Mental Health in Computer Science

An Investigation of How Mental Health Affects Learning Computer Science

KEITH NOLAN

A thesis submitted for the degree of

Doctor of Philosophy

Department of Computer Science

Maynooth University

October 2019

Supervisors: Dr. Susan Bergin & Dr. Aidan Mooney

Head of Department: Dr. Joseph Timoney

CONTENTS

1 INTRODUCTION 1
- 1.1 Motivation . . . 1
- 1.2 Objectives . . . 3
- 1.3 Definitions . . . 5
- 1.4 Research Method . . . 6
- 1.5 Thesis Overview . . . 7

2 RELATED LITERATURE 9
- 2.1 Mental Health in Computer Science . . . 9
  - 2.1.1 Research Questions . . . 10
  - 2.1.2 Method . . . 11
  - 2.1.3 Results - Background . . . 17
  - 2.1.4 Results . . . 19
  - 2.1.5 Discussion . . . 30
- 2.2 Conclusion . . . 31
- 2.3 Thesis Update . . . 32

3 STUDY BACKGROUND AND INSTRUMENTS 34
- 3.1 Participants . . . 34
- 3.2 Anxiety . . . 35
  - 3.2.1 Methods . . . 35
- 3.3 Sensors . . . 39
  - 3.3.1 Electrodermal Activity . . . 40
  - 3.3.2 Photoplethysmography . . . 42
- 3.4 Programming Questionnaire . . . 45
  - 3.4.1 Development . . . 45
  - 3.4.2 Validation . . . 46
  - 3.4.3 Sample Questions . . . 48

3.5 Additional Surveys 50
3.5.1 Programming Self-efficacy Questionnaire 50
3.6 Summary 50
3.7 Thesis Update 52
4 THESIS STUDY 1 55
4.1 Research Questions 55
4.2 Study Overview 56
4.2.1 Participant Profile 57
4.3 TE1-RQ1 : Anxiety and Performance 57
4.3.1 Required data 58
4.3.2 Analysis 60
4.4 TE1-RQ2 : Anxiety and Stress 64
4.4.1 Required data 64
4.4.2 Analysis 66
4.4.3 Discussion 70
4.5 TE1-RQ3 : Stress and Performance 70
4.5.1 Required Data 70
4.5.2 Analysis 71
4.5.3 Discussion 73
4.6 Summary of chapter 73
4.7 Thesis Update 75
5 THESIS STUDY 2 78
5.1 Background 79
5.1.1 Programming Self-efficacy 79
5.1.2 Confidence in responses 80
5.2 Research Questions 81
5.3 Changes to Study 1 protocol 82
5.4 Study Protocol 82
5.4.1 Participant Profile 83
5.5 VAL-RQ1 : Anxiety Levels 83

5.5.1 Required data 84
5.5.2 Discussion 84
5.6 VAL-RQ2 : Stress and Question Difficulty 85
5.6.1 Required data 85
5.6.2 Photoplethysmography 85
5.6.3 Discussion 86
5.7 VAL-RQ3 : Behavioural Responses 87
5.7.1 Required data 87
5.7.2 Discussion 88
5.8 TE2-RQ1 : Stress and Confidence 91
5.8.1 Required data 91
5.8.2 Confidence In Responses 92
5.8.3 Analysis 93
5.8.4 Discussion 95
5.9 TE2-RQ2 : Anxiety and Confidence 96
5.9.1 Required data 96
5.9.2 Analysis 96
5.9.3 Discussion 99
5.10 Summary of Chapter 100
5.11 Thesis Update 104
6 LARGE SCALE STUDY OF ANXIETY IN COMPUTER SCIENCE 107
6.1 Research Questions 107
6.2 Study Protocols 108
6.2.1 Surveys 108
6.2.2 Data Collection 109
6.2.3 Gender and Age Profile 109
6.3 Anxiety in CS1 students 110
6.3.1 Required Data 110
6.3.2 Analysis 110
6.3.3 Discussion 112

6.4 SUR-RQ2: Anxiety and Exam Result 112
6.4.1 Required Data 112
6.4.2 Analysis 112
6.4.3 Discussion 116
6.5 SUR-RQ3: Programming Self-efficacy and Anxiety 117
6.5.1 Required Data 117
6.5.2 Analysis 117
6.5.3 Discussion 121
6.6 SUR-RQ4: Programming Self-efficacy and Exam Performance 121
6.6.1 Required Data 121
6.6.2 Analysis 121
6.6.3 Discussion 122
6.7 SUR-RQ5: CS students Programming Self-efficacy 122
6.7.1 Required Data 122
6.7.2 Analysis 123
6.7.3 Discussion 123
6.8 Summary of Chapter 124
6.9 Thesis Update 126
7 GENDER DIFFERENCES IN COMPUTER SCIENCE 129
7.1 Motivation and Background 129
7.2 Research Questions 132
7.3 GD-RQ1 : Stress Responses Differences 133
7.3.1 Analysis 133
7.3.2 Discussion 137
7.4 GD-RQ2: State Anxiety and Trait Anxiety Differences . . . 138
7.4.1 Analysis 138
7.4.2 Discussion 140
7.5 GD-RQ3: Behavioural Responses Differences 141
7.5.1 Analysis 141

7.5.2 Discussion 144
7.6 GD-RQ4: Programming Self-Efficacy Difference 144
7.6.1 Analysis 144
7.6.2 Discussion 145
7.7 Summary of Chapter 145
7.8 Thesis Update 146
8 CONCLUSIONS AND FUTURE WORK 149
8.1 Conclusions 149
8.1.1 Objective 1 150
8.1.2 Objective 2 150
8.1.3 Objective 3 151
8.1.4 Objective 4 151
8.1.5 Further Contributions 152
8.2 Future Work 153
8.2.1 Anxiety and performance in Computer Science . . 153
8.2.2 Mental health of CS Students 154
8.2.3 Use of physiological sensors 154
Bibliography 155
Appendix 174
A PUBLICATIONS 174
B STATISTICAL TECHNIQUES 176
C ANALYSIS TECHNIQUES 181
C.1 Electrodermal Activity 181
C.2 Photoplethysomgram 182
D SYSTEMATIC LITERATURE REVIEW TABLE 184
D.1 Surveys used 184
D.2 Systematic Literature Review Table 185
E PROGRAMMING QUESTIONS 189
E.1 Programming Questions 189
F INTERNATIONAL COLLABORATION 192

F.1 Motivation for International Collaboration 192
F.2 Materials 194
F.3 Experimental Paradigms 194
F.3.1 Frontal Alpha Asymmetry 195
F.3.2 Relationship to attention to threat 195
F.3.3 Emotional Stroop Paradigm 196
F.3.4 Dot Probe Paradigm 198
F.4 Concept Studies 199
F.4.1 Paradigm Discussion 202
F.5 Behavioural Study 204
F.5.1 Experimental Protocol 204
F.6 Results 206
F.6.1 Participant Profile 206
F.6.2 Dot Probe Experiment 206
F.6.3 Emotional Stroop 209
F.7 Summary 211
G ETHICAL APPROVAL 213
G.1 Ethical Approval 213
H CONSENT AND INFORMATION SHEET 217
H.1 Sample Consent and Information Sheet 217
I PARTICIPANT QUESTIONNAIRE 222
I.1 Participant Questionnaire 222

LIST OF FIGURES

Figure 1.1 Flow Of Thesis 8

Figure 2.1 Number of papers collected each year 18

Figure 2.2 Logical Flow Of Thesis 33

Figure 3.1 EDA placement of sensors on the finger tips . . . 42

Figure 3.2 PPG setup 43

Figure 3.3 Heart Beats captured from PPG 44

Figure 3.4 How instrument and Materials Interact 52

Figure 3.5 Logical Flow Of Thesis 54

Figure 4.1 Scatter plot showing the cluster assignments for State anxiety, Trait anxiety, #Correct responses and Total Response Time. 62

Figure 4.2 Scatter plot showing the cluster assignments for State anxiety, Trait anxiety, SCRs and PPG data. . 68

Figure 4.3 Scatter plot showing the cluster assignments for SCRs and Total Response Time. 72

Figure 4.4 Logical Flow Of Thesis 77

Figure 5.1 Screenshot of confidence-in-response scale. 82

Figure 5.2 Clusters of SCRs Programming self-efficacy and Confidence-in-response values. 94

Figure 5.3 Clusters of participants State anxiety, Trait anxiety and Confidence-in-response values displayed on a scatter plot. 98

Figure 5.4 Logical Flow Of Thesis 106

Figure 6.1 Bar chart showing the first year averages of male and female state anxiety and trait anxiety and the normal College population levels marked in horizontal lines. 111
Figure 6.2 Results of clustering State anxiety and exam mark. 114
Figure 6.3 Clusters of participants Trait anxiety and Exam mark. 115
Figure 6.4 Chart showing Self-efficacy vs State anxiety clusters. 118
Figure 6.5 Chart showing Self-efficacy vs Trait anxiety clusters. 120
Figure 6.6 Chart showing Self-efficacy vs Exam mark and the trend line showing the linear relationship for the First Year survey. 122
Figure 6.7 Logical Flow Of Thesis 128
Figure 7.1 Logical Flow Of Thesis 148
Figure B.1 Elbow method showing the number of clusters . . 178
Figure B.2 Clusters of participants by State and Trait anxiety. 179
Figure F.1 Muse headband with sensors on forehead and resting on ears. 194
Figure F.2 Examples of the presentation of words in the Classic Stroop. 197
Figure F.3 Examples of the presentation of negative words in the Emotional Stroop. 197
Figure F.4 Flow of a typical trial in a Dot-Probe Experiment. 199

LIST OF TABLES

Table 2.1 Quality assessment of studies used in the systematic literature review. 16
Table 3.1 Samples of both the State and Trait forms with the possible answers. 39
Table 3.2 Core concepts shown in the experiment with N representing a new concept. 47
Table 3.3 Programming Self-efficacy questionnaire with possible answers. 51
Table 4.1 Age and gender profile of participants in study 1. 57
Table 4.2 Average State anxiety and Trait values for male and female participants and associated p-values. In addition, the Normal male and female score are presented in *italics*. 58
Table 4.3 All concepts with percentage breakdown of correct and incorrect responses. 59
Table 4.4 Average time taken in seconds (s) to respond to each question. 60
Table 4.5 Correlations of the factors 61
Table 4.6 Table showing the average State anxiety, Trait anxiety, #Correct responses, Total Response Time per cluster. 62
Table 4.7 Correlations of the factors 63
Table 4.8 Average number of Skin Conductance Responses across each question difficulty band for Study 1. . 65
Table 4.9 Average $ln(RMSSD)$ and Average Heart Rate values for each half in study 1. 66

Table 4.10	Correlations of the factors	67
Table 4.11	Table showing the average State anxiety, Trait anxiety, SCRs and PPG data per cluster.	69
Table 4.12	Correlations of the factors	69
Table 4.13	Correlations of the factors	71
Table 4.14	Table showing the average SCRs and Total Response Time per cluster.	73
Table 5.1	Age and gender profile of participants in Study 2.	83
Table 5.2	Average values from Study 1 Study 2 of State anxiety and Trait anxiety.	84
Table 5.3	Average number of SCRs across each question difficulty band.	85
Table 5.4	Average $ln(RMSSD)$ values and Average Heart Rate values for both Study 1 and Study 2.	86
Table 5.5	The results of each of the questions for Study 1 and Study 2.	88
Table 5.6	Average time taken in seconds (s) to respond correctly and incorrectly to each question for Study 2.	89
Table 5.7	Confidence and correctness in answers are outlined here. (CC - Correct and Confident, CNC - Correct and Not Confident, IC - Incorrect and Confident and INC - Incorrect and Not Confident)	92
Table 5.8	Correlations of SCRs, Programming self-efficacy and Confidence-in-response	93
Table 5.9	Average values of clusters	94
Table 5.10	Correlations of SCRs, Programming self-efficacy and Confidence-in-response clusters	95

Table 5.11 Correlations within clusters of State anxiety, Trait anxiety, programming self-efficacy and confidence-in-response. 97

Table 5.12 Average values of clusters of State anxiety, Trait anxiety Programming self-efficacy and confidence-in-response. 98

Table 5.13 Correlations within clusters of State anxiety, Trait anxiety, programming self-efficacy and confidence-in-response. 99

Table 6.1 Age and gender breakdown of students in both the First year. 110

Table 6.2 Average State anxiety and Trait anxiety levels. . . 110

Table 6.3 Properties of exam results. 113

Table 6.4 Correlations between State anxiety, Trait anxiety and Exam Mark within the study. 113

Table 6.5 Description of clusters of State anxiety and Exam Mark 114

Table 6.6 Description of clusters of Trait anxiety and Exam Mark 116

Table 6.7 Description of clusters of Self-efficacy and State anxiety 118

Table 6.8 Description of clusters of Self-efficacy and State anxiety 120

Table 6.9 Programming Self-efficacy broken into bands with counts of participants in each band 123

Table 7.1 Average number of SCRs across the studies and each question difficulty band categorised by gender133

Table 7.2 Cohen d values of the differences between the genders between EDA values 135

Table 7.3 Average ln(RMSSD) and Heart Rate values for each study half categorised by gender 136
Table 7.4 Average State anxiety and Trait anxiety values for male and female participants and associated p-values. N values in brackets () 139
Table 7.5 Cohen d values for State and Trait values 140
Table 7.6 Percentage of correct answers, average time taken in seconds (s) to respond correctly to each question, grouped by gender 142
Table 7.7 Percentage of correct answers, average time taken in seconds (s) to respond correctly to each question, grouped by gender 143
Table 7.8 Programming Self-Efficacy values (10-40 points) for male and female participants and associated p-values 145
Table D.1 Table of surveys used among studies. 184
Table D.2 Accepted studies 185
Table F.1 Age and gender profile of participants 206
Table F.2 Average number of correct responses with their associated average response times for the Dop-Probe. 208
Table F.3 Trials broken into 4 sets of 64 trials considered for Dot-Probe analysis 210
Table F.4 Average number of correct responses with their associated average response times for the Emotional Stroop. 211

ACRONYMS

ACM Association for Computing Machinery

ECG Electrocardiogram

EDA Electrodermal Activity

GSR Galvanic Skin Response

HR Heart Rate

HRV Heart rate variability

ITiCSE Innovation and Technology in Computer Science Education

IZOF Individual Zone of Optimum Functioning

MCQ Multiple-Choice Question

PPG Photoplethysmogram

RMSSD Root Mean Squared of Successive Differences

SCR Skin Conductance Response

STAI State Trait Anxiety Inventory

ABSTRACT

The mental health of third level students is potentially at an all-time low. Reports such as the *My World Survey*, the *My World Survey 2* and the *Union of students of Ireland Report* indicate that third level students in Ireland are suffering from mental health issues. For students, mental well-being is associated with effective learning, and their ability to navigate through university, coping with the challenges and stresses of student life. As such, this project attempted to investigate the effects that mental health factors such as stress and anxiety have on programming performance within a first-year Computer Science population.

