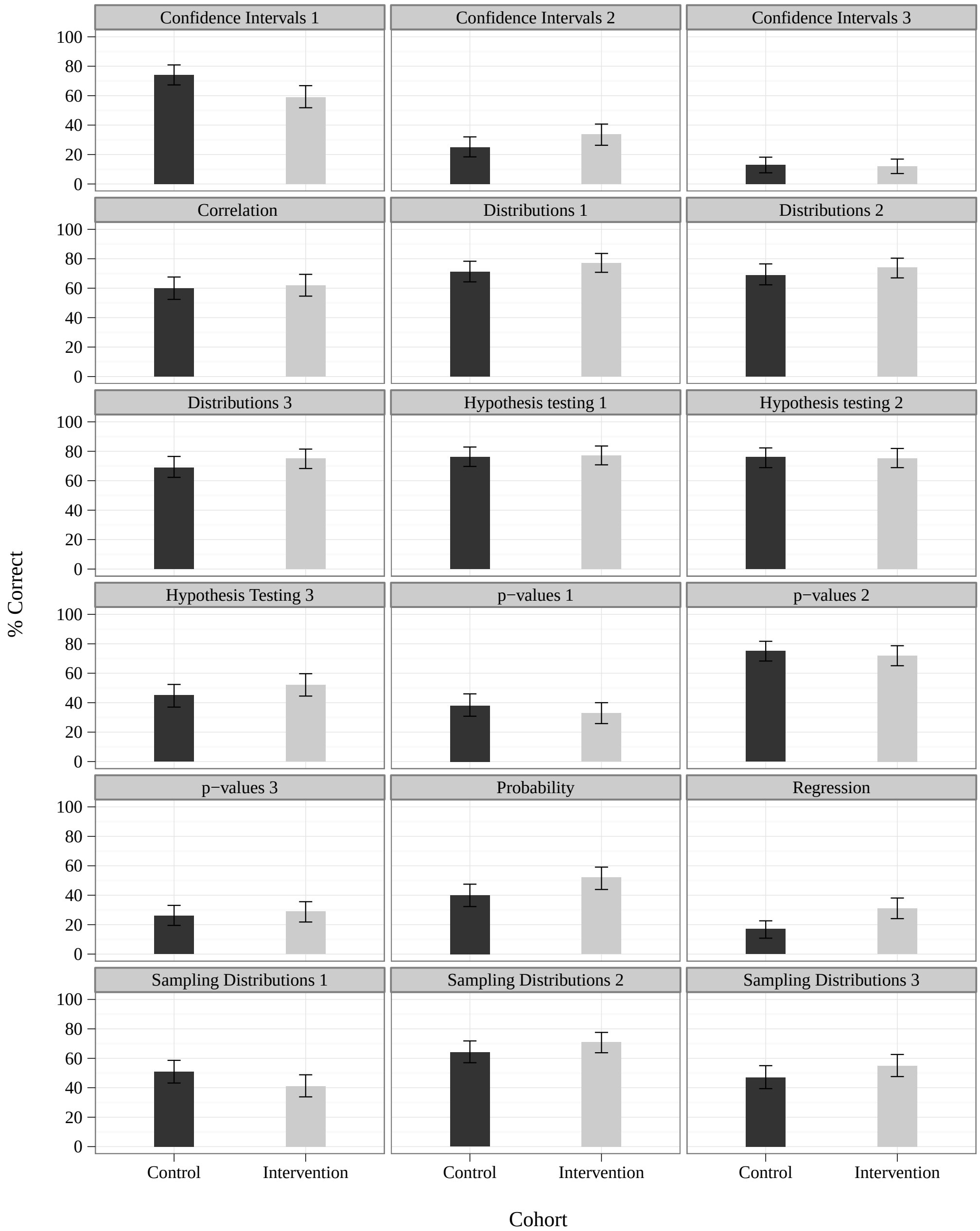


Figure 10.2: Proportion of correct responses between cohorts for each of the conceptual change questions. Error bars show 95% CI for proportions.

Table 10.5: Univariate and Multivariate Logistic Regression Models Predicting Correct Responses on Conceptual Change Questions

	Univariate Models					Multivariate Models[a]				
			95% *CI*					95% *CI*		
Question	B^b	*SE*	Lower	OR	Upper	B^b	*SE*	Lower	OR	Upper
Distributions I	0.32	0.25	0.83	1.37	2.25	0.48	0.28	0.94	1.62	2.80
Distributions II	0.21	0.25	0.76	1.23	2.00	0.28	0.27	0.79	1.32	2.23
Distributions III	0.27	0.25	0.81	1.31	2.13	0.30	0.27	0.80	1.35	2.28
Confidence Intervals I	-0.67**	0.24	0.32	0.51	0.82	-0.62*	0.25	0.33	0.54	0.87
Confidence Intervals II	0.41	0.25	0.93	1.50	2.43	0.48	0.26	0.98	1.61	2.65
Confidence Intervals III	-0.08	0.34	0.48	0.93	1.79	0.15	0.36	0.58	1.16	2.34
Probability	0.47*	0.22	1.03	1.60	2.49	0.59*	0.24	1.13	1.80	2.87
p-values I	-0.24	0.23	0.50	0.79	1.24	-0.20	0.24	0.51	0.82	1.32
p-values II	-0.16	0.25	0.52	0.85	1.39	-0.13	0.27	0.52	0.88	1.48
p-values III	0.13	0.25	0.70	1.13	1.84	0.12	0.26	0.68	1.12	1.86
Hypothesis testing I	0.06	0.26	0.63	1.06	1.77	0.05	0.28	0.61	1.05	1.82
Hypothesis testing II	-0.01	0.26	0.60	0.99	1.64	0.08	0.28	0.63	1.09	1.87
Hypothesis Testing III	0.30	0.22	0.87	1.35	2.08	0.41	0.24	0.95	1.51	2.39
Correlation	0.09	0.23	0.70	1.09	1.70	0.21	0.24	0.77	1.24	2.00
Sampling Distributions I	-0.39	0.22	0.44	0.68	1.05	-0.36	0.23	0.44	0.70	1.10
Sampling Distributions II	0.29	0.24	0.84	1.33	2.12	0.35	0.25	0.88	1.42	2.29
Sampling Distributions III	0.32	0.22	0.89	1.37	2.12	0.40	0.23	0.95	1.49	2.35
Regression	0.82**	0.27	1.33	2.26	3.85	0.87**	0.28	1.38	2.39	4.14

[a] Controlling for program and baseline multiple choice scores, [b] Control = 0, Intervention = 1

* $p < .05$, ** $p < .01$

10.3.4 Conceptual Change Questions Response Distributions for Significant Effects

The conceptual change questions associated with a statistically significant cohort effect were analysed to explore the resulting change in distributions of student responses. The response distributions for all 18 of the conceptual change questions are reported in Appendix B.5. Question response distributions for Confidence Intervals I, Probability, and Regression are reported here.

Confidence Intervals I

Figure 10.3 shows the response distributions between cohorts for the Confidence Intervals I (Q7) question. This question was associated with statistically significantly poorer statistical reasoning in the intervention cohort. Students with a good understanding of confidence intervals should have been able to identify answer (c) as the correct choice. This was the case in 117/158 (74.1%) of the control cohort, but dropped to 99/167 (59.3%) in the intervention cohort. This associated shift in response patterns revealed an increased tendency in the intervention cohort to select option (e). Assuming that the cognitive conflict-based activity for confidence intervals improved students' statistical reasoning, this was an unexpected associated change. It appeared that for some students, the intervention may have introduced an unexpected misconception. On checking responses to the other questions related to confidence intervals, the lack of an associated statistically significant improvement in Confidence Interval questions II and III suggested that overall the cognitive conflict-based confidence interval activity was largely ineffective. This is not surprising given the complexity of confidence interval theory and the well documented difficulty that students have with their interpretation (Fidler, 2006). It is important to note here that the confidence interval questions used as outcomes in this trial were adapted from the original CAOS items. This was done to reflect the course's emphasis on interpreting confidence intervals as estimates which will include the true population parameter a certain percentage of the time through the process of repeated sampling. The course is careful to avoid common interpretations such as "we are 95% confident that the true parameter is captured by this confidence interval" as students' commonly misinterpret the use of the word "confidence" to refer to the probabilistic location of the parameter (i.e. Bayesian Credible Interval) and not confidence in the procedure used to calculate the interval (Albert, 1997). Therefore, students were expected to choose answers that referred in some way to repeated sampling and the long run expected behaviour of confidence intervals.

Question 7. Imagine that there are 100 different researchers each studying the sleeping habits of patients with apnoea. Each researcher takes a random sample of size 50 from the same population. Each researcher is trying to estimate the mean hours of sleep that patients with apnoea get at night, and each one constructs a 95% confidence interval for the mean. With greatest likelihood, how many of these 100 confidence intervals will NOT capture the true mean?

a. None
b. 1 or 2
c. About 5
d. About half
e. 95 to 100

		Responses					
Cohort		a	b	**c**	d	e	Total
Control	N	7	13	117	11	10	158
	%	4.4	8.2	74.1	7.0	6.3	100.0
Intervention	N	9	12	99	17	30	167
	%	5.4	7.2	59.3	10.2	18.0	100.0
Total	N	16	25	216	28	40	325
	%	4.9	7.7	66.5	8.6	12.3	100.0

Figure 10.3: Confidence Intervals I (Q7) Response Distributions

Probability

Figure 10.4 shows the response distributions between cohorts of the Probability (Q17) conceptual change question. In the control cohort 63/158 (39.9%) of students correctly identified answer (a) as the most plausible sequence of sample proportions. In the intervention cohort, this proportion increased to 86/167 (51.5%). The intervention cohort was less likely to pick option (d) which indicated evidence of the equiprobability bias. However, a large proportion of students from both cohorts, 48/158 (30.4%) and 42/167 (25.1%) selected answer (b) which suggested students had a poor understanding of sampling variability.

Question 17. Imagine you have a Petri dish that contains thousands of cells with several different colours created by adding a culture. We know that the culture produces 35% yellow cells. Five students each take a random sample of 20 cells, one at a time, and record the percentage of yellow cells in their sample. Which sequence below is the most plausible for the percent of yellow cells obtained in these five samples?

a. 30%, 35%, 15%, 40%, 50%
b. 35%, 35%, 35%, 35%, 35%.
c. 5%, 60%, 10%, 50%, 95%.
d. Any of the above.

Cohort		Responses				Total
		a	b	c	d	
Control	N	63	48	10	37	158
	%	39.9	30.4	6.3	23.4	100.0
Intervention	N	86	42	12	27	167
	%	51.5	25.1	7.2	16.2	100.0
Total	N	149	90	22	64	325
	%	45.8	27.7	6.8	19.7	100.0

Figure 10.4: Probability (Q17) Response Distributions

Regression

Figure 10.5 shows the response distributions between cohorts for the Regression (Q37) question. This question tested students' understanding of the limitations of regression models and the misconception that regression models can be used to extrapolate beyond the range of the data. In the control cohort, only 26/156 (16.7%) of students identified (c) as the correct answer. However, in the intervention cohort this rose to 52/167 (31.1%). This indicated that the intervention cohort was associated with being less likely to incorrectly extrapolate beyond the range of a regression model.

10.4 Discussion

The aim of this trial was to evaluate the effect of brief cognitive conflict-based activities embedded in lectures on correcting common misconceptions in introductory statistics courses. This trial hypothesised that doing so would be associated with improved statistical reasoning as measured by performance on multiple choice exam questions

Question 37. The number of people living in Australian remote rural areas has declined steadily during the last century. Data gathered on the Australian remote rural populations (hundreds of thousands of people) from 1910 to 2000 were used to generate the following regression equation: Predicted Remote Rural Population = 1167 - .59 (YEAR).Which method is best to use to predict the number of people living in Australian remote rural areas in 2050?

a. Substitute the value of 2050 for YEAR in the regression equation, and compute the predicted farm population.

b. Plot the regression line on a scatterplot, locate 2050 on the horizontal axis, and read off the corresponding value of population on the vertical axis.

c. Neither method is appropriate for making a prediction for the year 2050 based on these data.

d. Both methods are appropriate for making a prediction for the year 2050 based on these data.

Regression (Q37)		Responses				
Cohort		a	b	**c**	d	Total
Control	N	42	31	26	57	156
	%	26.9	19.9	16.7	36.5	100.0
Intervention	N	41	14	52	60	167
	%	24.6	8.4	31.1	35.9	100.0
Total	N	83	45	78	117	323
	%	25.7	13.9	24.1	36.2	100.0

Figure 10.5: Regression (Q37) Response Distributions

relating to misconceptions targeted by the activities. This trial tracked performance on select multiple choice questions embedded in exams for a control and intervention cohort. The results of this trial found a weak statistically significant improvement associated with the intervention cohort after controlling for baseline multiple choice scores and students' program. This weak improvement was evident in only two out of eighteen conceptual change multiple choice questions or two out of the eight cognitive conflict activity interventions. Surprisingly, one question in the intervention cohort was associated with significantly poorer performance.

Cognitive conflict-based activities have been found to have positive effects for correcting common misconceptions about statistics. These studies have demonstrated moderate effect sizes which contrast with the small overall effect demonstrated by this trial. However, there are some major differences between this and previous studies. Previous studies have examined only a few misconceptions in isolation. This trial addressed a wide range of misconceptions throughout an entire semester. While the overall effect size estimate for total conceptual change scores may have been weak, the highest associated effect size estimate observed for the Regression question, $OR = 2.39$, was quite moderate given the brevity of the intervention. Other studies have typically intervened in longer sessions (e.g. Kalinowski et al., 2008; Liu et al., 2010). The overall associated trend suggested that 14 of 18 conceptual change questions were associated with improved, albeit not statistically significant, statistical reasoning in the intervention cohort. This weak trend towards improvement suggests that the brief activities need improvement and further evaluation.

Another major difference between this trial and previous studies is the delivery of cognitive conflict-based activities in lectures. Previous studies have typically embedded conceptual change activities in tutorial sessions or dedicated classes. Compared to lectures, these sessions are more likely to have higher attendance and student engagement as participation is often a course requirement. A lecture, on the other hand, is often not compulsory, nor is student participation a requirement. It is possible that the weaker effect observed in this trial was due to the delivery of the intervention activities during lectures. Anecdotally, course instructors reported consistent and high student attendance rates throughout the semester. The reason the intervention was embedded in lectures was to ensure that the same highly experienced lecturer could deliver the activities to all attending students. Having access to highly experienced tutors or providing the necessary training to new tutors is a serious limitation to delivering many educational interventions. However, investing time in this training to enhance the effect of interventions may outweigh the costs.

This trial also differed from previous research regarding the time to follow-up. A strength of this trial was that it utilised a more valid time frame for follow-up. Previous studies have typically measured outcomes immediately (Watson, 2002a, 2002b, 2007; Hirsch & O'Donnell, 2001; Liu et al., 2010) after intervention. Only Kalinowski et al. (2008) and Jazayeri et al. (2010) have included more meaningful five week follow-up periods. As the cognitive conflict activities in this trial were embedded throughout the semester, the exam follow-up time ranged from 3 weeks for the Regression activity to 12 weeks for the Distributions activity. The short-term follow-up periods used in previous research cannot be used to estimate temporal stability. However, the results from Kalinowski et al. (2008) and Jazayeri et al. (2010) suggest excellent stability up to five weeks for misconceptions relating to sampling variability and the misapplication of the modus tollens argument in hypothesis testing.

The effect of brief lecture-based cognitive conflict activities may be moderated by the complexity of the concept being targeted. The two activities that were associated with a statistically significant improvement, Probability and Regression, were relatively simpler concepts to correct when compared to concepts related to statistical inference (i.e. sampling variability, p-values, hypothesis testing and confidence intervals). This agrees with the observations of Limón (2001) who stated that conceptual change is a gradual process and where dramatic changes are required not much should be expected from only brief interventions. The positive results of Kalinowski et al. (2008) suggest that conceptual change can be achieved for correcting misconceptions related to hypothesis testing using much longer interventions (i.e. 45 minutes). This interpretation suggests that brief lecture-based intervention can be used to correct simpler misconceptions and more intensive tutorial-based interventions left for more complex and pervasive misconceptions. Both methods can serve a useful purpose in the introductory statistics course.

The associated significant reverse effect observed for the confidence interval question highlights the importance of carefully evaluating teaching interventions. Sometimes the interventions that are implemented, no matter how well designed, can have unwanted and unexpected effects. Incorporating the careful evaluation of interventions is an important and necessary step for statistics instructors. With these evaluation data in hand this activity can be reviewed and response patterns between cohorts can be analysed. Hypotheses can be formulated about the cause of the associated negative effect and adjustments made to be followed up in future cohorts.

The major methodological limitation of this trial is the cohort design. Possible cohort effects cannot be ruled out. For example, the intervention group may have simply been a more studious cohort of students. However, including baseline multiple choice scores on questions that were not expected to be influenced by the conceptual change interventions would have helped control for this possible confounding effect. Future studies may aim to evaluate cognitive conflict activities with more controlled experimental designs to further minimise the possibility of such effects.

The multiple choice questions used to evaluate students' conceptual change also have their limitations. This trial made the assumption that the student chose the correct answer because their statistical reasoning was correct and they did not have any major misconceptions. However, with multiple choice questions students can get the right answer for the wrong reason (Jolliffe, 2010). Short-answer questions overcome this issue as students are required to construct their answers, but the downside is the increased marking time and difficulty. Watson (2002a, 2002b, 2007) used interview techniques to great effect to assess students' conceptual understanding following cognitive conflict prompts, but had to rely on transcripts of interviews and specialised grading schemes to evaluate student responses. This method allowed the researchers to gain valuable insight into students' conceptualisation, but would be impractical to implement in a larger trial that evaluated hundreds of students. Both forced-choice and open-ended assessment formats have their place in evaluation research. The researcher advises against relying on the outcomes of one method over the other, but instead a convergence of evidence from both methods is what is required.

The use of the multiple choice questions used in this study also assumed that the questions evaluated statistical reasoning and that the selection of the “correct” answer was a valid measure of statistical understanding. This assumption may have been incorrect for the use of some questions. For example, re-considering Question 17 depicted in Figure 10.4, the use of the term “sequence” is problematic. If the exact sequence was important then the correct answer would not have been (a). However, if a sequence was counted, then option (a) becomes the correct choice because the variability in percentages is more in line with the expectations of sampling in a binomial experiment of this size. In hindsight, there were some issues with the validity of the statistical reasoning questions that may have confused students and led to unreliable measures of their statistical reasoning. The challenge of assessing the outcomes of statistics education remains a challenge, not only for instructors, but also for researchers. Great care must be exercised when selecting outcome measures for interventions even when using standardised instruments promoted as valid and reliable measures of learning outcomes, i.e. CAOS 4 (delMas et al., 2006, 2007).

Plans are in place to further develop, refine and continue the evaluation of the cognitive conflict-based activities in future cohorts. The aim will be to maximise their effectiveness on addressing the misconceptions covered in this trial and eventually will begin to include other misconceptions as needed. With continued monitoring and further optimisation the true potential effect of these activities will be achieved. Future research should also focus on understanding the factors that impact on the effectiveness of cognitive conflict-based interventions. This trial suggests that lecture-based cognitive conflict activities are more suited to correcting simpler misconceptions, whereas intensive tutorial-based interventions should be used to target more pervasive misunderstanding of difficult concepts (i.e. statistical inference). Studies by Liu et al. (2010) show that specialised computer assisted learning may prove to be another effective medium for producing conceptual change via cognitive conflict. The benefit of computer assisted methods is that it could be tailored to both short and intensive formats.

10.5 Conclusion

Poor reasoning about statistical concepts is often precipitated by a student’s misconceptions. If students’ statistical reasoning is to be developed appropriately then these misconceptions require the careful attention of instructors and treatment if necessary. Cognitive conflict has been used effectively throughout science education to correct misconceptions by promoting conceptual change. There is now a growing body of research suggesting that cognitive conflict used in statistics education is no exception. The outcomes of this trial provide further, but somewhat weaker, evidence to support the use of cognitive conflict interventions in the introductory statistics course. Given the widespread prevalence of misconceptions related to statistical concepts, statistics instructors have a high demand for theoretically valid and empirically verified interventions aimed at improving students' statistical reasoning. Furthermore, these interventions must be simple and practical to implement or widespread use will never be achieved. Much more research is needed before the efficacy of cognitive conflict strategies reach consensus, but for now it seems that the evidence is beginning to accumulate.

Part III

Experiential Learning for Developing Statistical Thinking

Chapter 11

Part III - Abstract

Project-based learning (PBL) has been a popular alternate assessment method implemented to actively engage students in statistics education. PBL has been theoretically proposed to enable the development of students' statistical thinking by engaging them in the entire data investigative process of statistical enquiry (MacGillivray & Pereira-Mendoza, 2011). A recent technological development of an online virtual environment, known as the *Island* (Bulmer, 2011), is further evidence of statistics education's increasing interest in PBL. The *Island* simulates a large human population that can be recruited for the purpose of conducting virtual scientific studies. However, the validity of using the *Island* for PBL requires further empirical verification as does the link proposed between PBL and the development of statistical thinking

There are two main objectives to Part III which were addressed in two separate studies, I and II. Study I evaluated student perceptions and experiences of using the *Island* for PBL in an online introductory statistics course (Chapter 13). Study I utilised an explanatory mixed-method design. Forty-two students who enrolled in an online postgraduate introductory biostatistics courses responded to an *Island* questionnaire which rated their level of agreement to three aspects of using the *Island* for PBL - *engagement*, *ease of use* and *contributes to understanding.* Students were also asked to provide qualitative comments and five students participated in semi-structured in-depth interviews. Qualitative feedback was analysed to explain the results from the quantitative questionnaire. The results of the quantitative survey in Study I demonstrated highly positive attitudes towards the use of the *Island* for PBL. Thematic analysis of qual-

itative comments and student interviews revealed that the *Island's* ability to engage students in the data investigative process of statistical enquiry may assist on improving students' statistical thinking.

Study II was an initial attempt to empirically test the proposed link between PBL using the *Island* and students' development of statistical thinking in a large introductory statistics course (Chapter 14). Study II randomly allocated 356 students enrolled in a large introductory statistics course for science students to either an experimental or observational course project using the *Island*. Students worked as individuals or in groups of up to three on a topic of their choosing. During an end of semester tutorial, students completed an open-ended short answer test of statistical thinking about experimental and observational studies. Students' performance on the test's subscales was linked back to their project type allocation. The results of Study II attempted to empirically verify this link by evaluating if project type allocation impacted on students' performance on the experimental and observational subscales of the test of statistical thinking. The results of this analysis found inconclusive evidence of a dependence between students' subscale performance and the types of projects allocated to them.

While students have highly positive attitudes towards the use of the *Island* for PBL, the proposed theoretical impact of PBL on statistical thinking remains to be seen. The assessment of statistical thinking and the implementation of evaluation research in statistics education continues to present major challenges to this important area of research. Future research should continue to evaluate the impact of PBL on the development of students' statistical thinking.

Publications

Reference to works in Part III should cite the following peer-reviewed paper that arose throughout the course of the dissertation. The outcomes from Study I were presented at the 2012 *International Association for Statistics Education (IASE) Roundtable Conference*, held in Cebu, Philippines (Baglin, Bedford, & Bulmer, 2012). Following this conference, a expanded version was invited and accepted for publication into a special edition of the *International Journal of Innovations in Science and Mathematics Education* (Baglin, Bedford, & Bulmer, n.d.).

Chapter 12

Part III - Introduction

12.1 Experiential Learning and Project-based Learning

Introductory statistics courses have adopted a large variety of assessment methods. Traditional exam-based assessment is now typically supplemented by alternate assessment methods such as individual or group projects, oral presentations, portfolios, reflective journals, minute papers, concept maps, written reports, critiques of news reports or articles, tutorial activities, formative assessment quizzes, and assignments (American Statistical Association, 2005; Garfield & Chance, 2000; Garfield & Gal, 1999a). This variety has arisen from the search for assessment practices that promote student learning (Garfield & Gal, 1999a) through active participation (MacGillivray, 2010). Individual and group project-based learning have been popular choices. As MacGillivray (2010) explains, projects aim to provide students with "experiential learning of the whole process of statistical enquiry" (p. 28). Experiencing data collection and analysis gets to the heart of statistics education and actively engages students in processes which connect learning with reality (Snee, 1993; Forster & MacGillivray, 2010).

According to Snee (1993), experiential learning is learning by doing. More specifically, experiential learning can be defined as "the process by which knowledge is created through the transformation of experience" (Kolb, 1984, p. 38). Project-based learning (PBL), not to be confused with problem-based learning, is inherently experiential. PBL is a pedagogical framework designed to engage students in learning through the investigations of authentic problems (Blumenfeld et al., 1991). As the students engage in activities they produce some type of product that aims to address the original question or problem (Blumenfeld et al., 1991). For example, students enrolled in an introductory statistics course might be presented with a research question, e.g. “Do males tend to have a smaller second finger:fourth finger ratio length compared to females?” PBL requires the students to actively gather and analyse data to answer the research question posed. The product of the project might be a report or poster presenting the students’ statistical analysis and findings. Even in this very simple example of PBL, students experience the entire process of statistical enquiry which is claimed to help develop students’ statistical thinking (MacGillivray, 2010; MacGillivray & Pereira-Mendoza, 2011; Snee, 1993).

12.2 Statistical Thinking

Statistical thinking is a difficult concept to define, and there is no single agreed upon definition. After statistical literacy and reasoning, statistical thinking is considered the highest order learning outcome of statistics education (Garfield, delMas, & Zieffler, 2010). Statistical literacy involves a basic understanding of statistical nomenclature and probability as a measure of uncertainty (Ben-Zvi & Garfield, 2005). It also includes the fundamental ability to manage, manipulate and present different representations of data (Ben-Zvi & Garfield, 2005). Statistical reasoning refers to the logic people apply in order to understand and interpret statistical information (Garfield & Chance, 2000). Chance (2002) concluded from a review of the literature that statistical thinking is largely an understanding of what a statistician does. Chambers (1993), and later Cameron (2009), are more specific, listing five categories of work characteristic of being a statistician. These include the following:

1. Preparing data, including planning, collection, organisation and validation
2. Analysing data, by models or other summaries
3. Presenting data in written, graphical or other form
4. Formulating a problem so that it can be addressed through statistical means
5. Carrying out research to develop new statistical methods

Statistical thinking also involves an understanding of research designs, including the need to experiment to establish causation, and how to choose appropriate pre-existing statistical procedures for a study (Ben-Zvi & Garfield, 2005). A good statistical thinker can also use this understanding to critique and evaluate statistical results of studies (Ben-Zvi & Garfield, 2005). Evidently, statistical thinking cannot be thought of as a single construct. Instead, statistical thinking is better understood as the way a statistician problem solves with data. Similar models have been adapted in statistic education to capture the essential features of what is referred to as the data investigative process. Wild and Pfannkuch (1999) adopt the PPDAC (Problem, Plan, Data, Analysis, Conclusions) model proposed by MacKay and Oldford (1994, as cited in Wild & Pfannkuch, 1999) for explaining this problem-solving approach (PSA). Marriott et al. (2009) use a similar framework, but with only four stages: Specify the problem and plan, collect data, process and represent data, interpret and discuss (PCPD).

Wild and Pfannkuch's 1999 statistical thinking paradigm also identify five types of fundamental thinking which they derived from interviews with students and practising statisticians. These types of thinking included recognizing the need for data, *transnumeration*, *consideration of variation*, *reasoning with statistical models*, and *integrating the statistical and contextual*. *Recognizing the need for data* refers to the understanding that data are necessary to meaningfully answer research questions as opposed to anecdotal and subjective experiences which are unreliable and misleading. Pfannkuch and Wild (2005) defined transnumeration as "changing representations to engender understanding" (p. 18). Transnumeration involves the process of gathering appropriate data and then transforming the data into information that leads to the understanding of a phenomenon under investigation. Consideration of variability is an understanding of the omnipresence of variability in data and how this variability leads to uncertainty (e.g. the use of samples). This requires an understanding of sources and types of variability as well as the knowledge to deal with it by ignoring, planning or controlling. Reasoning with statistical models refers to the understanding of the models that statisticians use. Statistical models include obvious methods such as regression, and also include more basic tools used for statistical reasoning such as summary statistics and graphical displays. Statistical models help researchers detect patterns in data amongst the noise of variability. Lastly, integrating the statistical and contextual refers to the ability to synthesise the context of a study with the knowledge gained from statistical models. As statistics captures representations of contextualized reality, the ability to gain knowledge from data requires contextual understanding. Wild and Pfannkuch's model of statistical thinking is the most comprehensive model of statistical thinking proposed and therefore a suitable foundation for guiding its assessment (for a detailed discussion see Pfannkuch & Wild, 1998, 2000, 2005; Wild & Pfannkuch, 1999).

Given the complexity of these definitions, statistical thinking is also challenging to assess. The ARTIST (Assessment Resource Tool for Improving Statistical Thinking) Project website (https://apps3.cehd.umn.edu/artist/index.html) provides enhanced traditional assessment items recommended for the assessment of statistical thinking (see Garfield, delMas, & Zieffler, 2010). These items are exemplars for how traditional assessment methods can be enhanced (Wild, Triggs, & Pfannkuch, 1997), but as Chance (2002) observes "evidence of statistical thinking lies in what students do spontaneously, without prompting or cue from the instructor" (p. 130). This observation suggests that assessing statistical thinking with traditional forced-choice assessment methods (e.g. multiple-choice) might be problematic. Watson (1997) argues that statistical thinking needs to be assessed in an open-ended format as forced-choice questions limit the ability for students to demonstrate their knowledge. Students might also get the right answer for the wrong reason (Jolliffe, 2010). Watson assessed statistical thinking on a hierarchy of skills involving a basic understanding of statistical terminology, the ability to embed the language and concepts of statistics into a wider context and the questioning of statistical claims. Open-ended formats require students to construct their answer which provides explicit insight into their understanding. Open-ended formats appear to be more in line with the types of fundamental statistical thinking proposed by Wild and Pfannkuch (1999). Wild and Pfannkuch's model is reminiscent of a statistical consultant cogitating over the statistical aspects of a project being discussed. What data are needed? How can data be obtained? What sources of variability must be controlled? How will the results of the project be analysed and communicated? What are the limitations? Can they be overcome? Evidence of statistical thinking will be embedded in the asking of these questions and their subsequent answers. This type of thinking is challenging to capture in forced-choice format assessment.

12.3 Project-based Learning for Statistical Thinking

Given that statistical thinking reflects the way a statistician problem solves with data, the most likely way to develop this outcome becomes obvious. However, traditional learning and assessment methods (e.g. lectures and exams), don't lend themselves easily to actively engaging students in the statistician's data investigative process. Fortunately, as MacGillivray and Pereira-Mendoza (2011) explain, project-based learning (PBL) can be effectively used for this purpose. As mentioned previously, Wild and Pfannkuch (1999) adopt the PPDAC (Problem, Plan, Data, Analysis, Conclusions) model proposed by MacKay and Oldford (1994, as cited in Wild & Pfannkuch, 1999) as a model for explaining the problem solving process. The PPDAC model acts as a framework for the delivery of PBL for engaging students in the the data investigative process, which ultimately targets the development of statistical thinking (MacGillivray & Pereira-Mendoza, 2011) and synthesises students' knowledge of statistics for real applications (MacGillivray, 1998). Note that it is possible to use PBL to engage students in only select elements of data investigations, e.g. only steps AC of PPDAC, however, the focus of PBL in this dissertation is specifically on engagement in the entire PPDAC data investigative process. Many statistics instructors have reported on the success of this approach.

Holmes (1997) incorporated a free-choice data collection and analysis project into a statistics course for secondary and college level students. Projects were incorporated due to the dissatisfaction with the outcomes of traditional assessment practices. Holmes reported that the projects helped put statistics in context, improved student engagement, provided students with valuable experience with real data and emphasised the practicality of statistics. G. Smith (1998) modified an introductory statistics course to incorporate a semester-long series of team projects which involved both written and oral reports. Working in teams of three, students completed six mini-projects throughout the semester requiring them to gather and analyse data. For example, one project involved students comparing the average sugar content of cereals displayed on the top, middle and bottom shelves of local grocery stores. Smith found an overwhelming positive attitude towards the use of the mini-projects and an improvement in students' grades on end of semester exams when compared to previous cohorts.

Potthast (1999) evaluated the use of cooperative learning activities on students' understanding of key statistical concepts. Two sections of an introductory course participated in the study. One section completed four cooperative learning experiences, such as, a take-home project on t-tests. The other section of the course did not engage in these activities. Results indicated that students who participated in the four cooperative learning activities scored significantly higher on two out of four mini-tests used as outcome measures for each cooperative learning experience. However, there was no statistically significant difference on average mini-test scores assessing students' understanding of t-tests covered in the project learning experience.

Carnell (2008) examined the effect of student-designed data collection projects on students' attitudes towards statistics. Carnell compared two non-randomly allocated sections of an introductory statistics course. One section completed a course project involving the collection and analysis of student-designed projects. The other section did not complete the projects. Students worked individually or in groups of up to four. A survey of attitudes towards statistics was given before and after the projects. This involved surveying students on the value of statistics, the difficulty of the course, interest in the subject, affect towards the course, perceived level of statistical competence and the amount of effort they exerted. Carnell found no statistically significant difference in changes of attitudes towards statistics between the project and no project groups.

Griffiths and Sheppard (2010) reported on the use of poster presentations of projects completed on a real-world data set. Students worked in groups of four to create a poster presentation that demonstrated key statistical analysis of a large health data set. The authors reported positive student feedback, but the impact of the project on students' understanding of statistics was not evaluated. Griffiths and Sheppard reported difficulties with finding project topics that interest all students. Fortunately, a recent development of an online virtual environment, known as the *Island* (Bulmer, 2010; Bulmer & Haladyn, 2011; Bulmer, 2005), designed for simulating scientific research design and data collection may overcome this problem.