This project had four objectives. First, was to examine the relationship between student anxiety and CS1 programming performance. Second, was to examine the relationship between student stress and CS1 programming performance. Third, was to examine the relationship between student anxiety and stress. Finally, was a review the data obtained throughout the project, to identify analyse and identify gender differences.

As an initial contribution of this project, a detailed systematic literature review on the role of anxiety in learning in Computer Science was carried out. No such review had previously been completed making this a timely addition to the field. As a second contribution, a novel study investigating the use of physiological sensors to investigate stress in an online MCQ examination with first-year Computer Science students was carried out. Findings suggest that there is a positive relationship between EDA and question difficulty. The third contribution was three studies on anxiety in Computer Science students, one containing a large sample (at least 65% of the CS1 cohort). Related to this was

the novel finding that Computer Science students are more anxious. In addition was the investigation on programming self-efficacy and confidence in answers and their relationship to anxiety, arousal and performance. Evidence on the importance of programming self-efficacy was found to re-validate previous findings. The final contribution was a novel study on gender differences in stress, anxiety and self-efficacy. The findings presented are novel, providing telling insights into the role that different factors have on mental health when learning to program.

ACKNOWLEDGMENTS

Firstly, I would like to thank my family, Geoff, Deirdre and Michelle, for their unwavering support and encouragement throughout my studies. For their constant support, picking me at the low points and celebrating all the small things, no matter how small. I truly can not thank you enough.

To my supervisors, Dr. Susan Bergin and Dr. Aidan Mooney, what can I say; your support though out the entire process, constantly asking the questions I did not want to answer and pushing me to think about things in a different light. For the long distance phone calls and constant encouragement, I could not have had a better supervising team. I hope to we get the opportunity to work together in the future.

To my fellow postgraduate students in the Computer Science department, thank you for putting up with my antics over the past 4 years. To those who participated in all my experiments, thank you. To my proof readers, Mark Noone, William Clifford and Toby Burns, your help and feedback is appreciated. For your support and advice through out the thesis, Keith Quille, thank you for your enthusiasm, help and support over the 4 years. Emlyn Hegarty, my "Ph.D person", what else can I say but my sincerest thank you for your friendship and support over the many years.

Finally, to Deirdre F, - thank you for your patience and encouragement throughout my studies. Thank you for the many days and nights motivating me, putting up with the many weekends lost to research and the time spent proof reading and providing comments on the thesis. I am looking forward to our next adventure together.

DECLARATION

I confirm that this is my own work and the use of all material from other sources has been properly cited and fully acknowledged.

Keith Nolan
October 2019

1 INTRODUCTION

1.1 MOTIVATION

A computational model designed to predict student success in introductory Computer Science(CS), named PreSS, was developed between 2003 and 2006. It could successfully predict student performance with 80% accuracy after minimal exposure to programming concepts (approximately one-quarter of the way through an introductory module) [10, 12]. The model was tested with 240 students over multiple years and in multiple institutions. The model used a number of factors to determine success. Of the 25 factors that were initially examined, the three main determinants of success were found to be programming self-efficacy, mathematical ability and the number of hours per week a student plays computer games. Accuracy was increased however when gender-specific models were developed as several of the factors varied significantly by gender.

The model was revisited a decade later and resulted in the development of a fully-automated web-based version known as PreSS# [83]. The new model improved on PreSS by removing the need for paper-based surveys and allowing for a real-time system of predicting success/failure with different visualisations. The model incorporated new factors such as age, student predictive grade and time spent on social media. Overall the predictive model did not improve dramatically, however, the real significance of the project was the confirmation that although the landscape had changed considerably over the 10 year period since Press was developed (student profile, new languages, changes in technology usage (e.g. social media growth)), the model could still predict performance with similarly high levels of accuracy. There has however been one notable constant since the development of the original Press model: failure and attrition rates in computer science have not changed.

The causes of the failure and attrition rates are likely complex and multifaceted and perhaps mental health is a contributing factor. Mental well-being is a significant concern worldwide with many studies and interventions (teaching techniques, support services) developed to support students [42]. In 2012 Headstrong, now Jigsaw, a registered charity aiming to improve young peoples mental health, conducted the *My World Survey* with over 8,200 young Irish adult participants. They found that in any given 100 students, irrespective of subject or discipline, 40 students suffer from depression[1] and 38 students suffer from anxiety[1] with the three main stressors identified as, college, money and work [31]. The study was replicated in 2019 with over 8,290 young adults and findings suggest that 26% of young adults were categorised as being in the severe to very severe anxiety category which is an 11% rise on the My World Survey 1 results [32]. More recently, the Union of Students of Ireland conducted a similar study of Irish University students with 3,340 participating [81]. Findings suggest that 38% of students have severe levels of anxiety, 30% of students have depression and 17% have some form of stress. These findings of the *My World Survey 1*, the *My World Survey 2* and the *Union of Students of Ireland report* indicates that in general student mental health is a real cause for concern. Thus, it seems valuable to investigate how mental health may relate to performance

1 Measured by DASS-21. The depression scale assesses dysphoria, hopelessness, devaluation of life, self-deprecation, lack of interest/involvement, anhedonia and inertia. The anxiety scale assesses autonomic arousal, skeletal muscle effects, situational anxiety, and subjective experience of anxious affect.

in Introductory Computer Science (CS1). As such, it seems reasonable that including students anxiety and stress as factors in models such as PreSS# could improve the accuracy of such models.

To do this however, appropriate instruments for measuring anxiety and stress would need to be determined and used in a CS environment. With respect to anxiety, several off-the-shelf instruments have been validated for collecting the anxiety of college students either prior of after an event. Gaining an insight into ones biological signals (to measure Stress and Heat Rate) has almost become commonplace in today's world. Physiological sensors have become more accessible to the public through smartwatches and other wearable technology. The use of these sensors in the modern-day classroom can allow educators to gain insight into how a student is engaging (physiologically) with a class and with the course material. This is valuable as when a student becomes stressed they begin to disengage with the material being presented [92, 114]. Identifying students who have disengaged in class due to becoming stressed and responding in a timely and appropriate fashion could make a considerable difference to the student and their learning.

1.2 OBJECTIVES

The objectives of this thesis are:

- To examine the relationship between student anxiety and CS1 programming performance.
- To examine the relationship between student stress and CS1 programming performance.
- To examine the relationship between student anxiety and stress.
- To review the data obtained throughout the project, to identify analyse and identify any gender differences.

To understand the relevance of these objectives, a brief explanation is provided.

To examine the relationship between anxiety and CS1 programming performance.

Anxiety has been well researched across numerous disciplines, however, little has been investigated with respect to student anxiety in Computer Science. If a relationship between anxiety and performance can be identified in this thesis, interventions designed to aid students in the future can be put into place.

To examine the relationship between stress and CS1 programming performance.

This project will investigate if there is a relationship between stress and performance. If such a relationship can be established, a real-time system could be developed which monitors and responds to stress signals to improve the chances of success.

To examine the relationship between anxiety and stress.

Given the scope of the project, it is hoped that consumer-grade sensor technology could be utilised to provide real-time information as to when a student starts to become anxious by potentially measuring their stress signal. This will reduce the need to administer lengthy questionnaires.

Review the data obtained throughout the project, to identify analyse and identify and gender differences.

Given that gender specific models have improved the accuracy of Press#, it seemed reasonable to investigate gender as part of this project. Given the breath of data that will be collected there is an opportunity to investigate the gender differences in anxiety, stress signals and performance might be in a computer science setting. By doing this, gender-specific supports could be developed which might reduce the gender gaps in CS.

1.3 DEFINITIONS

Terms that will be used throughout this thesis are defined in this section. Anxiety and stress are terms used synonymously, however they are not the same. The following definitions are used in this thesis:

- Anxiety - Anxiety is an emotion based on the appraisal of threat, an appraisal which entails symbolic, anticipating and uncertain elements [62]. There are two main types of anxiety: State anxiety and Trait anxiety. State anxiety is defined as an unpleasant emotional arousal in the face of threatening demands or dangers. A cognitive appraisal of threat is a prerequisite for the experience of this emotion [56]. Possible sources of State anxiety could be driving, flying, taking tests etc. Trait anxiety refers to the tendency to attend to, experience, and report negative emotions such as fears and worries across many situations [41]. Examples of Trait anxiety are harder to form given the inherent nature of Trait anxiety. The higher the Trait anxiety measure, the more susceptible one is to experience general anxiety, i.e. someone with high Trait anxiety might respond negatively to a stimulus whereas someone with low Trait anxiety may not respond at all.

- Physiological Change - A physiological change is a change in behavioural responses that people have little control of, for example, heart rate, sweat rate, skin temperature, breathing rate, blink rate etc.

- Emotional Arousal - Emotional arousal is a physiological change due to psychological activation. An example of this would be fright or excitement that is induced by an external stimulus that was not expected. It is difficult to understand if emotional arousal is positive or negative without knowing the external stimulus. For this project, it can be argued that the emotional arousal that students experience here will be stress. This is due to the nature of the experimental setup stress as there are in an unknown situation completing CS1 tasks.

- Stress - Stress is an imbalance between physical and psychological factors [38]. For example, a change in the environment around a person will cause a person to make an appraisal of the environment. If the appraisal of the situation is negative, pressure could be exerted on the person causing an internal representation of stress. An example of this might be a surprise test in a class. The sudden onset of the test will influence the psychological factors for the student and will cause stress.

1.4 RESEARCH METHOD

A systematic literature review will be undertaken. This will allow for the generation of a clear picture of what research has been conducted in the scope of Computer Science and mental health. Also, the systematic literature review will follow a strict set of protocols to produce an exhaustive search of the literature and allow for easy reproduction and verification.

Background research on off-the-shelf instruments suitable for use in this work will be carried out. This research will also help inform the design of the research protocol and any instruments that need to be developed for the project. These instruments would allow for the collection of empirical evidence through the use of surveys, physiological signals and performance metrics. Once these instruments are decided upon, the studies can take place.

It was decided that in order to conduct this research, three interconnected consecutive studies would be performed. To examine the relationships between anxiety, stress and performance an initial study will be carried out. The study will provide preliminary evidence that will be validated through a follow-up study. A final study will be carried out for further clarification or evidence as required. Following this, the outcomes of all studies will be investigated with respect to gender. It is expected for the project to follow the flow depicted in Figure 1.1 whereby anxiety, stress and Computer Science programming performance will be the main focus themes of the research project.

1.5 THESIS OVERVIEW

Chapter two of this thesis provides a detailed systematic review of the role of anxiety when learning to program. This chapter highlights the process of conducting a systematic review and describes the main findings of such. Chapter three focuses on the instruments and sensor technologies used throughout the thesis. The development and validation of a programming MCQ test are also outlined along with the sensor technologies used.

Chapter four describes the first study of this thesis which investigates if stress levels can be detected in an MCQ test. The relationship between self-reported anxiety measures and physiological data is explored. Chapter five builds on the work of the first study with the relationship between anxiety, confidence, and self-efficacy examined in more detail. Chapter six investigates the levels of anxiety in first-year CS. In addition to this, the programming self-efficacy of students was collected and its relationship with performance is examined. Chapter seven discusses the gender differences that were uncovered throughout the thesis experiments and explores how these gender differences may be mitigated. Chapter eight provides the conclusions of the thesis and outlines possible future work based on the findings of the thesis.

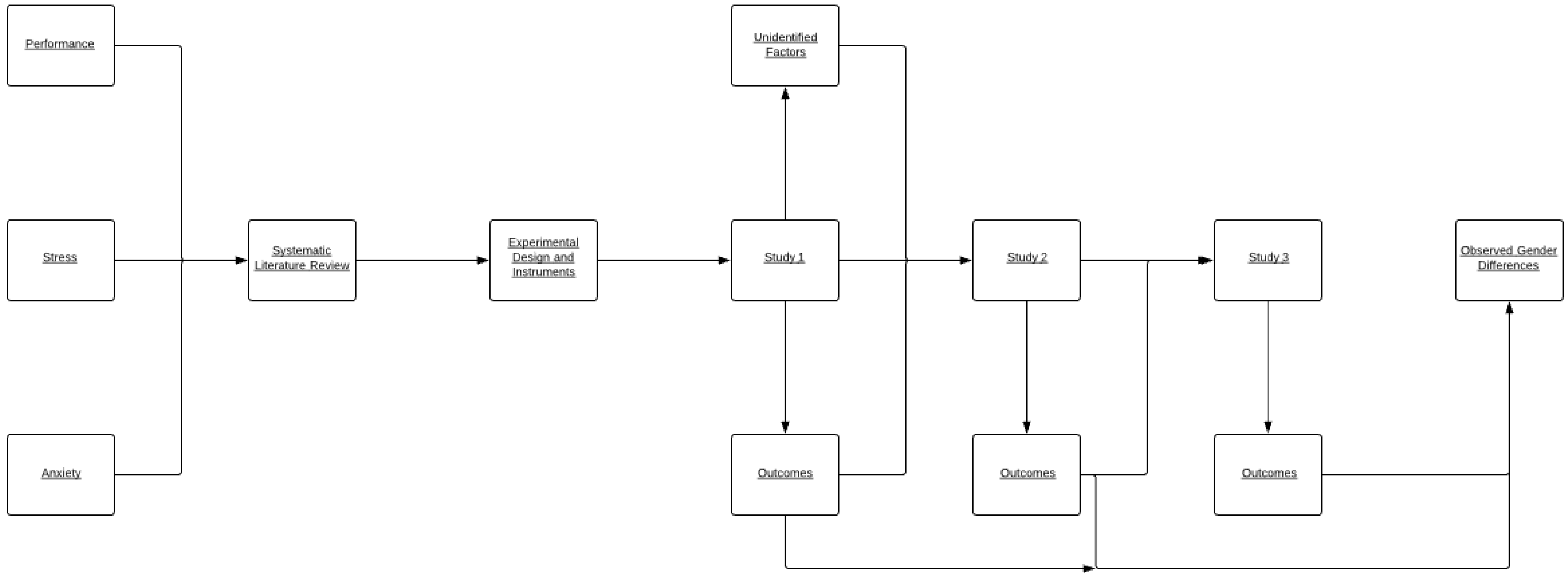


Figure 1.1: This is the flow of the project. Anxiety, Stress and Performance will inform a Systematic Literature Review which will inform Instruments and Materials. From there the studies will be conducted and outcomes will inform all objectives.

2

RELATED LITERATURE

While literature reviews have been conducted in the fields of anxiety and mental health in college students, these reviews have been broad and subject independent. This chapter describes a detailed systematic review of anxiety and stress in Computer Science. The research questions of the review, methodologies and findings are presented.

2.1 MENTAL HEALTH IN COMPUTER SCIENCE

At the commencement of this research, no previous review could be found that synthesised the state of the art on the relationship between anxiety and stress and the study of Computer Science. To that extent, it was decided that a through literature review should be conducted. A systematic review of the literature was chosen over a traditional narrative review as such an approach, although more involved, results in an unbiased, thorough and reproducible review through the application of a strict protocol (as described in Section 2.2). This is important as so little it is known in this space.

As a starting point to the systematic review, potential causes of anxiety and stress were examined and identified as follows:

1. Programming as a topic (the language, syntax, error generation, learning environment etc.) can lead to anxiety and stress.

2. Test anxiety is relevant as it is present in any discipline.

3. Computer anxiety, that is anxiety induced through the use of a computer, is important given the volume of computer usage.

4. Anxiety and Stress associated with mathematics given its strong relationship with CS and programming is also important (e.g. writing programs to determine prime numbers, greatest common divisor, factorial, etc).

Thus the goal was to review the role of anxiety and stress of students when learning to program by considering anxiety and stress associated with programming itself, mathematical concepts, computer usage and assessment.

2.1.1 RESEARCH QUESTIONS

The starting point of the review was to develop the research questions which would be addressed. Several research questions were defined to incorporate the breadth of sources of anxiety when learning to program. The defined questions were as follows:

LR-RQ1 Is there a relationship between anxiety and stress and learning to program (language, syntax, compilation etc.)

LR-RQ2 Is there a relationship between mathematical anxiety and learning to program?

LR-RQ3 Does computer usage cause stress and anxiety when learning to program?

LR-RQ4 Does test anxiety affect learning to programme and more broadly Computer Science students?

These questions will inform how the studies will be conducted and as well as informing the objectives set out in Chapter 1.

2.1.2 METHOD

Introduction

The systematic literature review carried out was based on Kitchenham's method as applied to software engineering [57]. This method of performing a review was chosen as the process is well documented. Kitchenham outlines how to identify the need for the review, how to develop a strict protocol to follow for the review and how to report the findings from the review.

The following steps are listed in the method:

- Identify the need for a systematic literature review and define your research questions.

  -Addressed in Section 1.1 and Section 2.1.1.

- Carry out an exhaustive search for studies.

  -Discussed in Section 2.1.2.2 and Section 2.1.2.3.

- Assess quality of accepted studies.

  -Discussed in Section 2.1.2.6.

- Extract data from accepted studies.

  -Discussed in Section 2.1.2.5.

- Compile background information on the studies.

  -Discussed in Section 2.1.3.

- Summarise and synthesise study results.

  -Discussed in Section 2.1.4.

The correct application of these steps leads to a rigorous, exhaustive and reproducible meta-review [57].

Search Terms

In this review, two primary search terms were used: *Anxiety* and *Programming*.

Given that the terms *"anxiety"*, *"emotional arousal"* and *"stress"* are often used synonymously, *"emotional arousal"* and *"stress"* were also used in addition to *"anxiety"* to increase the search scope. In addition to the above search terms, the following secondary search terms were used to narrow the number of results returned from the databases: *Learning*, *Mathematics*, *Computer*, *Test (Exam)*.

Resources Searched

An extensive search of five publication repositories was carried out between February 2015 and August 2019 using the search terms mentioned in Section 2.1.2.2. The repositories were: the ACM Digital Library, IEEE Xplore, ERIC, Science Direct and Google Scholar.

The ACM Digital Library (ACM DL) contains over 566,494 full-text papers. When searching the ACM DL with the primary search terms, "anxiety" returned 508 results and "programming" returned 118,649 results. As secondary search terms were added, the number of results returned further decreased. Given the low number of papers returned in the "anxiety" search, it was decided that all results would be screened for inclusion in the review.

The IEEE Xplore database contains over four million citations. It was searched using the same search criteria as the ACM DL. As there were only 98 results returned after searching "anxiety" and "programming", all papers were screened.

The ERIC database was then searched as the database is specifically for papers relating to education. The same search criteria used for the ACM DL and IEEE Xplore were employed. The search only returned papers that had previously been found in either the ACM DL or IEEE Xplore. Science Direct contains over 12 million citations relating to Physical Sciences and Engineering, Life Sciences, Health Sciences, and Social Sciences and Humanities. The database was searched to identify any other research related to our research questions, using the same search criteria. No additional papers were found. Google Scholar was used as a final search space to eliminate the likelihood that a relevant publication had been missed. No additional studies were found.

Document Selection

From searching the databases and referenced material, of the 500+ studies identified for possible inclusion, a total of ninety-three studies were identified based on their title alone to address some of the research questions. Full texts of those studies were then obtained. The abstracts for all ninety-three unique studies were then reviewed using the criteria outlined below and in Section 2.1.2.6 to exclude any studies that were not directly related to the research questions.

Following the methodology outlined in Kitchenhams procedure, inclusion and exclusion criteria were developed [57]. The criteria were as follows.

All texts were included that:

- potentially answered one or more research question.
- focused on anxiety or stress in programming.
- focused on anxiety or stress which related to either mathematics anxiety, computer anxiety or test anxiety.

All studies were excluded that:

- were in the form of a book or grey literature (opinion pieces, technical reports, blogs, presentation, etc.).

- related to primary or secondary school learning (one study was kept as it was deemed relevant due to the class group the study focused on [55]).

Data Extraction and Synthesis

From the ninety-three studies found based on the title, sixty-six of those were subsequently rejected after reviewing the abstract. The remaining papers were screened using the inclusion/exclusion criteria. A final list consisting of twenty-seven relevant papers that satisfied the inclusion criteria and informed the research questions was compiled. Some of these studies were only useful for one question while others addressed more than one research question.

Mendeley reference manager was used to record the reference details of each study. Along with this, a separate document was used to record additional results that Mendeley couldn't include e.g. a summary of the study. Extracted data from the twenty-seven studies is provided in Table D.2 in Appendix D.

Quality assessment

Each primary study was evaluated based on quality assessment criteria defined in Kitchenham's systematic literature reviews for software engineering [57]. The most relevant questions were taken from a set of 18 questions and applied to this review. These questions were:

- How credible are the findings?
- How well does the evaluation address its original aims and objectives?
- How well was the data collection carried out?
- How well can the route to any conclusions be seen?

- How adequately has the research process been documented?

A scoring system was developed to grade each of the studies. Each of the five questions has three different possible answers which are unique to each question. The possible answers are outlined below. The grading system was Yes = 1.0, Somewhat = 0.5, No = 0.0. The threshold for an accepted study was 3.0 The score for each study is shown in Table 2.1. The questions with the possible answers are as follows:

Question 1 How credible are the findings?

-Yes, the findings are very credible.

-Somewhat, the findings are partially credible.

-No, the findings are not credible.

Question 2 How well does the evaluation address its original aims and objectives?

-Yes, the evaluation addresses the original aims and objectives.

-Somewhat, the evaluation addresses the original aims and objectives implicit.

-No, the evaluation does not address the original aims and objectives.

Question 3 How well was the data collection carried out?

-Yes, the data collection was carried out well and outlined clearly.

-Somewhat, the data collection was carried out well but not outlined clearly.

-No, the data collection was not carried out well.

Question 4 How well can the route to any conclusions be seen?

-Yes, the route to the conclusion is seen.

-Somewhat, the route to the conclusion is implicit.

-No, the route to the conclusion can not be inferred.

Question 5 How adequately has the research process been documented?

-Yes, the research method is well documented.

-Somewhat, the research method is implicit.

-No, the research method can not be inferred.

Table 2.1: Quality assessment of studies used in the systematic literature review.

Reference	How credible are the findings?	How well does the evaluation address its original aims and objectives?	How well was the data collection carried out?	How well can the route to any conclusions be seen?	How adequately has the research process been documented?	Total
Baloglu et al. [4]	Yes	Somewhat	Yes	Somewhat	Somewhat	3.5
Chang [20]	Yes	Somewhat	Yes	Yes	Somewhat	4
Maurer [66]	Yes	Yes	Yes	Yes	Somewhat	4.5
Deloatch et al. [29]	Yes	Yes	Yes	Yes	Yes	5
Doyle et al. [33]	Somewhat	Somewhat	Yes	Somewhat	Somewhat	3
Connolly et al. [25]	Yes	Yes	Yes	Yes	Yes	5

Falkner et al. [35]	Yes	Yes	Yes	Yes	Yes	5
Kavakci et al. [55]	Yes	Yes	Yes	Yes	Yes	5
Macher et al. [65]	Yes	Yes	Yes	Yes	Somewhat	4.5
Chua et al. [22]	Yes	Yes	Yes	Yes	Somewhat	4.5
Scott et al. [97]	Yes	Yes	Yes	Somewhat	Yes	4.5
Todman et al. [107]	Yes	Yes	Yes	Yes	Yes	5
DeRaadt [85]	Yes	Yes	Yes	Yes	Yes	5
Fone [39]	Somewhat	Yes	Yes	Yes	Yes	4.5
Gerritsen et al. [40]	Yes	Yes	Yes	Yes	Yes	5
Guynes [44]	Yes	Yes	Yes	Yes	Yes	5
Hamer et al. [46]	Yes	Yes	Yes	Yes	Yes	5
Fenwick et al. [37]	Yes	Yes	Yes	Yes	Yes	5
Melin et al. [67]	Yes	Yes	Yes	Yes	Yes	5
Mills [68]	Yes	Yes	No	Yes	Yes	4
Ngai et al. [69]	Yes	Yes	Yes	Yes	Yes	5
Suraweera [105]	Yes	Yes	Yes	Yes	Yes	5
Owolabi et al. [78]	Somewhat	Yes	Yes	Somewhat	Somewhat	3.5
Vitasari et al. [109]	Yes	Yes	Yes	Yes	Yes	5
Blanchard et al. [93]	Yes	Yes	Yes	Yes	Yes	5
Deloatch et al. [30]	Yes	Yes	Somewhat	Yes	Yes	4.5
Dos Santos et al. [16]	Somewhat	Yes	Yes	Yes	Yes	5