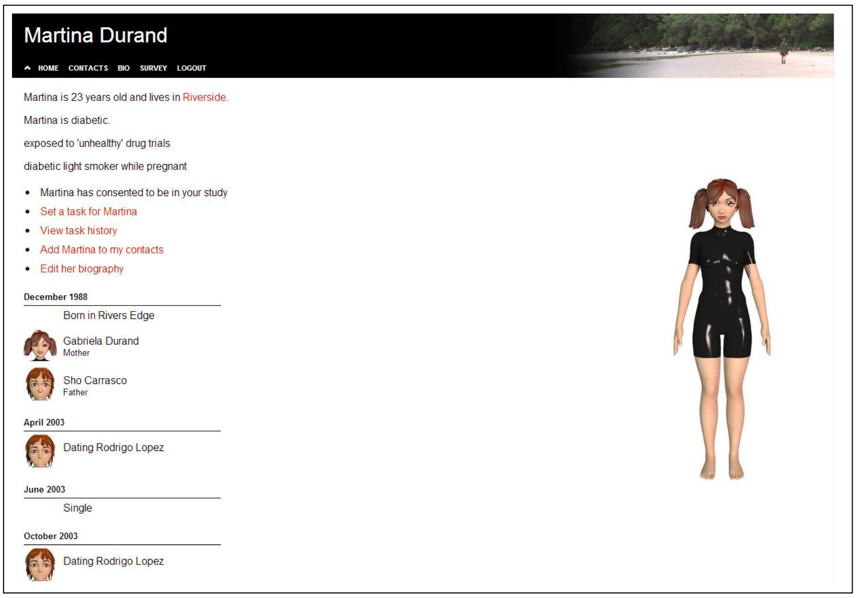
12.4 The *Island*

The *Island* was designed specifically to address the challenges of delivering individualised, authentic, realistic and engaging projects within the constraints of a large introductory statistics course (Bulmer, 2010; Bulmer & Haladyn, 2011). The *Island* (depicted in Figure 12.1) is a freely available online virtual environment accessed via a secure website interface (http://island.maths.uq.edu.au, request a login by emailing island@maths.uq.edu.au). Behind the website runs a complex, real-time, and realistic human population simulation. The *Island* is inhabited with virtual "Islanders" who each have their own unique name, personal history and virtual avatar (see Figure 12.2 a.). Islanders can be sampled and recruited for the purpose of scientific research by navigating between 39 towns (only 36 are shown on the map, see Figure 12.1). Each Islander occupies a house in these towns (see Figure 12.2 b.).

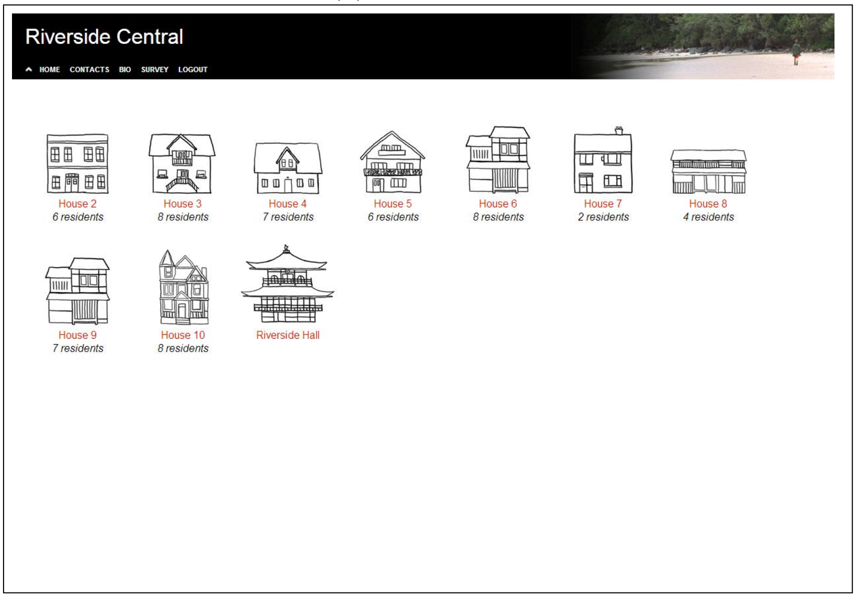
The current *Island* comprises of two different simulations. The first simulation seeded the current population from an initial shipwreck of 108 people in 1779. This simulation proceeds in monthly steps and probabilistically determines disease contraction, death, relationships (e.g. dating and marriage), pregnancy and relocation. Approximately 15,000 Islanders have existed (both living and dead) over the entire history of the simulation. At the time of publishing, the estimated population is in excess of 9,000. The town halls store information about birth, deaths and marriages. This archival information is perfect for epidemiological studies.

The second set of simulations control the various types of data that can be collected from the *Island*. These data are obtained by setting tasks for consenting Islanders. There are now in excess of over 200 different tasks available (See Figure 12.3 a. and b.). Task categories and examples include survey items (e.g. "How anxious do you feel right now?"), blood tests (e.g. cholesterol, glucose, and type), physiological measures (e.g. blood pressure, pulse rate, and spirometer), alcoholic drinks (e.g. red wine, beer and vodka), non-alcoholic drinks (e.g. green tea, water and coffee), food (e.g. chocolate, carrots and banana), injections (e.g. adrenaline, methamphetamine and morphine), tablets (e.g. aspirin, codeine and vitamin D), other drugs (e.g. cigarette, reefer and betel nut), mental tasks (e.g. IQ test, memory test and mental arithmetic),

Figure 12.1: The *Island* (Bulmer & Haladyn, 2011)



(a) An Islander



(b) The town and houses of Riverside

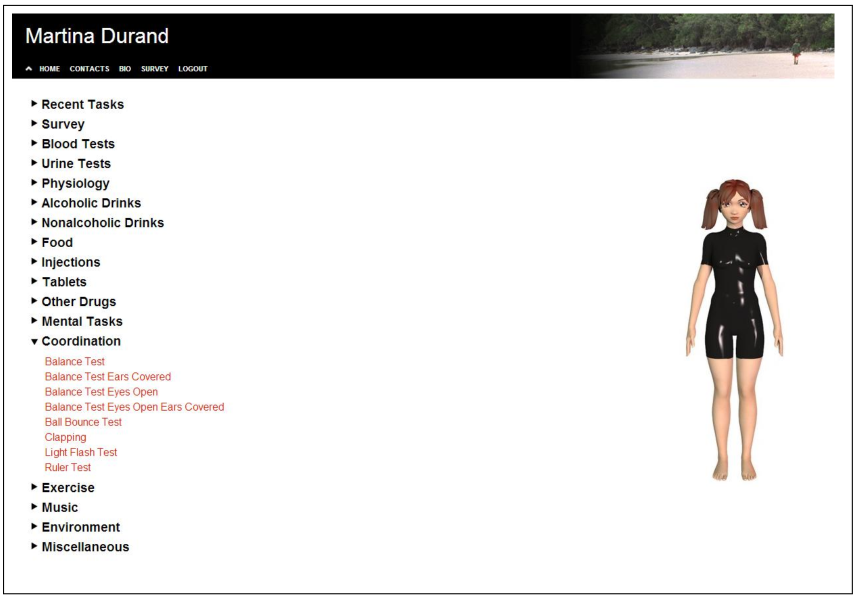
Figure 12.2: Town and Islanders

coordination (e.g. balance test, ruler test and light flash test), exercise (e.g. swimming, running and strength test), music (e.g. classical music, heavy metal music and play flute) and environment (e.g. nap, read book and sit). Biographical information for each Islander includes demographic information (e.g. age, gender, residency), medical records (e.g. smoking history, disease diagnosis), family tree and relationship history. The task simulations run in real time and most are based on mathematical models built from scientific literature. For example, Bulmer and Haladyn (2011) report there are statistical models governing the effect of caffeine on exercise, alcohol on blood pressure, ageing on body temperature, oxygen on cognitive performance, obesity on cholesterol, sleep on mental tasks, and smoking on blood pressure.

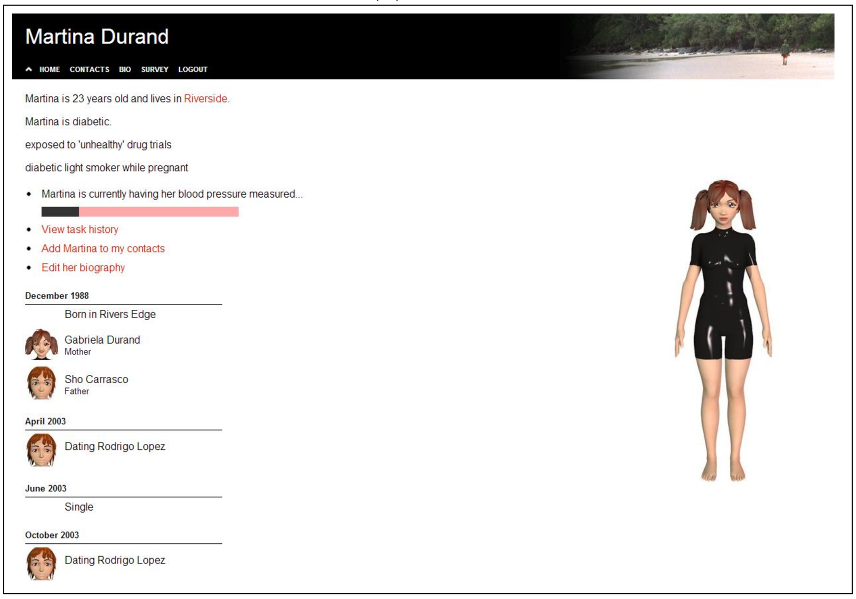
There are a number of key features to the Island that make it ideal for PBL in statistics education. The wide range of tasks and demographic information available on the *Island* allows students to self-select topics of interest to them. It also allows students to design and implement of wide variety of research designs including surveys, observational studies, case-control studies, correlational studies and experiments. The *Island* has been designed to give students an authentic research experience. Islanders may refuse consent, drop out during an experiment, lie about their age, get sick or fall asleep late at night. The *Island* does not provide a way to automatically sample Islanders. Thus, students must deal with the issues of sample size and sample selection. Interactions with Islanders cannot be automated. Students quickly realise the “cost” of research with the cost being students’ time. The *Island* does not provide students with data files or summarised data. Nor does it provide tools for data analysis. The aim here being to provide students with the experience of gathering raw data and preparing data files for analysis as they would in real-world research.

12.5 Rationale and Aims

While virtual simulation software aimed at enhancing student learning has been used in a wide variety of disciplines including statistics (e.g. Neumann, Neumann, & Hood, 2011), public health (e.g. Spinello & Fischbach, 2004), ecology (e.g. Stafford, Goodenough, & Davies, 2010), physiology (e.g. Dobson, 2009), and biology (e.g. Lin & Lehman, 1999), the *Island* is a relatively new instalment for statistics education and



(a) Tasks



(b)Measuring blood pressure

Figure 12.3: The *Island* Interface

distinguishes itself with its ambitious aim to realistically simulate an entire human population for the purpose of delivering project-based assessment in large classes. While Bulmer (2010) reported positive student feedback using the *Island* in a large introductory statistics courses and Linden, Baglin, and Bedford (2011) reported similar results from a course in the design and management of clinical trials, further studies are needed to validate the use of the *Island* for PBL in other educational contexts and student populations. Also, few studies have investigated establishing a link between PBL and students' development of statistical thinking. While students' perception of PBL is an important consideration, the proposed learning benefits of PBL require empirical verification. Theoretically, these methods of learning and assessment should work, but statistics education research must continue to rigorously evaluate its practices (delMas, 2002). Consequently, the aims for Part III were as follows:

1. To evaluate student perceptions and experiences of using the *Island* for PBL focusing at developing and assessing statistical thinking in an online introductory statistics course. While this is an indirect method for evaluating the effectiveness of the *Island* for PBL, understanding students' perceptions is an important initial step.

2. To empirically test the proposed link between PBL using the *Island* and students' development of statistical thinking in large introductory statistics courses. This would validate existing theoretical beliefs about the role of PBL in statistics education and also shed light on the development of statistical thinking itself.

Chapter 13

Part III - Study I

13.1 Aim of Study I

The aim of this study was to evaluate student perceptions and experiences of using the *Island* for semester long projects designed to develop and assess statistical thinking in an online introductory statistics course for masters' students. While this was an indirect method for evaluating the effectiveness of the *Island* for PBL, it served as an important initial step that was built upon in Study II.

13.2 The Course

The course in this study was an online introductory biostatistics course. It is largely taken by Masters of Laboratory Medicine students, a majority of whom are international temporary on-shore students. Other students that typically enrol in the course include students from Masters by Coursework of Statistics and Operations Research, Medical Science and Biotechnology. The course had been growing in popularity over the years. Masters students, who often have family and work commitments, are attracted by the flexibility of the online delivery. The course covered the usual introductory topics including descriptive statistics, probability, estimation, one-sample inference, two-sample inference, categorical data, non-parametrics, correlation and regression, basic epidemiology and one-way ANOVA. The course assessment was broken up into three parts: weekly on-going assessment, online tests and a major course project. The ongoing assessment (10%) consisted of weekly exercise submissions to ensure students are working through weekly content. The tests, which make up 60% of the course grade, involved a mid-semester test (15%), late semester test (15%) and final test during the exam period (30%).

In the years prior to 2011, the projects (weighted 30%) required students to find available data sets, either from their workplace or the internet, in order to complete a project demonstrating the application of statistical thinking using knowledge gained in the course. The inclusion of these projects aimed to enhance a student's statistical thinking by getting them to "do" statistics, i.e. experiential learning. The project was split between a research proposal due mid semester (5%) and development of a project presentation summary slideshow due at the end of the semester (25%). Students had the option to audio or video record commentary for the presentation. However, only a few students did so. Project presentations were marked utilising a rubric which rated students on levels of achievement (unacceptable, needs improvement, good and superior) across the following five criteria: 1) Topic Background, Rationale and Research Question, 2) Method, 3) Statistical Analysis and Presentation of Results, 4) Discussion and Conclusion, 5) Professionalism (spelling, grammar, references and visual appearance).

Project-based work prior to 2011 had been problematic. Approximately half of the students each semester were unable to find suitable data sets. To avoid disadvantaging these students, a number of pre-existing large biomedical data sets were provided. This created issues with authenticity, the possibility of collusion, and poor student engagement. By using pre-existing data, the students were also missing out on the planning and data collection stages of the PPDAC model of statistical enquiry. A better approach would involve conducting scientific research from the ground-up, from planning right through to data collection, analysis and reporting. However, doing so within the constraints of the online course was inconceivable prior to the *Island*.

Island-based projects replaced the pre-existing projects in both semesters of 2011. While students were still allowed to analyse data from their workplaces, this was only allowed with permission from the course lecturer. Remarkably, only one student in 2011 took up this offer. The *Island*-based projects required students to investigate a research topic of their choosing in order to demonstrate the application of a statistical technique covered in the course. The *Island* gave students access to an environment allowing them to choose from a large variety of topics whilst ensuring that each student's data was individualised and available online. The *Island*-based projects would also give students the experience of conducting an entire cycle of a simulated scientific study. Examples of the topics chosen by students are listed in Table 13.1. The topic diversity reflects a large degree of variability in what students perceived answerable in light of the data available. A wide variety of research designs were employed, including correlational, observational and experimental designs.

Table 13.1: Eight Examples of Student Project Topics

Project Title
Short Term Effects of Caffeine from Cola on Mental Acuity
Murder and Relationship Instability
The Effects of Eating Habits on Blood Pressure in Adults
The Relationship Between Sleep and Wellbeing
Association between Blood Type and Disease Mortality
Comparison of Natural and Synthetic Insulin
The Effect of Cocoa on Sensory Memory
Effect of Exercise on Anxiety and Endorphin Levels

13.3 Method

This study was funded by the RMIT College of Science, Engineering and Health 2011 under the Scheme for Teaching and Learning Research (STeLR). The grant used to fund this study also included the evaluation of the *Island* in a clinical trials course. Outcomes of this part of the grant are reported in Linden et al. (2011). This study refers only to the use of the *Island* for PBL in the online biostatistics course. Ethics approval for this project was provided by the RMIT College Human Ethics Advisory Network on the 22^{nd} December 2010 (Project No: A&BSEHAPP 87-10). A sample of 42 students from the Semester 1 and 2, 2011 offerings of the introductory biostatistics course participated in the evaluation of the *Island* project-based assessment. These students were recruited through email invitations sent at the end of the semester inviting them to complete an online questionnaire. Online versions of the study's plain language statement and consent forms are presented in Appendix C.1 and C.2). The participation rate across the semesters was 18/35 (51%) for first semester and 24/43 (56%) for second semester. The average age of the sample was 29 years ($SD = 3$). There were 15 (35.7%) males and 27 (64.3%) females. The sample was mostly on-shore international (28/42, 66.7%) students studying full-time (33/42, 78.6%).

An explanatory sequential mixed methods approach was used for evaluating student perceptions (Creswell & Plano Clark, 2011). This type of design involves first gathering quantitative data and then following up with qualitative methods to explain the quantitative results. In the quantitative phase of the research, students responded to an 18-item online questionnaire designed to evaluate student perceptions of using the *Island* (see Appendix C.2). Three specific aspects of using the *Island* were assessed using this questionnaire - *engagement*, *ease of use* and *contributes to understanding*. Each item was responded to on a seven point Likert scale ranging from (1) strongly disagree to (7) strongly agree. Agreement to an item was defined as a participant scoring an item as a 5, 6, or 7. Reliability of each subscale was measured using Cronbach's α which found that α =.79, .62 and .90 for engagement, ease of use and contributes to understanding respectively.

Following the quantitative questionnaire, two open-ended questions were included for qualitative feedback. These questions were (1) "Share at least one positive experience of using the *Island*" and (2) "Was there anything that you did not like about using the *Island* or you think needs improvement?" The second, qualitative phase used qualitative comments given in the questionnaire and five semi-structured in-depth interviews to assist in explaining the results of the quantitative questionnaire (see Appendix C.3. The interviews were conducted over telephone with five volunteer students. Qualitative comments and interview data were analysed using thematic analysis (Braun & Clarke, 2006). This method involved six steps: data familiarisation, initial coding, theme searching, theme revision, theme definition and naming, and reporting.

13.4 Results and Discussion

The descriptive statistics of the quantitative responses to the *Island* questionnaire are shown in Table 13.2. These quantitative results will be discussed alongside themes identified in the qualitative thematic analysis to help explain and expand upon the forced-choice responses. The themes will be discussed around the three domains of the *Island* questionnaire, engagement, ease of use and contributes to understanding.

Table 13.2: *Island* Questionnaire Descriptive Statistics (Both Semesters Combined)

Scale	Item	*M*	*SD*	Agree	%
Engagement (Cronbach's α = .79)					
	Enjoyed using for project	5.93	1.02	40	95.2%
	Enjoyed being in control of virtual study	5.71	1.11	37	88.1%
	Did not enjoy using for projects (R)	2.43	1.40	5	11.9%
	Felt immersed in virtual study	4.86	1.32	25	59.5%
	Recommend to other students	5.71	1.38	36	85.7%
	Positive experience overall	5.88	1.38	38	90.5%
Ease of Use (Cronbach's α = .62)					
	Easy to use	5.62	1.21	39	92.9%
	Difficult to use (R)	3.48	1.80	11	26.2%
	Learning to use was difficult (R)	2.21	1.26	4	9.5%
	More instructions needed (R)	4.45	1.80	24	57.1%
	Easy to conduct virtual scientific studies	5.48	1.29	34	81.0%
Contributes to Understanding (Cronbach's α= .90)					
	Better understanding of scientific research design	5.43	1.33	33	78.6%
	Appreciation for practical consideration of scientific research	5.55	1.31	35	83.3%
	Improved understanding of how data are collected	5.43	1.40	33	78.6%
	Better understanding of statistical analysis in scientific research design	5.50	1.44	35	83.3%
	Improved confidence with design, implementation and analysis of scientific studies	5.31	1.39	33	78.6%
	Experience with statistical issues that arise during research	5.76	1.30	36	85.7%
	Improved understanding of how scientific studies are analysed	5.74	1.25	36	85.7%

Note. N = 42, R = reversed item

The results from the *Island* Questionnaire showed a remarkable overall positive perception of using the *Island* for course projects (Table 13.2). For example, 38/42 (90.5%) of students agreed that using the *Island* for projects was an overall positive experience. Qualitatively, when eliciting from students the reasons behind the positive experience, the major theme that emerged was the *Island's* ability to immerse students. Two major themes emerged to explain this engagement – *realism* and *contextualisation*. By far the most powerful feature of the *Island* that appeared to immerse students was the *Island's* realism, "*It feels like a real Island*". The realism was aided by the *Island's* open-endedness. Students appreciated the wide range of tasks available that allowed them to individualise their project topics, although some students requested further additions. Students also liked how Islanders realistically reacted to various treatments which were the topic of their scientific studies, "*It was fun to see how individual 'islanders' reacted to the various tasks, and the selection of tasks available was extensive.*"

The realism behind the *Island* is an important feature, however, the *Island* is no substitute for "real" research experience. Bulmer and Haladyn (2011) discuss the tension between the *Island's* reality and fantasy. While many aspects of the *Island* are eerily realistic, many others are not so. For example, the *Island* population demographics does not reflect a real-world population, Islanders can live to unusually old ages, simulated models governing the effects of tasks (e.g. taking drugs) are not perfect, and many proposed models are yet to be implemented. Students sometimes express concern about not finding an expected association known to exist in the real world and question whether this will impact their grade. This point provides the perfect opportunity for instructors to discuss with students the nature of science, the importance of reporting false findings (file drawer effect), sampling variability, statistical power and scientific replication. The *Island* can act as a bridge between the artificial classroom environment and real-world research.

The *Island's* ability to contextualise the theory being covered in the course was also a very powerful way to captivate students in PBL. One student summarised this perfectly as follows:

> I didn't enjoy [Introductory Biostatistics] (I found it a chore) until we got to the Island: Suddenly I had a problem, and to solve it I had to learn about study design, sampling and sample sizes, statistical power, statistical methods etc. It was no longer a chore, but a mission.

This student may otherwise never have been engaged in the course had it not been for the use of *Island*-based projects. This response suggests the link between engagement with the *Island*-based projects and its impact on students' statistical thinking.

In terms of ease of use, there were some mixed perceptions. While students felt the *Island* was relatively easy to use (39/42, 92.9%), conflictingly, about a quarter (11/42, 26.2%) of students also reported that the *Island* was difficult to use. The fact that most students agreed that more instructions were needed (24/42, 57.1%) provides some explanation for this inconsistency. However, qualitative themes offered further explanation. Students agreed that using the *Island* made conducting scientific studies possible within the course, "*Using the Island I had the opportunity to conduct a full research without having the classical real problems which normally interfere with it (like costs and time)*". This theme related to ease of use was labelled *facilitates virtual studies*. On the other hand, a second theme, *time inconvenience*, revealed students felt that aspects of using the *Island* were too time consuming, "*Having to wait in 'real time' for data gathering is a bit frustrating - a bit too realistic!.*" Others suggested ways to overcome this by using task automation, "*It would have been great if we could schedule tasks in advance and the islanders then carry them out as per the schedule. It took me a lot of time having to manually instruct islanders to carry out a regular task.*" A few students also criticized the Islander's sleeping patterns, "*It took a very long time to administer the tasks I wanted, especially when islanders go to sleep at around 10.30pm!*". In summary, students felt that the *Island* made research a virtual reality; however, certain aspects of using the *Island* were perceived as being an unnecessary time nuisance.

Bulmer and Haladyn (2011) explain that the *Island's* ease of use is limited in many ways, but only by deliberate design. Bulmer and Haladyn wanted the *Island* to not only simulate a human population, but also simulate what it is like to conduct scientific research. They wanted students to experience recruitment, sampling, experimentation, data collection, data entry and statistical analysis, i.e. the PPDAC cycle. While they are quick to point out that *Island* research is still far easier than real world research, they do contend that the *Island* acts as an intermediate method of connecting research with statistical analysis. In the authors' opinion it would be a disservice to students to build the expectation that data collection is convenient and instantaneous. It would degrade the real world experience aspect of the *Island*. Regardless, instructors, who are probably all too aware, should anticipate that some students will not relish the hard work of gathering realistically simulated data.

Overall, there was vast agreement in students' perception that project-based work on the *Island* had a positive impact on students' understanding of scientific research design, data collection, and statistical analysis, i.e. their statistical thinking. Encouragingly, 36/42 (85.7%) of students agreed that using the *Island* for project-based work had improved their understanding of how scientific studies are analysed. Qualitative responses provided clues as to how the *Island*-based projects may have assisted. Many respondents expressed the view that the *Island*-based projects improved their understanding by putting statistical analysis within a context or by helping them to "*apply what has been learnt*". This sub-theme of contributes to understanding was labelled *learning by doing*. The projects also helped students in thinking about the bigger picture of statistics in scientific research, "*It gave a whole rounded picture of the collection of your data set*". The *Island* gave them an appreciation for practical issues, e.g. time, and the difficulties that can arise. The *Island* helped put statistical analysis in perspective and in doing so, students seemed to gain a deeper understanding, "*I got a chance to understand my statistics and I used what I've learned on the Island. I think it is a great experience having time on that wonderful place. I really recommend the Island for new students to conducting further research with different topics.*" This theme was called *putting it all together*. One particular student also believed that the *Island* had improved their confidence in their ability to conduct scientific research. Before using the *Island*, this student explained that they were dreading the commencement of their Master research project. However, after one project on the *Island*, the student admitted that they were now looking forward to getting started.

Not all students seemed to benefit. One highly experienced student working in the marketing industry found the *Island*-projects of no direct benefit. They explained that the concepts and activities completed in the *Island* projects encompass what they do on a day-to-day basis. This drawback may be re-interpreted as validation of the real-world applicability of *Island*-based projects. A few students appeared to have missed some important points. For example, one student was surprised when they unknowingly experienced natural biological variability, "*sometimes the participants change their answers at the same day. For example; when you ask about cholesterol; the result will be for the first time 155 and the second time will be 160 or something*". Another student expressed disappointment that not all Islanders wanted to fill out their survey, "*Some people in the villages don't do the survey*".

From the instructor's perspective, the use of *Island*-based projects had a number of benefits. Individualisation of topics created great diversity, whereas in the past, diversity was lacking. This made marking the projects far more enjoyable, but somewhat more difficult to compare between students. Clear marking rubrics were helpful in this respect. The *Island*-based projects felt more authentic due to the individualisation and diversity of topics. Student activity logs available to instructors from the *Island* made it possible to confirm students had collected the data presented in their projects. The students' data sets were also a good source for examples and assessment items to be used in the future. From an assessment perspective, the projects provided unique insight into the students' ability to think statistically by getting them to carry out scientific research design and analysis from the ground-up.

The results reported in this study on students' experience and perceptions of using the *Island* for project-based assessment in an online introductory statistics course suggest that students perceived using the *Island* as being engaging, relatively easy to use and beneficial to the development of their statistical thinking. A limitation to this conclusion was the response rate. A positive response bias cannot be ruled out. However, these results were consistent with findings from a similar study by Linden et al. (2011) which used the same questionnaire and had a 91% response rate. The qualitative comments used to explore the students' experience were obtained from the qualitative questions in the questionnaire and through five semi-structured in-depth interviews. As these comments were provided by volunteers, the extent to which these comments represent all students is unknown.

The results of this study suggest that the *Island*, in and of itself, does not develop a student's ability to think statistically. The *Island* acts as a virtual playground for students to experience the PPDAC cycle. It is through this experience of learning by doing that students become motivated to question, learn and understand the statistical concepts related to what they are doing. This is how *Island*-based projects are hypothesised to help develop students' statistical thinking. This study suggests that multiple design factors of the *Island* work together to achieve the level of engagement required to facilitate this development. In conclusion, according to students perceptions the *Island*-based projects were a valid approach to the delivery of PBL in an online introductory statistics course. This finding was also consistent with other studies on *Island*-based projects (Bulmer, 2010; Linden et al., 2011).

13.5 Conclusion

Despite these positive findings, more research is required for understanding how *Island*-based projects can improve assessment methods and student learning outcomes. Studies which map specific learning outcomes to the use of *Island*-based projects would validate the proposed education benefits of its use and lead to a better understanding of the development of statistical thinking in the introductory statistics course. A second study was designed to address this aim.

Chapter 14

Part III - Study II

14.1 Aim of Study II

Study I found overall high student satisfaction in using the *Island* for PBL. Qualitative evidence suggested that the *Island's* ability to engage students in statistical enquiry may have a positive impact on the development of students' statistical thinking. Therefore, the aim of Study II was to attempt to empirically associate project-based learning outcomes with measures of statistical thinking outcomes in an introductory statistics course. If the *Island*-based projects develop students' ability to think statistically then differences would be expected to exist between students who engage in different types of projects. This study manipulated the type of research project a student conducted on the *Island* in order to see if it would have an impact on their ability to think statistically about their project type relative to students who engaged in other types of projects. Specifically, this study allocated students to conduct either experimental or observational study designs. Observational study designs also included correlational designs. The major difference between these types of designs boils down to the deliberate manipulation of an independent variable. Each type of research design requires a unique type of statistical thinking.

It was hypothesized that students' performance on a test of statistical thinking about experimental and observational study scenarios would depend on the type of projects completed on the *Island*. Specifically, students allocated to experimental projects would outperform students who completed observational studies on statistical thinking about experiments and vice versa for statistical thinking about observational studies.

14.2 Method

14.2.1 The Course and Participants

This study obtained institutional ethics approval from the University of Queensland's Behavioural & Social Science Ethical Review Committee on 25^{th} January 2012 (Project No. 2011001393). The study was embedded within a large introductory statistics course for undergraduate science students. The 12-week course was composed of three hours of lectures and two hours of tutorials each week. Topics covered in the course include design of experiments and ethical research, exploratory data analysis, probability, and statistical inference. Assessment included weekly quizzes (15%), a paper review (15%), a major project (20%) and an end of semester exam (50%). The course was offered over two semesters. This study was conducted in the first semester offering. Project-based learning was used in the course to engage students in the PPDAC cycle with the aim of developing their statistical thinking. Projects were completed individually or in groups of up to three. While students were allocated a type of study design to use for their project, i.e. experimental or observational, students were permitted to propose their own research topics which was enabled by the open-ended nature of the *Island*. The 20% weighting for the project was split into 5% for a short research proposal, submitted at the end of Week 7, and 15% for a report in the style of a conference abstract that gave a summary of their methods and results, submitted at the end of semester. Both the proposals and the reports were marked by tutors with feedback provided.

Students were approached to participate in this study during a regular tutorial following a lecture on ethics in scientific research. An online version of a participant information sheet (see Appendix C.4) and consent form (see Appendix C.5) was provided to students. The recruitment was deliberately conducted in this tutorial to provide students with an illustration of obtaining informed consent. However, in retrospect, this may have impacted negatively on the overall participation rate as the ethics lecture content covered many examples of unethical scientific conduct. This may have made students extra cautious of the level of risk posed by this study. There were a total of 574 students enrolled in the first semester course of which 356 (62%) consented to have their data recorded. After students chose an individual project or formed groups they were randomly allocated to complete either an observational or experimental project on the *Island*. Following allocation, students were permitted to change groups prior to submitting their project proposals. This created an unanticipated imbalance in the proportion of experimental and observational projects. A much higher proportion of students swapped from observation to experimental projects (48/160, 30%) rather than vice versa (14/196, 7%, see Figure 14.1). Of the 367 consenting students, 126 (35%) completed observational projects and 239 (65%) completed experimental projects. Of the 126 students who completed observational projects, 102 (81%) and 103 (82%) finished the study by completing the observational and experimental subscales of the Test of Statistical Thinking respectively (TST, see Figure 1). Of the 230 students who completed experimental projects, 186 (81%) and 190 (83%) completed the TST observational and experimental subscales (see Figure 14.1).

14.2.2 Test of Statistical Thinking

A Test of Statistical Thinking (TST) was developed to measure students' statistical thinking about experimental and observational study designs. The TST was completed online by all students in the course during a tutorial session towards the end of the semester and following completion of the *Island*-based projects. When designing the TST exercises, the first step was to define the proposed learning outcomes associated with engagement in the *Island*-based projects. These outcomes were linked to Wild and Pfannkuch's types of thinking (see Table 14.1). Chance's 2002 assessment mantra, "assess what you value" (p. 10) was also kept in mind. This meant that the TST would assess the most pertinent outcomes of the course.

The final version of the TST (see Appendix C.6) used in this study presented students with two research scenarios, one relating to an observational study and one relating to an experimental study. The observational study explored the association between high protein diets and body fat percentage and the experimental study explored caffeine consumption and attention in lectures. Each scenario required students

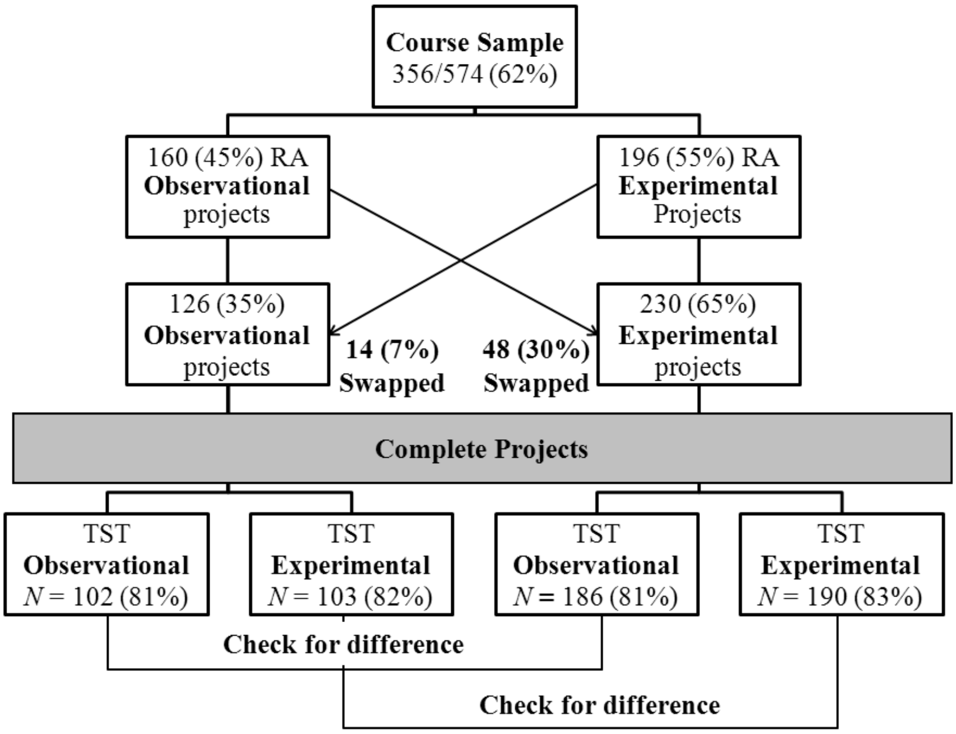


Figure 14.1: Study flow-chart. RA = Randomly allocated, TST = Test of Statistical Thinking

Table 14.1: Mapping Wild and Pfannkuch's 1999 Fundamental Components of Statistical Thinking to *Island*-based Project Outcomes

Type of Thinking	*Island*-based Learning Outcome
1. The recognition of the need for data	Students will understand the role of data in answering research questions by engaging in the data investigative cycle (PPDAC)
2. Transnumeration - Identifying and transforming appropriate data into representations of a model that leads to understanding. This occurs at multiple stages – obtaining data to answer a research question and transforming data (e.g. descriptive statistics, plots) to convey meaning and understanding (Pfannkuch & Wild, 2000)	The *Island*-based projects require students to design a study and decide how data can be gathered to address a specific research question. Students must transnumerate the variables being investigated. This requires students to judge the appropriateness of measures selected both from a perspective of reliability/validity and practicality (e.g. time). Students use their data to convince others of their finding. This leads them to explore ways to best represent their data in a meaningful way that illustrates their findings (e.g. descriptive statistics and graphical displays).
3. Consideration of variation - Knowledge and understanding comes with uncertainty due to the omnipresence of variation.	By collecting real data, students experience the issue of drawing inferences about populations using samples. Sample variability leads to uncertainty and requires the use of statistical models to find signals in the presence of noise. Hypothesis testing is a statistical method used to judge the presence of a signal amongst noise when taking random samples from a population. Students learn the importance of this consideration by conducting hypothesis testing on their data.
4. Reasoning with statistical models - Understanding of statistical models, how they relate to research design, and how they contribute to understanding	The *Island*-projects require students to develop a research design that will address a particular research question. Students must relate their data to a suitable statistical model that will allow the student to address the research question. Different research designs/types of data require different statistical models. Students learn to apply the proper statistical models in different research scenarios. Students also learn how to reason with different models within the context of their projects.
5. Integrating the statistical and contextual - Integrating and interpreting statistics within the context of the problem	Students apply statistical models within a research context. These statistical models aim to address the original research question. Students learn to synthesise statistical analysis within the research context in order to summarise what knowledge has been gained (Pfannkuch & Wild, 2005).

to propose the design and analysis of a study by addressing six fundamental questions that were associated with Wild and Pfannkuch's (1999) fundamental statistical thinking types. This involved selecting a sample, defining variables, identifying appropriate summary statistics and graphical displays, proposing a suitable statistical test based on the nature of the variables selected, justifying the use of hypothesis testing and anticipating expected results if a positive outcome was found. Each question used a short-answer format. While it would have been more economical to use a multiple-choice format, the aim was to have students construct their own answers, similar to what they were required to do for the projects and in line with the recommendations of Smith (1998).

A marking scheme was developed to aid in grading student responses (see Appendix C.7). The scheme marked students on a scale ranging from High (3 points) to Poor (0) for each question. All attempts by consenting students were marked by the same assessor who was blinded to the students' project allocation. At the same time a regular tutor in the course used the same scheme to mark attempts by non-consenting students. Following grading, all students were given feedback on their responses and provided with exemplar responses to compare with their answers. This study only reports the data from the 62% of consenting students.

A Principal Components Analysis (PCA) using an eigenvalue greater than one approach for component selection was performed on the 12 questions of the TST. Varimax rotation was used for component rotation. The aim of the PCA was to test whether the questions loaded into experimental and observational subscales. The PCA extracted four components which explained a total of 70.66% of the variability in TST scores (Table 14.2). The first component, labelled *Experimental*, explained a total of 21.36% of the variability in TST scores and was composed of questions 2, 3, 4 and 6 of the experimental scenario. The second component, labelled *Observational*, was composed of questions 2, 3, 4 and 6 from the observational scenario and explained a further 21.21% of variability in TST scores. The third factor, labelled *Hypothesis Testing*, explained 15.57% of the variability in TST scores and was composed of questions 5 from both the observational and experimental scenarios. The final fourth component, labelled *Sampling*, explained 12.52% of variability in TST scores and was made up of question 1 from both scenarios. For the purpose of evaluation, only the total scores for the Experimental and Observational component subscales were used to compare the project conditions. These subscales assessed students' statistical thinking about gathering appropriate data, summarising and communicating the selected data, selecting an appropriate statistical test and envisaging a positive result for both experimental and observational scenarios.

Table 14.2: Principal Components Analysis of the Test of Statistical Thinking

	Components			
	Experimental	Observational	Hypothesis Testing	Sampling
Eigenvalues	4.97	1.26	1.15	1.10
% Variance Explained	41.44	10.52	9.560	9.140
Exp4	**0.763**	0.196	0.258	0.121
Exp3	**0.762**	0.270	0.072	0.140
Exp2	**0.747**	0.119	0.087	0.172
Exp6	**0.693**	0.340	0.188	0.041
Obs2	0.138	**0.777**	-0.025	0.208
Obs3	0.283	**0.755**	0.092	0.088
Obs4	0.231	**0.717**	0.190	0.125
Obs6	0.235	**0.685**	0.343	0.009
Obs5	0.139	0.171	**0.887**	0.129
Exp5	0.242	0.140	**0.882**	0.094
Exp1	0.268	0.023	0.081	**0.844**
Obs1	0.067	0.312	0.138	**0.795**
Cronbach's α	0.817	0.803	0.858	0.667

14.3 Results

Intercorrelations amongst study variables and descriptive statistics between project conditions are shown in Table 14.3. Swapping project conditions was positively correlated with group size, but not with project marks or TST scores. Group size and project marks were positively correlated. TST scores for the observational and experimental subscales were positively correlated with each other and with the project marks. Descriptive statistics show on average that the observational project condition had smaller group sizes and project marks, but had higher TST scores for both the observational and experimental subscales when compared to students in the experimental project condition.

Table 14.3: Descriptive Statistics for Project Types and Intercorrelations Between Variables

Variable		1.	2.	3.	4.	5.	6.	7.
1. Swapped[1]		-	.43**	.08	0.01	.01	.02	0.03
2. Group Size			-	.24**	-0.04	-.04	-.01	0.00
3. Project Mark				-	.20**	.26**	.14*	0.13
4. TST Observational					-	.58**	.40**	.38**
5. TST Experimental						-	.43**	.40**
6. TST Hypothesis Testing							-	.29**
7. TST Sampling								-
Observational Project	M		1.63	8.24	10.32	10.73	3.95	4.9
	SD		0.81	2.14	2.72	2.71	1.65	1.31
	N	14/126 (11%)	126	120	104	105	103	103
Experimental Project	M		2.03	8.75	10.17	10.31	3.68	4.87
	SD		0.84	2.04	3.15	3.09	1.52	1.21
	N	48/230 (21%)	230	224	187	190	185	185

[1] No = 1, Yes = 2, * $p < .05$, ** $p < .01$, TST = Test of Statistical Thinking.

Two one-way analysis of covariance (ANCOVA) models were used to evaluate the impact of project condition allocation on observational and experimental TST scores. The models controlled for swapping, group size and project marks. The results of these models are shows in Table 14.4. Figure 14.2 displays the adjusted means and 95% confidence intervals (CI) estimated using the ANCOVA models reported in Table 14.4.

The overall ANCOVA model predicting TST observational scores was statistically significant, $F(4, 294) = 31.67, p = .006, \eta^2 = .05$. However, after controlling for all other variables in the model, there was no statistically significant difference between the observational and experimental project conditions, $F(1, 283) = 0.736, p = .39, \eta^2 = .004$. The only significant predictor was found to be project marks, $F(1, 283) = 13.69, p < .001, \eta^2 = .05$. Swapping, $F(1, 283) = 0.622, p = .43, \eta^2 = .002$, and group size, $F(1, 283) = 1.50, p = .22, \eta^2 = .005$, were not statistically significant covariates (Table 14.4).

For TST experimental scores, the overall ANCOVA model was also statistically significant, $F(4, 288) = 6.48, p < .001, \eta^2 = 0.08$. However, once again, no significant difference was found between observational and experimental conditions after controlling for covariates, $F(1, 288) = 2.08, p = .15, \eta^2 = 0.007$. Swapping, $F(1, 288) =$

Table 14.4: ANCOVA Models Predicting TST Scores for Observational and Experimental Subscales

	TST Observational Scores					
Parameters	B	95% CI	SE	t	p	η^2
Swapping[1]	0.381	(-0.57, 1.332)	0.483	0.789	0.431	0.002
Group Size	-0.288	(-0.75, 0.174)	0.235	-1.226	0.221	0.005
Project Mark	0.327	(0.153, 0.501)	0.088	3.700	<0.001	0.046
Project Condition[2]	0.323	(-0.419, 1.066)	0.377	0.858	0.392	0.003
Observational Project Mean[3]	10.43	(9.843, 11.019)	N = 103			
Experimental Project Mean[3]	10.10	(9.677, 10.538)	N = 185			
	TST Observational Scores					
Swapping[1]	0.42	(-0.506, 1.346)	0.471	0.892	0.373	0.003
Group Size	-0.342	(-0.793, 0.108)	0.229	-1.496	0.136	0.008
Project Mark	0.418	(0.249, 0.586)	0.086	4.884	<0.001	0.076
Project Condition[2]	0.525	(-0.192, 1.242)	0.364	1.442	0.15	0.007
Observational Project Mean[3]	10.81	(10.235, 11.374)	N = 103			
Experimental Project Mean[3]	10.28	(9.865, 10.694)	N = 185			

[1] No = 1, Yes = 2, [2] Observational = 1, Experimental = 2
[3] Adjusted for swapping, group size and project mark.

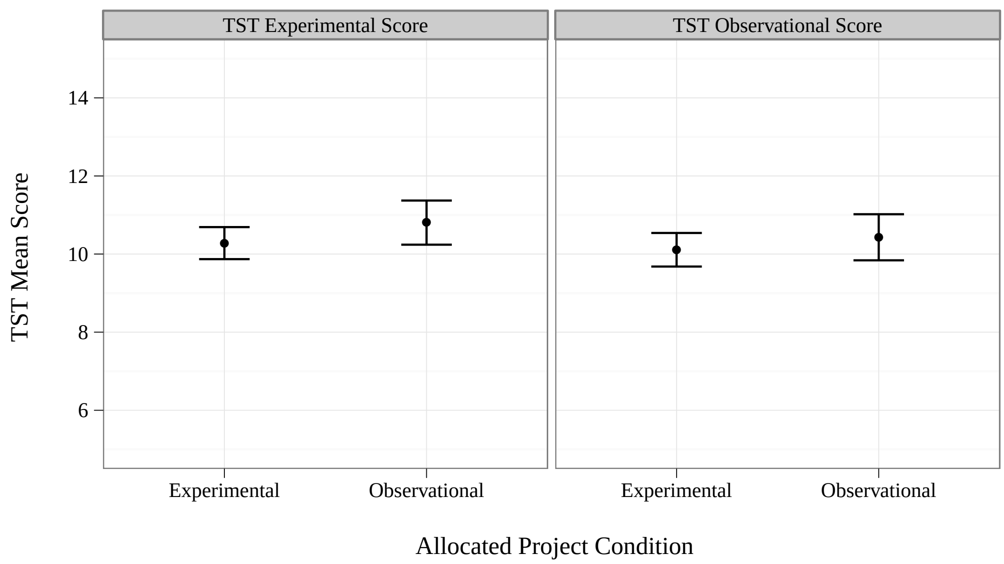


Figure 14.2: Adjusted Means with 95% CI for TST Experimental and Observational Subscale Scores Between Project Condition Allocation

$0.80, p = .37, \eta^2 = .003$, and group size, $F(1, 288) = 2.24, p = .14, \eta^2 = .008$, did not reach statistical significance, but project mark was a significant positive predictor, $F(1, 288) = 23.86, p < .001, \eta^2 = 0.08$ (Table 14.4).

14.4 Discussion

The aim of Study II was to evaluate the impact of project-based learning using the *Island* on students' statistical thinking. In this study individuals and groups of students were randomly allocated to complete projects on the *Island* using either an experimental or observational study design. This study hypothesised that statistical thinking about the students' respective study design would be enhanced above and beyond students who completed the alternate study design. Following the submission of projects, students completed a test of statistical thinking about experimental and observational studies. The results of this study failed to find any evidence of a statistically significant link between project allocation and statistical thinking outcomes. These findings suggest that regardless of the type of research design engaged in for *Island*-based project work, subsequent performance on measures of statistical thinking about different types of research design was not enhanced above other students who completed an alternate research design.

There are a number of possible interpretations which may explain the findings on this initial research. One possibility is that students gained comparable knowledge of both major types of research designs from regular course content (e.g. lectures, notes and tutorials). This interpretation would suggest that the *Island*-based projects provided no added benefit to students' statistical thinking as measured by the TST. This finding would be consistent with Potthast (1999) who found a take home project had no effect on a group of students' understanding of t-tests when compared to a control group. Another possibility is that the experimental and observational studies shared too much in common, e.g. gathering a sample, quantifying variables and selecting similar analysis, as did the the scenarios in the TST. A lack of divergent validity between the constructs of experimental and observational studies may completely explain these results.

Another possibility is that the learning outcomes from the *Island*-based project work are different to the outcomes that were measured in this study. Perhaps the outcomes of PBL in statistics education are more practical (e.g. knowing how to sample, enter data into a data file and manage a study) and less conceptual (e.g. understanding the role of hypothesis testing and the reason for random samples). On the other hand, many outcomes that were assessed by the TST would be hypothesized to be directly enhanced by PBL on the *Island* (e.g. defining and measuring appropriate variables, selecting summary statistics and graphical displays, choosing an appropriate statistical test that fits with the type of variables measured and anticipating the nature of positive results). These outcomes dominated the scoring of the TST and therefore any benefit conveyed by the *Island*-based projects should have been evident.

There are a number of other issues which also impact on any conclusions attained from this study. This study did not include a no-project control group due to ethical reasons. It is possible that the *Island* did in fact benefit students' statistical thinking, but did so for thinking about both types of research designs. For example, in order for students to effectively implement an observational research design, perhaps these students heavily researched observational studies and contrasted it with an experiment. While there is no way to test this hypothesis in the current study, a future study design could include students randomly allocated to a no-project condition. Then a more direct evaluation of the proposed benefits of project-based learning on statistical thinking could be made. Regardless, there was still a logical reason to predict that the nature of study design used in projects might link to indicators of students' statistical thinking about different research designs. However, the results of this study failed to find such evidence.

Methodologically, this study's strengths were in its careful a-priori design, the use of randomisation to minimise the probability of confounding and the use of a large sample size to rule out issues with statistical power. The study was limited by the number of students changing groups (swapping) after random allocation and the self-assessment nature of the TST. As this study kept a record of which students' swapped groups, a swapping covariate was included in the statistical model. This variable aimed to control for a possible confounding caused by systematic difference in swappers and non-swappers (e.g. weaker students may have swapped to the experimental designs because they perceived them as being easier projects). Due to ethical constraints, the TST was administered towards the end of the semester during a regular tutorial. Students were given a participation mark for completing the test, but students' effort on this test may have been reduced as it was not summative. Future research should embed measures of statistical thinking in exams where students are most likely to produce their best effort. Nonetheless, anecdotal evidence from course instructors indicated that students typically exhibited high levels of engagement in tutorial exercises. The 64% consent rate must also be raised as a limitation. It is possible that consent rates were lowered due to students' overestimating the ethical risk posed by this project. Thus, the effect of a sampling bias cannot be ruled out.

The validity of the TST must also be discussed. This test has not been used in previous research. To maximise its potential validity, the development of the TST was guided by Wild and Pfannkuch's 1999 model of statistical thinking. The short-answer format ensured that students demonstrated and constructed their answers in line with recommendation by G. Smith (1998). The discriminant validity between observational and experimental statistical thinking subscale was supported by principal components analysis. However, further testing of its validity in different samples is required. While many features of observational and experimental research designs bring with them unique aspects of statistical thinking, the foundations of both types of research are very similar. The same statistical summaries and tests can be used for either type. Regardless, if project-based assessment improved students' statistical thinking, a student should be able to demonstrate a deeper level of statistical thinking about the research design they completed their project on. The TST was designed to target key indicators of statistical thinking pertinent to the *Island*-based projects' course learning outcomes. The TST was not designed to measure all aspects of statistical thinking, nor was it designed to be a suitable measure for all introductory statistics courses.

Another general limitation to the nature of the*Island*-based projects employed in Study II was the bivariate focus. As most real-world research questions and data investigations are multivariate, the *Island*-based projects completed by students were limited in real-world applicability. This is not a limitation of the *Island* itself, as multivariate data collection is possible, but a limitation of topics covered in the introductory statistics course and the need to align assessment to the content. Regardless, these projects were successful in providing students with their first, but somewhat simplified, experience of the entire data investigative cycle. Students need to start with a foundation before being thrown into the multivariate "deep end" of real-world research. The *Island*-based projects used in this study aimed to provide that foundation.

The variables that are able to be measured on the *Island* are fixed before the projects commence, although it is possible to request additions provided enough time is given to the creator (Bulmer, 2011). The inclusion of hundreds of different variables creates a relatively open-ended experience, however, there are still limitations to the virtual environment. This may be perceived as a limitation to the Island's design, but it could be argued that things are no different for real world research where practical and financial constraints limit the collection of data. The focus on applied human research also means that the *Island* won't suit all student disciplines, but but that was never the intention of the Island. The *Island* was created to overcome the major practical and ethical issues of conducting research on human participants.

14.5 Conclusion

Project-based learning has a well-documented body of literature reporting students' widespread positive attitudes towards its use in introductory statistics courses. This is reason enough to incorporate this popular form of alternate learning and assessment. However, PBL's popularity is heavily founded on the premise that it will help students develop statistical thinking by engaging them in the data investigative process (MacGillivray & Pereira-Mendoza, 2011). Studies that have aimed to empirically verify this proposed link have been lacking, hence the rationale for this study. While the research hypotheses of this initial study were not supported, the conclusions must be regarded as being inconclusive due to a number of limitations. Regardless, the outcomes do highlight major challenges related to the assessment of statistical thinking and lessons learnt for evaluation research related to statistics thinking. These challenges must be addressed in order to establish a relationship between PBL and statistical thinking in the introductory statistics course. Doing so will not only assess the merit of PBL for developing statistical thinking, but will also lead to a better understanding of the very development of statistical thinking itself.

Conclusion

Each separate part of this dissertation considered a different learning outcome in statistics education with the goal in mind that these outcomes could be improved with the application of learning theory-based methods. Specifically, this dissertation considered the development of the important technological skill of operating statistical packages, the correction of common misconceptions of statistical concepts and the development of students' ability to think statistically. These outcomes were broadly tied to the hierarchical model of the major learning outcomes of statistics education that includes the notions of statistical literacy, reasoning and thinking.

Due to the complexity of statistical knowledge and diversity of the learning outcomes targeted in each part, three learning theory-based methods were employed. These methods included active-exploratory training for statistical package skills, cognitive conflict activities for correcting misconceptions and experiential learning for the development of statistical thinking. The main rationale behind each part was the need for empirical evidence in statistics education that validates the use of these methods in the statistics classroom.

Technological skill now pervades all modern notions of statistical literacy (Gould, 2010). In Part I, an active-exploratory training approach, known as Error-management training (EMT), was compared to conventional guided training (GT) for the development of analogical and adaptive transfer of statistical packages skills. Two major experiments were conducted across two cohorts of an introductory statistics course where psychology students were trained to operate the statistical package *SPSS*. Across the two experimental trials no evidence was found as to the superiority of either training approach on measures of training transfer. The outcomes from both trials highlighted a strong dependency between the ability to use statistical technology and statistical knowledge itself. The major conclusions from this part are that training methods don't appear to impact the development of statistics technological skill and that other factors need to be explored in order to help foster the development of these vital skills.

Common misconceptions of statistical concepts can lead to poor statistical reasoning. Statistics instructors must be conscious of these misunderstandings and have effective strategies for preventing and correcting misconceptions. In Part II, the effect of short lecture-based conceptual change activities for correcting common statistical misconceptions were evaluated in a prospective cohort study. The results found a weak statistically significant reduction in measures of statistical misconceptions in a cohort of students given the conceptual changed-based activities one year when compared to a control year that did not get the activities. This effect was present after controlling for covariates. However, as the effect was much weaker compared to previous studies, the brief format may have had only a limited effect. More ingrained or hard to change misconceptions are most likely to require more intensive intervention.

Part III of the dissertation evaluated the use of the online virtual environment, known as the *Island*, for engaging students in the entire data investigative cycle of empirical enquiry. Specifically, students engaged in experiential project-based learning (PBL) on the *Island* with the aim of developing their ability to think statistically. The first study evaluated student feedback of using the *Island* for PBL in an online postgraduate introductory statistics course. Feedback given by the students was highly positive and provided qualitative evidence to support the effect of project-based learning on the development of statistical thinking. In a second experimental study, the proposed link between PBL and statistical thinking was evaluated in an undergraduate introductory statistics course. Undergraduate science students were randomly allocated as individuals or small groups to complete either an observational or experimental project using the *Island*. After the projects were completed, the same students completed a test of statistical thinking, designed for the purpose of the study, about observational and experimental studies. The experiment hypothesised that students' performance on the observational and experimental sub-scales of the test of statistical thinking would be related to the type of project randomly allocated to students on the *Island*. The results found no evidence of such a relationship. However, this was most likely due to the major similarities between observational and experimental studies that was assessed in the test of statistical thinking and the more general challenge of assessing statistical thinking for the purpose of evaluation research. Further research is needed to better understand the learning outcomes developed by PBL in statistics education and how statistical thinking can be better evaluated for the purposes of empirical research.

The main outcomes from each part highlight the significant challenges of statistics education research and the significant work that lies ahead in providing empirical verification of learning theory-based methods. Designing and implementing evaluative studies in real-world statistics courses exemplify the intricate, complex, and multifaceted nature of statistical knowledge. Multiple learning methods must be utilised to provide a robust learning experience, especially in the introductory statistics course that will lay the foundation. Assessing the outcomes of these learning experiences remains as much a challenge for instructors as it is for statistics education researchers. This dissertation aimed to answer a number of important and practical research questions that face statistics instructors. However, the answers to these questions won't come easily. Statistics education researchers must continue to build their body of knowledge so that future students can benefit from the best that a statistics education can offer.

Publications Arising

The following publications have arisen as a result of the work performed in this dissertation.

Baglin, J., & Da Costa, C. (n.d.). Applying a theoretical model for explaining the development of technological skills in statistics education. *Technology Innovations in Statistics Education*, In review.

Baglin, J., & Da Costa, C. (2010). An experimental study comparing strategies of learning how to use statistical software packages in introductory statistics courses. In H. MacGillivray & B. Phillips (Eds.), *Proceedings of the Seventh Australian Conference on Teaching Statistics, December 2010*. Fremantle, Western Australia. Retrieved from http://opax.swin.edu.au/~3420701/OZCOTS2010/OZCOTS2010_paper_Baglin.pdf

Baglin, J., & Da Costa, C. (2012a). An experimental study evaluating error management training for learning to operate a statistical package in an introductory statistics course: Is less guidance more? *International Journal of Innovation in Science and Mathematics Education*, *20*, 48–67. Retrieved from http://ojs-prod.library.usyd.edu.au/index.php/CAL/article/view/5809/6529

Baglin, J., & Da Costa, C. (2012b). A theoretical framework for developing statistical package competence as an outcome of statistical literacy. In *Proceedings of the International Association of Statistics Education 2012 Roundtable Conference: Technology in Statistics Education: Virtualities and Realities, July 2012*. Cebu City, Philippines.

Baglin, J., & Da Costa, C. (2012c). Students' thoughts and perceptions of training to use statistical packages in introductory statistics courses: A qualitative study.

In H. MacGillivray & B. Phillips (Eds.), *Proceedings of the Eighth Australian Conference on Teaching Statistics, July 2012*. Adelaide, South Australia.

Baglin, J., & Da Costa, C. (2013). Comparing training approaches for technological skills development in introductory statistics courses. *Technology Innovations in Statistics Education*, *7*. Retrieved from http://www.escholarship.org/uc/item/1hc308sv

Baglin, J., Da Costa, C., Ovens, M., & Bablas, V. (2011). An experimental study comparing two different training strategies on how to use statistical software packages in an introductory statistics course. In M. Sharma, A. Yeung, T. Jenkins, E. Johnson, G. Rayner & J. West (Eds.), *Proceedings of the Australian Conference on Science & Mathematics Education (17th Annual UniServe Science Conference), September 2011* (pp. 162 –168). Melbourne, Australia.

Baglin, J., Bedford, A., & Bulmer, M. (n.d.). Students' experiences and perceptions of using a virtual environment for project-based assessment in an online introductory statistics course. *Technology Innovations in Statistics Education*, In review.

Baglin, J., Bedford, A., & Bulmer, M. (2012). Students' experiences and perceptions of using a virtual environment for project-based assessment in an online introductory statistics course. In *Proceedings of the International Association of Statistics Education 2012 Roundtable Conference: Technology in Statistics Education: Virtualities and Realities, July 2012*. Cebu City, Philippines. Retrieved from http://icots.net/roundtable/docs/Thursday/IASE2012_Baglin_Bedford_Bulmer.pdf

References

Albert, J. (1997). Teaching Bayes' rule: A data-oriented approach. *The American Statistician*, *51*, 247–253.

Allen, P., & Bennett, K. (2008). *SPSS for the health and behavioural sciences.* South Melbourne, VIC: Thomson.

American Statistical Association. (2005). *Guidelines for assessment and instruction in statistics education.* Washington, DC: Author. Retrieved from http://www.amstat.org/education/gaise/GAISECollege.htm

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., ... Wittrock, M. C. (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: Allyn & Bacon.

Baloglu, M. (2002). Psychometric properties of the statistics anxiety rating scale. *Psychological Reports*, *90*, 315–325.

Bell, B. S., & Kozlowski, S. W. J. (2002). Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology*, *55*, 267–306.

Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology*, *93*, 296–316.

Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory model. *Journal of Experimental Education*, *57*, 247–261.

Benson, K., & Hartz, A. J. (2000). A comparison of observational studies and randomized, controlled trials. *New England Journal of Medicine*, *342*, 1878–1886.

Ben-Zvi, D., & Garfield, J. B. (2005). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3–15). New York: Kluwer Academic Publishers.

Ben-Zvi, D., & Garfield, J. B. (2008). Introducing the emerging discipline of statistics education. *School Science and Mathematics*, *108*, 355–361.

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician*, *51*, 241–246.

Bloom, B. S., Masia, B. B., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals.* Chicago: Susan Fauer Company, Inc.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, *26*, 369–398.

Bowman, A., & Gilmour, W. H. (1998). Computer-based learning in statistics: A problem solving approach. *The Statistician*, *47*, 349–364.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In P. Mussen, J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: cognitive development* (4th ed., pp. 77–166). New York: Harvard University Press.

Bulmer, M. (2005). Virtual humans for teaching and learning statistics. In I. MacColl (Ed.), *Evolution: Universities, students and technology: Proceedings of the Apple University Consortium Academic and Developers Conference, September 2005* (pp. 1–8). Hobart, Tasmania.

Bulmer, M. (2010). Technologies for enhancing project assessment in large classes. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics, July 2010.* Ljubljana, Slovenia. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_5D3_BULMER.pdf

Bulmer, M. (2011). Island. Retrieved from http://island.maths.uq.edu.au/access.php?/index.php

Bulmer, M., & Haladyn, J. K. (2011). Life on an Island: A simulated population to support student projects in statistics. *Technology Innovations in Statistics Education*, *5*. Retrieved from http://escholarship.org/uc/item/2q0740hv

Cameron, M. (2009). Training statisticians for a research organisation. In *Proceedings of the 57th Session of the International Statistical Institute, August 2009*. Durban, South Africa. Retrieved from http://isi.cbs.nl/iamamember/CD8-Durban2009/A5Docs/0276.pdf

Carnell, L. J. (2008). The effect of a student-designed data collection project on attitudes toward statistics. *Journal of Statistics Education*, *16*. Retrieved from http://www.amstat.org/publications/jse/v16n1/carnell.pdf

Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*, 98–113.

Chambers, J. M. (1993). Greater or lesser statistics: a choice for future research. *Statistics and Computing*, *3*, 182–184.

Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, *10*. Retrieved from http://www.amstat.org/publications/jse/v10n3/chance.html

Chance, B. L., Ben-Zvi, D., Garfield, J. B., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, *1*. Retrieved from http://escholarship.org/uc/item/8sd2t4rr

Chance, B. L., & Garfield, J. B. (2001). New approaches to gathering data on student learning for research in statistics education. In *Proceedings of the 53rd Session of the International Statistical Institute, August 2001*. Seoul, Korea. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/4/751.pdf

Chillarege, K. A., Nordstrom, C. R., & Williams, K. B. (2003). Learning from our mistakes: Error management training for mature learners. *Journal of Business and Psychology*, *17*, 369–385.

Cobb, G. W. (1992). Teaching statistics. In L. A. Steen (Ed.), *Heeding the call for change: suggestions for curricular action* (pp. 3–43). Washington, DC: Mathematical Association of America.

Cohen, S., Smith, G., Chechile, R. A., Burns, G., & Tsai, F. (1996). Identifying impediments to learning probability and statistics from an assessment of instructional software. *Journal of Educational and Behavioural Statistics*, *21*, 35–54.

Concato, J., Shah, N., & Horwitz, R. I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine*, *342*, 1887–1892.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Cruise, R. J., Cash, R. W., & Bolton, D. L. (1985). Development and validation of an instrument to measure statistical anxiety. In *American Statistical Association 1985 Proceedings of the Section on Statistical Education* (pp. 92–97). Las Vagas, Nevada.

Debowski, S., Wood, R. E., & Bandura, A. (2001). Impact of guided exploration and enactive exploration on self-regulatory mechanisms and information acquisition through electronic search. *Journal of Applied Psychology*, *86*, 1129–1141.

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, *62*, 119–142.

delMas, R. (2002). Statistical literacy, reasoning, and learning: A commentary. *Journal of Statistics Education*, *10*. Retrieved from http://www.amstat.org/publications/jse/v10n3/delmas_discussion.html

delMas, R., Garfield, J. B., Ooms, A., & Chance, B. L. (2006). Assessing students' conceptual understanding after a first course in statistics. In *Proceedings of the Annual Meetings of the American Educational Research Association, April 2006.* San Francisco, CA.

delMas, R., Garfield, J. B., Ooms, A., & Chance, B. L. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, *6*, 28–58.

Dobson, J. L. (2009). Evaluation of the virtual physiology of exercise laboratory program. *Advances in Physiology Education*, *33*, 335–342.

Dormann, T., & Frese, M. (1994). Error training: Replication and the function of exploratory behavior. *International Journal of Human-Computer Interaction*, *6*, 365–372.

Fidler, F. (2006). Should psychology abandon p-values and teach CIs instead? Evidence-based reforms in statistics education. In A. Rossman & B. L. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, July 2006* (pp. 1 –6). Salvador, Brazil. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/17/5E4_FIDL.pdf

Finch, S. (1998). Explaining the law of large numbers. In *Proceedings of the Fifth International Conference on the Teaching of Statistics, June 1998* (pp. 732–737). Singapore. Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/2/Topic6n.pdf

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, *28*, 161–186.

Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology*, *83*, 218–233.

Forster, M., & MacGillivray, H. (2010). Student discovery projects in data analysis. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics, July 2010*. Ljubljana, Slovenia. Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_4G2_FORSTER.pdf

Francis, G. (2007). *Introduction to SPSS for Windows: v15.0 and 14.0 with notes for Studentware*. Frenchs Forest, NSW: Pearson Education Australia.

Frese, M., Albrecht, K., Altmann, A., Lang, J., Papstein, P. V., Peyerl, R., … Wendel, R. (1988). The effects of an active development of the mental model in the training process: Experimental results in a word processing system. *Behaviour & Information Technology*, *7*, 295–304.

Frese, M., Brodbeck, F., Heinbokel, T., Mooser, C., Schleiffenbaum, E., & Thiemann, P. (1991). Errors in training computer skills: On the positive function of errors. *Human-Computer Interaction*, *6*, 77–93.

Furnham, A., & Chamorro-Premuzic, T. (2004). Personality and intelligence as predictors of statistics examination grades. *Personality and Individual Differences*, *37*, 943–955.