2.1.3 RESULTS - BACKGROUND

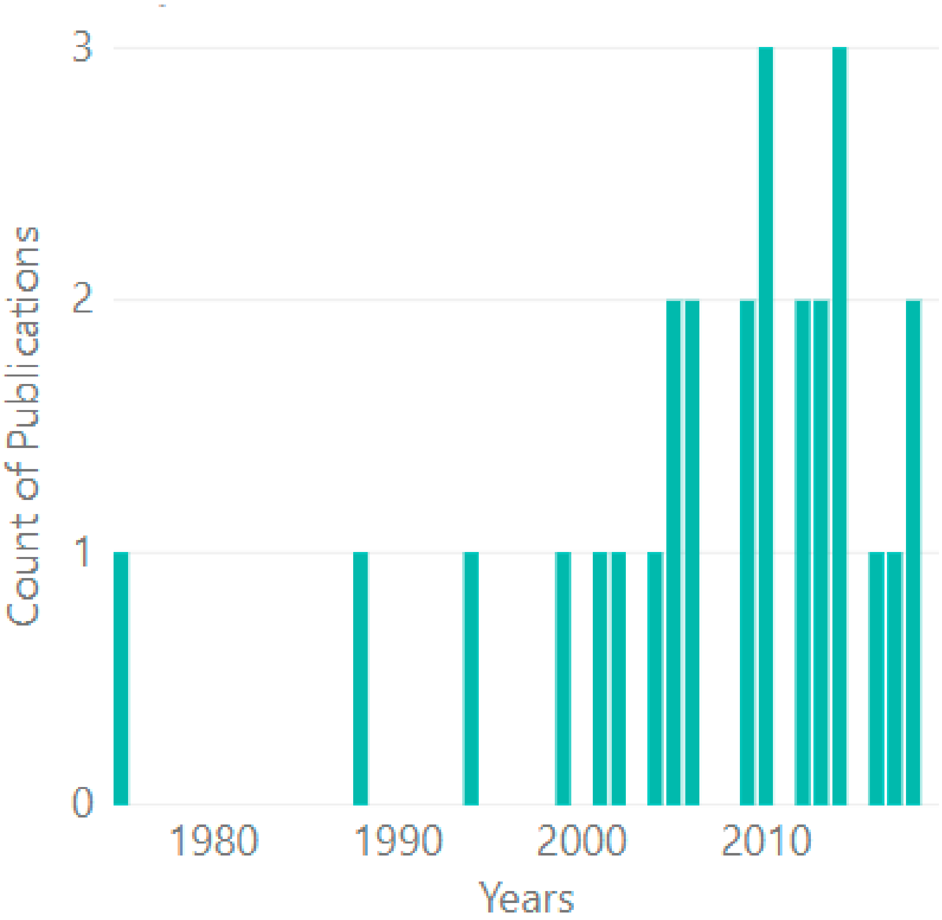
Types of studies

Of the 27 papers accepted, 82% of the studies were empirical studies and interviews. These studies were evidence-based studies where data was collected largely by questionnaires, however, some were experimental. Literature reviews on computer anxiety accounted for 19% of the accepted papers.

Temporal view of publications

The distribution of the primary studies throughout the years is shown in Figure 2.1. As can be seen, there is an increase in the number of publications after 2005, showing a growing interest in the research area.

Figure 2.1: Number of papers collected each year



Data sources

All studies chosen for this review were either published in conference proceedings or journals. Table D.2 shows the distribution of primary studies derived from their publication channels.

2.1.4 RESULTS

This section reports the findings from the literature review for each of the research questions outlined in Section 2.1.1. Although an extensive review was carried out, only 27 studies were found to address the research questions.

Although the number of studies is not large, they do make valuable contributions. The studies either have a large number of participants or are longitudinal in nature. They highlight interesting implications of anxiety in learning to program over time.

LR-RQ1: Is there a relationship between anxiety and stress and learning to program (language, syntax, compilation etc.)?

This research question was informed by nine studies: Connolly et al. [25], Guynes [44], Chang [20], Scott et al. [97], Melin et al. [67], Falkner et al. [35], Gerritsen et al.[40], Hamer et al.[46] and Ngai et al. [69].

Connolly et al. through a longitudinal study investigated anxiety when studying CS. Specifically, the study investigated the variance of anxiety amongst undergraduate computing students, with an emphasis on learning to program during their first year. This study was conducted over two years and 86 students participated. The study was set up in two parts, where students took a questionnaire at the start of their first year and then again at the end of the first year. From the study, two important factors were investigated: 1) computer self-efficacy and 2) state of anxiety.

Computer self-efficacy[1] was measured across 11 questions. In the pre-survey, 23% of students claimed they were *"unsure"* when asked if they would be able to learn a programming language. This was before they had any experience on the course. Such a finding is perhaps under-

1 Ones confidence in ones computing ability

standable in that programming is a new subject but a cause of concern given that people already have a negative perception about/related to their programming ability.

The cognitive, emotional and psychological states of anxiety[2] that students face in programming situations was also examined. Before the semester began, 44% of students reported that they did not feel relaxed when using a computer let alone programming. However, it was noted that the student's sense of worry and stress did diminish by the end of the year.

One of the key findings of the study is the student's perception of their ability to learn how to program. Connolly gathered feedback from the participants of the study and found evidence that novice programmers were computer-phobic; for example *"I'm afraid I'll wreck the program/hard drive"*. For novice programmers, receiving any sort of programming error can be a source of displeasure which could lead to stress and anxiety. Coupled with this, Connolly et al. also found that confidence and self-efficacy affect learning to program. It was found that the lower the confidence and self-efficacy, the harder it was for a student to complete a programming task correctly.

Guynes investigated the impact of system response time on State anxiety. Eighty-six participants took part in the experiment in which they had to edit a file containing 28 errors. In their analysis, Guynes reported that there was a statistically significant relationship between State anxiety and system response time (α=0.05,p=0.0155) [44].

Chang investigated if there was a relationship between anxiety and programming-task complexity and how this relates to programming skills [20]. The study consisted of 307 participants and measured perceived task complexity and self-reported anxiety levels using the Computer Attitude Scale. Results were based on three different levels of

2 States of anxiety was defined in this study as Worry, Happiness, Stress and Distractibility.

programming task complexity from easy to hard. Results showed that there was a significant relationship between perceived programming task complexity and anxiety levels, that is, as perceived programming task complexity increased, so to perceived anxiety levels.

Scott et al. hypothesised that students programming practice behaviour is negatively impacted by anxiety [97]. The original intention of Scott et al.'s study was to assemble and validate an instrument to assess self-belief in CS1. Two hundred and thirty-nine students participated in this study. Of note, they found in terms of programming anxiety students often worry when completing debugging tasks and they would start to feel nervous stressed when they try to find and fix programming bugs.

Melin et al. investigated how project orientated work affects learning [67]. The project orientated work was incorporated into the course. A total of 60 CS students participated in the course for over 15 weeks. The biggest worry for students was that their grade would be affected by other students who didn't do their share of the work. Students worries were alleviated by the introduction of a clear marking scheme. By the end of the course, students felt more confident about their programming skills.

Group work is becoming more popular in programming. With group work, students work is constantly being scrutinised by peers. While the student's work is not being formally assessed, the fact that their peers are assessing the work can cause anxiety. Falkner et al. investigated how collaborative activities may introduce stress and anxiety for students [37]. In their study, 10 students participated in an interview. The goal of the interview was to understand from the students perspective 1) the purpose of collaborative activities, 2) whether collaborative activities are perceived as positive or negative experiences and 3) how relationships between students within the groups worked. They concluded that students were stressed and anxious when working in groups. This is due to students not working as a group but rather as individuals when completing tasks.

Gerritsen et al. investigated the effects that pressure and stress can have on a learner [40]. In their study, they investigated physiological signals during high-stress activities. They had a total of 21 participants and found that during high-stress moments, the perception of the complexity of the task can define how hard a task is for that person.

Hamer et al. reported on a large scale study of 1500 students on the topic of peer assessment [46]. They reported that peer assessment is a source of anxiety to students as the mark received from different peers may be vastly different depending on the relationship to those peers.

Ngai et al. conducted a study which aimed to see if self-assessment helps to reduce student stress and anxiety. Thirteen participants took part in the study. The participants were asked to 1) assess their ability level and 2) self-grade their programming task. Results showed that with a clear assessment criteria students stress was reduced and anxiety was elevated [69].

From this review, it can be concluded that **there is weak evidence of a relationship between anxiety and stress and learning to program**. When learning how to program a multitude of factors can contribute to feelings of stress and anxiety in students such as 1) receiving errors, 2) task complexity, and, 3) collaborative learning. Receiving any sort of error for the program just written can be a source of displeasure which could lead to increased feelings of stress and anxiety. Not only is programming a source of stress and anxiety - system response time and the program task affects State anxiety.

LR-RQ2: Is there a relationship between mathematical anxiety and learning to program?

Of the 27 studies selected for this study, seven studies informed this research question. The studies are Owolabi et al. [78], Suraweera [105], Fone [39], Mills [68], Macher et al.[65], Vitasari et al.[108] and Baloglu et al.[4].

Owolabi et al. investigated the relationship between mathematical anxiety and programming anxiety by surveying students studying both computer science and mathematics [78]. They found a positive correlation between Mathematical anxiety and Computer anxiety ($r = 0.272$). While the correlation of $r = 0.272$ is not a strong relationship, this correlation indicated that there is a slight relationship and should be investigated further. In addition, they found a significant correlation between mathematical anxiety and computer programming achievement ($\alpha = 0.01$, $r = 0.450$) [78]. Similar to the correlation between mathematical anxiety and Computer Anxiety, the correlation of $r = 0.450$ indicates a positive but moderate relationship and the $\alpha = 0.01$ shows that the correlation is significant which would show that the chances of obtaining such a correlation by chance are less than five times out of 100.

Suraweera investigated the concept of Discrete Mathematics being taught by the Mathematics department in his institution. He noted that students were not understanding the material and subsequently could not apply the material in CS leading to feelings of stress. He designed a framework to enhance the teaching and learning of Discrete Mathematics. This meant that Discrete Mathematics was now being taught by the Computer Science department. After putting this into practice, students reported feeling more confident in their ability and less stressed and anxious [105].

Fone argued for reducing mathematical overheads (proofs and hard to follow methods) to reduce unnecessary mathematical and programming anxiety [39]. The concept of neural networks is one that is rooted in Mathematics. Fone used Microsoft Excel to demonstrate to a class of 21 students the operation of neural networks. Following this single demonstration, the student's ability to program a neural network improved and reduced their reported programming and mathematical anxiety [39].

Macher et al. were interested in how self-efficacy and different learning strategies can influence mathematical learning [65]. As part of this study, 147 students participated (112 females, 35 males). Questionnaires on mathematics and trait anxiety, deep-level strategies, self-concept and interest in mathematics were administered. An interesting finding was that students with higher levels of Trait anxiety appeared to experienced higher levels of mathematical anxiety ($r = 0.541$). Along with mathematical anxiety, it was found that mathematical self-concept and an interest in mathematics are both negatively related to mathematical anxiety ($r = -0.246, r = -0.403$ respectively at $\alpha < 0.01$)[65].

Mills noted that students were writing programs and not following a particular algorithm. This led to students not knowing if the program that they wrote would compile. In this study, Mills outlined a mathematical technique that demonstrates how to know you have written a program semantically correct [68]. He discusses how if the students follow a systematic approach to writing a program it can aid in reducing feelings of anxiety.

Vitasari et al. investigated the role that mathematical anxiety has on academic success. Vitasari et al. conducted a study with 770 students [108]. The study aimed to investigate the psychological barriers that students encounter when they are performing a mathematics task. They found that Mathematics is perceived as a difficult subject (t=72.414, p=0.000). Baloglu et al. conducted a study on 759 third level students to investigate the differences which exist in Mathematics anxiety. They asked the students to do some basic Mathematics questions such as multiplication and division. Following the questions, the students completed a survey to capture Mathematical anxiety and stress. Results from the survey show that basic mathematics questions still induce mathematical anxiety and stress[4].

From this review, it can be concluded that **there is a negative relationship between mathematical anxiety and learning to program**. Many of the concepts that are taught in CS have a basis in Mathematics. When students receive assignments that have a strong basis in Mathematics, they can find it hard to draw the link between what is being asked and how to complete the assignment.

Having a certain level of Mathematics in CS is necessary. As part of introductory CS courses, Discrete Mathematics is often taught as a part of the course, something that is generally delivered by a Mathematics Department. This practice has led to unsatisfactory results. These poor results cause students to become more anxious in their ability. This could be because there is sometimes a disconnect between theoretical and applied applications. In CS, there is a tendency to use applied mathematics and so if the CS department was to teach the Discrete Mathematics course it could be tailored to CS and inherently more applied.

One finding, while only a single non repeated study that appears to stand out was the experiment conducted by Fone [39]. The method of demonstrating a concept in a simpler/ more familiar environment appeared to both increase understanding and reduce anxiety. Perhaps if this this practice is generalised and adapted of other concepts, this can be a helpful teaching tool.

LR-RQ3: Does computer usage cause stress and anxiety when learning to program?

Five papers were found that were relevant to this question namely Doyle et al. [33], Chua et al. [22], Maurer [66], Todman et al. [107] and Dos Santos et al [93]. Many topics taught in CS involve the use of a computer and even the basic interaction with a computer, using a virtual learning environment, for example, may be enough to make a student anxious or stressed[93]. Computer anxiety is defined as the "negative emotions and cognitions evoked in actual or imaginary interaction with computer-based technology" [22].

A study by Doyle et al. investigated computer anxiety felt by CS students. In particular, they focussed on how computer anxiety is directly related to self-efficacy and prior computer experience [33]. Computer experience can include any computer courses previously completed, computer training, computer gaming experience etc [33]. The study involved 163 participants (32 female, 131 male) across 4 different years in University. Students were asked to fill out a questionnaire and interestingly a strong inter-dependence between computer anxiety, self-efficacy and computer experience was found [33]. Also, they found that final year CS students are still anxious ands stressed when it comes to completing a computer task.