Garfield, J. B. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, *10*, 58–69. Retrieved from http://www.amstat.org/publications/jse/v10n3/garfield.html

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*, 22–38.

Garfield, J. B., & Ben-Zvi, D. (2005). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 397–409). New York: Kluwer Academic Publishers.

Garfield, J. B., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*, 372–396.

Garfield, J. B., & Chance, B. L. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning*, *2*, 99–125.

Garfield, J. B., delMas, R., & Zieffler, A. (2010). Assessing important learning outcomes in introductory tertiary statistics courses. In P. Bidgood, N. Hunt & F. Jolliffe (Eds.), *Assessment methods in statistical education: an international perspective* (pp. 75–86). Chichester, WS: John Wiley & Sons.

Garfield, J. B., & Gal, I. (1999a). Assessment and statistics education: Current challenges and directions. *International Statistical Review*, *67*, 1–12. Retrieved from http://www.jstor.org/stable/1403562

Garfield, J. B., & Gal, I. (1999b). Teaching and assessing statistical reasoning. In L. Stiff (Ed.), *Developing mathematical reasoning in grades k-12* (pp. 207–219). Reston, VA: National Council Teachers of Mathematics.

Garfield, J. B., Hogg, B., Schau, C., & Whittinghill, D. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education*, *10*. Retrieved from http://www.amstat.org/publications/jse/v10n2/garfield.html

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but where afraid to ask. In D. Kaplan (Ed.), *The sage handbook of quantitative methodology for social sciences* (pp. 391–408). Thousand Oaks, CA: Sage.

Gould, R. (2010). Statistics and the modern student. *International Statistical Review*, *78*, 297–315.

Greenland, S., & Finkle, W. D. (1995). A critical look at methods for handling missing covariates in epidemiologic regression analyses. *American Journal of Epidemiology*, *142*, 1255–1264.

Griffiths, P., & Sheppard, Z. (2010). Assessing statistical thinking and data presentation skills through the use of a poster assignment with real-world data. In P. Bidgood, N. Hunt & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 47–56). Chichester, WS: John Wiley & Sons.

Gully, S. M., Payne, S. C., Koles, K., & Whiteman, J. A. K. (2002). The impact of error training and individual differences on training outcomes: An attribute-treatment interaction perspective. *Journal of Applied Psychology*, *87*, 143–155.

Haller, H., & Krauss, S. (2002). Misinterpretations of significance : A problem students share with their teachers? *Methods of Psychological Research Online*, *7*, 1–20. Retrieved from http://www.metheval.uni-jena.de/lehre/0405-ws/evaluationuebung/haller.pdf

Hassad, R. A. (2012). Faculty attitude toward technology-assisted instruction for introductory statistics in the context of educational reform. In *Proceedings of the International Association of Statistics Education 2012 Roundtable Conference: Technology in Statistics Education: Virtualities and Realities, July 2012.* Cebu City, Philippines. Retrieved from http://icots.net/roundtable/docs/Thursday/IASE2012_Hassad.pdf

Heimbeck, D., Frese, M., Sonnentag, S., & Keith, N. (2003). Integrating errors into the training process: The function of error management instructions and the role of goal orientation. *Personnel Psychology*, *56*, 333–361.

Hesketh, B. (1997). Dilemmas in training for transfer and retention. *Applied Psychology*, *46*, 317–339.

Hirsch, L. S., & O'Donnell, A. (2001). Representativeness in statistical reasoning: Identifying and assessing misconceptions. *Journal of St*, *9*. Retrieved from http://www.amstat.org/publications/jse/v9n2/hirsch.html

Holmes, P. (1997). Assessing project work by external examiners. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 153–164). IOS Press.

Hunter, J. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*, 340–362.

International Association for Statistical Education. (2012). Recent doctoral dissertations in statistics education. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/dissertations/dissertations.php

Jazayeri, M., Lai, J., Fidler, F., & Cumming, G. (2010). Understanding of sampling variability: Confronting students' misconception in relation to sampling variability. In H. MacGillivary & B. Phillips (Eds.), *Proceedings of the Seventh Australian Conference on Teaching Statistics, December 2010.* Fremantle, Western Australia.

Jolliffe, F. (1998). What is research in statistics education? In *Proceedings of the Fifth International Conference on the Teaching of Statistics, July 1998* (pp. 801–806). Singapore.

Jolliffe, F. (2003). Towards a database of research in statistical education. *Statistics Education Research Journal*, *2*, 47–58.

Jolliffe, F. (2010). Assessing statistical thinking. In P. Bidgood, N. Hunt & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 71–74). Chichester, WS: John Wiley & Sons.

Kalinowski, P., Fidler, F., & Cumming, G. (2008). Overcoming the inverse probability fallacy: A comparison of two teaching interventions. *Methodology*, *4*, 152–158.

Kamouri, A. L., Kamouri, J., & Smith, K. H. (1986). Training by exploration: Facilitating the transfer of procedural knowledge through analogical reasoning. *International Journal of Man-Machine Studies*, *24*, 171–192.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74*, 657–690.

Kanfer, R., Ackerman, P. L., & Heggestad, E. D. (1996). Motivational skills & self-regulation for learning: A trait perspective. *Learning and Individual Differences*, *8*, 185–209.

Karoly, P. (1993). Mechanisms of self-regulation: A systems view. *Annual Review of Psychology*, *44*, 23–52.

Keith, N., & Frese, M. (2005). Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology*, *90*, 677–691.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology*, *93*, 59–69.

Keith, N., Richter, T., & Naumann, J. (2010). Active/exploratory training promotes transfer even in learners with low motivation and cognitive ability. *Applied Psychology: An International Review*, *59*, 97–123.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice-Hall.

Kozlowski, S. W. J., & Bell, B. S. (2006). Disentangling achievement orientation and goal setting: Effects on self-regulatory processes. *Journal of Applied Psychology*, *91*, 900–916.

Kozlowski, S. W. J., Gully, S. M., Brown, K. G., Salas, E., Smith, E. M., & Nason, E. R. (2001). Effects of training goals and goal orientation traits on multidimensional training outcomes and performance adaptability. *Organizational Behavior and Human Decision Processes*, *85*, 1–31.

Lazar, J., & Norcio, A. (2003). Training novice users in developing strategies for responding to errors when browsing the web. *International Journal of Human-Computer Interaction*, *15*, 361–377.

Lecoutre, B. (2006). Training students and researchers in Bayesian methods. *Journal of Data Science*, *4*, 207–232.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, *11*, 357–380.

Linden, M., Baglin, J., & Bedford, A. (2011). Teaching clinical trial design and management using an online virtual environment. In M. Sharma, A. Yeung, T. Jenkins, E. Jonhson, G. Rayner & J. West (Eds.), *Proceedings of the Australian Conference on Science & Mathematics Education (17th Annual UniServe Science Conference), September 2011* (pp. 107 –113). Melbourne, Australia. Retrieved from http://escholarship.library.usyd.edu.au/journals/index.php/IISME/article/viewFile/4812/5570

Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, *36*, 837–858.

Liu, T. (2010). Developing simulation-based computer assisted learning to correct students' statistical misconceptions based on cognitive conflict theory, using "correlation" as an example design basis of SALS. *Educational Technology & Society*, *13*, 180–192.

Liu, T., Lin, Y. C., & Kinshuk. (2010). The application of Simulation-Assisted Learning Statistics (SALS) for correcting misconceptions and improving understanding of correlation. *Journal of Computer Assisted Learning*, *26*, 143–158.

MacGillivray, H. (1998). Developing and synthesizing skills for real situations through student projects. In *Proceedings of the Fifth International Conference on the Teaching of Statistics, July 1998.* Singapore. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/2/Topic8n.pdf

MacGillivray, H. (2010). Variety in assessment for learning statistics. In P. Bidgood, N. Hunt & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 21–34). Chichester, WS: John Wiley & Sons.

MacGillivray, H., & Pereira-Mendoza, L. (2011). Teaching statistical thinking through investigative projects. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education:*

A Joint ICM/IASE Study: The 18th ICMI Study (pp. 109–120). New York, NY: Springer Science+Business Media B. V.

Marriott, J., Davies, N., & Gibson, L. (2009). Teaching, learning and assessing statistical problem solving. *Journal of Statistics Education*, *17*. Retrieved from http://www.amstat.org/publications/jse/v17n1/marriott.pdf

Martocchio, J. J. (1994). Effects of conceptions of ability on anxiety, self-efficacy, and learning in training. *Journal of Applied Psychology*, *79*, 819–825.

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *60*, 48–58.

McDaniel, M. A., & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. *Ethics & Behavior*, *7*, 129–159.

Mills, J. D. (2003). SPSS textbooks: A review for teachers. *Statistics Education Research Journal*, *2*, 59–70.

Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, *51*, 254–261.

Neumann, D. L., Neumann, M. M., & Hood, M. (2011). Evaluating computer-based simulations, multimedia and animations that help integrate blended learning with lectures in first year statistics. *Australasian Journal of Educational Technology*, *27*, 274–289.

Nolan, D., & Temple Lang, D. (2010a). Computing in the statistics curricula. *The American Statistician*, *64*, 97–107. Retrieved from http://www.tandfonline.com/doi/abs/10.1198/tast.2010.09132

Nolan, D., & Temple Lang, D. (2010b). Integrating computing and data technologies into the statistics curricula. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics, July 2010*. Ljubljana, Slovenia. Retrieved from http://icots.net/8/cd/pdfs/invited/ICOTS8_9B2_NOLAN.pdf

Nordstrom, C. R., Wendland, D., & Williams, K. B. (1998). "To err is human": An examination of the effectiveness of error management training. *Journal of Business and Psychology*, *12*, 269–282.

Onwuegbuzie, A. J., & Seaman, M. A. (1995). The effect of time constraints and statistics test anxiety on test performance in a statistics course. *Journal of Experimental Education*, *63*, 115–124.

Oswald, P. A. (1996). Classroom use of the personal computer to teach statistics. *Teaching of Psychology*, *23*, 124–126.

Ottaviani, M. G. (2005). Research into statistics education as a discipline. *International Statistical Review*, *73*, 207–209.

Pfannkuch, M., & Wild, C. (1998). Investigating the nature of statistical thinking. In *Proceedings of the Fifth International Conference on the Teaching of Statistics, June 1998* (pp. 461–467). Singapore. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/2/Topic4f.pdf

Pfannkuch, M., & Wild, C. (2000). Statistical thinking and statistical practice: Themes gleaned from professional statisticians. *Statistical Science*, *15*, 132–152. Retrieved from http://www.jstor.org/stable/2676728

Pfannkuch, M., & Wild, C. (2005). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17–46). New York: Kluwer Academic Publishers.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*, 211–227.

Potthast, M. J. (1999). Outcomes of using small-group cooperative learning experiences in introductory statistics courses. *College Student Journal*, *33*, 34–42.

Pretorius, T. B., & Norman, A. M. (1992). Psychometric data on the Statistics Anxiety Scale for a sample of South African students. *Educational and Psychological Measurement*, *52*, 933–937.

Richter, T., Naumann, J., & Groeben, N. (2000). Attitudes toward the computer: Construct validation of an instrument with scales differentiated by content. *Computers in Human Behavior*, *16*, 473–491.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10*. Retrieved from http://www.amstat.org/publications/jse/v10n3/rumsey2.html

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*, 450.

Rybowiak, V., Garst, H., Frese, M., & Batinic, B. (1999). Error Orientation Questionnaire (EOQ): Reliability, validity, and different language equivalence. *Jounrnal of Organizational Behavior*, *20*, 527–547.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, *103*(484), 1334–1344.

Shrive, F. M., Stuart, H., Quan, H., & Ghali, W. A. (2006). Dealing with missing data in a multi-question depression scale: A comparison of imputation methods. *BMC Medical Research Methodology*, *57*. doi:doi:10.1186/1471-2288-6-57

Skinner, B. F. (1968). *The technology of teaching*. New York: Appleton-Century-Crofts New York.

Smith, B. (2003). Using and evaluating resampling simulations in SPSS and Excel. *Teaching Sociology*, *31*, 276–287.

Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education*, *6*. Retrieved from http://www.amstat.org/publications/jse/v6n3/smith.html

Snee, R. D. (1993). What's missing in statistics education. *The American Statistician*, *47*, 149–154.

Spinello, E. F., & Fischbach, R. (2004). Problem-based learning in public health instruction: A pilot study of an online simulation as a problem-based learning approach. *Education for Health*, *17*, 365–373.

Stafford, R., Goodenough, A. E., & Davies, M. S. (2010). Assessing the effectiveness of a computer simulation for teaching ecological experimental design. *Bioscience Education*, *15*. Retrieved from http://www.bioscience.heacademy.ac.uk/journal/vol15/beej-15-1.pdf

Stevens, C. K., & Gist, M. E. (1997). Effects of self-efficacy and goal-orientation training on negotiation skill maintenance: What are the mechanisms? *Personnel Psychology*, *50*, 955–978.

Tabernero, C., & Wood, R. E. (1998). Implicit theories versus the social construal of ability in self-regulation and performance on a complex task. *Organizational Behavior and Human Decision Processes*, *78*, 104–127.

Tremblay, P. F., Gardner, R. C., & Heipel, G. (2000). A model of the relationships among measures of affect, aptitude, and performance in introductory statistics. *Canadian Journal of Behavioural Science*, *32*, 40–48.

Vere-Jones, D. (1995). The coming of age of statistical education. *International Statistical Review*, *63*, 3–23.

Vere-Jones, D. (1998). Background influences on the development of statistical education. In *Proceedings of the Fifth International Conference on the Teaching of Statistics, June 1998.* Retrieved from http://iase-web.org/documents/papers/icots5/Keynote4.pdf

Watson, J. M. (1997). Assessing statistical thinking using the media. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107–121). IOS Press. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/assessbk/chapter09.pdf

Watson, J. M. (2002a). Creating cognitive conflict in a controlled research setting: Sampling. In *Proceedings of the Sixth International Conference on the Teaching of Statistics, July 2002.* Cape Town, South Africa. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/1/6a1_wats.pdf

Watson, J. M. (2002b). Inferential reasoning and the influence of cognitive conflict. *Educational Studies in Mathematics*, *51*, 225–256.

Watson, J. M. (2007). The role of cognitive conflict in developing students' understanding of average. *Educational Studies in Mathematics*, *65*, 21–47.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*, 223–265.

Wild, C., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society*, *174*, 247–295. doi:10.1111/j.1467-985X.2010.00678.x

Wild, C., Triggs, C., & Pfannkuch, M. (1997). Assessment on a budget: Using traditional methods imaginatively. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 205–220). IOS Press. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/assessbk/chapter16.pdf

Wood, R. E., Kakebeeke, B., Debowski, S., & Frese, M. (2000). The impact of enactive exploration on intrinsic motivation, strategy, and performance in electronic search. *Applied Psychology*, *49*, 263–283.

Appendices

Appendix A

Part I

A.1 Pilot Questionnaire

QUESTIONNAIRE

Before we proceed, please fill out the following information.

Participant Information

1. **What is your age?**

2. **What is the name of the undergraduate program you are enrolled in?**

3. **Gender?** ☐ **Male** ☐ **Female**

4. **International or domestic student?** ☐ **International** ☐ **Domestic**

5. **Full-time or part-time?** ☐ **Full-time** ☐ **Part-time**

Participant Responses - Prior to Training

Instructions - Please rate your **confidence in learning the skills necessary** while you're in this session to successfully complete the following tasks where: **(1) no confidence at all** and **(10) complete confidence**. (Circle your response)

1. To use the statistical package to compute basic descriptive statistics (e.g. compute means and standard deviations)

No confidence at all									Complete confidence
1	2	3	4	5	6	7	8	9	10

2. To use the statistical package to create basic graphical displays of data (e.g. bar graphs and scatter plots)

No confidence at all									Complete confidence
1	2	3	4	5	6	7	8	9	10

3. To use the statistical package to conduct basic statistical inference (e.g. generate *p*-values)

No confidence at all									Complete confidence
1	2	3	4	5	6	7	8	9	10

The tutorial will continue on the next page...

Participant Responses – Following the Training

Instructions - Please rate your **confidence in current ability** after the first session to successfully complete the following tasks where: **(1) no confidence at all** and **(10) complete confidence**. (Circle your response)

1. To use the statistical package to compute basic descriptive statistics (e.g. compute means and standard deviations)

No confidence at all									Complete confidence
1	2	3	4	5	6	7	8	9	10

2. To use the statistical package to create basic graphical displays of data (e.g. bar graphs and scatter plots)

No confidence at all									Complete confidence
1	2	3	4	5	6	7	8	9	10

3. To use the statistical package to conduct basic statistical inference (e.g. generate *p*-values)

No confidence at all									Complete confidence
1	2	3	4	5	6	7	8	9	10

Instructions -Please rate the level of anxiety you experience in relation to the following statements where a **(1) indicates no anxiety** and a **(10) indicates very strong anxiety**. (Circle your response)

1. Using a computer statistics package to run statistical analysis

No anxiety									Very strong anxiety
1	2	3	4	5	6	7	8	9	10

2. Interpreting statistical analysis from a computer statistics package

No anxiety									Very strong anxiety
1	2	3	4	5	6	7	8	9	10

Overall, please rate the difficulty of this tutorial, where a **1** indicates the easiest tutorial you have ever completed and a 10 indicates the most difficult tutorial you have ever completed.

Tutorial Difficulty

Very Easy									Very Difficult
1	2	3	4	5	6	7	8	9	10

Once you have finished answering the questions, please let the tutor know. The tutor will need to quickly check your work before leaving.

Thanks for your participation!

A.2 Pilot Plain Language Statement

RMIT UNIVERSITY

RMIT Human Research Ethics Committee

School of Mathematics and Geospatial Sciences

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2283
Fax +61 3 9925 2454

INVITATION TO PARTICIPATE IN A RESEARCH PROJECT
PROJECT INFORMATION STATEMENT

Project Title

Comparing Methods of Teaching Statistics Software Packages in Computer Tutorials

Investigators

- Mr. James Baglin (Lead Investigator: BAppSc, Psychology – Honours; PhD Candidate, Statistics, james.baglin@rmit.edu.au, 9925 6118)
- Dr. Cliff Da Costa (Project Supervisor: Associate Professor, SMGS, RMIT University)

Dear student,

You are invited to participate in a research project being conducted by RMIT University. This information sheet describes the project in straightforward language, or 'plain English'. Please read this sheet carefully and be confident that you understand its contents before deciding whether to participate. If you have any questions about the project, please ask one of the investigators.

Who is involved in this research project? Why is it being conducted?

This project is being conducted as a part of a PhD in Statistics by James Baglin. The project is being supervised by Dr. Cliff Da Costa who is an Associate Professor in the School of Mathematics and Geospatial Sciences. The project has been approved by the RMIT Human Research Ethics Committee (HREC). The study is being conducted to evaluate different methods of teaching statistical software packages.

Why have you been approached?

You have been approached to participate in the study because you are a RMIT university student over the age of 18, have completed (or currently about to complete) an introductory statistics course and have not been formally taught to use the statistical software package SPSS. If you are not over the age of 18, we do not require you to participate.

What is the project about? What are the questions being addressed?

This study is being conducted to investigate the different methods of teaching statistical software packages. We want to determine the most effective method available. This will help improve the learning outcomes in computer tutorials where statistical software packages are taught.

If I agree to participate, what will I be required to do?

By participating, you we will ask you to come to the university and attend a 1 hour tutorial in the computer lab. Please note that involvement in this study has no association with any courses you complete. In this session we will get you to complete some statistical analysis using the statistical package SPSS and also to answer some quick questions on how your session went. That will be the extent of your involvement. As part of your participation, we will also ask your permission to access the grade you received in your introductory statistics course. However, it is completely voluntary whether you choose to do so. But please remember that your involvement in this study and the subsequent collection of your grades and responses will be kept strictly confidential according to Australian privacy laws and university guidelines. Here at RMIT University, we take all legal and ethical matters relating to your confidentiality very seriously.

What are the risks or disadvantages associated with participation?

There are very few risks associated with your participation in this project. The most prominent risk being that your responses and grades will be known by the lead investigator. However, the lead investigator will never disclose, use or publish this sensitive information. In the event that you have concerns about your participation in the study you are encouraged to contact the lead investigator, James Baglin (Email: james.baglin@rmit.edu.au Ph: 9925-6118).

What are the benefits associated with participation?

Your participation in this project will go a long way in improving the methods by which statistical software packages are taught to future students. By participating, you will also be entered into a raffle for an awesome prize (The winner selects from either an IPod Touch, a Nintendo DS or a $200 Gift Voucher) to show our appreciation for your time.

What will happen to the information I provide?

The information gathered from this study will be kept strictly confidential. Only the lead investigator will have access to your identifying information. Your personal information will only be used to schedule sessions and access your grade. Your personal information will never be used or given to anyone else for any other purpose. This study is also only interested in looking at trends and not individual responses. You will never be identified as being a participant in this study.

Summarised and aggregated results from this study will appear in future reports and peer-reviewed publications. The investigators would be more than happy to send you a copy provided they have been completed. Just contact the lead investigator with your request.

What are my rights as a participant?

As a participant in this study you are ensured basic ethical rights. This includes the right to withdraw from the study at any given time without prejudice, the right to have any unprocessed data removed and destroyed provided it can be reliably identified, and the right to have any questions answered at any time. You can exercise your ethical rights by contacting the lead investigator.

Whom should I contact if I have any questions?

If you have any questions relating to the study please contact the lead investigator James Baglin (Email: james.baglin@rmit.edu.au Ph: 9925 6118).

If you have any concerns pertaining to the ethical conduct of this study you can directly contact the RMIT Human Research Ethics Committee (HREC) by telephone (03) 9925 2251.

Yours sincerely,

J.Baglin

James Baglin

BAppSc (Psych – Hons)
PhD Candidate (Statistics)
RMIT University, Plenty Road
Bundoora VIC 3083
Ph: 9925 6118
Email: james.baglin@rmit.edu.au

Clifford Da Costa

Dr. Cliff Da Costa

PhD (Statistics)
RMIT University, Plenty Road
Bundoora VIC 3083
Ph: 9925 6114
Email: cliff.dacosta@rmit.edu.au

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001.
Details of the complaints procedure are available at: http://www.rmit.edu.au/rd/hrec_complaints

A.3 Pilot Consent Form

RMIT UNIVERSITY

School of Mathematics and Geospatial Sciences

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2283
Fax +61 3 9925 2454

CONSENT TO PARTICIPATE IN A RESEARCH PROJECT

Project Title

Comparing Methods of Teaching Statistics Software Packages in Computer Tutorials

Investigators

- Mr. James Baglin (Lead Investigator: BAppSc, Psychology – Honours; PhD Candidate, Statistics, james.baglin@rmit.edu.au, 9925 6118)
- Dr. Cliff Da Costa (Project Supervisor: Associate Professor, SMGS, RMIT University)

It is very important that you consent to participate in this study. By signing the following line, you declare that you have read and fully understood the plain language statement given to you and you consent to participate in the study.

Portfolio	Science, Engineering and Health
School of	Mathematics and Geospatial Sciences
Project Title	Comparing Methods of Teaching Statistics Software Packages in Computer Tutorials
Lead Investigator	James Baglin (james.baglin@rmit.edu.au, Ph. 9925 6118)
Supervisor	Cliff Da Costa (cliff.dacosta@rmit.edu.au, Ph. 9925 6114)

I have received a statement explaining my involvement in this project.

1. I consent to participate in the above project, the particulars of which - including details of the interviews or questionnaires - have been explained to me.

I authorise the investigator or his or her assistant to interview me or administer a questionnaire.

2. I acknowledge that:
   (a) Having read Plain Language Statement, I agree to the general purpose, methods and demands of the study.
   (b) I have been informed that I am free to withdraw from the project at any time and to withdraw any unprocessed data previously supplied.
   (c) The project is for the purpose of research and/or teaching. It may not be of direct benefit to me.
   (d) The privacy of the personal information I provide will be safeguarded and only disclosed where I have consented to the disclosure or as required by law.
   (e) The security of the research data is assured during and after completion of the study. The data collected during the study may be published, and a report of the project outcomes will be provided to the School of Mathematics and Geospatial Sciences. Any information which will identify me will not be used.

Participant's Consent

(Participant's Signature) **(Date)**

Participant's First Name **Participants Last Name**

(Participant's Student Number)

A.4 Pilot Training Exercise Examples

Example of Pilot Exercise – Active-exploratory
Tutorial Activity 3 – Split the File and Selecting Cases

Many situations arise in data analysis when you need to compare groups or split analysis between groups. SPSS has two very useful features for doing this.

Task 3.1	Get descriptive statistics for current mean salary for the three Employment categories. Use the **Split File** command. What is the current mean salary for administrative employees?
Hint	Use **Data→Split File** by **Employment Category** before running descriptive statistics. Ensure you **turn off the Split File before continuing with other analysis**!
Answer:	

Task 3.2	What is the mean current salary of female managers?
Hint	Use **Data→Select Cases** command and then click **If** to set up a filter. Type in `gender = 'f' & jobcat = 3` to select only female managers in the dataset. Now do the descriptives. Ensure you **turn off the Select cases before continuing with other analysis**!
Answer:	

Task 3.2	What is the mean current salary of male managers?
Hint	Use the **Data→Select Cases** command and change the filter to look at only male managers. You will need to change the filter to do this.
Answer:	

Example of Pilot Exercise – Guided
Tutorial Activity 3 – Split the File and Selecting Cases

Many situation arise in data analysis when you need to compare groups or split analysis between groups. SPSS has two very useful features for doing this. The first feature is the **Split File** Command.

The **Split file** command separates your analysis between a grouping variable. For example, if you wanted to split your analysis between males and females or different age categories. To use the **Split File** feature to split our analysis between gender:

1. Click **Data**
2. Select **Split File**
3. Select **Compare Groups**
4. Move "**Gender**" into the "**Group based on**" box
5. Click **OK** to complete apply the split

These steps are summarised in Figure 8.

Now run some simple descriptive and you will see that the Split File command has been enabled. Ensure you turn this feature off when not needed!

Activity

Use the **Split File** command to split the dataset by Employment category. Then run descriptive statistics to find the current mean salary of administrative employees.

Your Answer:

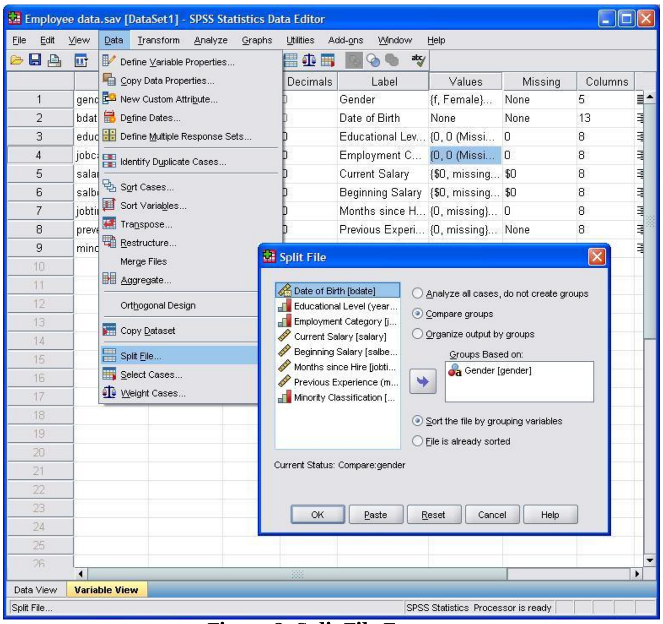


Figure 8: Split File Feature

Turn off the Split File before continuing.
The **Select Cases** feature is useful for conditionally selecting a group of cases from the dataset. For example, what if we wanted to know the current mean salary of female managers? We can use the **Select Cases** command to find out.

1. Click **Data**
2. Select **Select Cases**
3. Select **If Condition is satisfied** and the click **If**
4. Build the select condition by typing in: `gender = 'f' & jobcat = 3`
5. Click **Continue**
6. Click **OK** to complete apply the filter

These steps are summarised in Figures 9 and 10

The line of code tells SPSS to select all the cases that are female and have a job category of manager (which was coded as a 3 in the dataset). Once you apply this filter, any case that does not satisfy the condition is crossed out and excluded from future analysis. Be warned though. Make sure you turn **Select Cases** off when you no longer need it. This is done by following these steps:

1. Click **Data**
2. Select **Select Cases**
3. Select **All cases**
4. Click **OK**

Activity

Use the **Select Cases** command to find the mean current salary of male managers.

Your Answer:

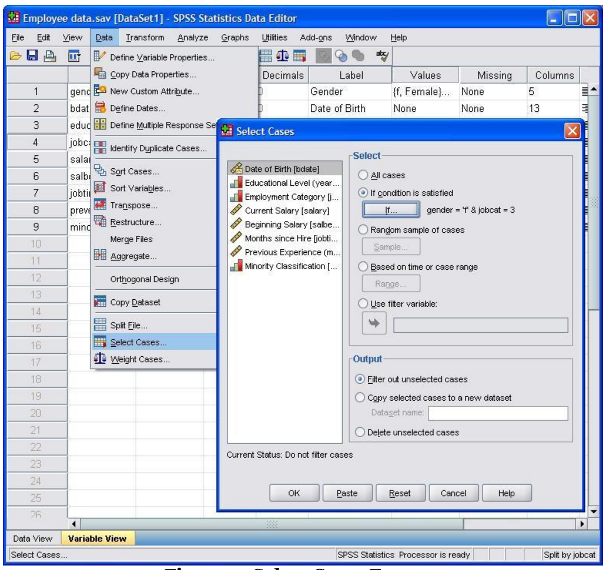


Figure 9: Select Cases Feature

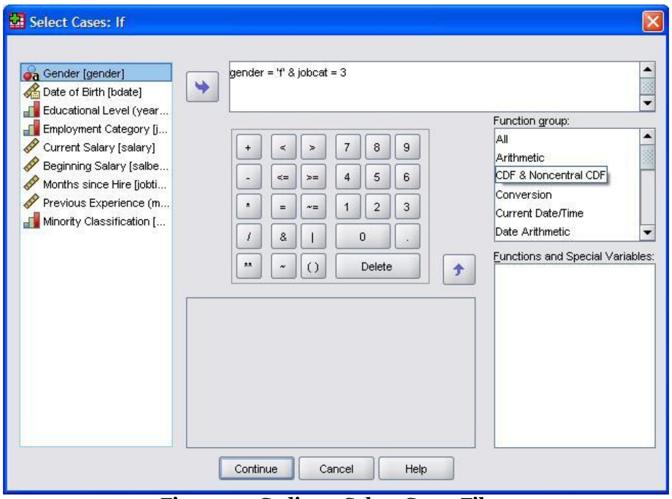


Figure 10: Coding a Select Cases Filter

Turn the select cases command off before continuing.

A.5 Trial I - Consent Form

RMIT Human Research Ethics Committee **HREC Form 2b**

Prescribed Consent Form For Persons Participating In Research Projects Involving Questionnaires or Disclosure of Personal Information

Portfolio	Science, Engineering and Health
School of	Mathematical and Geospatial Sciences
Name of participant:	**First Name:**
	Last Name:
	Student No:
Project Title:	Comparing Different Strategies of Learning How to Use Statistical Packages in Introductory Statistics Courses

Name(s) of investigators:	(1)	James Baglin	Phone:	9925 6118
	(2)	Dr. Cliff Da Costa	Phone:	9925 6114

1. I have received a statement explaining the questionnaire involved in this project.
2. I consent to participate in the above project, the particulars of which - including details of the questionnaires - have been explained to me.
3. I authorise the investigator or his or her assistant to administer a questionnaire.
4. I acknowledge that:

   (a) Having read Plain Language Statement, I agree to the general purpose, methods and demands of the study.
   (b) I have been informed that I am free to withdraw from the project at any time and to withdraw any unprocessed data previously supplied.
   (c) The project is for the purpose of research and/or teaching. It may not be of direct benefit to me.
   (d) The privacy of the personal information I provide will be safeguarded and only disclosed where I have consented to the disclosure or as required by law.
   (e) The security of the research data is assured during and after completion of the study. The data collected during the study may be published, and a report of the project outcomes will be provided to SMGS. Any information which will identify me will not be used.
   (f) Following the completion of the MATH1275/MATH1276 course, my grade will be recorded for the purposes of this study

Participant's Consent

Participant: ______________________ ***Date:*** ____________

(Signature)

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001. The telephone number is (03) 9925 2251. Details of the complaints procedure are available from the above address.

Human Research Ethics Committee, January 2011

A.6 Trial I - PLS

RMIT UNIVERSITY

RMIT Human Research Ethics Committee

School of Mathematics and Geospatial Sciences

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2283
Fax +61 3 9925 2454

INVITATION TO PARTICIPATE IN A RESEARCH PROJECT

PROJECT INFORMATION STATEMENT

Project Title

Comparing Different Strategies of Learning How to Use Statistical Packages in Introductory Statistics Courses

Investigators

- Mr. James Baglin (Lead Investigator: BAppSc, Psychology – Honours; PhD Candidate, Statistics, james.baglin@rmit.edu.au, 9925 6118)
- Dr. Cliff Da Costa (Project Supervisor: Associate Professor, SMGS, RMIT University)

Dear student,

You are invited to participate in a research project being conducted by RMIT University. This information sheet describes the project in straightforward language, or 'plain English'. Please read this sheet carefully and be confident that you understand its contents before deciding whether to participate. If you have any questions about the project, please ask one of the investigators.

Who is involved in this research project? Why is it being conducted?

This study is being conducted to investigate different strategies of learning to use the statistical package SPSS. We want to determine the most effective strategy available. This will help us design better introductory statistics courses and help better prepare you to use SPSS in the future. This study is being conducted as a part of a PhD in Statistics by James Baglin. The project is being supervised by Dr. Cliff Da Costa who is an Associate Professor in the School of Mathematics and Geospatial Sciences. The project has been approved by the RMIT Human Research Ethics Committee (HREC).