Chua et al. conducted a review of 10 studies that report on potential correlates of computer anxiety (gender, age, computer experience, locus of control, cognitive appraisal, math anxiety, communication apprehension, computer course structure, and learning styles) [22]. The review reports on the relationship between computer anxiety, age, gender and computer experience. They found that correlates such as computer anxiety and computer experience are inversely related.

Maurer conducted a literature review on computer anxiety and its correlates [66]. The review consists of 38 studies. In the review, Maurer discusses different correlates such as experience, gender, age, academic major, etc. Maurer reports that computer experience is a correlate of computer anxiety but still requires further research.

Dos Santos et al. also conducted a systematic literature review on Computer Anxiety and interaction [93]. They examined all the papers that used the Computer Anxiety Rating Scale. The review consisted of 111 studies. Findings from the review would suggest that poor User Interface design can contribute to a persons Computer Anxiety.

In the above reviews, gender was investigated as a correlate. While gender is not considered a strong correlate [22, 66], Todman et al. have suggested that perhaps biological gender is not a variable in computer anxiety but psychological gender is. When examining psychological gender, each person would identify with one category: masculine, feminine, androgynous or indifferent. A study with 138 CS students was conducted by Todman et al. and it was found that students who have a more feminine identity experience a greater sense of computer anxiety. Given the conflicting information and inconclusive nature about gender being a correlate, it is hard to say that gender is a factor of computer anxiety and this is an area that should be further researched [22, 66, 107].

From this review, it can be concluded that **computer usage can cause anxiety when learning to program**. The relationship identified between computer experience and anxiety appears to be the strongest. Results from studies show that computer anxiety can be reduced through computer experience in a CS course but it depends on the type of experience which the user is exposed to during a course. For example, if the student is computer facing for the entire course, their experience would be greater than the experience gained by a student who is computer faced for part of the course. Intuitively one might expect that the more experience you have studying a CS course the less anxious you should be. However, final year CS students are still anxious when it comes to completing a computer task.

LR-RQ4: Does test anxiety affect learning to program and more broadly Computer Science students?

Five studies were found to inform this research question: Deloatch et al. [29, 30], DeRaadt [85] Kavakci et al. [55] and Blanchard et al.[16].

Test anxiety is an unpleasant state associated with the feeling of tension and apprehension, worrisome thoughts and the activation of the autonomic nervous system when an individual faces evaluative achievement demanding situations [55].

Deloatch et al. investigated how exam modality relates to students perceptions of test anxiety and performance during programming exams [29]. A survey was administered to measure student perception of test anxiety of paper-based exams and online exams. Three hundred and ninety-one students participated in this survey. After analysing the results, 22% of students (n=61, $\bar{x}$=4.26, SD=1.51) perceived high test anxiety for paper-based exams while 23% of students (n=64, $\bar{x}$=4.15, SD=1.67) experience high test anxiety for online exams.

De Raadt proposed a method of allowing students to create cheat sheets for exams. Eighty-nine students took part in the experiment. While exam marks did improve marginally, each student that created a cheat sheet reported that their levels of test anxiety reduced before and during the exam [85].

Fenwick et al. trialled a novel method of exam revision which consists of a 24-second technical description of a concept relating to the course and a clear summary that anyone without a technical background could understand in 7 seconds called a 24/7 lecture [37]. A total of 100 participants took part in the experiment. For the student to succeed in creating the 24/7 lecture the student had to completely understand the topic. Responses from the students showed that not only were they more confident in the concept they discussed, but they felt less anxious about the exam[37].

Kavakci et al. investigated the variables that are related to students planning to take University entrance exams [55]. The aim was to identify the predictors of test anxiety. A total of 436 students participated in the study. They found that 48% of students experienced test anxiety.

In a separate study by Deloatch et al. investigating the effects of supportive comments on social media before an exam on Test anxiety [30]. In the study, 1,235 students took part in this mass scale study and were in different years of study. 94% of the students were enrolled in a Data Structures course. Before sitting a programming task, the students took the trait form of the STAI before and after asking for supportive comments on their social media accounts. Findings suggest that supportive comments reduce test anxiety. Also, Blanchard et al. investigated the effect of auto-generated social media encouragement on test anxiety [16]. In a study consisting of 27 students, similar to Deloatch et al, findings suggest that test anxiety is reduced. However, when students knew if the message was auto-generated, the levels of anxiety did not reduce to the same degree as a message from a real person.

From this review, it can be concluded that **test anxiety does affect learning to programming**. While there was no mention of programming in general, it can be argued that all CS courses involve a degree of programming and so the exams associated with the courses will have a degree of programming. In recent years, computer-based exams have become more prevalent. The impact of the modality used to assess students in programming is currently unknown and future research is required,however, it seems that online assessment reduced students anxiety and stress. However, what is known is that supportive messages from peers before an exam.

2.1.5 DISCUSSION

The review found evidence on the relationship between programming anxiety and programming-task complexity and how this relates to programming skills. A significant relationship between perceived task complexity and self-perceived anxiety levels has been identified. Also, a longitudinal study which investigated the variance of anxiety amongst undergraduate computing students was described. It was found that students have low levels of self-belief when conducting programming tasks. This is compounded by evidence that students are leaving university as anxious programmers and going into industry lacking confidence in their ability [9].

While it is known that programming is difficult, with the introduction of group work, students appear to be anxious about their work being examined by their peers [87].

Mathematical anxiety was also examined due to the close relationship between programming and Mathematics. Students are anxious about the teaching methods used [78]. In addition to the teaching methods, self-efficacy was identified as a key factor when learning Mathematics. Consequently, by attempting to increase a students self-efficacy this would, in turn, reduce Mathematical anxiety. In addition, the learning strategies employed by students influences levels of mathematical anxiety.

Given the high availability of technological devices (Smartphones, smartwatches, portable devices, computers and laptops etc.) in today's climate, students should have high exposure to these devices in an educational setting. One would assume that students who choose CS as a degree choice would not fear to interact with a computer. However, even after a four-year degree, students still feel anxious when working on a computer.

Testing and assessment can induce anxiety. While tests and assessment are different, the anxiety that is experienced is categorised as test anxiety. Anxiety in assessment is inevitable, however, educators are now beginning to change the modality of how programming is being assessed. One study has observed that online assessment marginally reduces anxiety in students when programming is being assessed however the differences between paper-based assessment and online assessment are still unclear [97]. Anxiety in testing, regardless of discipline, is also inevitable. One method that can reduce students text anxiety is requesting messages of support from their peers on their social media accounts [16, 30].

The findings here can inform the teaching and learning of programming and help us to be mindful of the role of anxiety and its implications in learning [100].

2.2 CONCLUSION

This review makes several valuable contributions. Firstly, the meta-review provides insight and promotes awareness of the anxiety of our learners. This can be used to improve our teaching and learning methods and assessment decisions. Secondly, given the concerns for the mental health of our students, this review has identified how little focus has been given to such an important area.

There is a need for more research to be carried out in this area. In particular, two types of studies would be very valuable: 1) further empirical studies (using qualitative and quantitative methods) that build upon the findings identified in this review. This would allow for the re-validation of all previous findings. 2) There is a need for more experimental studies that involve real-time physiological measurement of anxiety using sensor technology. For example, measuring heart rate or electrodermal activity during ecologically valid programming tasks. Doing this during the completion of a task may give us an understanding of the specific source(s) of anxiety or stress. This is particularly important now as wearable technologies have become widely available and are potentially an untapped resource in teaching and learning.

2.3 THESIS UPDATE

To provide a clear picture of this project, the objectives of this thesis are laid out below:

- To examine the relationship between anxiety and CS1 programming performance.
- To examine the relationship between stress and CS1 programming performance.
- To examine the relationship between anxiety and stress.
- To review the data obtained throughout the project, to identify analyse and identify any gender differences.

This chapter set out to conduct to conduct a review of the literature surrounding Computer Science and mental health. The review was systematic in nature which allowed for a robust and reproducible review. It was shown that CS students suffer from stress and anxiety.

Figure 2.2 shows the updated thesis flow and how the systematic literature review implicates the next sequence of work. The following chapter, Chapter 3, will discuss the study instruments and materials.

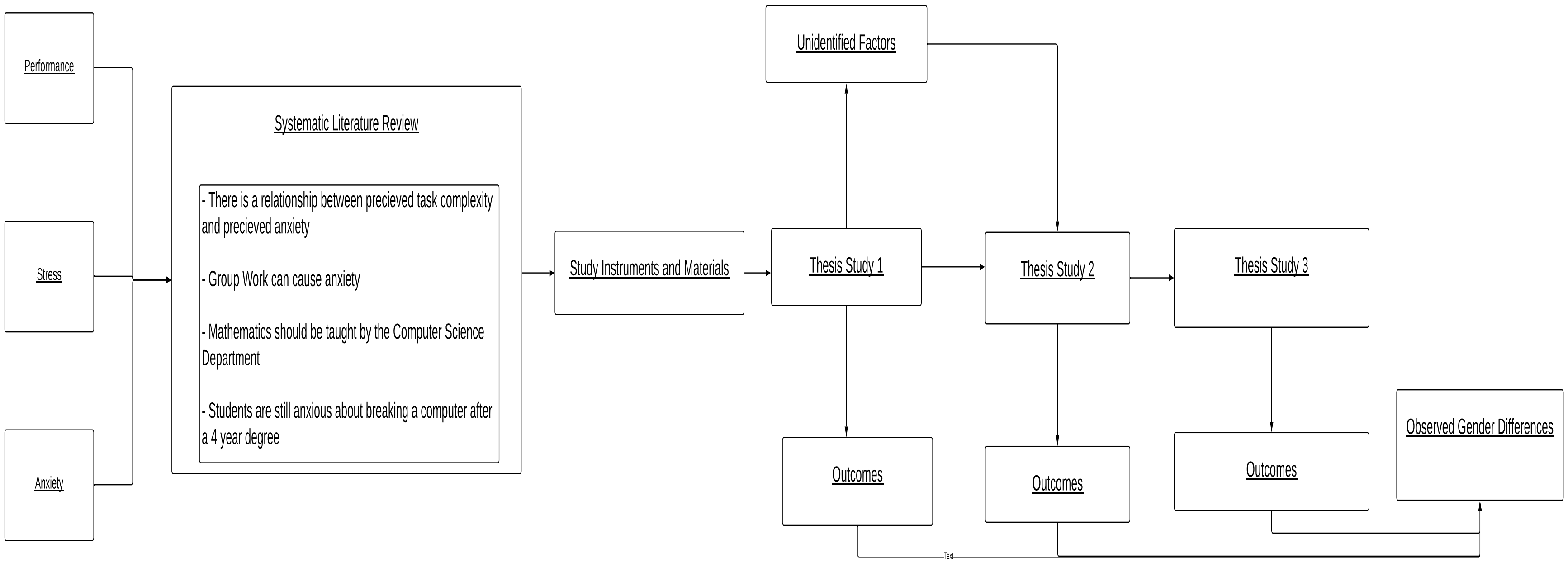


Figure 2.2: Update of the research methodology. Outcomes of the related literature will inform both Chapter 3 and the studies performed in Chapters 4,5 and 6.

3

STUDY BACKGROUND AND INSTRUMENTS

This chapter describes the different instruments that were used and/or developed for use in this research project. The chapter begins by describing the participants in this project. Following this, methods to determine anxiety are described. Concluding the chapter, the development and validation of an MCQ test for use in Chapters 4 and 5 is discussed.

3.1 PARTICIPANTS

Participants who took part in this thesis were taking (Study 1 and Study 2) or had just completed (Study 3) the CS1 module at Maynooth University. This cohort of participants was chosen given they were of focus in the PreSS and PreSS# studies. CS1 is the introduction to computer programming module. Students in this module, in general, have no previous formal study of CS or experience of programming. The module runs over twelve weeks and consists of three hours of lectures, three hours of labs and six hours of independent study per week. The module covers programming fundamentals in the Java programming language, typically delivered in the following order:

- Variables.
- Types.
- Expressions and Assignment.
- Simple I/O.

- Conditional and iterative control structures (if statements and while loops).
- Strings and string processing.
- Arrays.
- Other fundamentals such as problem-solving and computer architecture.

This module structure is similar to the proposed Java Programming 1 course outlined in the ACM Curriculum Guidelines for Undergraduate Programs in Computer Science [26].

3.2 ANXIETY

A significant part of this project was to review the anxiety of CS students. To do this, multiple methods of capturing anxiety were investigated. These methods included the use of psychometric methods, physiological methods and self-assessed surveys. This section will discuss the various methods, the final choices and the justifications for these choices.

3.2.1 METHODS

Psychometric methods – Frontal Alpha Asymmetry

Frontal asymmetry is the average differences between brain activity in the frontal areas of the brain. Frontal Alpha Asymmetry is an area that has been studied extensively as part of research on emotional and motivational processes, specifically, right and left sides brain differences in alpha power. Frontal Alpha Asymmetry was initially detected by

Davidson et al. and validated by Hagemann et al. when investigating different biomarkers of personality [28, 45]. They discovered that people with increased left-frontal alpha power were found to process information positively compared to people processing the information on the right-hand side of the brain where a more negative processing mode was observed [28, 45].

Feelings of withdrawal have been linked to right frontal EEG activity when the person is resting and also in the face of new emotionally threatening situation [24, 48]. This bias is evident in healthy children and adults [23], individuals with increased temperament, given their high negative emotional state or individuals with anxiety and depression [36], and individuals with a current or past history of mood disorder [2]. In contrast, greater left frontal EEG activity has been linked to approach tendencies, involving both positive emotions, such as joy [34], and negative emotions, such as anger [47]. Paradigms such as the Emotional Stroop Paradigm and the Dot-Probe Paradigm were chosen as experimental paradigms as they measure attention to threat which has a strong relationship to anxiety. As part of an International Collaboration investigated the viability of using these experimental paradigms with a CS student population to measure anxiety was undertaken. The results were disappointing and psychometric tests were ruled out as a viable method of collecting participant anxiety. The collaboration is outlined detailed in Appendix F.

Surveys – State Trait Anxiety Inventory

The STAI has been the survey of choice for many clinicians to aid in the diagnosis of anxiety. First introduced by Spielberger, this scale has been used to measure self-reported State anxiety and Trait anxiety in both high-school and college students since the late 1960s. As of 2014, over 14,000 studies have been published citing the use of the STAI [102]. The

STAI is routinely used as a clinical survey in diagnosing anxiety and is arguably the most commonly used tool in the evaluation of anxiety, with 12 language versions available [101]. The survey contains 40 questions, 20 relating to State anxiety and 20 relating to Trait anxiety. The survey is graded on a 4-point Likert scale with values from 1–4. Given that the STAI is graded on a 4-point scale, the lowest possible score is 20 (obtained by supplying the value of 1 for all 20 questions) and the highest possible score is 80 (obtained by supplying the value of 4 for all 20 questions). The lower that somebody scores, the less their level of anxiety and conversely, the higher the overall score the greater the level of anxiety. In this research, the STAI was used to gather the self-assessed anxiety level of a participant before the commencements of each experiment. Due to the STAI being under copyright, the full questionnaire can not be disclosed, however, some of the questions within the STAI are outlined in Table 3.1 along with the possible responses to the questions.

At this point, it should be noted that the STAI has been normalised with a population in the United States of America whereas this work is focused on Irish University students. The normalisation population consisted of 855 college students enrolled in introductory psychology courses at the University of South Florida. Although no date range is specified given that the students are studying introductory psychology it is reasonable to assume that a large proportion are school leavers (18 to 22) with a small proportion of mature students (23+) comparable to the studies in this project. The gender breakdown is 324 Males and 531 Females. The alpha reliability's of the normalised results were calculated and were reported at a $\alpha = 0.91$ and $\alpha = 0.90$ indicating extremely strong reliability. Although the age range is likely similar, culture and background may not be, and as such, comparisons between our findings and that of the instruments normalised population need to be interpreted accordingly. In addition, other surveys were considered. These surveys were

- Computer Anxiety Rating Scale [50]:

  The Computer Anxiety Rating Scale was developed in 1987 with the aim of reliably measuring anxiety levels when interacting with a computer. The scale related to math anxiety, test anxiety and State anxiety. Given that the scale focused on multiple types of anxiety, it would be hard to determine the cause of the anxiety.

- Short Computer Anxiety Scale [63]:

  The Short Computer Anxiety Scale was developed to reduce the number of questions usually asked to determine Computer anxiety. The 6-item scale focused on confidence using computers rather than anxiety.

- Computer Programming Anxiety Questionnaire [25]:

  The Computer Programming Anxiety Questionnaire was developed capture the levels of programming anxiety in Irish students. The questionnaire consisted of questions relating to student demographics, goal orientation, experience in gaining computing skills, sense of control, computer self-concept, and state of anxiety in computing situations. The questionnaire has not been validated across different institutions and other studies.

While each of these surveys has their own merits, however, none of the surveys can distinguish between State anxiety and Trait anxiety and so it would be difficult to understand if the student is anxious due to the task at hand or if they are more prone to have an anxious disposition.

Table 3.1: Samples of both the State and Trait forms with the possible answers.

State Form	Not At All	Somewhat	Moderately So	Very Much So
I feel calm	1	2	3	4
I feel upset	1	2	3	4
I feel nervous	1	2	3	4
Trait Form	Not At All	Somewhat	Moderately So	Very Much So
I feel satisfied with myself	1	2	3	4
I wish I could be as happy as others seem to be	1	2	3	4
I have disturbing thoughts	1	2	3	4

3.3 SENSORS

The use of sensor technology in the project allowed for the capture of stress through the use of physiological signals while completing CS1 activities. It is a known fact that EDA and PPG are indicators of Stress [18]. In today's market, there are multiple options for complete sensor technology suites. One such suite is the Biopac system in which medical-grade technology is employed. The suite of sensors has been extensively used and tested in a range of different disciplines. The only drawback to the Biopac system is it is a fully wired system and can be very intrusive as sensors would have to be attached to a participants chest. For this project, an as-close-to authentic system is wanted. This means the system has to be non-intrusive and potentially wireless.

The Shimmer 3 GSR+ was used to gather physiological signals during all thesis experiments. The Shimmer 3 GSR+ contains both a Photoplethysmogram sensor and a Galvanic Skin Resistance sensor. The Shimmer is wireless and allows the wearer a somewhat unrestricted range of motion. The sensor technology is described in detail in Section 3.3.1 and Section 3.3.2.

3.3.1 Electrodermal Activity

Electrodermal Activity (EDA) is one of the most commonly used measures for a physiological response, with studies focusing on a variety of tasks from measuring attention to predicting abnormal behaviours such as lying[17]. Electrodermal Activity, otherwise known as Galvanic Skin Response (GSR), is the measure of the electrical current that the skin conducts between two points. Activation of the sympathetic nervous system (SNS) is very common with a magnitude of situations capable of inducing a large SNS activation. These activation's are of interest as they are indicators of arousal. The SNS is one of three primary divisions of the autonomic nervous system with the others being the parasympathetic nervous system and the enteric nervous system.

The SNS's primary function is to control the body's fight or flight response, however, it is constantly maintaining homeostasis[1]. The SNS controls how much a person sweats and depending on the situation presented, the person may sweat more or less. This situation could be

1 The ability or tendency to maintain internal stability in an organism to compensate for environmental changes.

classified as either excitement or stress. There are two main types of sweat glands in the human body:

1. **eccrine glands** which are the major sweat glands of the human body.
2. **apocrine glands** which are scent glands, and their secretions usually have an odour.

The eccrine glands are mainly involved in emotional responses (excitement or stress) and therefore for this research are the glands that are of most interested in recording. The EDA signal is composed of two main components: a slowly varying baseline level, known as skin conductance level (SCL) and a skin conductance responses (SCRs), which include reactions to specific eliciting stimuli. SCRs are of interest in emotional quantification, as they provide a measure of the level of arousal and engagement of an individual, in response to stimuli in their environment. This response will either be characterised as excitement or stress, depending on the individual's appraisal of the situation. Individual SCRs are often characterised by metrics such as their latency, amplitude, rise time and recovery time, and the interpretation of these metrics to provide insights into the emotional state is a very active area of research [3].

As EDA is the conductance of the skin, two electrodes are attached to the tip of the fingers to measure conductance between the points of contact as shown in Figure 3.1. The tips of the fingers are chosen as there is a high concentration of sweat glands. While Figure 3.1 shows the sensors on the first two fingers, this can be altered to any finger if needed. A small current is applied to the electrodes and the conductance is measured between them. EDA is used in this study to determine when a student becomes aroused, that is when a student begins to react to a situation and a stress response (heart or sweat rate increases) occurs, and to determine if that arousal is constructive or destructive.

Figure 3.1: The EDA electrodes are placed on the tips of the fingers as there is a high concentration of sweat gland on the top of the fingers.

3.3.2 PHOTOPLETHYSMOGRAPHY

A photoplethysmogram (PPG) is a sensor that detects changes in the volume of blood flow by measuring the difference in the light reflected into the sensor. While a PPG can be used to measure different values, it is often used to obtain a pulse measurement. The PPG uses a pulse oximeter, which uses a light source to illuminate the skin while a second photo-sensor measures the changes in light absorption. Figure 3.2 shows the structure of the PPG, one part of the sensor illuminates the skin through the use of a Light Emitting Diode (LED) and a light sensor captures the level of light that is reflected. The difference between the output of the light sensor is used to determine the captured PPG values. This value will change every time the heart beats and pumps blood to the periphery, and the outcome of this is observed in Figure 3.3 and the experiments described in this thesis.

Every time the heart beats, a different amount of light is absorbed by the blood meaning the amount of light reflected on to the sensor also changes. This pattern of changing values will create a waveform

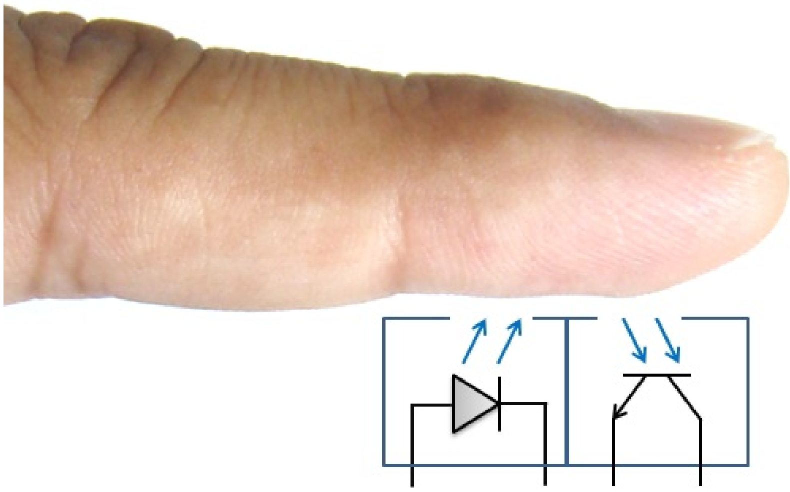


Figure 3.2: A light-emitting diode illuminates the skin while a photo-sensor records the reflected light.

that mimics the beats of the heart. Figure 3.3 shows four beats of the heart, which are represented by the blue line. The red vertical lines indicate two types of activity within the heart, namely systole and diastole. Systole is the event of the lower chambers of the heart pushing blood through the body. This is indicated by a steep increase in PPG values. This is a fast reaction. A diastole event is slower and longer-lasting. The event occurs when the top two chambers of the heart fill with blood and so the pressure in the arteries is lowered. The dicrotic notch in Figure 3.3 shows the beginning of the diastole event. The time between the first two red lines in Figure 3.3 indicate this diastole event. The time between the second and third red lines indicate the systole event and is much faster.

As well as having valuable data from the PPG for plotting heartbeats, other valuable information is also available. The heart rate can be estimated by capturing the number of peaks (equivalent to the number of heart beats) and running a sliding window of 60 seconds over the data while counting the number of beats. Taking the average of the sliding windows, the average Heart Rate is calculated. A deeper analysis can

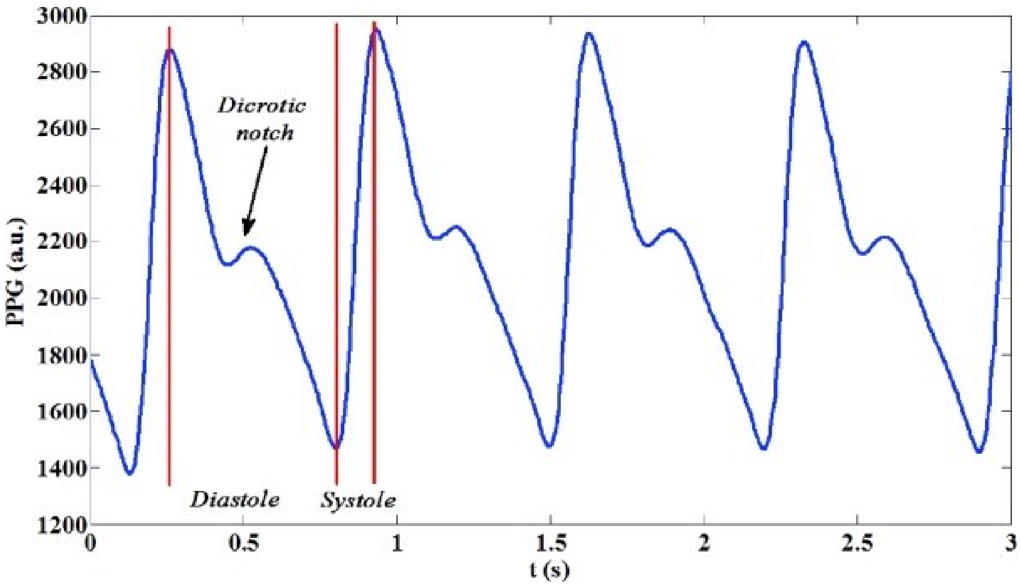


Figure 3.3: This is a PPG waveform that outlines the activity of the heart in blue. The time between the beats can be calculated by taking the time of each peak and subtracting the previous one from it.

be carried out on the raw PPG data. Heart Rate Variability (HRV), the variation in the time interval between heartbeats, is a known indicator of the interplay between the sympathetic nervous system[2] and parasympathetic[3]. This interplay can provide an insight into the fight-flight[4] response in the body which can be an indicator of stress. If there is more variability in the beat to beat data, the person could have perceived a threat, and so the systems that control the beating of the heart begin to fight for control. Using this knowledge, HRV measures can be used as

2 The sympathetic nervous system's primary process is to stimulate the body's fight-or-flight response. It is, however, constantly active at a basic level to maintain homeostasis

3 The parasympathetic system conserves energy as it slows the heart rate, increases intestinal and gland activity, and relaxes sphincter muscles in the gastrointestinal tract.