Why have you been approached?

As part of your regular involvement in MATH1275/MATH1276, you will participate in computer labs where you will learn to use the statistical package SPSS. SPSS will be used extensively throughout your undergraduate psychology career. It is important that you learn how to use this statistical package effectively.

What is the project about? What are the questions being addressed?

This study is being conducted to investigate different strategies of learning to use the statistical package SPSS. We want to determine the most effective strategy available. To do this we would like all students enrolled in MATH1275/MATH1276 to participate.

If I agree to participate, what will I be required to do?

By agreeing to participate, you will attend an allocated lab in which you will be trained to use the statistical package SPSS. The training strategies in these labs will differ. If you do not agree to be allocated by the course lecturer and instead would like to allocate yourself, please let the lecturer or tutor know. You will be allowed to allocate yourself to a lab provided you can be accommodated in the lab you choose (i.e. there are enough spaces available).

You will also be asked to fill out a few short questionnaires and carry on with the course as usual by attending scheduled computer labs. In fact, you may even forget that you are participating in an ongoing study. It is really that simple. We will then record how you progress through the semester by recording your computer lab work. Whether you chose to participate or not will have no impact on your mark. Participation is strictly voluntary. You may also withdraw from the study at any time. Information gathered in these labs will not be graded. Computer labs are only marked on participation.

We will also ask your permission to access the final grade for the course. Please remember that your involvement in this study and the subsequent collection of your grades and responses will be kept strictly confidential according to Australian privacy laws and university guidelines. We take our participant's privacy and confidentiality very seriously.

What are the risks or disadvantages associated with participation?

There are very few risks associated with your participation in this project. The most prominent risk being that your responses and grades will be known by the lead investigator. However, the lead investigator will never disclose, use or publish this sensitive information. In the event that you have concerns about your participation in the study you are encouraged to contact the lead investigator, James Baglin (Email: james.baglin@rmit.edu.au Ph: 9925-6118).

What are the benefits associated with participation?

There are no direct benefits associated with participation. However, your participation in this project will go a long way in improving the methods by which statistical packages are taught in future courses.

What will happen to the information I provide?

The information gathered from this study will be kept strictly confidential. Only the lead investigator and the project supervisor will have access to your identifying information. Your personal information will only be used to access your grade. Your personal information will never be used or given to anyone else for any other purpose. This study is only interested in looking at trends and not individual responses. You will never be identified as being a participant in this study. Data will be stored on a password protected RMIT computer and questionnaires will be locked in filing cabinets. Data and questionnaires will only be accessible by the lead investigator. Identifiable data will be destroyed after 5 years.

Summarised and aggregated results from this study will appear in future reports and peer-reviewed publications. The investigators would be more than happy to send you a copy provided they have been completed. Just contact the lead investigator with your request.

What are my rights as a participant?

As a participant in this study you are ensured basic ethical rights. This includes the right to withdraw from the study at any given time without prejudice, the right to have any unprocessed data removed and destroyed provided it can be reliably identified, and the right to have any questions answered at any time. You can exercise your ethical rights by contacting the lead investigator.

What other issues should I be aware of before deciding whether to participate?

If you are unable to contact the lead investigator or feel that you cannot talk or raise concerns with the lead investigator, you may contact the School of Mathematical and Geospatial Sciences Head of School, Professor John Hearne (Email: john.hearne@rmit.edu.au, Ph. 9925 2283).

Whom should I contact if I have any questions?

If you have any questions relating to the study please contact the lead investigator James Baglin.

Yours sincerely,

J. Baglin

James Baglin
PhD Candidate (Statistics)
RMIT University, Plenty Road
Bundoora VIC 3083
Ph: 9925 6118
Email: james.baglin@rmit.edu.au

Clifford Da Costa

Dr. Cliff Da Costa
PhD (Statistics)
RMIT University, Plenty Road
Bundoora VIC 3083
Ph: 9925 6114
Email: cliff.dacosta@rmit.edu.au

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001. Details of the complaints procedure are available at: http://www.rmit.edu.au/governance/complaints/research
HREC No: BSEHAPP 48 – 10 Date: 21/2/2011 Version No. 1

A.7 Trial I - Questionnaire Psychometrics

PCA Item Component Loadings for Scales used in SPT

Scale Items	Loading
Self-efficacy (Cronbach's $\alpha = .83$)	
1. To use the statistical package to compute descriptive statistics (e.g. compute means and standard deviations)	.888
2. To use the statistical package to create graphical displays of data (e.g. bar graphs and scatter plots	.895
3. To use the statistical package to conduct statistical inference (e.g. generate p-values)	.798
Anxiety (Cronbach's $\alpha = .74$)	
1. I felt tense when training to use SPSS	.796
2. I felt pressured when training to use SPSS	.857
3. I feel anxious when I need to use SPSS outside of training (e.g. Using SPSS for other courses)	.718
4. I feel relaxed when using SPSS outside of training (e.g. Using SPSS for other courses) (R)	.598
Metacognition (Cronbach's $\alpha = .91$)	
1. I revised my approach for completing statistical procedures in SPSS to deal with more complex tasks	.683
2. While completing statistical procedures in SPSS, I monitored how well I was learning to use SPSS by seeing how easy it was for me to complete each task	.654
3. I thought carefully about how well I completed previous statistical procedures in SPSS before moving onto other tasks	.675
4. As I practiced statistical procedures in SPSS, I evaluated how well I was learning to use SPSS by seeing how easy it was for me to complete each task	.730
5. When my methods were not successful for completing statistical procedures in SPSS, I experimented with different approaches for completing the procedure	.626

continued on next page

continued from previous page

Scale Items	Loading
6. I chose to concentrate more when conducting statistical procedures in SPSS to improve areas of weakness identified in previous tasks	.869
7. I chose to dedicate more effort to new statistical procedures in SPSS that would help me to learn more about the program	.617
8. As I practised statistical procedures in SPSS of different difficulty levels, I changed how I approached the task	.639
9. I tried to monitor closely the statistical procedures in SPSS where I needed the most practice	.737
10. I noticed where I made the most mistakes in SPSS during the computer labs and focused on improving those areas	.806
11. I put more effort into SPSS procedures that I found most difficult	.745
12. I used my ability to complete previous statistical procedures in SPSS to revise how I would approach future tasks	.694
Performance Utility (Cronbach's $\alpha = .94$)	
1. SPSS will be a useful tool for doing my statistical analysis	.885
2. It will be useful to have SPSS handy when I am working with or studying statistics	.909
3. I would welcome having my other courses use SPSS when doing statistic	.899
4. Many tasks, such as descriptive statistics, will be easier and faster using SPSS	.909
5. I will be able to do many statistical tasks more smoothly with SPSS than without it	.921
6. For a substantial part of my studies, SPSS will be a useful tool	.839
7. It would be difficult to imagine doing statistics without SPSS	.732
Error-framing - Learning from Errors (Cronbach's $\alpha = .86$)	
1. From my errors, I have learned a lot about how to work with SPSS	.790
2. When an error occurred, it was an important piece of information for using SPSS	.767
3. My errors have shown me what I can do better in SPSS	.833
4. Errors were helpful for me to improve my work with SPSS	.919
Error-framing - Error Strain (Cronbach's $\alpha = .80$)	
1. When I made a mistake in SPSS, I lost my temper and got angry about it	.443

continued on next page

continued from previous page

Scale Items	Loading
2. I was afraid of making errors when learning to use SPSS	.804
3. While working with SPSS, I was worried I could do something wrong	.843
4. When I made an error in SPSS, I was ashamed	.800
5. It was stressful to me when I made an error in SPSS	.774
Emotional Control (Cronbach's $\alpha = .88$)	
1. When difficulties arose during computer labs I did not allow myself to lose my composure	.779
2. When difficulties arose during computer labs I purposely continued to focus myself on the task	.791
3. When difficulties arose during computer labs I calmly considered how I could continue with the task	.692
4. When difficulties arose during computer labs I allowed myself to be distracted by worrisome thoughts (R)	.664
5. When difficulties arose during computer labs I let myself become distracted (R)	.786
6. When difficulties arose during computer labs I let myself be sidetracked from the task (R)	.738
7. When difficulties arose during computer labs I was able to focus all my attention on the task	.640
8. When difficulties arose during computer labs l was able to motivate myself to continue	.831
Active Exploration - Guided (Cronbach's $\alpha = .42$)	
1. I used step-by-step instructions when learning to use SPSS (R)	.692
2. I copied how other students completed tasks in SPSS (R)	.297
3. When I was unsure about how to complete a task in SPSS, I would immediately ask the tutor/or a friend for help	.600
Active Exploration - Active (Cronbach's $\alpha = .69$)	
4. I explored the features of SPSS without much instruction by changing options or trying different analyses in order to complete each lab exercise	.755
5. I tried to discover how to operate SPSS myself without any instruction	.842
6. I actively explored SPSS during computer labs	.671

A.8 Trial I - Qualitative Interview Schedule

Comparing Different Strategies of Learning How to Use Statistical Packages in Introductory Statistics Courses: Qualitative Interviews Questions

Participant:________________________ Condition:____________________

Interview Schedule

During MATH1275/MATH1276, you trained to use the statistical package SPSS during computer labs. I would like to ask you some questions relating to your experience of the computer lab training as well as your perceptions of using SPSS. Are there any questions before we begin the interview?

1. Do you think it was important to learn SPSS as part of your statistics course? Why or why not? (Performance utility)
2. What was your overall attitude of the SPSS training in the computer labs? Was it positive or negative? Why? (Intrinsic motivation)
3. How confident do you feel in your ability to use SPSS? Did training improve your confidence? Why or why not? (self-efficacy)
4. How did you find the difficulty of the training? Was it hard or easy? Why or why not (perceived difficulty)
5. During training, what were the typical emotions you experienced? Were these emotions distracting or beneficial to training? If they were distracting, did you use any strategies to overcome them? What were these strategies? (Anxiety, emotional control)
6. Did you find yourself relying on the help of the tutor of fellow students to complete each lab? If so, what types things did you get help with? (Manipulation checks)
7. When you made and error in SPSS or encountered a problem, what did you typically think to yourself? (Error framing)
8. How did typically work through problems in the labs? Did you use any strategies? (Emotional control – error framing, Meta-cognition, exploration)
9. Do you think the SPSS training prepared you for the self-assessment tasks? Why or why not? (Analogical and adaptive transfer)
10. If you were asked to do a statistical analysis in SPSS that was not covered during training, do you think you could figure it out for yourself? How would you go about figuring out how? (Adaptive transfer)
11. Do you think the SPSS training has prepared you for using SPSS outside of training, e.g. in other courses for lab reports and other assignments? Why or why not? (Transfer)
12. In an ideal world, what do you think would be the perfect way to train to use SPSS? (Expectations)
13. **EMT Only:** During training you were presented with some sayings regarding errors. Can you remember these sayings and if so can you recite them? (Manipulation check)
14. Is there anything else you would like to add in regards to the statistical package training?

A.9 Trial I - Qualitative Interview PLS

RMIT UNIVERSITY

RMIT Human Research Ethics Committee

School of Mathematics and Geospatial Sciences

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2283
Fax +61 3 9925 2454

INVITATION TO PARTICIPATE IN A RESEARCH PROJECT

PROJECT INFORMATION STATEMENT

Project Title

Comparing Different Strategies of Learning How to Use Statistical Packages in Introductory Statistics Courses: Qualitative Interviews

Investigators

- Mr. James Baglin (Lead Investigator: BAppSc, Psychology – Honours; PhD Candidate, Statistics, james.baglin@rmit.edu.au, 9925 6118)
- Dr. Cliff Da Costa (Project Supervisor: Associate Professor, SMGS, RMIT University)

Dear student,

You are invited to participate in a research project being conducted by RMIT University. This information sheet describes the project in straightforward language, or 'plain English'. Please read this sheet carefully and be confident that you understand its contents before deciding whether to participate. If you have any questions about the project, please ask one of the investigators.

Who is involved in this research project? Why is it being conducted?

This study is being conducted to investigate different strategies of learning to use the statistical package SPSS. We want to determine the most effective strategy available. This will help us design better introductory statistics courses and help better prepare you to use SPSS in the future. This study is being conducted as a part of a PhD in Statistics by James Baglin. The project is being supervised by Dr. Cliff Da Costa who is an Associate Professor in the School of Mathematics and Geospatial Sciences. The project has been approved by the RMIT Human Research Ethics Committee (HREC).

Why have you been approached?

As part of your regular involvement in MATH1275/MATH1276, you will participate in computer labs where you will learn to use the statistical package SPSS. You have also consented to participate in the first stage of this research earlier in the semester. We now want to talk to you about your experience in the computer labs.

What is the project about? What are the questions being addressed?

This study is being conducted to investigate different strategies of learning to use the statistical package SPSS. We want to understand your experience of the computer labs sessions in order to help design better learning strategies in future courses.

If I agree to participate, what will I be required to do?

By agreeing to participate, you will be interviewed by a researcher about your experience in the computer labs. The interview will take place at a time and place convenient to you. During the interview you will respond to questions about your experience of the computer labs. The interview should only take approximately 30 minutes of your time. The interview will be voice recorded for later transcription and analysis. Participation in the interview is strictly voluntary. You may also withdraw from the interview at any time. Your interview recording and transcript will be kept strictly confidential according to Australian privacy laws and university guidelines. We take our participant's privacy and confidentiality very seriously.

All participants in the interviews will receive a free Hoyts movie ticket to show our appreciation for your time.

What are the risks or disadvantages associated with participation?

There are very few risks associated with your participation in this project. The most prominent risk being that your personal information will be known by the lead investigator. However, the lead investigator will never disclose, use or publish this sensitive information. In the event that you have concerns about your participation in the study you are encouraged to contact the lead investigator, James Baglin (Email: james.baglin@rmit.edu.au Ph: 9925-6118).

What are the benefits associated with participation?

There are no direct benefits associated with participation. However, your participation in this project will go a long way in improving the methods by which statistical packages are taught in future courses.

What will happen to the information I provide?

The information gathered from the interview will be kept strictly confidential. However, any information that you provide can be disclosed only if (1) it is to protect you or others from harm, (2) a court order is produced, or (3) you provide the researchers with written permission. You will never be identified as being a participant in this study. Interview recordings will be stored on a password protected RMIT computer and transcripts will be locked in filing cabinets. Interviews and transcripts will only be accessible by the lead investigator. Identifiable data will be destroyed after 5 years.

Data gathered from the transcripts of interviews will appear in future reports and peer-reviewed publications. All data reported from the interviews will be de-identified for the purpose of reporting. The investigators would be more than happy to send you a copy of future reports provided they have been completed. Just contact the lead investigator with your request.

What are my rights as a participant?

As a participant in this study you are ensured basic ethical rights. This includes the right to withdraw from the study at any given time without prejudice, the right to have any unprocessed data removed and destroyed provided it can be reliably identified, and the right to have any questions answered at any time. You can exercise your ethical rights by contacting the lead investigator.

What other issues should I be aware of before deciding whether to participate?

If you are unable to contact the lead investigator or feel that you cannot talk or raise concerns with the lead investigator, you may contact the School of Mathematical and Geospatial Sciences Head of School, Professor John Hearne (Email: john.hearne@rmit.edu.au, Ph. 9925 2283).

Whom should I contact if I have any questions?

If you have any questions relating to the study please contact the lead investigator James Baglin.

Yours sincerely,

J. Baglin	Clifford Da Costa
James Baglin	Dr. Cliff Da Costa
PhD Candidate (Statistics)	PhD (Statistics)
RMIT University, Plenty Road	RMIT University, Plenty Road
Bundoora VIC 3083	Bundoora VIC 3083
Ph: 9925 6118	Ph: 9925 6114
Email: james.baglin@rmit.edu.au	Email: cliff.dacosta@rmit.edu.au

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001. Details of the complaints procedure are available at: http://www.rmit.edu.au/governance/complaints/research
HREC No: BSEHAPP 48 – 10 Date: 10/03/2011 Version No. 1

A.10 Trial II - PLS

RMIT UNIVERSITY

RMIT Human Research Ethics Committee

School of Mathematical and Geospatial Sciences

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2283
Fax +61 3 9925 2454

INVITATION TO PARTICIPATE IN A RESEARCH PROJECT

PROJECT INFORMATION STATEMENT

Project Title

Comparing Different Strategies of Learning How to Use Statistical Packages in Introductory Statistics Courses

Investigators

- Mr. James Baglin (Lead Investigator: BAppSc, Psychology – Honours; PhD Candidate, Statistics, james.baglin@rmit.edu.au, 9925 6118)
- Dr. Cliff Da Costa (Project Supervisor: Associate Professor, SMGS, RMIT University)

Dear student,

You are invited to participate in a research project being conducted by RMIT University. This information sheet describes the project in straightforward language, or 'plain English'. Please read this sheet carefully and be confident that you understand its contents before deciding whether to participate. If you have any questions about the project, please ask one of the investigators.

Who is involved in this research project? Why is it being conducted?

This study is being conducted to investigate different strategies of learning to use the statistical package SPSS. We want to determine the most effective strategy available. This will help us design better introductory statistics courses and help better prepare you to use SPSS in the future. This study is being conducted as a part of a PhD in Statistics by James Baglin. The project is being supervised by Dr. Cliff Da Costa who is an Associate Professor in the School of Mathematics and Geospatial Sciences. The project has been approved by the RMIT Human Research Ethics Committee (HREC).

Why have you been approached?

As part of your regular involvement in MATH1275/MATH1276, you will participate in computer labs where you will learn to use the statistical package SPSS. SPSS will be used extensively throughout your undergraduate psychology career. It is important that you learn how to use this statistical package effectively.

What is the project about? What are the questions being addressed?

This study is being conducted to investigate different strategies of learning to use the statistical package SPSS. We want to determine the most effective strategy available. To do this we would like all students enrolled in MATH1275/MATH1276 to participate.

If I agree to participate, what will I be required to do?

By agreeing to participate, all you need to do is carry on with the course. As part of your course you will attend computer labs where you will be trained to use the statistical package SPSS. You will also be asked to fill out a few short questionnaires at the beginning and end of semester. In fact, you may even forget that you are participating in an ongoing study. It is really that simple. We will then record how you progress through the semester by recording your computer lab work. Whether you chose to participate or not will have no impact on your mark. Participation is strictly voluntary. You may also withdraw from the study at any time. Information gathered in these labs will not be graded. Computer labs are only marked on participation.

We will also ask your permission to access the final grade for the course. Please remember that your involvement in this study and the subsequent collection of your grades and responses will be kept strictly confidential according to Australian privacy laws and university guidelines. We take our participant's privacy and confidentiality very seriously.

What are the risks or disadvantages associated with participation?

There are very few risks associated with your participation in this project. The most prominent risk being that your responses and grades will be known by the lead investigator. However, the lead investigator will never disclose, use or publish this sensitive information. In the event that you have concerns about your participation in the study you are encouraged to contact the lead investigator, James Baglin (Email: james.baglin@rmit.edu.au Ph: 9925-6118).

What are the benefits associated with participation?

There are no direct benefits associated with participation. However, your participation in this project will go a long way in improving the methods by which statistical packages are taught in future courses.

What will happen to the information I provide?

The information gathered from this study will be kept strictly confidential. Only the lead investigator and the project supervisor will have access to your identifying information. Your personal information will only be used to access your grade. Your personal information will never be used or given to anyone else for any other purpose. This study is only interested in looking at trends and not individual responses. You will never be identified as being a participant in this study. Data will be stored on a password protected RMIT computer and questionnaires will be locked in filing cabinets. Data and questionnaires will only be accessible by the lead investigator. Identifiable data will be destroyed after 5 years.

Summarised and aggregated results from this study will appear in future reports and peer-reviewed publications. The investigators would be more than happy to send you a copy provided they have been completed. Just contact the lead investigator with your request.

What are my rights as a participant?

As a participant in this study you are ensured basic ethical rights. This includes the right to withdraw from the study at any given time without prejudice, the right to have any unprocessed data removed and destroyed provided it can be reliably identified, and the right to have any questions answered at any time. You can exercise your ethical rights by contacting the lead investigator.

What other issues should I be aware of before deciding whether to participate?

If you are unable to contact the lead investigator or feel that you cannot talk or raise concerns with the lead investigator, you may contact the School of Mathematical and Geospatial

Sciences Head of School, Professor John Hearne (Email: john.hearne@rmit.edu.au, Ph. 9925 2283).

Whom should I contact if I have any questions?

If you have any questions relating to the study please contact the lead investigator James Baglin.

Yours sincerely,

J. Baglin

James Baglin
PhD Candidate (Statistics)
RMIT University, Plenty Road
Bundoora VIC 3083
Ph: 9925 6118
Email: james.baglin@rmit.edu.au

Clifford Da Costa

Dr. Cliff Da Costa
PhD (Statistics)
RMIT University, Plenty Road
Bundoora VIC 3083
Ph: 9925 6114
Email: cliff.dacosta@rmit.edu.au

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001. Details of the complaints procedure are available at: http://www.rmit.edu.au/governance/complaints/research
HREC No: BSEHAPP 48 – 10 Date: 02/11/2011 Version No. 2

A.11 Trial II - Consent Form

RMIT Human Research Ethics Committee **HREC Form 2b**

Prescribed Consent Form For Persons Participating In Research Projects Involving Questionnaires or Disclosure of Personal Information

Portfolio	Science, Engineering and Health
School of	Mathematical and Geospatial Sciences
Name of participant:	**First Name:**
	Last Name:
	Student No:
Project Title:	Comparing Different Strategies of Learning How to Use Statistical Packages in Introductory Statistics Courses

Name(s) of investigators:	(1)	James Baglin	Phone:	9925 6118
	(2)	Dr. Cliff Da Costa	Phone:	9925 6114

1. I have received a statement explaining the questionnaire involved in this project.
2. I consent to participate in the above project, the particulars of which - including details of the questionnaires - have been explained to me.
3. I authorise the investigator or his or her assistant to administer a questionnaire.
4. I acknowledge that:

   (a) Having read Plain Language Statement, I agree to the general purpose, methods and demands of the study.
   (b) I have been informed that I am free to withdraw from the project at any time and to withdraw any unprocessed data previously supplied.
   (c) The project is for the purpose of research and/or teaching. It may not be of direct benefit to me.
   (d) The privacy of the personal information I provide will be safeguarded and only disclosed where I have consented to the disclosure or as required by law.
   (e) The security of the research data is assured during and after completion of the study. The data collected during the study may be published, and a report of the project outcomes will be provided to SMGS. Any information which will identify me will not be used.
   (f) Following the completion of the MATH1275/MATH1276 course, my grade will be recorded for the purposes of this study

Participant's Consent

Participant: ____________________ ***Date:*** __________

(Signature)

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001. The telephone number is (03) 9925 2251. Details of the complaints procedure are available from the above address.

Human Research Ethics Committee, January 2012

A.12 Trial II - Performance Utility

Adapted Items (Richter et al., 2000)	Loading
Performance Utility (Cronbach's $\alpha = .88$)	
1. SPSS will be a useful tool for doing my statistical analysis	.828
2. It will be useful to have SPSS handy when I am working with or studying statistics	.852
3. I would welcome having my other courses use SPSS when doing statistic	.785
4. Many tasks, such as descriptive statistics, will be easier and faster using SPSS	.841
5. I will be able to do many statistical tasks more smoothly with SPSS than without it	.840
6. For a substantial part of my studies, SPSS will be a useful tool	.774
7. It would be difficult to imagine doing statistics without SPSS	.584

A.13 Trial II - *SPSS* Certification Tasks

Student Name:______________________________

Student No: _____________

SPSS Certification Assessment Task

MATH1275/1276 Semester 1, Week 12, 2012
Version A

Instructions

- Read these instructions carefully before beginning.
- Log into *WebLearn* and download the **"GSS 2010 Condensed.sav"** data file from under the Assignments tab. You will use this data file to complete each assessment task.
- On the next page, you will find SPSS output from 6 different analyses. Your job will be to use your knowledge of SPSS to **replicate this output as closely as possible**.
- This will be done under **exam conditions**. No talking and no assistance.
- You may use a copy of the **SPSS Quick Guide** for assistance. You can download a copy from under the Assignments tab in *WebLearn.*
- Your performance on this task will be graded on how closely you can replicate the output. The closer you get, the higher your competency will be graded.
- This lab is worth 5% participation. To get this grade you must demonstrate that you have attempted **at least 4 tasks** to the best of your ability. However, you should try to attempt all six.
- Don't worry if you cannot replicate the exercises perfectly. They have been designed to challenge you. Some tasks are more challenging than others. Try to get as close as possible.
- Don't take too long on any one exercise. Come back to difficult exercises if you have time.
- **Copy your single closest replication's** output of each exercise into a *Word* document. Label each output with the exercise number it refers to.
- **Only include your closest replication!** If you include more than one per exercise, only the first one will be assessed.
- Save the *Word* file containing your closest replications using your **Student Number**, e.g. "3110740.doc". **Upload** this word file under **Assignments** in ***WebLearn***.
- Ask the tutor if you have any questions.
- Answer the questions on page 2.
- Before leaving, have a tutor check that you have uploaded the document correctly.
- Hand this sheet back to the tutor with your name and student number printed at the top before leaving.

Turn to the next page to begin.

<u>Once you have finished the exercises</u>, come back and answer the following questions about your perceptions of the SPSS Certification Assessment Task. (Circle your responses)

1. Overall, how difficult did you find the exercises?

1. Extremely Easy	2.	3.	4.	5.	6.	7. Extremely Difficult

2. Overall, what level of anxiety did you experience while completing the exercises?

1. No anxiety	2.	3.	4.	5.	6.	7. Extreme Anxiety

3. To what extent did you feel the training and practice that you completed during the semester prepared you for these exercises?

1. Not at all	2.	3.	4.	5.	6.	7. Complete preparation

Turn to the next page to begin.

Exercise 1

Replicate this table showing the descriptive statistics of highest year of school completed between males and females.

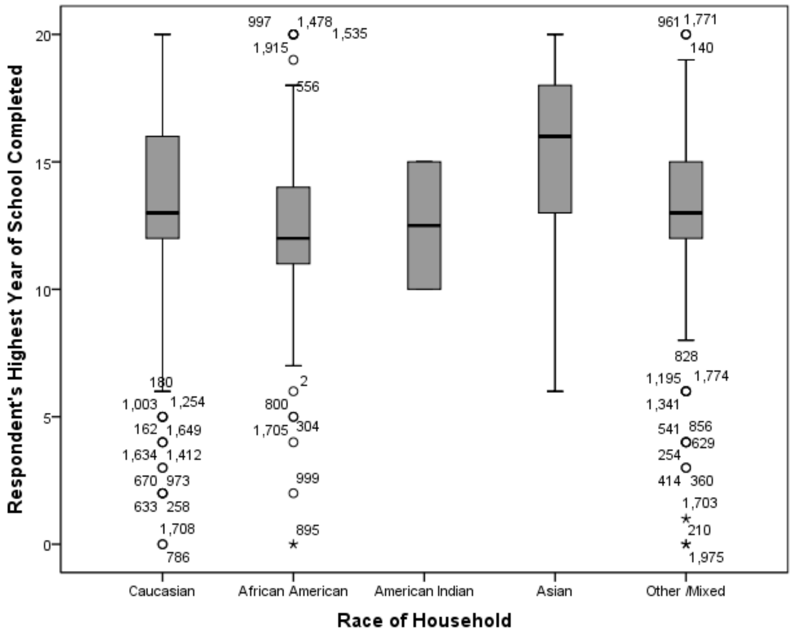
Report

Respondent's Highest Year of School Completed

Respondent's Gender	Mean	Median	N	Std. Deviation
Male	13.45	13.00	889	3.258
Female	13.47	13.00	1150	3.064
Total	13.46	13.00	2039	3.149

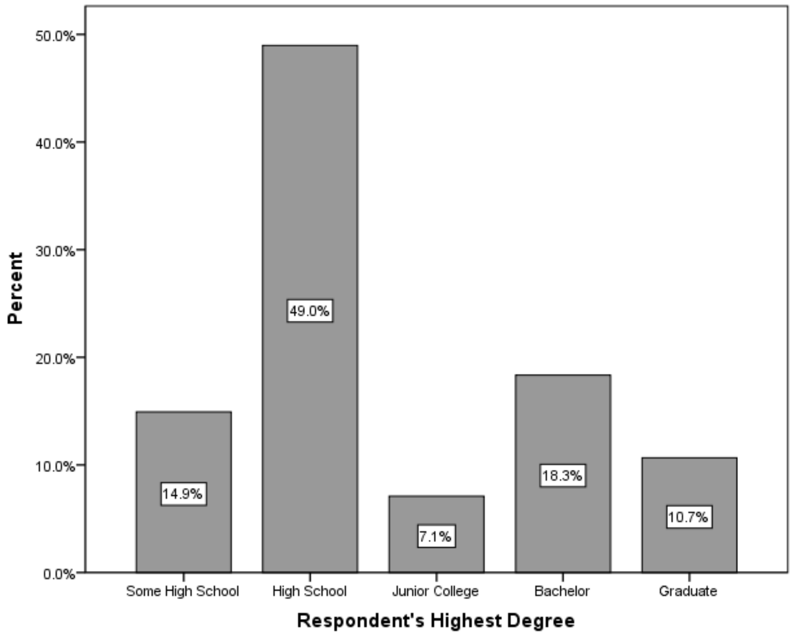
Exercise 2

Replicate this plot showing the distribution of highest year of school completed across race of household.



Exercise 3

Replicate the plot below showing the distribution of the highest level of education obtained by the sample.



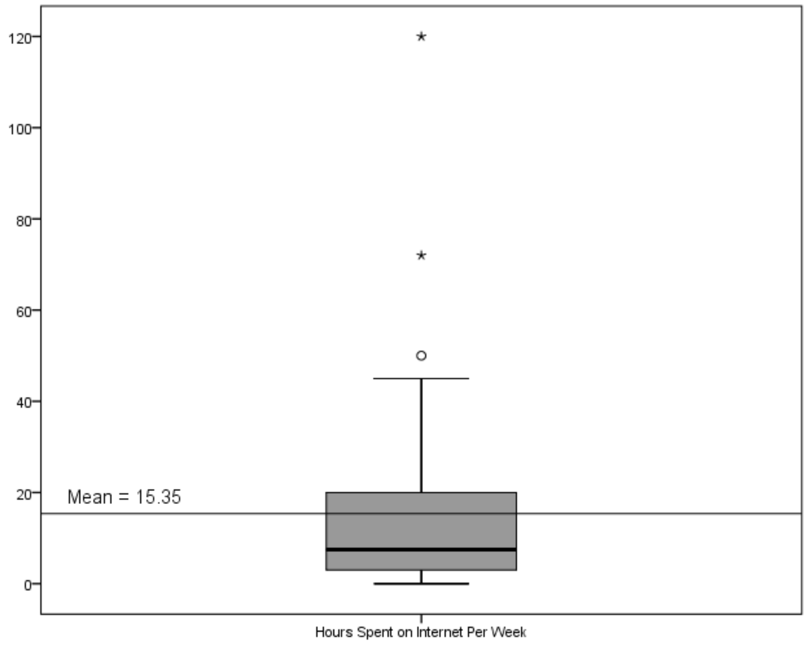
Exercise 4

Replicate the following custom table summarising the demographic characteristics of the survey sample.

			Respondent's Gender		
			Male	Female	Total
Respondent's Age	Mean		48	48	48
	SD		17	18	18
	N		890	1151	2041
Respondent's Highest Year of School Completed	Mean		13	13	13
	SD		3	3	3
	N		889	1150	2039
Estimated Family Income	Mean		50230	42015	45697
	SD		41175	37766	39532
	N		809	996	1805
Race of Household	Caucasian	Count	689	860	1549
		%	77.6%	74.8%	76.0%
	African American	Count	108	182	290
		%	12.2%	15.8%	14.2%
	American Indian	Count	0	2	2
		%	.0%	.2%	.1%
	Asian	Count	20	26	46
		%	2.3%	2.3%	2.3%
	Other /Mixed	Count	71	79	150
		%	8.0%	6.9%	7.4%

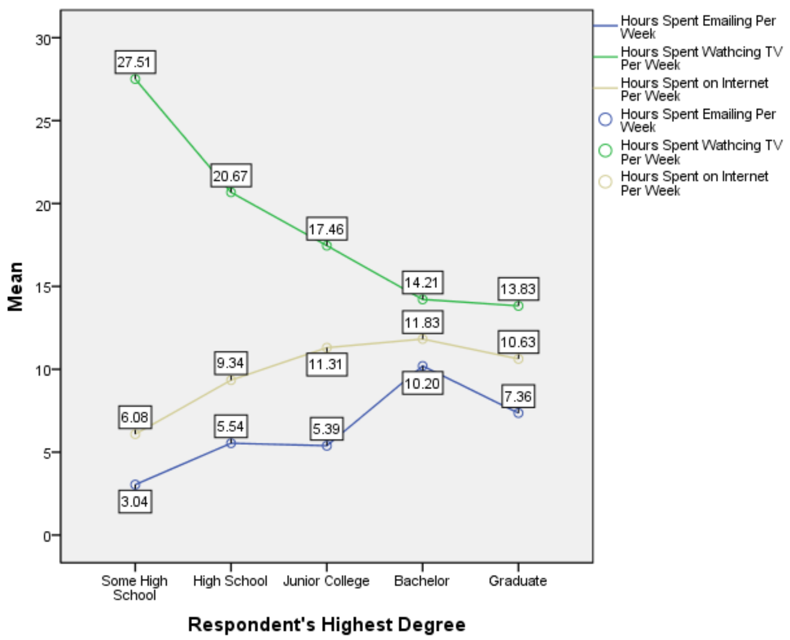
Exercise 5

Replicate the plot below showing the distribution of hours spent on the internet per week for males under the age of 25. The plot includes a reference line showing the location of the mean in comparison to the median.



Exercise 6

Replicate this plot showing the mean hours per week that respondents across different levels of education spent watching TV, using email and using the internet. The hours spent watching TV per week variable was calculated using the hours spent watching TV per day variable.



End of Assessment

Print your name on the front of this handout and answer the questions on page 2.

<u>Return this handout to a tutor before leaving.</u>

Ensure you upload your *Word* doc containing your best replications under the Assignments tab in *WebLearn*

SPSS Certification Assessment Task

MATH1275/1276 Semester 1, Week 12, 2012
Version B

Instructions

- Read these instructions carefully before beginning.
- Log into *WebLearn* and download the **"GSS 2010 Condensed.sav"** data file from under the Assignments tab. You will use this data file to complete each assessment task.
- On the next page, you will find SPSS output from 6 different analyses. Your job will be to use your knowledge of SPSS to **replicate this output as closely as possible**.
- This will be done under **exam conditions**. No talking and no assistance.
- You may use a copy of the **SPSS Quick Guide** for assistance. You can download a copy from under the Assignments tab in *WebLearn*.
- Your performance on this task will be graded on how closely you can replicate the output. The closer you get, the higher your competency will be graded.
- This lab is worth 5% participation. To get this grade you must demonstrate that you have attempted **at least 4 tasks** to the best of your ability. However, you should try to attempt all six.
- Don't worry if you cannot replicate the exercises perfectly. They have been designed to challenge you. Some tasks are more challenging than others. Try to get as close as possible.
- Don't take too long on any one exercise. Come back to difficult exercises if you have time.
- **Copy your single closest replication** output of each exercise into a *Word* document. Label each output with the exercise number it refers to.
- **Only include your closest replication!** If you include more than one per exercise, only the first one will be assessed.
- Save the *Word* file containing your closest replications using your **Student Number**, e.g. "3110740.doc". **Upload** this word file under **Assignments** in ***WebLearn***.
- Ask the tutor if you have any questions.
- Answer the questions on page 2.
- Before leaving, have a tutor check that you have uploaded the document correctly.
- Hand this sheet back to the tutor with your name and student number printed at the top before leaving.

Turn to the next page to begin.

<u>Once you have finished the exercises</u>, come back and answer the following questions about your perceptions of the SPSS Certification Assessment Task. (Circle your responses)

1. Overall, how difficult did you find the exercises?

1.	2.	3.	4.	5.	6.	7.
Extremely Easy						Extremely Difficult

2. Overall, what level of anxiety did you experience while completing the exercises?

1.	2.	3.	4.	5.	6.	7.
No anxiety						Extreme Anxiety

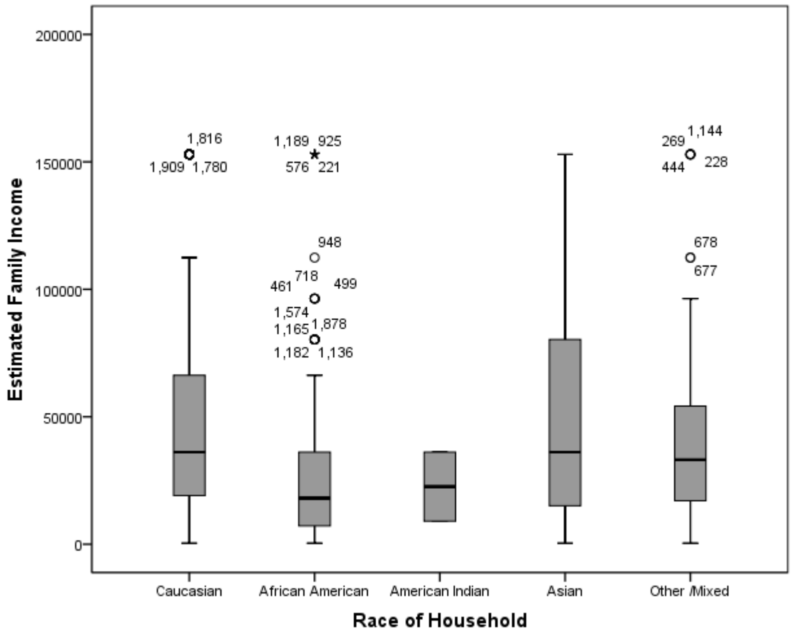
3. To what extent did you feel the training and practice that you completed during the semester prepared you for these exercises?

1.	2.	3.	4.	5.	6.	7.
Not at all						Complete preparation

Turn to the next page to begin.

Exercise 1

Replicate this plot showing the distribution of family income across race of household.



Exercise 2

Replicate this table showing descriptive statistics for respondent's age between males and females.

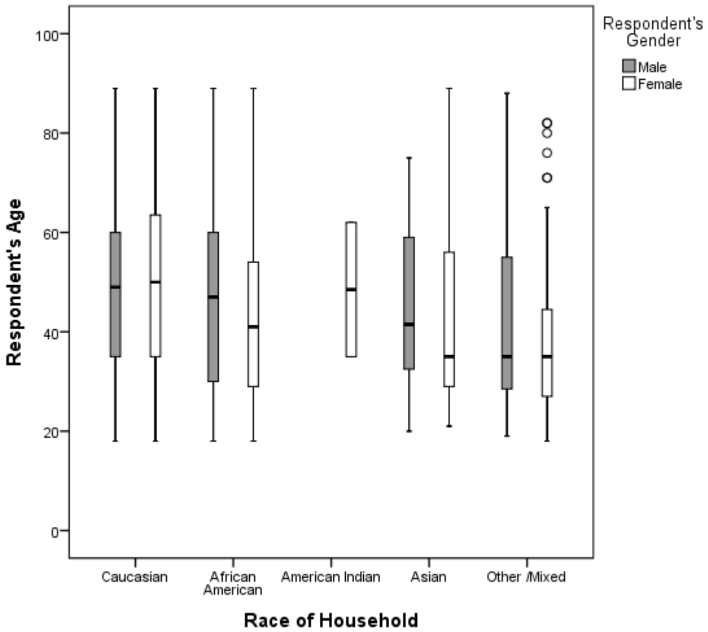
Report

Respondent's Age

Respondent's Gender	Mean	Median	N	Std. Deviation
Male	47.78	47.00	890	17.044
Female	48.11	47.00	1151	18.159
Total	47.97	47.00	2041	17.678

Exercise 3

Replicate this plot below showing the distribution of the sample's age across race of household and gender.



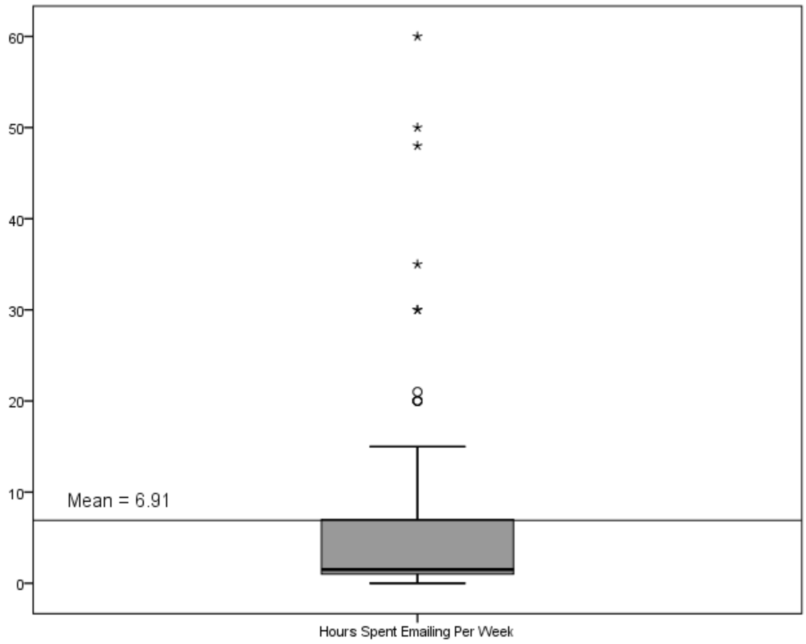
Exercise 4

Replicate the following custom table summarising the demographic characteristics of the survey sample.

			Respondent's Gender		
			Male	Female	Total
Race of Household	Caucasian	Count	689	860	1549
		%	44.5%	55.5%	100.0%
	African American	Count	108	182	290
		%	37.2%	62.8%	100.0%
	American Indian	Count	0	2	2
		%	0.0%	100.0%	100.0%
	Asian	Count	20	26	46
		%	43.5%	56.5%	100.0%
	Other /Mixed	Count	71	79	150
		%	47.3%	52.7%	100.0%
Respondent's Age	Mean		48	48	48
	SD		17	18	18
	N		890	1151	2041
Respondent's Highest Year of School Completed	Mean		13	13	13
	SD		3	3	3
	N		889	1150	2039
Estimated Family Income	Mean		50230	42015	45697
	SD		41175	37766	39532
	N		809	996	1805

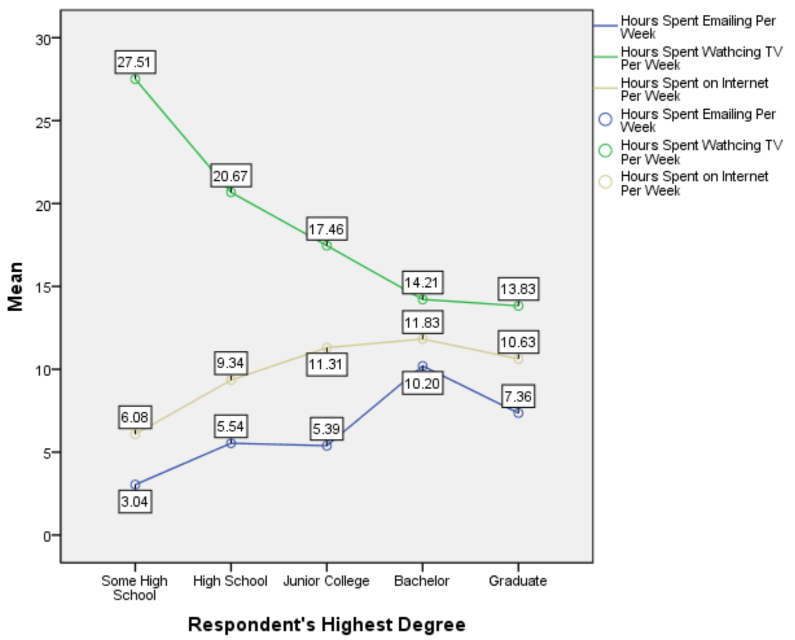
Exercise 5

Replicate the plot below showing the distribution of hours spent emailing per week for females under the age of 25. The plot includes a reference line showing the location of the mean in comparison to the median.



Exercise 6

Replicate this plot showing the mean hours per week that respondents across different levels of education spent watching TV, using email and using the internet. The hours spent watching TV per week variable was calculated using the hours spent watching TV per day variable.



End of Assessment

Print your name on the front of this handout and answer the questions on page 2.

<u>Return this handout to a tutor before leaving.</u>

Ensure you upload your *Word* doc containing your best replications under the Assignments tab in *WebLearn*

Student Name:_____________________________

Student No: _____________

SPSS Certification Assessment Task

MATH1275/1276 Semester 1, Week 12, 2012
Version C

Instructions

- Read these instructions carefully before beginning.
- Log into *WebLearn* and download the **"GSS 2010 Condensed.sav"** data file from under the Assignments tab. You will use this data file to complete each assessment task.
- On the next page, you will find SPSS output from 6 different analyses. Your job will be to use your knowledge of SPSS to **replicate this output as closely as possible**.
- This will be done under **exam conditions**. No talking and no assistance.
- You may use a copy of the **SPSS Quick Guide** for assistance. You can download a copy from under the Assignments tab in *WebLearn.*
- Your performance on this task will be graded on how closely you can replicate the output. The closer you get, the higher your competency will be graded.
- This lab is worth 5% participation. To get this grade you must demonstrate that you have attempted **at least 4 tasks** to the best of your ability. However, you should try to attempt all six.
- Don't worry if you cannot replicate the exercises perfectly. They have been designed to challenge you. Some tasks are more challenging than others. Try to get as close as possible.
- Don't take too long on any one exercise. Come back to difficult exercises if you have time.
- **Copy your single closest replication** output of each exercise into a *Word* document. Label each output with the exercise number it refers to.
- **Only include your closest replication!** If you include more than one per exercise, only the first one will be assessed.
- Save the *Word* file containing your closest replications using your **Student Number**, e.g. "3110740.doc". **Upload** this word file under **Assignments** in ***WebLearn***.
- Ask the tutor if you have any questions.
- Answer the questions on page 2.
- Before leaving, have a tutor check that you have uploaded the document correctly.
- Hand this sheet back to the tutor with your name and student number printed at the top before leaving.

Turn to the next page to begin.

<u>Once you have finished the exercises</u>, come back and answer the following questions about your perceptions of the SPSS Certification Assessment Task. (Circle your responses)

1. Overall, how difficult did you find the exercises?

1. Extremely Easy	2.	3.	4.	5.	6.	7. Extremely Difficult

2. Overall, what level of anxiety did you experience while completing the exercises?

1. No anxiety	2.	3.	4.	5.	6.	7. Extreme Anxiety

3. To what extent did you feel the training and practice that you completed during the semester prepared you for these exercises?

1. Not at all	2.	3.	4.	5.	6.	7. Complete preparation

Turn to the next page to begin.

Exercise 1

Replicate this table showing the descriptive statistics of the respondent's highest year of school completed across race of household.

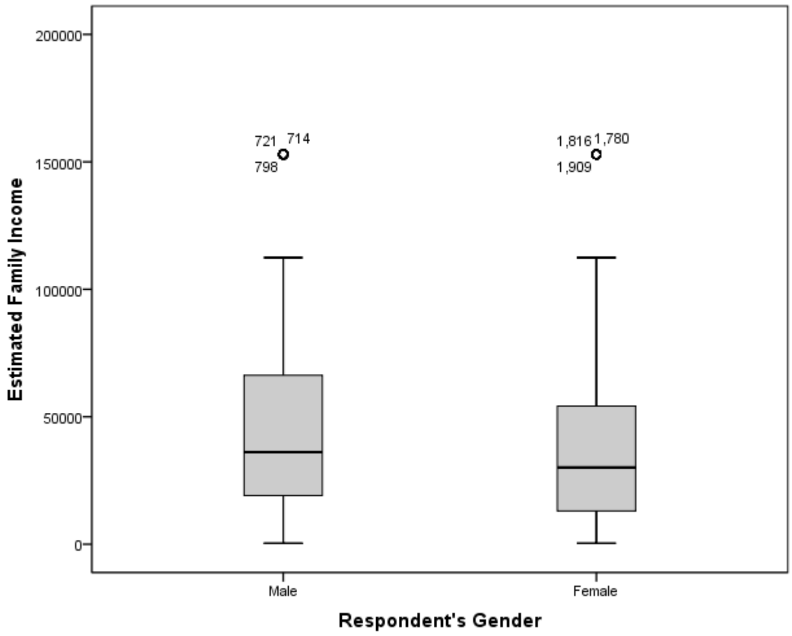
Report

Respondent's Highest Year of School Completed

Race of Household	Mean	Median	N	Std. Deviation
Caucasian	13.65	13.00	1545	3.083
African American	12.72	12.00	290	2.804
American Indian	12.50	12.50	2	3.536
Asian	15.20	16.00	45	3.488
Other /Mixed	12.45	13.00	150	3.746
Total	13.46	13.00	2032	3.145

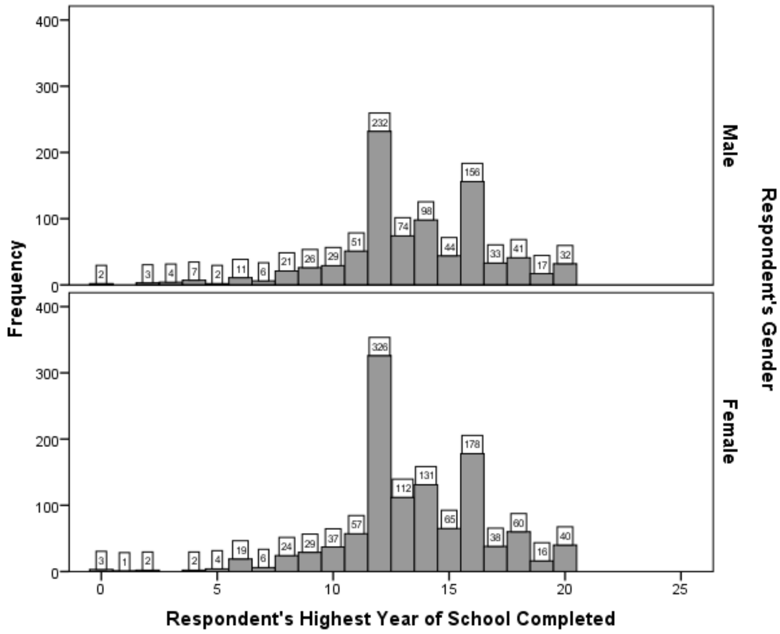
Exercise 2

Replicate this plot showing the distribution of household income across gender.



Exercise 3

Replicate this plot below showing the distribution of the sample's highest year of school completed between gender.



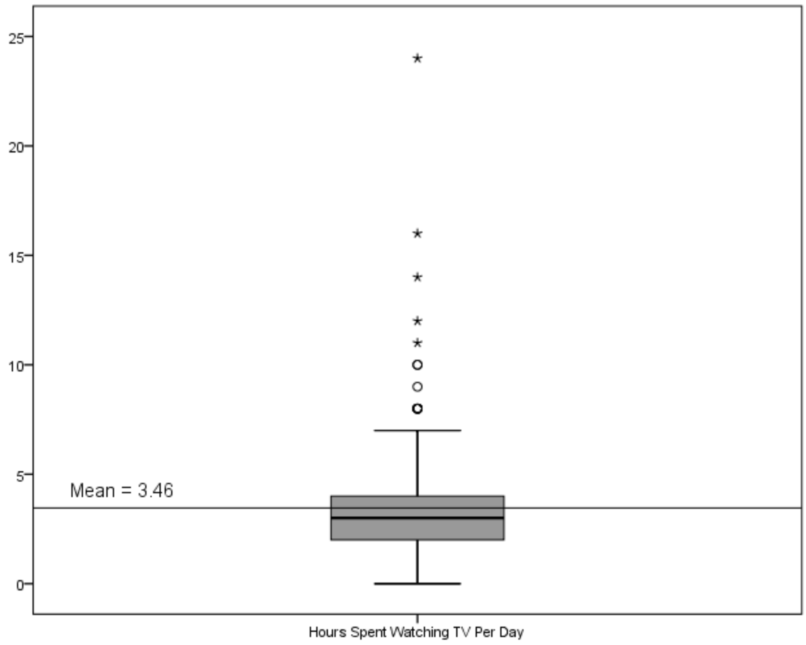
Exercise 4

Replicate the following custom table summarising the demographic characteristics of the survey sample.

			Respondent's Gender								
			Male			Female			Total		
			M	SD	N	M	SD	N	M	SD	N
Respondent's Age	Race of Household	Caucasian	49	17	689	50	18	860	49	18	1549
		African American	47	18	107	43	17	180	45	17	287
		Asian	46	17	20	44	21	26	45	19	46
		Other /Mixed	41	18	71	38	16	79	40	17	150
Respondent's Highest Year of School Completed	Race of Household	Caucasian	14	3	687	14	3	858	14	3	1545
		African American	12	3	108	13	3	182	13	3	290
		Asian	16	3	20	15	4	25	15	3	45
		Other /Mixed	12	4	71	13	3	79	12	4	150
Estimated Family Income	Race of Household	Caucasian	52746	41558	625	46442	39024	740	49328	40312	1365
		African American	32283	35420	95	25557	26320	161	28053	30129	256
		Asian	58857	42931	19	50459	54819	23	54258	49392	42
		Other /Mixed	50633	39196	68	30468	25931	66	40701	34712	134

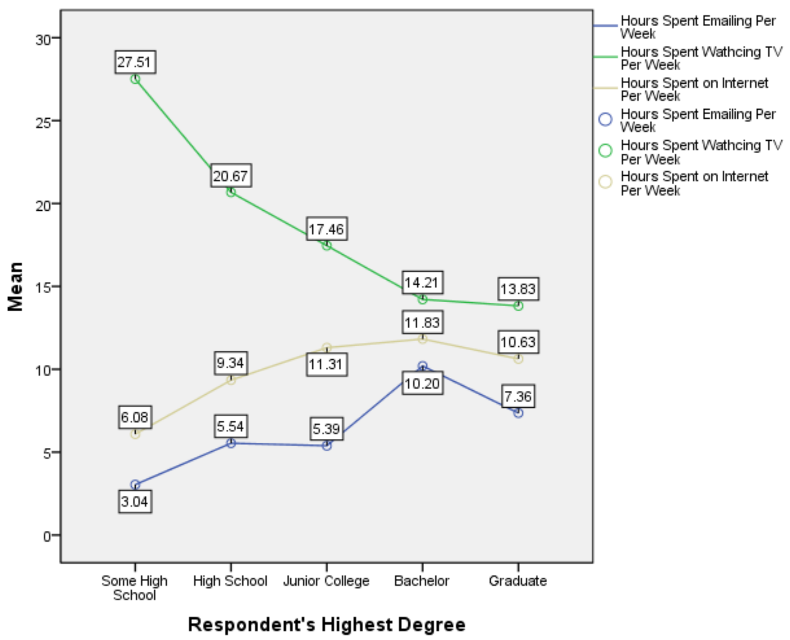
Exercise 5

Replicate the plot below showing the distribution of hour spent watching TV per day for females over the age of 60. The plot includes a reference line showing the location of the mean in comparison to the median.



Exercise 6

Replicate this plot showing the mean hours per week that respondents across different levels of education spent watching TV, using email and using the internet. The hours spent watching TV per week variable was calculated using the hours spent watching TV per day variable.



End of Assessment

Print your name on the front of this handout and answer the questions on page 2.

Return this handout to a tutor before leaving.

Ensure you upload your *Word* doc containing your best replications under the Assignments tab in *WebLearn*

8

A.14 Trial II - *SPSS* Certification Task Scoring

A.14.1 Version A

Version A Scoring Code

Question	Description/Criteria	Marks
1a	Compare means used with correct variables – schooling and gender	/1
	Median added	/2
	Median inserted between Mean and *N*	/1
2a	Boxplot with correct variables – highest year of school completed by race	/2
3a	Created bar chart	/1
	Y axis shows %	/1
	Correct variables used	/1
	Value labels added	/2
4a	Age, year of schooling, family income and race of household included	/1
	Table split by gender	/2
	Total column included	/1
	Column % included for categorical variables	/2
	SD and valid N included	/1
	Mean, SD and N relabelled	/2
	Statistics positioned as rows	/1
5a	Correctly selected cases (Males < 25 years) or (Select < 25 & Split file)	/2
	Create boxplot of filtered hours spent on Internet	/1
	Add reference line for mean	/2
	Add label for reference line	/2
	Labels removed	/2
6a	Hours watching TV per week converted to hours per week	/2
	New variable labelled correctly	/1
	Line plot with highest degree on x axis	/1
	Multiple lines for each variable on one plot	/1
	Markers added	/1
	Labels added	/2
	Adaptive Transfer Total[1]:	/32

[1] Only questions 3 - 6 were included for adaptive transfer scores.

A.14.2 Version B

Version B Scoring Code

Question	Description/Criteria	Marks
1b	Boxplot with correct variables – Income by race	/2
2b	Compare means used with correct variables – age and gender	/1
	Median added	/2
	Median inserted between Mean and N	/1
3b	Y axis shows respondent's age	/1
	X axis shows race of household	/1
	Clustered by male and female	/2
	Labels removed	/1
4b	Age, year of schooling, family income and race of household included	/1
	Table split by gender	/1
	Total column included	/2
	Column % included for categorical variables	/2
	SD and valid N included	/1
	Mean, SD and N relabelled	/2
	Statistics positioned as rows	/1
5b	Correctly selected cases (Females < 25 years) or (Select < 25 & Split file)	/2
	Created boxplot of filtered hours spent emailing	/1
	Add reference line for mean	/2
	Add label for reference line	/2
	Labels removed	2
6b	Hours watching TV per week converted to hours per week	/2
	New variable labelled correctly	/1
	Line plot with highest degree on x axis	/1
	Multiple lines for each variable on one plot	/1
	Markers added	/1
	Labels added	/2
	Adaptive Transfer Total[1]:	/32

[1] Only questions 3 - 6 were included for adaptive transfer scores.

A.14.3 Version C

Version C Scoring Code

Question	Description/Criteria	Marks
1c	Compare means used with correct variables – schooling and race	/1
	Median added	/2
	Median inserted between Mean and N	/1
2c	Boxplot with correct variables – income by gender	/2
3c	Created histogram of highest year of school completed	/1
	Panelled by gender	/2
	X axis starts at 0	/1
	Labels added	/1
4c	Age, year of schooling, family income included	/1
	Row variables split by race of household	/2
	Table split by gender in columns	/1
	Total column included	/2
	SD and valid N included	/1
	Mean, SD and N relabelled	/2
	Statistics positioned as columns	/1
5c	Correctly selected cases (Females > 60 years) or (Select > 60 & Split file)	/2
	Create boxplot of filtered hours spent watching TV per day	/1
	Add reference line for mean	/2
	Add label for reference line	/2
	Labels removed	/2
6c	Hours watching TV per week converted to hours per week	/2
	New variable labelled correctly	/1
	Line plot with highest degree on x axis	/1
	Multiple lines for each variable on one plot	/1
	Markers added	/1
	Labels added	/2
	Adaptive Transfer Total[1]:	/32

[1] Only questions 3 - 6 were included for adaptive transfer scores.

Appendix B

Part II

B.1 Cognitive Conflict Study PLS and Consent

RMIT UNIVERSITY

RMIT Human Research Ethics Committee

School of Mathematics and Geospatial Sciences

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2283
Fax +61 3 9925 2454

INVITATION TO PARTICIPATE IN A RESEARCH PROJECT

PROJECT INFORMATION STATEMENT

Project Title

Evaluating the Outcomes of an Introductory Course in Statistics

Investigators

- Mr. James Baglin (Lead Investigator: BAppSc, Psychology – Honours; PhD Candidate, Statistics, james.baglin@rmit.edu.au, 9925 6118)
- Dr. Cliff Da Costa (Project Supervisor: Associate Professor, SMGS, RMIT University)

Dear student,

You are invited to participate in a research project being conducted by RMIT University. This information sheet describes the project in straightforward language, or 'plain English'. Please read this sheet carefully and be confident that you understand its contents before deciding whether to participate. If you have any questions about the project, please ask one of the investigators.

Who is involved in this research project? Why is it being conducted?

This project is being conducted as a part of a PhD in Statistics by James Baglin. The project is being supervised by Dr. Cliff Da Costa who is an Associate Professor in the School of Mathematics and Geospatial Sciences. The project has been approved by the RMIT Human Research Ethics Committee (HREC). The study is being conducted to evaluate the outcomes of a newly developed introductory statistics course.

Why have you been approached?

You have been approached to participate in the study because you are a RMIT university student over the age of 18 and enrolled in an introductory course in statistics. If you are not over the age of 18, we do not require you to participate.

What is the project about? What are the questions being addressed?

This study is being conducted to investigate the outcomes of a recently developed introductory course in statistics. We need you, the student, to help us in determining whether this new course is effective in teaching statistics.

If I agree to participate, what will I be required to do?

By participating, all you will be required to do is fill in a questionnaire. The questionnaires will only take between 10 to 15 minutes to complete. We need your responses so that we can measure the effectiveness of the course and make improvements in the future. We also wish to match your questionnaires together with your grades at the end of the semester. This will require you to give us permission to record your name and grades. However, it is completely voluntary whether you choose to do so. But please remember that your name, responses, and grades will be kept strictly confidential according to Australian privacy laws and university guidelines.

What are the risks or disadvantages associated with participation?

There are very few risks associated with your participation in this project. The most prominent risk being that your responses and grades will be known by the lead investigator who will be involved in the course. However, the lead investigator will not analyse the results until you have finished the course and received your official grade. Therefore, the lead investigator will have no idea if you have participated in the project until you have finished. In the event that you have concerns about your participation in the study you are encouraged to contact the lead investigator, James Baglin (Email: james.baglin@rmit.edu.au Ph: 9925-6118).

What are the benefits associated with participation?

Your participation in this project will go a long way in improving the delivery of introductory statistics courses to future cohorts of students. By participating, you will also be entered into a raffle for two Hoyts movie tickets to show our appreciation for your time.

What will happen to the information I provide?

The information gathered from this study will be kept strictly confidential. Only the lead investigator will have access to your identifying information. Your personal information will only be used in the process of matching responses to the questionnaire with your grades. Your personal information will never be used or given to anyone else for any other purpose. This study is also only interested in looking at trends and not individual responses. You will never be identified as being a participant in this study.

Summarised and aggregated results from this study will appear in future reports and peer-reviewed publications. The investigators would be more than happy to send you a copy provided they have been completed. Just contact the lead investigator with your request.

What are my rights as a participant?

As a participant in this study you are ensured ethical rights. This includes the right to withdraw from the study at any given time without prejudice, the right to have any unprocessed data removed and destroyed provided it can be reliably identified, and the right to have any questions answered at any time. You can exercise your ethical rights by contacting the lead investigator.

Whom should I contact if I have any questions?

If you have any questions relating to the study please contact the lead investigator James Baglin (Email: james.baglin@rmit.edu.au Ph: 9925 6118).

Yours sincerely,

J.Baglin

James Baglin
PhD Candidate (Statistics)
RMIT University, Plenty Road
Bundoora VIC 3083
Ph: 9925 6118
Email: james.baglin@rmit.edu.au

Clifford Da Costa

Dr. Cliff Da Costa
PhD (Statistics)
RMIT University, Plenty Road
Bundoora VIC 3083
Ph: 9925 6114
Email: cliff.dacosta@rmit.edu.au

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001. Details of the complaints procedure are available at: http://www.rmit.edu.au/governance/complaints/research

HREC No: BSEHAPP 64 – 09 Date: 27/11/2009 Version No. 1

RMIT Human Research Ethics Committee **HREC Form 2b**

Prescribed Consent Form For Persons Participating In Research Projects Involving Questionnaires or Disclosure of Personal Information

Portfolio Science, Engineering and Health
School of Mathematical and Geospatial Sciences

Name of participant:

Your First Name:		**Please Fill in your details**
You Last Name:		
Your Student No:		

Project Title: **Evaluating the Outcomes of an Introductory Course in Statistics**

Name(s) of investigators: (1) James Baglin Phone: 9925 6118
(2) Dr. Cliff Da Costa Phone: 9925 6114

1. I have received a statement explaining the questionnaire involved in this project.
2. I consent to participate in the above project, the particulars of which - including details of the questionnaires - have been explained to me.
3. I authorise the investigator or his or her assistant to administer a questionnaire.
4. I acknowledge that:

   (a) Having read Plain Language Statement, I agree to the general purpose, methods and demands of the study.
   (b) I have been informed that I am free to withdraw from the project at any time and to withdraw any unprocessed data previously supplied.
   (c) The project is for the purpose of research and/or teaching. It may not be of direct benefit to me.
   (d) The privacy of the personal information I provide will be safeguarded and only disclosed where I have consented to the disclosure or as required by law.
   (e) The security of the research data is assured during and after completion of the study. The data collected during the study may be published, and a report of the project outcomes will be provided to SMGS. Any information which will identify me will not be used.
   (f) Following the completion of the MATH1238 course, my grade will be recorded for the purposes of this study

Participant's Consent

Participant: ______________________ ***Date:*** __________ SIGN HERE
(Signature)

Participants should be given a photocopy of this consent form after it has been signed.

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001. The telephone number is (03) 9925 2251. Details of the complaints procedure are available from the above address.

Human Research Ethics Committee, January 2011

B.2 Multiple Choice Exam Questions

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

The adapted CAOS items have been withheld from the online version of this dissertation. This was done to maintain the integrity of the CAOS items which may be used for assessment purposes in other statistics courses.

Contact the author to request a copy of the adapted items.

B.3 Adapted Multiple Choice Items

Adapted Multiple Choice Items

MC Exam[a]	COAS[b]	Scale	CC Question.	Modifications
Q1	1	B		Same as original
Q2	2	B		Added choice d)
Q3	N/A	B		Replaced CAOS 4 Question 6
Q4	3	CC	Distributions I	Same as original
Q5	4	CC	Distributions II	Same as original
Q6	5	CC	Distributions III	Same as original
Q7	N/A	CC	Confidence Intervals I	Adapted from ARTIST Scale - Tests of Significance Question 4 Replaced CAOS Question 7
Q8	N/A	CC	Sampling Distributions I	New question, replaced Question 9
Q9	8	B		Same as original
Q10	20	B		Same as original
Q11	N/A	B		Replaced CAOS 4 Question 10
Q12	11	B		Same as original
Q13	12	B		Same as original
Q14	13	B		Same as original

continued on next page

continued from previous page

MC Exam[a]	COAS[b]	Scale	CC Question.	Modifications
Q15	14	B		Reworded context of question
Q16	15	B		Reworded context of question
Q17	17	CC	Probability	Reworded context of question
Q18	18	B		Same as original
Q19	19	B		Same as original
Q20	N/A	B		Replaced CAOS Question 16
Q21	N/A	B		Replaced CAOS Question 20
Q22	N/A	B		Replaced CAOS Question 21
Q23	23	B		Same as original
Q24	24	B		Same as original
Q25	25	CC	p-values I	Same as original
Q26	26	CC	p-values II	Same as original
Q27	27	CC	p-values III	Same as original
Q28	28-31	CC	Confidence Intervals II	Combined CAOS Questions 28 - 31 together
Q29	N/A	CC	Hypothesis Testing I	Replaced Question 29
Q30	N/A	CC	Hypothesis Testing II	Replaced Question 30
Q31	22	CC	Correlation	Reworded
Q32	32	B		Reworded
Q33	N/A	B		New question, replaced CAOS Question 31

continued on next page

continued from previous page

MC Exam[a]	COAS[b]	Scale	CC Question.	Modifications
Q34	34	CC	Sampling Distributions II	Same as original
Q35	35	CC	Sampling Distributions III	Same as original
Q36	36	B		Same as original
Q37	39	CC	Regression	Reworded
Q38	N/A	CC	Confidence Intervals III	New question, Replaced Question 37
Q39	N/A	B		Adapted from ARTIST Website (Item ID = Q0464), replaced CAOS Question 38
Q40	40	CC	Hypothesis Testing III	Same as original

[a] See Appendix B.2, [b] Available from https://apps3.cehd.umn.edu/artist/index.html (delMas et al., 2007)

B.4 Cognitive Conflict-based Activity Slides

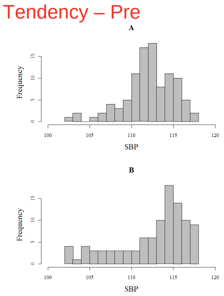
Central Tendency - Pilot

Measures of Central Tendency – Pre

Two histograms showing the distribution of 100 SBP readings from two samples are shown to the right. The descriptive statistics calculated from one of these samples are as follows.