4 The fight-or-flight response is a physiological reaction that occurs in response to a perceived harmful event, attack, or threat to survival. In a life-threatening situation, it will cause a person to either run from a situation of fight it.

an indicator of stress [60]. Various algorithms can be used to calculate HRV and these will be discussed in Section 4.4.1.2.

3.4 PROGRAMMING QUESTIONNAIRE

As this project was focused on a relationship between anxiety, stress and performance, a robust, validated method to assess CS1 knowledge was required. To do this an instrument to test participants knowledge and allow for an investigation of performance and stress at an individual concept level was created.

3.4.1 DEVELOPMENT

The programming comprehension exam was designed in-house and each question was subject to the following constraints:

- Multiple choice in nature.
  - There were four possible answers.
  - There was only one correct answer.
  - One "None of these" answer.
- Always had a clear output i.e. there was no hidden challenges or tricks in the question.

To allow for fine-grained analysis each question contained only one new concept, for example, a Loop, a Conditional Statement or a String. This allowed the responses to be analysed both individually and collectively so that the most likely concept causing difficulty could be identified. In addition to this, as physiological data was being collected throughout the experiment, if there were any changes detected in the physiological signals, an attempt could be made to correlate them to a specific concept.

3.4.2 VALIDATION

Thirteen questions were developed in total, and initially, the questions were categorised into their difficulty level as either easy, medium, or hard. To further ensure that the questions developed were of good and sound quality seven postgraduate research students (Research Masters/PhD candidates) in the Computer Science department in Maynooth University were recruited to:

1. Review all questions to get a sense of the range of concepts being asked.
2. Answer each question to ensure that the correct answer was identifiable.
3. Rate each question on a scale of 1 (easy) to 9 (hard) in terms of difficulty.

Results from the question reviews showed that for all of the initial 13 questions, all the postgraduate students' ratings were similar across all questions. The difficulty scores given to each of the questions by the reviewers were averaged. Following this, the questions were ranked in order of difficulty based on these scores. Of the thirteen questions, one question was removed from the potential pool of questions as one postgraduate student got the question wrong and so it may have been too hard for novice learners. Two questions were removed as they were "too long" compared to the other questions and so would not fit on the presentation screen. One question that had been labelled by the authors as "easy" was labelled as "medium" by several of the postgraduate reviewers. This question was also removed from the potential pool of questions as there was a difference in opinion between the authors and reviewers. This resulted in nine peer-validated questions for the experiment (see Section 3.4.3 for samples and see Appendix E for full question set).

Each question builds on the previous question with the first being the easiest and the ninth being the hardest, thus, the first three questions (Q1, Q2, Q3) were categorised as easy, the middle three questions (Q4, Q5, Q6) were categorised as medium, and the last three questions (Q7, Q8, Q9) were categorised as hard. Table 3.2 shows the nine core concepts examined in the experiment. The table also outlines the new concept, depicted by the use of "N" for that particular question and what other concepts were contained in each question (by the use of "x"). These core concepts were chosen as they were the concepts that the participants would have been exposed to in the CS1 course.

Table 3.2: Core concepts shown in the experiment with N representing a new concept.

Concepts	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
System output	N	x	x	x	x	x	x	x	x
String variables		N	x						x
String Concatenation			N						
If-else if-else statements				N					
Nested if-else statements					N				
While Loop						N	x		
Substring							N	x	
If-else with Substring								N	
Nested While Loop									N

As this experiment took place late in the CS1 course, participants had more exposure to certain concepts that were covered earlier in the module. For example, the participants would be very familiar with system output as they would use this in (virtually) every program that they write. Comparing this to concepts such as substring or nested loops, participants would have only been introduced to them towards the end of CS1. Following the validation of the questions, the expectation was that all participants should get the easy questions correct, most would get the medium questions correct and only some would get the hard questions correct.

3.4.3 SAMPLE QUESTIONS

The following questions are examples from the pool of accepted questions. Question 1 is an easy question, Question 4 is a medium question and Question 7 is a hard question. The answer to the question is highlighted beside the correct choice. The full set of questions are contained in Appendix E.

```
// Question1 Difficulty: Easy
public class q{
   public static void main(String [] args){
      System.out.println("Hello World!");
   }
}
// A : hello world!
// B : Hello world!
// C : Hello World! *Correct*
// D : None of These
```

```
// Question4 Difficulty : Medium
public class q{
   public static void main(String [] args){
      int x = 5;
      if(x<5){
```

```
            System.out.println("<");
        }
        else if(x>5){
            System.out.println(">");
        }
        else{
            System.out.println("=");
        }
    }
}
// A : <
// B : >
// C : = *Correct*
// D : None of These
```

```
// Question7 Difficulty : Hard
public class q{
    public static void main(String [] args){
        int count = 1;
        String x = new String("Hello World!");
        String newX = new String("");
        while (count< 12){
            newX = newX + x.substring(count, count+1);
            count+=2;
        }
        System.out.println(newX);
    }
}
// A : el ol! *Correct*
// B : Hello World!
// C : HloWrd
// D : None of These
```

3.5 ADDITIONAL SURVEYS

Two additional surveys were used to collect information from participants taking part in the studies. Bergin's Programming Self-Efficacy questionnaire was used to determine the participant's self-belief in their programming ability. A background survey was also used to collect information such as gender, age, competency of English, dominant hand and eyesight levels (see Appendix I).

3.5.1 PROGRAMMING SELF-EFFICACY QUESTIONNAIRE

As part of this project, the programming self-efficacy questionnaire was included. Created and validated by Bergin et al. [11] and re-validated multiple times by Quille et al. [84] the questionnaire is becoming widely accepted. The questionnaire was derived from the Rosenberg Self - Esteem questionnaire [89] and was adapted by Bergin to apply to programming. The programming self-efficacy questionnaire consists of 10 questions and has been shown to have a high inter-item and test-retest reliability [11, 13]. The ten questions on the survey are outlined in Table 3.3 along with the possible answers.

3.6 SUMMARY

The surveys used throughout this project were well cited with high test and re-test metrics. In terms of their effectiveness, they are extremely simple to administer and calculate. The STAI questionnaire, a gold standard questionnaire was used to capture the student's self-reported State anxiety and Trait anxiety. Bergin's Programming Self-Efficacy questionnaire, now widely used, was used to capture the programming self-efficacy of the students.

Table 3.3: Programming Self-efficacy questionnaire with possible answers.

	Not At All True Of Me	Somewhat True of Me	True Of Me	Very True Of Me
On a whole I am satisfied with my programming progress	1	2	3	4
At times I think that I am no good at all at programming	1	2	3	4
I feel that I have a number of good programming qualities	1	2	3	4
I am able to complete programming tasks as well as most other students in my class	1	2	3	4
I feel that I do not have much programming ability to be proud of	1	2	3	4
I certainly feel useless at programming at times	1	2	3	4
I feel that I am a person of worth, at least on a plane with other programmers in my class	1	2	3	4
I wish I could have more respect for my programming ability	1	2	3	4
All in all, I am inclined to feel that I am a failure at programming	1	2	3	4
I take a positive attitude towards my programming ability	1	2	3	4

The sensor technology that was used in this project is of a high standard. The Shimmer 3 GSR+ allows the capture of both PPG and EDA simultaneously. Also, the Shimmer 3 GSR+ is wireless which allows the participants more freedom while wearing the sensors, adding a degree of validity to the project.

Developing and validating a programming multiple-choice questionnaire which was concise and yet examined the range of concepts in CS1 was a challenge. The method used when creating the questions allowed us to identify possible problem concepts which in turn can help to inform educators on where to spend time when teaching. Figure 3.4 shows how each of the surveys and instruments interacts with the objectives outlined in Chapter 1.

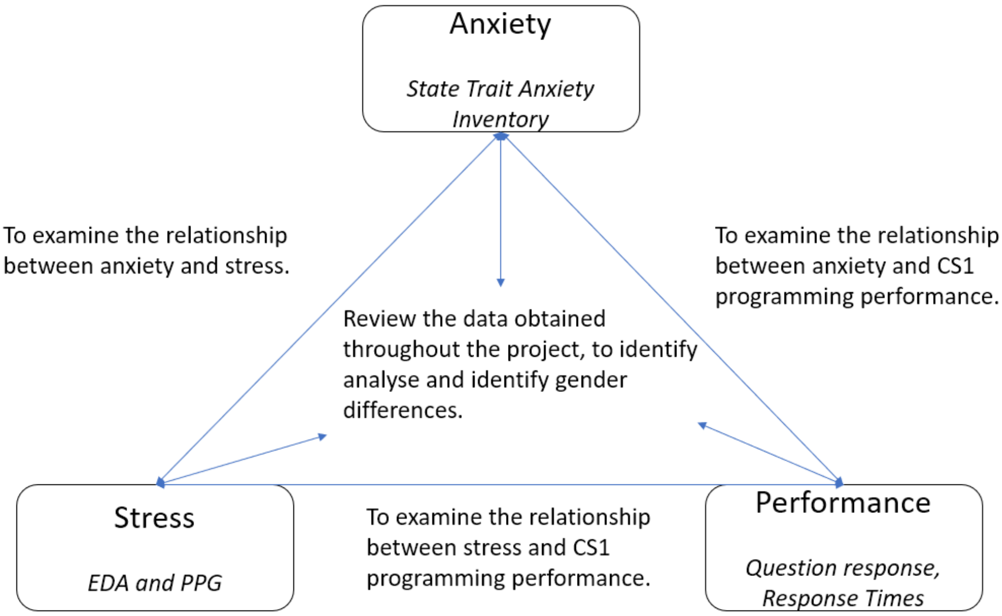


Figure 3.4: How instrument and Materials Interact.

3.7 THESIS UPDATE

As noted in Chapter 1, the objectives of this thesis are:

- To examine the relationship between anxiety and CS1 programming performance.
- To examine the relationship between stress and CS1 programming performance.
- To examine the relationship between anxiety and stress.

- To review the data obtained throughout the project, to identify analyse and identify any gender differences.

This chapter set out to identify the instruments and materials that would be used throughout this project. Each of the materials and instruments are designed to capture one of the following variables: 1) anxiety, 2) stress and 3) performance.

- Anxiety - State Trait Anxiety Inventory
- Stress - Measured by EDA and PPG
- Performance - Custom-made MCQ test

Figure 3.5 shows the updated thesis flow. At this point, the background literature has been gathered and the materials and instruments have been identified. Chapter 4 will outline the initial study which will attempt to answer the objectives of this project.

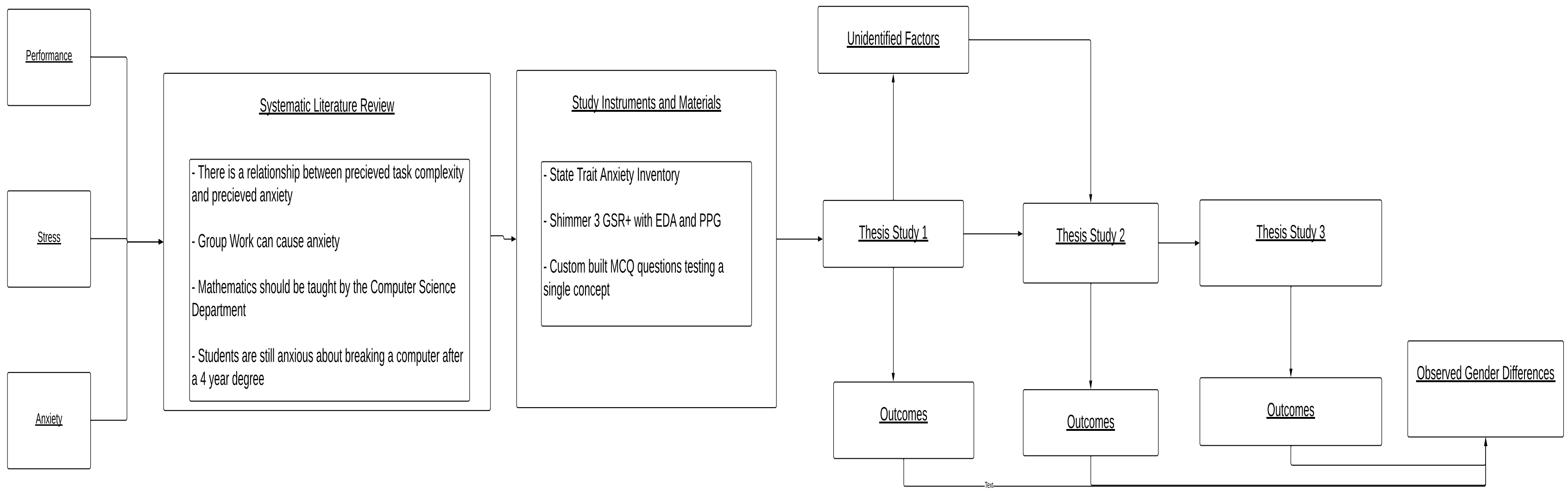


Figure 3.5: Update of the research methodology. Outcomes of this chapter will be included in studies 1, 2 and 3.

4 THESIS STUDY 1

This chapter outlines the first study, (Study 1), designed to gain insights on the objectives described in Chapter 1. This chapter begins by describing the research questions for Study 1. Then the study protocol is described. Findings and analysis for each of the research questions are presented and the chapter concludes by outlining the contribution the study makes to the thesis objectives.