Statistic	
Mean	113
Median	115
SD	4.26
Min	102
Max	118

Which sample have these statistics most likely come from?



School of Mathematical and Geospatial Sciences 1

Measures of Central Tendency – Pre

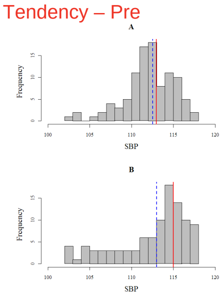
- Responses and Answer

School of Mathematical and Geospatial Sciences 2

Measures of Central Tendency – Pre

Statistic	A
Mean (---)	112.6
Median (—)	113
SD	3.03
Min	102
Max	118

Statistic	B
Mean (---)	113
Median (—)	115
SD	4.26
Min	102
Max	118



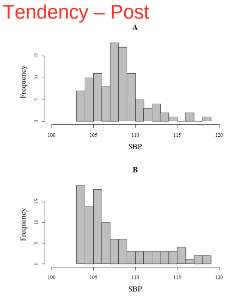
School of Mathematical and Geospatial Sciences 3

Measures of Central Tendency – Post

Two histograms showing the distribution of 100 SBP readings taken from two samples are shown to the right. The descriptive statistics calculated from one of these samples are as follows.

Statistic	
Mean	108.4
Median	108
SD	3.03
Min	103
Max	119

Which sample have these statistics most likely come from?



School of Mathematical and Geospatial Sciences 4

Measures of Central Tendency – Post

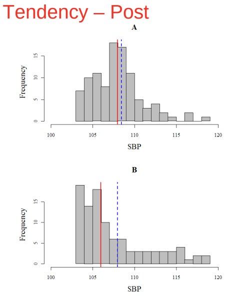
- Responses and Answer

School of Mathematical and Geospatial Sciences 5

Measures of Central Tendency – Post

Statistic	A
Mean (---)	108.4
Median (—)	108
SD	3.03
Min	103
Max	119

Statistic	B
Mean (---)	108
Median (—)	106
SD	4.26
Min	103
Max	119



School of Mathematical and Geospatial Sciences 6

1

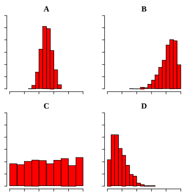
Distributions

Distributions – Pre 1

The histograms of four different random variables are shown to the right.

Which histogram is most likely to represent a distribution of **tests scores for a very difficult test**?

1. Histogram A
2. Histogram B
3. Histogram C
4. Histogram D



Distributions – Pre 1

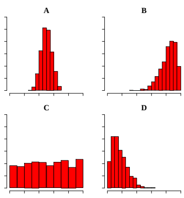
- Responses and Answer

Distributions – Pre 2

The histograms of four different random variables are shown to the right.

Which histogram is most likely to represent a distribution of **vertical jump heights** measured on a random sample of males?

1. Histogram A
2. Histogram B
3. Histogram C
4. Histogram D



Distributions – Pre 2

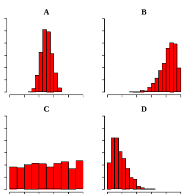
- Responses and Answer

Distributions – Pre 3

The histograms of four different random variables are shown to the right.

Which histogram is most likely to represent a distribution of the **last digit of a credit card number**?

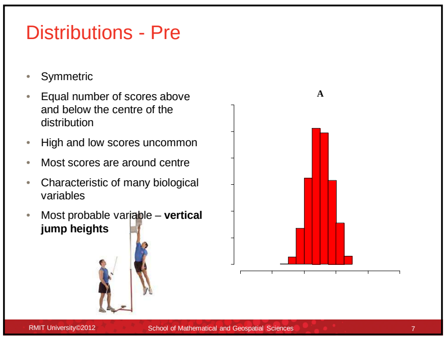
1. Histogram A
2. Histogram B
3. Histogram C
4. Histogram D

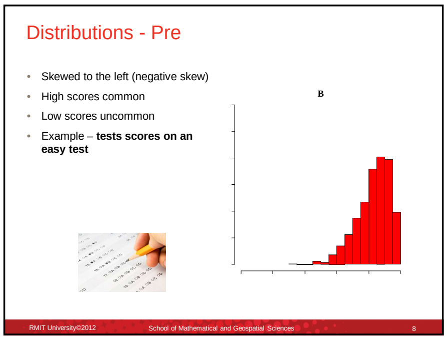


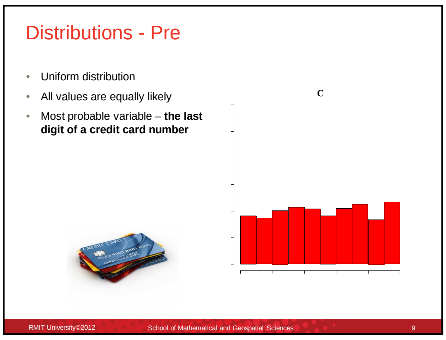
Distributions – Pre 3

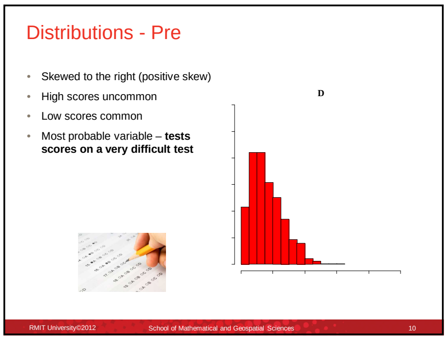
- Responses and Answer

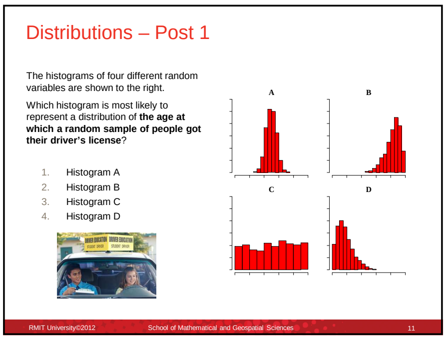
Distributions

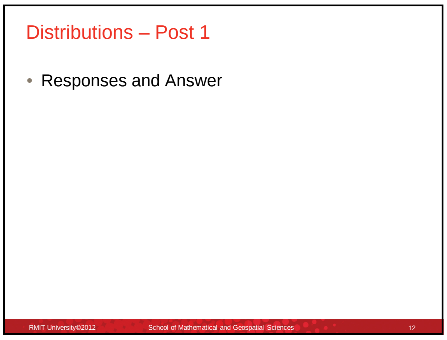




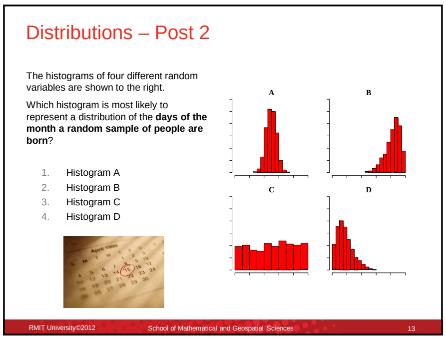


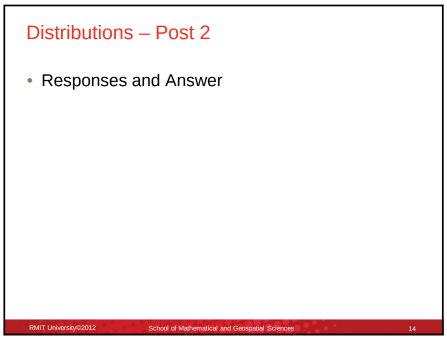


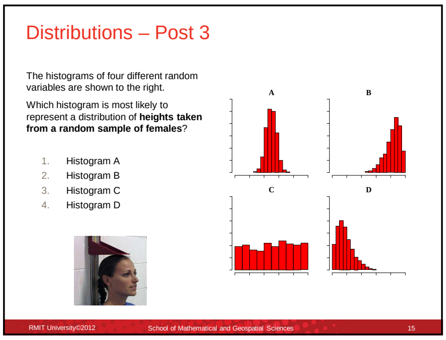


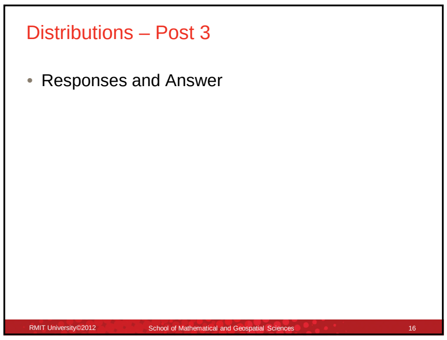


Distributions









Probability

Probability – Pre

Researchers know that at any time during winter, 10% of the population will have the common cold. Five different researchers randomly select 20 people from the population and record the percentage of people in their sample who have a cold. Which sequence below is the most plausible for the percentage of people with colds in each of the researchers' samples?

1. 15%, 10%, 15%, 5%, 20%
2. 10%, 10%, 10%, 10%, 10%
3. 30%, 80%, 60%, 5%, 10%
4. All the above are equally likely

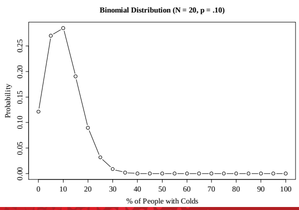
Probability – Pre

- Responses and Answer

Measures of Central Tendency – Pre

1. 15%, 10%, 15%, 5%, 20%

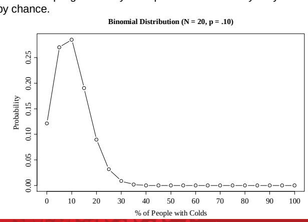
Plausible sampling variation. The sample percentages vary expectedly around 10% by chance.



Measures of Central Tendency – Pre

2. 10%, 10%, 10%, 10%, 10%

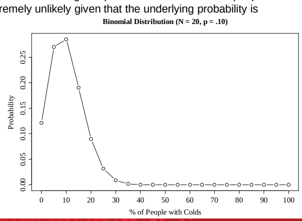
- Implausible **absence** of sampling variability. Samples should naturally vary around 10% just by chance.



Measures of Central Tendency – Pre

3. 30%, 80%, 60%, 5%, 10%

- Implausible sample results. Getting samples with 80% and 60% of people having colds is extremely unlikely given that the underlying probability is 10%



Measures of Central Tendency – Pre

4. All the above are equally likely

- As we have shown, not all outcomes are equally likely.

Probability

Probability – Post

According to the Australian Bureau of Statistics, 25% of the Australian adult population is obese. Five different research teams randomly sample 30 Australians each and record whether or not each person was obese. Which sequence below is the most plausible for the percentage of obese people in each of the research teams' samples?

1. 25%, 5%, 50%, 15%, 70%
2. 10%, 20%, 30%, 35%, 20%
3. 25%, 25%, 25%, 25%, 25%
4. All the above are equally likely

 School of Mathematical and Geospatial Sciences 7

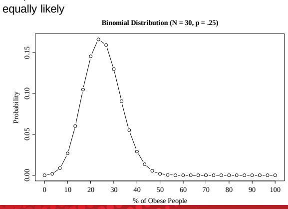
Probability – Post

- Responses and Answer

 School of Mathematical and Geospatial Sciences 8

Probability – Post

1. 25%, 5%, 50%, 15%, 70%
2. **10%, 20%, 30%, 35%, 20%**
3. 25%, 25%, 25%, 25%, 25%
4. All the above are equally likely



 School of Mathematical and Geospatial Sciences 9

p-values

p-values – Pre

A research article reports the results of a study looking at the association between diabetes and high glycaemic index (GI) foods. The researchers suspect that people with diabetes would be more likely to consume diets rich in high GI foods when compared to people without diabetes. The researchers conduct a Chi-square test of association. The results find that $X^2(df = 3) = 8.416$, $p = 0.038$. Which of the following statements best defines the *p*-value of this study.

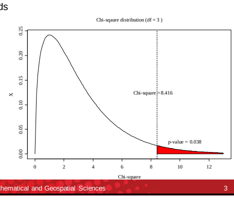
1. The probability of there being no association between diabetes and high GI foods
2. The probability of getting the result in this study, or one more extreme, assuming that there was no association between diabetes and high GI foods.
3. The probability that there is an association between diabetes and high GI foods
4. The probability of the researchers' results occurring by chance

p-values – Pre

- Responses and Answer

p-values – Definition

- A ***p*-value** tells us the probability of observing a sample result, or one more extreme, under the assumption that the Null hypothesis is true
- We can write this as $Pr(D|H_0)$ where D = Data and H_0 is the Null hypothesis
- Therefore, there was a .038 probability of observing a sample Chi-square statistic of 8.416, or one more extreme, under the assumption that there was no relationship between diabetes and high GI foods



p-values – Pre

1. **The probability of there being no association between diabetes and high GI foods**

   – This answer implies that the *p*-value is the probability of the Null hypothesis being true given the observed sample result.
   – If we wrote this out, we would write $Pr(H_0|D)$ – the probability of the Null hypothesis given the data
   – However, as we will see, $Pr(H_0|D)$ is not the same as $Pr(D|H_0)$

p-values – Pre

- Consider the following scenario where:
  – H = Hanged
  – D = Dead

The *p*-value is equivalent to $Pr(D|H)$ – The probability of a person being dead given that they were hanged

- This probability would be very high (e.g. 90%)

Let's assume we want to know $Pr(H|D)$ – The probability of a person being hanged given that they were dead

- This probability would be very low (e.g. 1%)

***Pr(D|H)* and *Pr(H|D)* are not the same probabilities.**

These probabilities are not interchangeable, just as the *p*-value, $Pr(D|H_0)$, is not the same as $Pr(H_0|D)$

p-values – Pre

2. **The probability of getting the result in this study, or one more extreme, assuming that there was no association between diabetes and high GI foods.**

   – This answer is correct
   – This answer implies that the *p*-value is the probability of observing a sample result (data), assuming the Null hypothesis (No association) is true
   – That's to say, $Pr(D|H_0)$

p-values

p-values – Pre

3. **The probability that there is an association between diabetes and high GI foods**

   – This answer implies that the *p*-value is the probability of the Alternate hypothesis being true given the observed sample result.
   – If we wrote this out, we would write $Pr(H_A|D)$ – the probability of the Alternate hypothesis being true given the data observed
   – However, as stated previously, the *p*-value gives us $Pr(D|H_0)$.
   – The *p*-value cannot be used as probability of the alternate hypothesis because it is calculated based on the assumption that the Null hypothesis is true.
   – $Pr(H_A|D)$ and $Pr(D|H_0)$ are not equivalent

p-values – Pre

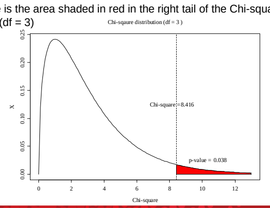
4. **The probability of the researchers' results occurring by chance.**

   – There are two things wrong with this answer
   – This answer implies that the *p*-value is an exact probability of a study's results occurring by chance. We know that the *p*-value is the probability of a study's results, <u>or one more extreme</u>, assuming the Null hypothesis is true
   – It also fails to acknowledge that the *p*-value is based on the assumption of the Null hypothesis being true.

p-values – Revision

Remember, the *p*-value in this scenario is the probability of observing a Chi-square statistic, or one more extreme, assuming there was no association between diabetes and high GI foods.

The *p*-value is the area shaded in red in the right tail of the Chi-square distribution (df = 3)



p-values – Post

A researchers investigates the association between watching TV and engaging in regular physical activity. The researchers find that as people watch more TV, they are less likely to engage in regular physical activity. The researchers report a *p*-value of the association that they tested. The *p*-value was 0.01. Which of the following statements best defines the *p*-value of this study.

1. The probability of getting the result in this study, or one more extreme, assuming that there was no association between watching TV and engagement in physical activity.
2. The probability of the researcher's results occurring by chance
3. The probability that there is an association between watching TV and engagement in physical activity.
4. The probability of there being <u>no</u> association between watching TV and engagement in physical activity.

p-values – Post

- Responses and Answer

Hypothesis Testing

Hypothesis Testing – Pre

The following analogy models the logic of hypothesis testing. In a criminal trial evidence is presented before a jury to determine whether or not an accused is guilty of a crime. On rare occasions, misleading evidence is presented to the jury that falsely convicts an innocent person. The Null hypothesis of the criminal trial is that the accused is innocent. The alternate hypothesis is that the accused is guilty. If a jury rejects the Null hypothesis, which of the following statements is true?

1. The accused is definitely guilty and should be convicted
2. The jury decides that the accused is guilty, but there is still a possibility that the accused is innocent
3. The accused is innocent and should be acquitted
4. The accused is most likely innocent, but they could be guilty

Hypothesis Testing – Pre

- Responses and Answer

Hypothesis Testing – The Logic Explained

- **Hypothesis Testing Logic**
  - Start by assuming that the Null hypothesis (H_0) is true
  - We set a level of unusualness that we want our data to achieve before we are comfortable rejecting H_0
    - This is called the significance level (α), e.g. 0.05.
  - We then calculate the probability of obtaining the sample data, or data more extreme, assuming H_0 is true
    - This is called the *p*-value
  - We reject H_0 when the data is considered unusual under the assumption that the Null hypothesis is true (i.e. $p < \alpha$). However, even after rejecting H_0, there is a still a small probability that H_0 might be true.
  - We fail to reject H_0 when the data is considered typical under the assumption that the Null hypothesis is true (i.e. $p > \alpha$). However, even after failing to reject H_0, there is a still a probability that H_0 might be false.

Hypothesis Testing – The Logic Explained

- **Jury Trial Analogy**
  - Null hypothesis: Accused is innocent
  - We set a burden of proof that evidence must achieve before the jury can decide the accused is guilty
    - Beyond a reasonable doubt (Significance level α)
  - The jury weighs up evidence by considering how likely the evidence presented is assuming the accused is innocent
    - Probability of evidence given that the accused is innocent (*p*-value)
  - Jury rejects H_0 (Accused is found guilty) when the evidence presented is considered unlikely to have been found under the assumption that the accused is innocent (i.e. $p < \alpha$). However, even after rejecting H_0, there is a still a possibility that the accused is innocent. False or misleading evidence may have been presented.
  - Jury fails to reject H_0 (Accused is found not guilty) when the evidence fails to reach the burden of proof of beyond a reasonable doubt (i.e. $p > \alpha$). However, even after failing to reject H_0, there is a still a probability that the accused is guilty. Perhaps not enough evidence was available to be presented to the jury.

Hypothesis Testing – Pre

1. **The accused is definitely guilty and should be convicted**

- This answer implies that the evidence presented before the jury is definitive proof of the accused's guilt.
- However, there is always a small probability that the evidence is false or misleading
- Therefore, there is no way to be 100% sure that the accused is guilty

Hypothesis Testing – Pre

2. **The jury decides that the accused is guilty, but there is still a possibility that the accused is innocent**

- This answer is correct
- The jury makes a decision to reject the Null hypothesis as the burden of proof was met. This support the decision that the accused is guilty.
- More importantly, this answer acknowledges that there is still a possibility that the accused could still be innocent (e.g. false or misleading evidence was presented)

Hypothesis Testing

Hypothesis Testing – Pre

3. **The accused is innocent and should be acquitted**

- The jury rejected the idea that the accused was innocent, and decided that the evidence pointed to a guilty verdict
- This answer is not logical
- Also, a jury can only decide whether the evidence points to the accused being guilty (reject H_0) or not guilty (fail to reject H_0)

Hypothesis Testing – Pre

4. **The accused is most likely innocent, but they could be guilty**

- This answer implies that the jury failed to reject the Null hypothesis that the accused was innocent. In other words, that the jury failed to be convinced of the accused's guilt.
- As the question clearly states that the jury rejected the Null hypothesis, this answer cannot be correct.

Hypothesis Testing – Post

A scientist uses a statistical test to determine whether or not a new drug is effective. The statistical test can sometimes fail to detect an effective drug. The Null hypothesis is that the drug is not effective. The alternate hypothesis is that the drug is effective. If the scientist fails to reject the Null hypothesis based on the results of the statistical test, which of the following statements is true?

1. There is statistically significant evidence that the drug is definitely effective and should be recommended for use
2. There is statistically significant evidence that the drug is not effective and should not be used
3. The scientist decides that there is not enough statistical evidence to support the effectiveness of the new drug, but there is still a probability that it might be effective
4. The scientist decides that there is statistically significant evidence that the new drug is effective, but there is still a probability that it might actually be ineffective

Hypothesis Testing – Post

- Responses and Answer

Confidence Intervals

Confidence Intervals – Pre

Researchers conduct a study looking at the effect of eating a high protein diet on stroke prevention in the elderly. The study compared the incidence of stroke over a two year period between two groups of people. One group eats a high protein diet, whereas the other group eats a regular diet. At the end of the study, the researchers calculate the results using *RR*. The results show that *RR* = 0.59 and a 95% confidence interval to be (0.34, 1.01). Which of the following interpretations of this confidence interval is most correct?

1. We are 95% certain that each person's risk of stroke in the high protein diet group was .34 to 1.01 times the same risk in the regular diet group.
2. The true population *RR* is between (0.34, 1.01) with 95% probability.
3. We would expect about 95% of all possible *RR* from this population to between .34 and 1.01.
4. If this study was repeated many times, 95% of the *CI*s calculated from these studies would capture the true population *RR*. The 95% *CI* (0.34, 1.01) is an example of one of these intervals.

Confidence Intervals – Pre

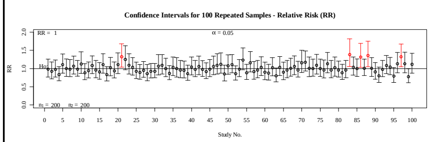
- Responses and Answer

Confidence Intervals - Definition

- **Definition**
  - A 95% confidence interval (*CI*) is an interval estimate for a population parameter, based on a sample statistic, where if many repeated samples of a certain size *n* were drawn from the population, and a *CI* for each sample statistic was calculated, 95% of these intervals would capture the true population parameter, whereas the other 5% would not.
  - Let's explore this definition…

Confidence Intervals - Explained

- The plot below shows 100 confidence intervals for 100 studies using 100 random samples taken from a population where *RR* = 1, i.e. the Null hypothesis (Ho:) is true.
- We can see that, as expected, 5/100 *CI*s fail to capture Ho: *RR* = 1
- This 5% is the Significance Level α = 0.05
- The other 95/100 (95%) *CI*s capture Ho: *RR* = 1



Confidence Intervals – Pre

1. **We are 95% certain that each person's risk of stroke in the high protein diet group was .34 to 1.01 times the same risk in the regular diet group.**

- Incorrect
- The 95% *CI* for *RR* is based on a <u>sample</u> summary statistic.
- *RR* = 0.59 means that the risk of stroke in the high protein diet group was 0.59 lower than the same risk in the regular diet group.
- The 95% *CI* of *RR* (0.34, 1.01) is calculated around *RR* = 0.59
- *RR* and the 95% *CI* of *RR* <u>do not</u> relate to an <u>individual's</u> risk.
- They relate to a group's or population's risk.
- Also, what does "95% certain" mean? This is ambiguous. Remember that confidence intervals are based on repeated sampling.

Confidence Intervals – Pre

2. **The true population RR is between (0.34, 1.01) with 95% probability.**

- Incorrect
- There are two possibilities when we calculate a 95% confidence interval.
  1. The *CI* captures the true population parameter *RR*
  2. The *CI* <u>does not</u> capture the true population parameter *RR*
- In reality we don't ever know which outcome is true because we don't know the value of the true population parameter. We can only estimate it. That's the whole reason behind conducting the study!
- Therefore, it does not make sense to say that the true population parameter lies between a confidence interval with 95% probability.
- The *CI* either captures the true population parameter, i.e. *Pr = 1.*
- Or, it fails to capture the true population parameter, i.e. *Pr = 0.*
- We don't know either way, so avoid definitions like this.

Confidence Intervals

Confidence Intervals – Pre

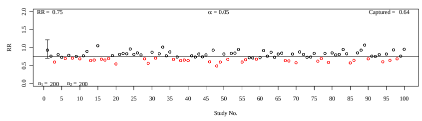
3. **We would expect about 95% of all possible *RR* from this population to between .34 and 1.01.**

– Incorrect
– This answer implies that that if we repeated this study many times, 95% of sample *RR*s will fall between the sample 95% *CI* (0.34, 1.01) calculated in this study.
– This is unlikely as the next slide will demonstrate.

Confidence Intervals – Pre

– Look at the plot below where the true population *RR* = .75.
– The first sample from this population finds *RR* = 0.93, 95% *CI* (0.71, 1.22).
– This is far away from *RR* = .75 due to normal sampling error.
– We can see that if 99 more studies were conducted, only 64% of these studies' *RR* were captured by the 1st study's 95% *CI*.
– Therefore, answer 3 is unlikely to be true

Proportion of 100 Repeated Studies' RR that are Captured by the 1st Study's RR 95% CI



Confidence Intervals – Pre

4. **If this study was repeated many times, 95% of the *CI*s calculated from these studies would capture the true population *RR*. The 95% *CI* (0.34, 1.01) is an example of one of these intervals.**

– This answer is the most correct
– This answer is very close to the correct definition of a confidence interval
– It acknowledges the central concept of repeated sampling in the definition of a confidence interval

Confidence Intervals – Post

You are reading through an epidemiological study looking at the association between gastric cancer and coffee consumption. The study found that people with gastric cancer were no more likely to consume coffee on a daily basis when compared to controls who did not have gastric cancer, *OR* = 0.92, 95% *CI* (0.77, 1.12). Which of the following statements best defines the 95% confidence interval reported in this study?

1. The true population *OR* is somewhere between (0.77, 1.12) with 95% probability
2. We are 95% confident that the true population parameter is captured within the interval (0.77, 1.12)
3. The 95% *CI* (0.77,1.12) is an interval estimate of a parameter based on a sample statistic. The theory of confidence intervals predicts that if we were to repeat this study many times, 95% of the *CI* calculated for each study would capture the true population *OR*.
4. We would expect about 95% of all possible sample *OR* from this population to between .77 and 1.12.

Confidence Intervals – Post

- Responses and Answer

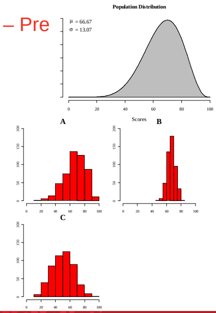
Sampling Distributions

Sampling Distributions – Pre

A population distribution of test scores is shown in the top graph. The population has a mean of 66.67 and standard deviation of 13.07.

Which histogram do you think represents a single random sample of 500 scores from the population distribution?

1. Histogram A
2. Histogram B
3. Histogram C
4. All histograms are plausible samples of *N* = 500



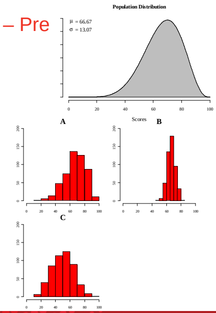
Sampling Distributions – Pre

- Responses and Answer

Sampling Distributions – Pre

Which histogram do you think represents a distribution of 500 random samples' means of size *N* = 5?

1. Histogram A
2. Histogram B
3. Histogram C
4. All histograms are possible sampling distributions of 500 sample means where *N* = 5

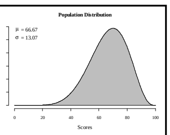


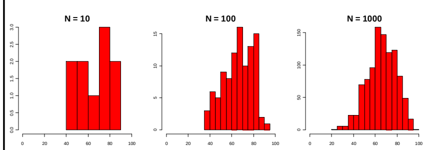
Sampling Distributions – Pre

- Responses and Answer

Sample Distribution

- The larger the random sample size *N*, the more the sample distribution will look like the population distribution.
- Why? Because larger random samples are more representative of the population.





Sampling Distributions

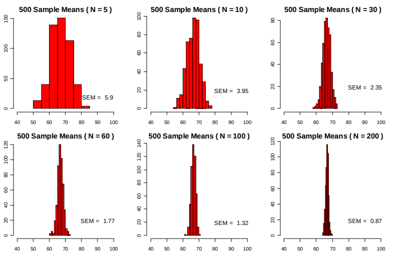
How to make a **sampling** distribution:

1. Take a random sample of size *N* from the population
2. Calculate a sample statistic, e.g. mean
3. Put the sample back into the population
4. Repeat steps 1 – 3 many times, each time recording the sample statistic, e.g. mean

- If you plotted all the sample statistics, you would be looking at a **sampling** distribution of that statistic, e.g. a **sampling** distribution of the mean.
- A **sample** distribution can be constructed by only plotting the data from step 1.
- Let's look at what happens to the variability in a **sampling** distribution as the size of the samples drawn from the population increase.

Sampling Distributions

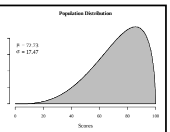
Sampling Distributions - Pre



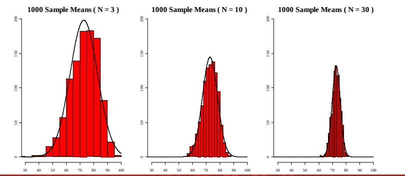
Sampling Distributions

- We can see that larger the sample size, the smaller the sampling distribution's variation.
- The standard deviation of a sampling distribution is called the standard error ($SE = \frac{\sigma}{\sqrt{n}}$).
- Why? Because larger random samples provide more accurate and precise estimates of population parameters. This reduces sampling error.
- Now let's look at what happens to the shape of a sampling distribution as the sample size *N* increases.

Sampling Distributions - CLT



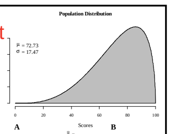
- Normal curves have been overlayed on each plot.
- Despite the non-normal population distribution, the sampling distribution begins to approximate a normal distribution as the sample size increases.
- This is known as the central limit theorem (CLT)



Sampling Distributions - Summary

- As a random sample's size *N* **increases**, the better that random sample will **approximate** it's population distribution.
- **Sampling distributions** are hypothetical distributions of sample statistics taken from many repeated random samples of size *N*.
- As a random sample's size *N* **increases**, a sampling distribution's **standard error decreases**, ($SE = \frac{\sigma}{\sqrt{n}}$).
- As sample size *N* increases, a sampling distribution will begin to approximate a normal distribution regardless of the shape of the underlying population distribution (**Central limit theorem**).

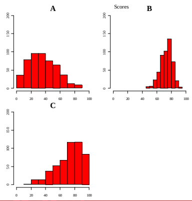
Sampling Distributions – Post



A population distribution of test scores is shown in the top graph. The population has a mean of 72.73 and standard deviation of 17.47.

Which histogram do you think represents a single random sample of 500 scores from the population distribution?

1. Histogram A
2. Histogram B
3. Histogram C
4. All histograms plausible samples of *N* = 500



Sampling Distributions – Post

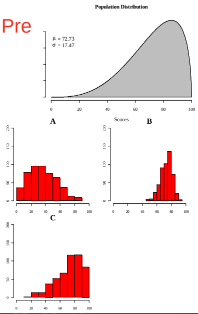
- Responses and Answer

Sampling Distributions

Sampling Distributions – Pre

Which histogram do you think represents a distribution of 500 random samples' means of size $N = 5$?

1. Histogram A
2. Histogram B
3. Histogram C
4. All histograms are possible sampling distributions of 500 sample means where $N = 5$



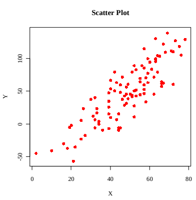
Sampling Distributions – Post

- Responses and Answer

Correlation

Correlation – Pre

- This scatter plot shows the relationship between an X and Y variable. Which of the following statements is true in relation to this scatter plot?
  1. As X increases, X causes a decrease in Y
  2. As X increases, X causes an increase in Y
  3. The relationship between X and Y is negative
  4. None of the above

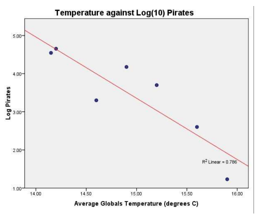


Correlation – Pre

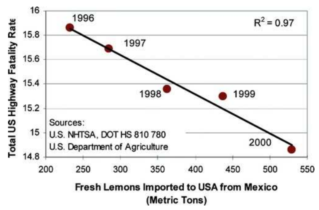
- Responses and Answer

Correlation – Pre

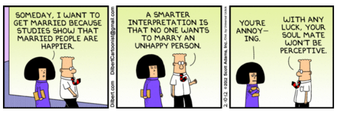
- Correlation does not equal causation!



Correlation ≠ Causation



Correlation ≠ Causation

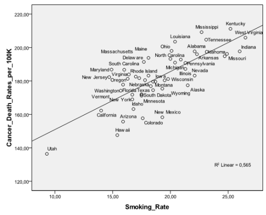


Correlation ≠ Causation

Correlation

Correlation ≠ Causation

- Or does it?



© RMIT University 2012 School of Mathematical and Geospatial Sciences 7

Correlation – Summary

- Just because two variables are correlated, it does not necessarily mean one causes a change in the other. At least three possible explanations exist:
  - X causes Y
  - Y causes X
  - Both Y and X are caused by Z
- Correlation can provided evidence that supports causal relationships, but correlation can never be regarded as proof of causation.

© RMIT University 2012 School of Mathematical and Geospatial Sciences 8

Correlation – Post

- Researchers report a strong positive correlation between pocket money and drug use which was found in a random survey of Australian adolescents. Which of the following is the best interpretation of this relationship?

1. The positive relationship suggests that giving adolescents pocket money leads to adolescent drug use.
2. The positive relationship suggests that adolescent drug use is higher in adolescents who are given pocket money.
3. The positive relationship suggests that adolescent drug use can be reduced by ensuring adolescents are given pocket money to reduce boredom and drug use.
4. None of the above are valid interpretations.

© RMIT University 2012 School of Mathematical and Geospatial Sciences 9

Correlation – Post

- Responses and Answer

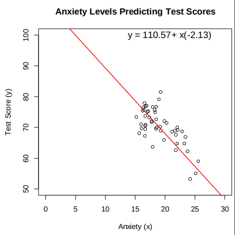
© RMIT University 2012 School of Mathematical and Geospatial Sciences 10

Regression

Regression – Pre 1

- Fifty people have their anxiety level measured prior to taking a test. The anxiety score can range from 0 – 30. High scores indicate high anxiety. The sample's test scores are recorded and plotted on a scatter plot with their anxiety score. A linear regression line is fitted to the data. According to the regression line, predict what a person with an anxiety score of 20 would score on the test.

1. 50
2. 110
3. 67
4. Cannot say for sure



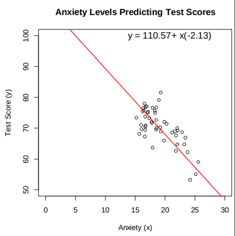
Regression – Pre 1

- Responses and Answer

Regression – Pre 2

- Using the same regression in the previous question, predict what a person with an anxiety score of 10 would score on the test.

1. 90
2. 110
3. 100
4. Cannot say for sure

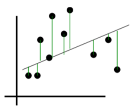


Regression – Pre 2

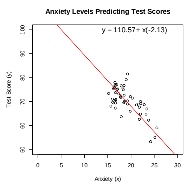
- Responses and Answer

Regression – Pre

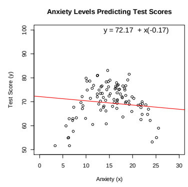
- Avoid extrapolating outside the range of your predictor (x) variable
- Because we don't have data for people who scored 10 on the anxiety (x) scale, we don't have any data to help predict what their test score will be.
- Let's see what happens if we sample some people with low anxiety and try to fit a linear regression



Regression - Pre



This is what we started with. Mainly, highly anxious people.

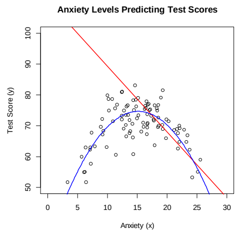


Now we thrown in some people who're very relaxed. Perhaps too relaxed. The linear relationship disappears.

Regression

Regression - Pre

- In this example, a non-linear regression (blue line) provides a better fit to test scores.
- Prediction with the linear regression (red) outside the range of data would have been very inaccurate.
- **Take home message: Avoid extrapolating regression beyond the range of your data!**



 School of Mathematical and Geospatial Sciences 7

Regression – Post

- A private health insurer surveys their members on how many hours per week they spend exercising. The survey reveals that their members report exercising an average of 2 hours per week with times ranging anywhere between 0 to 10 hours. A researcher finds a negative relationship between the member's average hours spent exercising per week and the annual monetary amount of benefits they claim on their policy. A regression of this relationship reveals that:

Annual Amount claimed $ = 700 – 19.9(Exercise hours per week)

Which of the following methods is appropriate for predicting the annual claim amount for an elite athlete who spends 20 hours per week exercising.

1. Plot the regression line on a scatter plot, look up 20 hours on the X axis and read off the annual claim amount value from the Y axis.
2. Enter 20 hours into the regression equation above and calculate the predicted annual claim amount.
3. Both these methods are suitable.
4. Neither of these methods are suitable.

 School of Mathematical and Geospatial Sciences 8

Regression – Post

- Responses and Answer

 School of Mathematical and Geospatial Sciences 9

B.5 Conceptual Change Question Response Patterns

Distributions I (Q4)		Responses					
Cohort		a	b	**c**	d	e	Total
Control	N	15	5	114	25	1	160
	%	9.4	3.1	71.3	15.6	0.6	100.0
Intervention	N	10	10	129	18	0	167
	%	6.0	6.0	77.2	10.8	0.0	100.0
Total	N	25	15	243	43	1	327
	%	7.6	4.6	74.3	13.1	0.3	100.0

Distributions II (Q5)		Responses				
Cohort		**a**	b	c	d	Total
Control	N	111	26	1	22	160
	%	69.4	16.3	0.6	13.8	100.0
Intervention	N	123	12	5	27	167
	%	73.7	7.2	3.0	16.2	100.0
Total	N	234	38	6	49	327
	%	71.6	11.6	1.8	15.0	100.0

Distributions III (Q6)		Responses				
Cohort		a	b	c	**d**	Total
Control	N	23	13	13	111	160
	%	14.4	8.1	8.1	69.4	100.0
Intervention	N	15	22	5	125	167
	%	9.0	13.2	3.0	74.9	100.0
Total	N	38	35	18	236	327
	%	11.6	10.7	5.5	72.2	100.0

Confidence Intervals I (Q7)		Responses					
Cohort		a	b	**c**	d	e	Total
Control	N	7	13	117	11	10	158
	%	4.4	8.2	74.1	7.0	6.3	100.0
Intervention	N	9	12	99	17	30	167
	%	5.4	7.2	59.3	10.2	18.0	100.0
Total	N	16	25	216	28	40	325
	%	4.9	7.7	66.5	8.6	12.3	100.0

Sampling Distributions I (Q8)		Responses				
Cohort		a	**b**	c	d	Total
Control	N	23	82	41	13	159
	%	14.5	51.6	25.8	8.2	100.0
Intervention	N	27	69	60	9	165
	%	16.4	41.8	36.4	5.5	100.0
Total	N	50	151	101	22	324
	%	15.4	46.6	31.2	6.8	100.0

Probability (Q17)		Responses				
Cohort		**a**	b	c	d	Total
Control	N	63	48	10	37	158
	%	39.9	30.4	6.3	23.4	100.0
Intervention	N	86	42	12	27	167
	%	51.5	25.1	7.2	16.2	100.0
Total	N	149	90	22	64	325
	%	45.8	27.7	6.8	19.7	100.0

p-values I (Q25)		Responses		
Cohort		**a**	b	Total
Control	N	61	98	159
	%	38.4	61.6	100.0
Intervention	N	55	112	167
	%	32.9	67.1	100.0
Total	N	116	210	326
	%	35.6	64.4	100.0

p-values II (Q26)		Responses		
Cohort		a	**b**	Total
Control	N	40	120	160
	%	25.0	75.0	100.0
Intervention	N	47	120	167
	%	28.1	71.9	100.0
Total	N	87	240	327
	%	26.6	73.4	100.0

p-values III (Q27)		Responses		
Cohort		a	**b**	Total
Control	N	118	42	160
	%	73.8	26.3	100.0
Intervention	N	119	48	167
	%	71.3	28.7	100.0
Total	N	237	90	327
	%	72.5	27.5	100.0

Confidence Intervals II (Q28)		Responses					
Cohort		a	b	c	d	**e**	Total
Control	N	15	26	53	25	40	159
	%	9.4	16.4	33.3	15.7	25.2	100.0
Intervention	N	19	19	56	17	56	167
	%	11.4	11.4	33.5	10.2	33.5	100.0
Total	N	34	45	109	42	96	326
	%	10.4	13.8	33.4	12.9	29.4	100.0

Hypothesis Testing I (Q29)		Responses				
Cohort		a	b	c	**d**	Total
Control	N	10	24	4	122	160
	%	6.3	15.0	2.5	76.3	100.0
Intervention	N	9	21	8	129	167
	%	5.4	12.6	4.8	77.2	100.0
Total	N	19	45	12	251	327
	%	5.8	13.8	3.7	76.8	100.0

Hypothesis Testing II (Q30)		Responses				
Cohort		a	b	c	**d**	Total
Control	N	11	22	6	121	160
	%	6.9	13.8	3.8	75.6	100.0
Intervention	N	21	20	0	126	167
	%	12.6	12.0	0.0	75.4	100.0
Total	N	32	42	6	247	327
	%	9.8	12.8	1.8	75.5	100.0

Correlation (Q31)		Responses				
Cohort		**a**	b	c	d	Total
Control	N	96	9	37	18	160
	%	60.0	5.6	23.1	11.3	100.0
Intervention	N	103	10	40	13	166
	%	62.0	6.0	24.1	7.8	100.0
Total	N	199	19	77	31	326
	%	61.0	5.8	23.6	9.5	100.0

Sampling Distributions II (34)		Responses			
Cohort		**a**	b	c	Total
Control	N	103	39	18	160
	%	64.4	24.4	11.3	100.0
Intervention	N	118	31	18	167
	%	70.7	18.6	10.8	100.0
Total	N	221	70	36	327
	%	67.6	21.4	11.0	100.0

Sampling Distributions III (Q35)		Responses			
Cohort		a	**b**	c	Total
Control	N	20	75	64	159
	%	12.6	47.2	40.3	100.0
Intervention	N	22	92	53	167
	%	13.2	55.1	31.7	100.0
Total	N	42	167	117	326
	%	12.9	51.2	35.9	100.0

Regression (Q37)		Responses				
Cohort		a	b	**c**	d	Total
Control	N	42	31	26	57	156
	%	26.9	19.9	16.7	36.5	100.0
Intervention	N	41	14	52	60	167
	%	24.6	8.4	31.1	35.9	100.0
Total	N	83	45	78	117	323
	%	25.7	13.9	24.1	36.2	100.0

Confidence Intervals III (Q38)		Responses					
Cohort		**a**	b	c	d	e	Total
Control	N	20	21	25	19	70	155
	%	12.9	13.5	16.	12.3	45.2	100.0
Intervention	N	20	18	28	25	75	166
	%	12.0	10.8	16.9	15.1	45.2	100.0
Total	N	40	39	53	44	145	321
	%	12.5	12.1	16.5	13.7	45.2	100.0

Hypothesis Testing III (Q40)		Responses				
Cohort		a	**b**	c	d	Total
Control	N	64	71	13	11	159
	%	40.3	44.7	8.2	6.9	100.0
Intervention	N	56	87	11	13	167
	%	33.5	52.1	6.6	7.8	100.0
Total	N	120	158	24	24	326
	%	36.8	48.5	7.4	7.4	100.0

Appendix C

Part III

C.1 Study I - Plain Language Statement and Consent

RMIT UNIVERSITY

RMIT Human Research Ethics Committee

School of Mathematics and Geospatial Sciences

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2283
Fax +61 3 9925 2454

INVITATION TO PARTICIPATE IN A RESEARCH PROJECT
PROJECT INFORMATION STATEMENT

Project Title

Evaluating the use of an Innovative Online Virtual Environment for Authentic Student Assessment of Scientific Research Design and Statistical Analysis

Investigators

- Mr. James Baglin (Lead Investigator: BAppSc, Psychology – Honours; PhD Candidate, Statistics, SMGS, RMIT University, james.baglin@rmit.edu.au, 9925 6118)
- Dr. Matthew Linden (Co-investigator: Senior Lecturer, SMS, RMIT University)

Dear student,

You are invited to participate in a research project being conducted by RMIT University. This information sheet describes the project in straightforward language, or 'plain English'. Please read this sheet carefully and be confident that you understand its contents before deciding whether to participate. If you have any questions about the project, please ask one of the investigators.

What is this research all about?

This study is being conducted to evaluate the effectiveness of the online tool the "Island" for teaching and assessing key competencies in experimental design and management. The Island is an online virtual environment created to give students the ability to design, conduct and analyse virtual experiments. We would like you to share your experiences of using the Island in your course. This will help us determine the effectiveness of using online virtual environments in science courses and help us decide if these environments should be used in future courses. This study is a joint project between the School of Mathematics and Geospatial Sciences and the School of Medical Science. The study has been approved by the RMIT Human Research Ethics Committee (HREC), but we still need your permission to survey you and use your results from this class to assess the effectiveness of the software as a teaching and learning tool.

Why me?

You have been approached because you are enrolled in a course that will be piloting the use of the Island and you are over the age of 18.

If I agree to participate, what will I be required to do?

Agreeing or not to participate will have no impact on the nature of your assessment in this course. The Island is part of your course and all students will use it. By agreeing to participate you will respond to a 10-15 minute feedback survey on how useful you found the software, you may participate in a focus group discussion on how it could be improved, and you agree to allow the researchers named above to use your course assessment (exam) results to determine if this software is an effective teaching tool. Your deidentified examination results will be compared to previous students who did not use the Island. Whether you chose to participate or not will have no impact on your mark. Participation is strictly voluntary and you may also withdraw from the study at any time.

Focus group discussions will be organised following exams and during a time that is convenient to participants. Discussion will take approximately 40-60 minutes and will be voice recorded. All data will be deidentified and your involvement in this study and the subsequent collection of your assessment and questionnaire responses will be kept strictly confidential according to Australian privacy laws and university guidelines.

What are the risks?

There are very few risks associated with your participation in this study. The most prominent risk being that your survey responses, focus group discussion and examination responses will be known by the lead investigator. However, the lead investigator will never disclose, use or publish this sensitive information for any other purpose not outlined in this information sheet. In the event that you have concerns about your participation in the study you are encouraged to contact the lead investigator, James Baglin (Email: james.baglin@rmit.edu.au Ph: 9925-6118) or the Human Research Ethics Committee directly (contact details at the bottom of this page).

What are the benefits?

There are no direct benefits associated with participation. However, your participation in this project will help Universities determine the merit of using the Island in future courses.

What will happen to the information I provide?

The information gathered from this study will be kept strictly confidential. Only the lead investigator will have access to your identifying information, survey responses or academic results. Your personal information will never be used or given to anyone else for any other purpose, except under the following circumstances. Any information that you provide can be disclosed only if (1) it is to protect you or others from harm, (2) a court order is produced, or (3) you provide the researchers with written permission".Your data will be securely kept at RMIT for 5 years after the completion of the study.

Summarised and aggregated results from this study will appear in future reports and peer-reviewed publications. You will be provided with a summary of the findings at the completion of the study.

What are my rights?

You have the right to withdraw from the study at any given time without prejudice, the right to have any unprocessed data removed and destroyed, and the right to have any questions answered at any time. You can exercise your ethical rights by contacting the lead investigator.

C.2 Study I - Questionnaire

RMIT UNIVERSITY

RMIT Human Research Ethics Committee

School of Mathematics and Geospatial Sciences

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2283
Fax +61 3 9925 2454

INFORMED CONSENT

Prescribed Consent Form For Persons Participating In Research Projects Involving Interviews, Questionnaires or Disclosure of Personal Information

Portfolio:	Science, Engineering and Health		
School of	Mathematics and Geospatial Sciences and Medical Science		
Name of participant:			
Project Title:	Evaluating the use of an Innovative Online Virtual Environment for Authentic Student Assessment of Scientific Research Design and Statistical Analysis		
Name(s) of investigators:	(1) James Baglin	Phone:	9925 6118
	(2) Dr. Matthew Linden	Phone:	9925 7898

1. I have received a statement explaining the interview/questionnaire involved in this project.
2. I consent to participate in the above project, the particulars of which - including details of the interviews or questionnaires - have been explained to me.
3. I authorise the investigator or his or her assistant to administer a questionnaire.
4. I acknowledge that:
   (a) Having read Plain Language Statement, I agree to the general purpose, methods and demands of the study.
   (b) I have been informed that I am free to withdraw from the project at any time and to withdraw any unprocessed data previously supplied.
   (c) The project is for the purpose of research and/or teaching. It may not be of direct benefit to me.
   (d) The privacy of the personal information I provide will be safeguarded and only disclosed where I have consented to the disclosure or as required by law.
   (e) The security of the research data is assured during and after completion of the study. The data collected during the study may be published, and a report of the project outcomes will be provided to all consenting participants. Any information which will identify me will not be used.

Participant's Consent

Participant: ______________________ ***Date:*** ______________

(Signature)

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001. The telephone number is (03) 9925 2251.
Details of the complaints procedure are available from the above address.

If you consent to participate, please fill out the following short questionnaire.

Please fill out the following information. Circle responses when the option permits.

1. Age: ______________________

2. Gender	**Male**	**Female**
3. Course	**ONPS2304**	**MATH1300**
4. Residency	**International**	**Domestic**
5. Load	**Full-time**	**Part-time**

Thinking about the Island, please rate your level of agreement to the following statements where: (1) strongly disagree and (7) strongly agree. (Circle your response)

Strongly Disagree	Disagree	Slightly disagree	Undecided	Slightly agree	Agree	Strongly Agree
6. I enjoyed using the Island for my course project.						
1	2	3	4	5	6	7
7. The Island was easy to use.						
1	2	3	4	5	6	7
8. The Island gave me a better understanding of scientific research design.						
1	2	3	4	5	6	7
9. I enjoyed being in control of my own virtual scientific study.						
1	2	3	4	5	6	7
10. I found it difficult to use the Island to conduct a virtual scientific study.						
1	2	3	4	5	6	7
11. The Island gave me a greater appreciation of the practical considerations of conducing scientific studies (e.g. planning data collection, getting samples, and managing time).						
1	2	3	4	5	6	7
12. I did not enjoy using the Island to conduct virtual scientific studies.						
1	2	3	4	5	6	7
13. Learning to use the Island was difficult.						
1	2	3	4	5	6	7
14. The Island helped me to improve my understanding of how scientific data is collected.						
1	2	3	4	5	6	7
15. I found myself immersed in my virtual scientific study.						
1	2	3	4	5	6	7
16. I wish there was more instructions for learning to use the Island as I felt it was initially difficult to learn.						
1	2	3	4	5	6	7
17. The Island gave me a better understanding of the role of statistical analysis in scientific research.						
1	2	3	4	5	6	7
18. I would recommend the Island to other students who complete this course.						
1	2	3	4	5	6	7
19. The Island made it easy to conduct virtual scientific studies.						
1	2	3	4	5	6	7
20. The Island contributed to my confidence in designing, conducting and analysing future scientific studies.						
1	2	3	4	5	6	7
21. The Island gave me experience in dealing with statistical issues that arise during the course of scientific research (e.g. sample size, selecting an appropriate statistical test, managing data, missing values, etc.).						
1	2	3	4	5	6	7
22. The Island improved my understanding of how scientific studies are analysed statistically.						
1	2	3	4	5	6	7
23. Overall, using the Island to conduct virtual scientific studies was a positive experience.						
1	2	3	4	5	6	7
Strongly Disagree	**Disagree**	**Slightly disagree**	**Undecided**	**Slightly agree**	**Agree**	**Strongly Agree**

Please turnover to continue the questionnaire.

Qualitative Feedback - Positives

Share at least one positive experience of using the Island.

Qualitative Feedback - Negatives

Was there anything that you did not like about using the Island or you think needs improvement? Explain your answers.

Please turnover to answer one more quick question.

Interviews

Would you like to share more of your experiences and thoughts about using the Island in an interview? Interviews will be arranged at a time and place convenient to you. If would like to participate, please leave your contact information below. A researcher will contact you at a later date to arrange an interview.

1. Name ______________________

2. Contact Number ______________________

3. Preferred Email ______________________

C.3 Study I - Semi-structured Interview Schedule

Interview Schedule ONPS2304

Participant: ___________________

Semi-structured Focus Group/Interview Schedule

What did you enjoy about using the Island?

How difficult did you find using the Island? What factors impacted that difficulty?

How do you think the Island has helped you understand the design, management and analysis of clinical trials?

Did the Island make you feel like a scientist conducting a clinical trial? If so, what was it about the Island that you think made your feel this way?

Were there any surprising things you learnt while using the Island?

How did using the Island for your project impact on your study habits?

If you could change or improve anything on the Island what would it be?

C.4 Study II - Participant Information Sheet

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

RMIT UNIVERSITY

School of Mathematical and Geospatial Sciences

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 (0) 3 9925 2283
Fax +61 (0) 3 9925 2454

School of Mathematics and Physics

University of Queensland
Brisbane 4072
Australia

Tel. +61 (0) 7 3346 7681
Fax +61 (0) 7 3365 3328

Participant Information Sheet

Evaluating Project-based Work in an Online Virtual Environment for Improving Students' Statistical Thinking

You are invited to participate in a joint research project being conducted by the University of Queensland and RMIT University. This information sheet describes the project in straightforward language, or 'plain English'. Please read this sheet carefully and be confident that you understand its contents before deciding whether to participate. If you have any questions about the project, please contact one of the investigators.

Investigators

Dr Michael Bulmer
Senior Lecturer, School of Mathematics and Physics, University of Queensland
m.bulmer@uq.edu.au
07 3365 7905

Mr James Baglin
PhD Candidate, School of Mathematical and Geospatial Sciences, RMIT University
james.baglin@rmit.edu.au
03 9925 6118

Dr Cliff Da Costa
Associate Professor, School of Mathematical and Geospatial Sciences, RMIT University
cliff.dacosta@rmit.edu.au
03 9925 6114

This project is being conducted as a part of a PhD in Statistics by James Baglin at RMIT University under the supervision of Dr Cliff Da Costa. Participants are being recruited from students undertaking STAT1201 at the University of Queensland. Dr Michael Bulmer is the course coordinator for STAT1201.

What is this study about?

This study is being conducted to evaluate the effectiveness of project work conducted with an online virtual environment (the Island) in improving understanding of quantitative research methods and analysis.

What will I be required to do?

By participating you will be required to participate in STAT1201 as you normally would, including the project work component. At the end of semester you will complete a tutorial quiz that measures statistical thinking about quantitative research methods. We wish to match your project information (including group allocation) with your responses to the tutorial quiz. This will require you to give us permission to record this course data for the purpose of research. However, it is completely voluntary whether you choose to do so. But please remember that your data will be deidentified and kept strictly confidential according to Australian privacy laws and university guidelines.

What are the risks or disadvantages associated with participation?

There are very few risks associated with your participation in this project. The most prominent risk is that your project information and quiz responses will be known by the investigators. However, the course coordinator will de-identify your data before passing it onto the other investigators and it will never be shared with anyone outside the project.

What are the benefits associated with participation?

There are no direct benefits associated with participation in this study. However, by participating you will be helping us improve our methods of teaching and learning in quantitative research design and analysis courses.

What will happen to the information I provide?

The information gathered from this study will be kept strictly confidential and will not be released to a third party. Your data will be securely kept at RMIT for 5 years after the completion of the study.

Summarised and aggregated results from this study will appear in future reports and peer-reviewed publications. You can obtain copies of these reports by contacting the lead investigator once they have been completed.

What are my rights as a participant?

As a participant in this study you are ensured ethical rights. This includes the right to withdraw from the study at any given time without prejudice, the right to have any unprocessed data removed and destroyed provided it can be reliably identified, and the right to have any questions answered at any time. You can exercise your ethical rights by James Baglin (james.baglin@rmit.edu.au, 03 9925 6118).

Contacts

This study has been approved by the UQ Behavioural & Social Sciences Ethical Review Committee (reference number 2011001393).

If you would like to discuss your participation in the study you are encouraged to contact James Baglin (james.baglin@rmit.edu.au, 03 9925 6118).

If you would like to speak to an officer of the University not involved in the study, you may contact the UQ Ethics Officer on 07 3365 3924 and quote reference number 2011001393.

C.5 Study II - Consent Form

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

RMIT UNIVERSITY

School of Mathematical and Geospatial Sciences

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 (0) 3 9925 2283
Fax +61 (0) 3 9925 2454

School of Mathematics and Physics

The University of Queensland
St Lucia, Brisbane 4072
Australia

Tel. +61 (0) 7 3346 7681
Fax +61 (0) 7 3365 3328

PARTICIPANT CONSENT FORM

Prescribed Consent Form For Persons Participating In Research Projects Involving Disclosure of Personal Information

School of	Mathematics and Physics (University of Queensland) Mathematical and Geospatial Science (RMIT University)
Name of participant:	
Student Number:	
Project Title:	Evaluating Project-based Work in an Online Virtual Environment for Improving Students' Statistical Thinking

Name(s) of investigators:

(1)	Dr. Michael Bulmer (University of Queensland)	Phone:	07 3365 7905
(2)	James Baglin (RMIT University)	Phone:	03 9925 6118
(3)	Dr. Cliff Da Costa (RMIT University)	Phone:	03 9925 6114

1. I have received a Participant Information Sheet explaining my participation in this project.
2. I consent to participate in the above project, the particulars of which have been explained to me.
3. I authorise the investigator or his or her assistant to record the information outlined in the Participant Information Sheet for the purpose of this project.
4. I acknowledge that:
   (a) Having read the Participant Information Sheet, I agree to the general purpose, methods and demands of the study.
   (b) I have been informed that I am free to withdraw from the project at any time and to withdraw any unprocessed data previously supplied.
   (c) The project is for the purpose of research and/or teaching. It may not be of direct benefit to me.
   (d) The privacy of the personal information I provide will be safeguarded and only disclosed where I have consented to the disclosure or as required by law.
   (e) The security of the research data is assured during and after completion of the study. The data collected during the study may be published, and a report of the project outcomes will be provided to all consenting participants. Any information which will identify me will not be used.

Participant's Consent

Participant: ______________________________ ***Date:*** ____________________
(Signature)

This study adheres to the Guidelines of the ethical review process of The University of Queensland. Whilst you are free to discuss your participation in this study with project staff (Dr. Michael Bulmer, Ph: 07 3365 7905), if you would like to speak to an officer of the University not involved in the study, you may contact the University's Ethics Officer on 07 3365 3924.

C.6 Study II - Test of Statistical Thinking

Observational Study

1. Suppose you need to conduct an observational/correlational study that will determine if there is statistical evidence of association/relationship between eating a diet high in protein and body fat percentage. Explain how you would design, conduct and analyse the results your study by addressing each of the following points.

I. Explain how you would obtain a sample for your study.
II. Explain what data you need to gather to answer the research question and how you would go about obtaining it.
III. Based on the data that you proposed to gather in II, explain how you would plan to summarise and present the results of the study.
IV. Which statistical test would you use to perform hypothesis testing based on the data that you proposed to gather in II and summarise in III? Justify your choice of test.
V. In your own words, explain why it is important to perform hypothesis testing for this study.
VI. Assume at the end of the study you find evidence of an association/relationship. Explain what you expect your summary data and hypothesis testing results to look like

Experiment

2. Suppose you need to conduct an experiment that will determine if caffeine consumption prior to a lecture helps to improve university student's attention. Explain how you would design, conduct and analyse the results your experiment by addressing each of the following points.

I. Explain how you would obtain a sample for your experiment.
II. Outline how you would design and conduct the experiment to obtain the required data to address the research question.
III. Based on the data that you proposed to gather in II, explain how you would plan to summarise and present the results of the experiment.
IV. Which statistical test would you use to perform hypothesis testing based on the data that you proposed to gather in II and summarise in III? Justify your choice of test.
V. In your own words, explain why it is important to perform hypothesis testing on the data from this experiment.
VI. Assume at the end of the experiment you find evidence that caffeine improves attention. Explain what you expect your summary data and hypothesis testing results to look like.

C.7 Study II - TST Grading Scheme - Observational

Q.1	High (3)	Moderate (2)	Low (1)	Poor (0)
I	The student will discuss the goal of getting a representative sample through some sort of random sampling technique.	The student will discuss the use of random sampling technique, but does not explain the goal of getting a representative sample of the population.	The student simply states that they need a sample and/or they will use statistical power analysis to get an appropriate sample size.	Student does not address sampling.
II	The student discusses how they will go about getting data on protein intake (e.g. food diaries) and body fat % (e.g. BMI).	The student correctly explains which variables they need to get data on and give some insight in how that data might be obtained.	The student only explains one variable that they need to obtain.	The student describes an experiment where they manipulate who gets a high protein diet. The student does not identify the data that needs to be gathered.
III	Based on the data that they gather in II, the student selects appropriate descriptive statistics and/or graphical displays that would effectively communicate the results of their study.	The student presents only one descriptive summary or graph that would effectively communicate the results of the study. However, other summaries or plots that could enhance the presentation of their results could be included.	The student selects an appropriate summary, but does not explain how that summary would be used.	The summary statistics or plot selected does not flow from the data proposed to be gathered in II. The summary statistic or plot selected is inappropriate. The student just lists a “table” or “plot”.
IV	Based on the data that they gathered in II and the summaries they explained in III, the student selects an appropriate statistical test and explains why this test suits their study.	The student selects an appropriate statistical test that suits the data they selected in II and summarized in III, however they cannot adequately explain why it is appropriate.	The student selects the right test based on the data they proposed to gather in II and summarize in III, perhaps by coincidence, but it is apparent they lack insight as to why.	The student selects an inappropriate test that does not suit the data they proposed to gather in II and summarize in III.

continued on next page

continued from previous page

Q.1	High (3)	Moderate (2)	Low (1)	Poor (0)
V	The student discusses the issue of drawing inference about populations using samples. They then describe that hypothesis testing is an attempt to address this issue by considering the likelihood of observing a sample results under the null hypothesis.	The student discussed something around the concept of their sample results occurring by chance. They do not identify the issues of drawing inferences about populations using samples.	They student merely states that hypothesis testing is to check for statistical significance, or to providence evidence to reject the null hypothesis.	Any response that does not fit into the other categories. E.g. to prove my results, to prove the research hypothesis, to prove that other factors did not account for an association etc.
VI	The student describes how their summary statistics and plots will look taking into account their responses to II, IV. They also explain how the test they chose in IV will be statistically significant.	The student only explains how either their summary statistics or hypothesis testing results will look, not both.	The student is able to recall that $p < 0.05$, but not attempt is made to link this with their descriptive statistics or the specific statistical test they chose.	The student describes a result that would be regarded evidence against an association ($p > .05$) or only describes what information they would consider, e.g. "I would look for a trend in the plots and the p-value of the statistical test.

Study II - TST Grading Scheme - Experiment

Q.2	High (3)	Moderate (2)	Low (1)	Poor (0)
I	The student will discuss the goal of getting a representative sample of students through some sort of random sampling technique.	The student will discuss the use of random sampling technique, but does not explain the goal of getting a representative sample of the population.	The student simple states that they need a sample and/or they will use statistical power analysis to get an appropriate sample size.	Student does not address sampling.
II	The student discusses the design of an experiment by addressing random allocation, the use of a control (placebo group) and manipulation of caffeine as an independent variable. The student proposed a suitable way of measuring attention, e.g. end of lecture quiz scores.	The student discusses the design of an experiment by using a control (placebo group) and manipulation of caffeine as an independent variable. However, they do not address random allocation. The student proposed a suitable way of measuring attention, e.g. end of lecture quiz scores.	The student describes some notion of an experiment, but does not provide enough detail. They do not explicitly state how they will measure attention/or the measure of attention is not suitable.	The student describes an observation research design. The student does not describe an experiment or does not describe how the data will be collected.
III	Based on the data that they gather in II, the student selects appropriate descriptive statistics and/or graphical displays that would effectively communicate the results of their study.	The student presents only one descriptive summary or graph that would effectively communicate the results of the study. However, other summaries or plots that could enhance the presentation of their results could be included.	The student selects an appropriate summary, but does not explain how that summary would be used.	The summary statistics or plot selected does not flow from the data proposed to be gathered in II. The summary statistic or plot selected is inappropriate. The student just lists a "table" or "plot".
IV	Based on the data that they gathered in II and the summaries they explained in III, the student selects an appropriate statistical test and explains why this test suits their study.	The student selects an appropriate statistical test that suits the data they selected in II and summarized in III, however they cannot adequately explain why it is appropriate.	The student selects the right test based on the data they proposed to gather in II and summarize in III, perhaps by coincidence, but it is apparent they lack insight as to why.	The student selects an inappropriate test that does not suit the data they proposed to gather in II and summarize in III.

continued on next page

continued from previous page

Q.2	High (3)	Moderate (2)	Low (1)	Poor (0)
V	The student discusses the issue of drawing inference about populations using samples. They then describe that hypothesis testing is an attempt to address this issue by considering the likelihood of observing a sample results under the null hypothesis.	The student discussed something around the concept of their sample results occurring by chance. They do not identify the issues of drawing inferences about populations using samples.	They student merely states that hypothesis testing is to check for statistical significance, or to providence evidence to reject the null hypothesis.	Any response that does not fit into the other categories. E.g. to prove my results, to prove the research hypothesis, to prove that other factors did not account for a difference etc.
VI	The student describes how their summary statistics and plots will look taking into account their responses to II, IV. They also explain how the test they chose in IV will be statistically significant.	The student only explains how either their summary statistics or hypothesis testing results will look, not both.	The student is able to recall that $p < 0.05$, but not attempt is made to link this with their descriptive statistics or the specific statistical test they chose.	The student describes a result that would be regarded as evidence against an effect for caffeine ($p > .05$) or only describes what information they would look at, e.g. ”I would look for a trend in the plots and the p-value of the statistical test.