4.1 RESEARCH QUESTIONS

As described in Chapter 1, the *My World Survey*, the *My World Survey 2* and the *Union of Students of Ireland Report* found that anxiety is a major concern for students in third-level education [31, 32, 81]. Greater insight on in the moment stress and anxiety of first year CS students could enable interventions to be put in place. This chapter explores the use of PPG and EDA sensors to measure stress and self-reported anxiety is measured using the STAI. To this end, the following research questions are addressed:

TE1-RQ1: Can a relationship between anxiety (as measured by the STAI) and CS1 programming performance be found?

TE1-RQ2: Can a relationship between anxiety (as measured by the STAI) and stress (as measured by PPG and EDA) be found?

TE1-RQ3: Can a relationship between stress (as measured by PPG and EDA) and CS1 programming performance be found?

4.2 STUDY OVERVIEW

Participants in this study were studying CS1 at the Department of Computer Science, Maynooth University and volunteered freely to take part in this study. No payment or favourable treatment was offered in return. Ethical approval was sought and granted to carry out this research (see Appendix G, Reference Number: BSRESC-2015-017). The study was carried out in November 2016.

The researcher and a single participant were present in the room for the duration of the study. The researcher was out of view from the participant and stayed in the room solely to ensure the study ran smoothly. Participants were seated at a desk with a monitor on it. On the desk, the participants had a keyboard, a mouse, a sheet of blank paper and a pen. Participants were instructed to read an information sheet, provided to them by the researcher, describing the study before commencement. Upon completion, if they had any issues or questions they were encouraged to ask for clarification. They were then asked to sign a consent form. A background demographic survey (see Appendix I) and the STAI (described in Section 3.2.1.2), were then given to the participant.

After this initial stage, the sensors (as described in Section 3.3) were placed solely on the non-dominant hand of the participant. A short 30-second baseline measurement was taken at the beginning of the study to ensure the sensors were functioning and recording properly and comfortable to wear. During this baseline measurement, the Shimmer 3 GSR+ was calibrated using its onboard software. The participant was encouraged during this time to stay focused on a cross presented to them on the screen.

After obtaining the baseline measurement, the participants started the MCQ test described in Section 3.4. All questions were presented, evenly counterbalanced in groups of Easy, Medium or Hard. Within each difficulty band, questions were always shown in the same order as outlined in Section 3.4. This was to ensure that there was no confounding effect. The participant was instructed to answer each question by using the mouse to click on their chosen answer. Each participant was provided with a pen and paper and told they were allowed to doodle. The participants were asked by the researcher to consider each question carefully and told that there were no time constraints.

4.2.1 PARTICIPANT PROFILE

Forty-two participants (30 male, 12 female) participated in this study. Table 4.1 presents the age and gender profiles of the participants. The majority of participants in the study are male and are between the ages of 17–19.

Table 4.1: Age and gender profile of participants in study 1.

Age	Male (N=30)	Female (N=12)
17–19	22 (74%)	10 (84%)
20–22	4 (13%)	1 (8%)
23+	4 (13%)	1 (8%)

4.3 TE1-RQ1 - AN A RELATIONSHIP BETWEEN ANXIETY (AS MEASURED BY THE STAI) AND CS1 PROGRAMMING PERFORMANCE BE FOUND?

The following sections will first present the data that will be required to answer the question. Then the analysis of the data is detailed followed by a discussion.

4.3.1 REQUIRED DATA

State Anxiety and Trait Anxiety

The self-reported State anxiety and Trait anxiety averages gathered using the STAI (discussed in Section 3.2.1.2) are presented in Table 4.2. The results are broken down by gender as there are different scoring metrics in the STAI guidelines for grading male and female participants. Gender differences will be explored in more detail in Chapter 8. The State anxiety averages for both male and female participants are close to normal values (36.47 and 38.76 male and female). Interestingly, the Trait anxiety levels are considerably higher than the normal values as outlined in the STAI manual with normal values being 38.30 and 40.40 for male and female participants respectively as opposed to the significantly higher values of 52 for males and 56.83 for females as found here.

Table 4.2: Average State anxiety and Trait values for male and female participants and associated p-values. In addition, the Normal male and female score are presented in *italics*.

	Male	Female	p-value	*Normal Male*	*Normal Female*
State	38.96	40.25	0.44	*36.47*	*38.76*
Trait	52	56.83	0.32	*38.30*	*40.40*

Question Responses

All participants answered all nine questions (Appendix E. Table 4.3 shows the percentage breakdown of responses (correct and incorrect) for each question along with the primary concept examined. It is apparent that some concepts are more challenging than others as a decrease in performance can be observed (Q6 - While Loop, Q7 - Substring, Q8 - If-else with Substring and Q9 - Nested While Loop). Question 7 appears to be an outlier with the lowest correct response rate and the highest total time taken. This is an interesting findings as Question 7 was labelled as the easiest of the hard questions.

Table 4.3: All concepts with percentage breakdown of correct and incorrect responses.

	Concept(s)	Correct (N)	Incorrect (N)
Q1	System output	100% (42)	0% (0)
Q2	String variables	95% (40)	5% (2)
Q3	String Concatenation	69% (29)	31% (13)
Q4	If-else if-else	100% (42)	0% (0)
Q5	Nested if-else	88% (37)	12% (5)
Q6	While Loop	55% (23)	45% (19)
Q7	Substring	26% (11)	74% (31)
Q8	If-else with Substring	57% (24)	43% (18)
Q9	Nested While Loop	43% (18)	57% (24)

Response Times

To investigate the variation in response time, correct and incorrect response times were examined, with a breakdown presented in Table 4.4. It is apparent that some concepts are more challenging than others as there is a large increase in response times (Q6 - While Loop, Q7 - Substring, Q8 - If-else with Substring and Q9 - Nested While Loop). This increase in Response Time seems reasonable when the decrease in Correct responses was observed in Table 4.3.

Table 4.4: Average time taken in seconds (s) to respond to each question.

	Correct	Incorrect
Question 1	14.64	0
Question 2	13.8	12.82
Question 3	21.22	15.65
Question 4	20.37	0
Question 5	24.07	25
Question 6	53.11	70.18
Question 7	95.97	75.65
Question 8	39.58	32.28
Question 9	71.53	66.11

4.3.2 ANALYSIS

State anxiety and Trait anxiety were investigated as individual components with respect to the following performance responses:

- Correct responses, and,
- Response Times

These comparisons were made to investigate if a relationship between the participant's self-reported anxiety and their actions and reactions during the MCQ test could be found. As an initial step, correlation tests were used to compare relationship across all factors. Table 4.5 presents these correlations.

Table 4.5 illustrates a strong relationship between State anxiety and Trait anxiety. This relationship is already well documented [102]. All other relationships are weak with correlations less than r=+/-0.35.

Table 4.5: Correlations of the factors

	State Anxiety	Trait Anxiety	#Correct Responses	Response Times
State Anxiety	1			
Trait Anxiety	0.747175	1		
#Correct Responses	-0.05694	-0.10115	1	
Response Times	-0.11523	0.122075	0.157553	1

Given the lack of direct relationships, a multivariate clustering algorithm was used to further examine the relationship between all the factors. The use of multivariate clustering allows for the investigation of the variables in an overarching model which can allow for the partial association of variables which may not be evident at a bi-variate level. The clustering algorithm used was a k-means clustering algorithm and factors were reduced using PCA. The k-means algorithm was chosen as it has a good performance when clustering data that is both categorical and numerical data and is easily reproducible [1]. These algorithms are described in Appendix C.

The algorithm returned 4 clusters. These clusters are shown on a Scatter plot in Figure 4.1. The averages of each of the clusters are outlined in Table 4.6.

Figure 4.1: Scatter plot showing the cluster assignments for State anxiety, Trait anxiety, #Correct responses and Total Response Time.

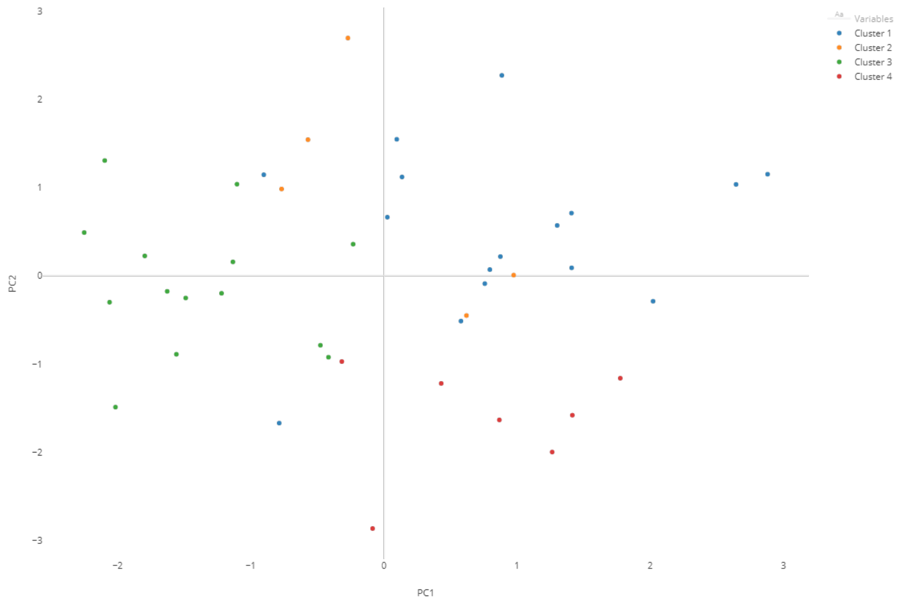


Table 4.6: Table showing the average State anxiety, Trait anxiety, #Correct responses, Total Response Time per cluster.

	State Anxiety	Trait Anxiety	#Correct Responses	Total Response Times
Cluster 1	54.63	67.94	5.94	352.36
Cluster 2	22.4	62.6	4.8	312.5
Cluster 3	21.29	19	6.71	318.78
Cluster 4	52.57	82.29	7.57	368.55

To examine the relationship between the clusters, correlations between the average value of each of the clusters was examined with Table 4.7 presenting the results of the correlations.

Table 4.7: Correlations of the factors

	State Anxiety	Trait Anxiety	#Correct Responses	Response Times
State Anxiety	1			
Trait Anxiety	0.7319	1		
#Correct Responses	0.4459	0.0437	1	
Total Response Time	0.9495	0.6912	0.6745	1

As can be seen in Table 4.7, there is a strong relationship between State anxiety, Trait anxiety and Total Response Time. The relationship indicates that as either State anxiety or Trait anxiety increases, so too does Total Response Time. Further analysis was conducted into bi-variate relationships, however, there were no new relationships found other to the ones reported here.

Discussion

Evidence of a a relationship between self-reported anxiety and Total Response Time in performance in CS programming was found through this study, as well as a slight relationship between State anxiety and #Correct Responses. The evidence of a relationship between both State and Trait anxiety and Total Response Time seems reasonable as there is a known relationship between attention to threat and response time. If a participant experienced the questions as a threat then this would explain the increase in response times.

4.4 Can a relationship between anxiety (as measured by the STAI) and stress (as measured by PPG and EDA) be found?

4.4.1 Required Data

Skin Conductance Response

Through out the study, sweat responses or SCRs was collected using the EDA sensor. An algorithm created in MIT Media Lab was used to calculate the SCRs and is described in more detail in Appendix C [106]. Any participant that had SCRs outside two standard deviations of the mean was removed as they were considered outliers within the data. This resulted in the removal of two males and one female from the study analysis leaving 39 valid sets of data.

The number of SCRs over the entire study and across the question difficulty band (Easy – Hard) was examined where a high number of SCRs indicates a high level of stress and a low number of SCRs indicates a low level of stress. Table 4.8 describes the average SCR's across the *Easy*, *Medium* and *Hard* questions. As can be observed in Table 4.8, there is a considerable increase in SRCs as the difficulty band get harder. This pattern was observed in the majority of participants. This phenomenon has been recorded in other areas of literature such as public speaking, however, this is a novel finding in terms of assessment in an MCQ test situation.

Photoplethysmography

A Photoplethysmogram was used during the experiment to capture the heart beat-to-beat data. From this data, factors such as heart rate and the Root Mean Square of the Successive Differences (RMSSD) could be

Table 4.8: Average number of Skin Conductance Responses across each question difficulty band for Study 1.

Question Difficulty	SCRs
Easy	2.23
Medium	4.95
Hard	8.64

calculated. The RMSSD is a measure of Heart Rate Variability and has been found to correlate to emotional arousal [21]. Methods on how these measures are calculated are detailed in Appendix C.

The PPG signals were analysed over the first and second halves of the study rather than individual questions or question bands as the time frame for reliable HRV measures were too short at the per question level. Halves were chosen by taking the total run time of the study and dividing by 2. All data was rearranged to follow the same flow: Easy Questions, followed by Medium questions, followed by Hard questions. It is important to note that the data is not evenly split in terms of what questions were included in each halves response time. One set of participant data had to be removed from the PPG data as sections of the data were lost during recording. Table 4.9 presents the average RMSSD data over the two halves of the study.

From Table 4.9 it can be seen that there is little to no difference in the PPG data across the study. Thus it appears that there is no noticeable (obvious) difference in the PPG signal based on the level of difficulty of the questions. Thus, this measure of Heart Rate did not indicate that participants became more stressed as the study progressed.

Table 4.9: Average $ln(RMSSD)$ and Average Heart Rate values for each half in study 1.

	Average
Half 1 ln (RMSSD)	4.99
Half 2 ln (RMSSD)	5.00
Average Heart Rate Half 1	85
Average Heart Rate Half 2	85.41

4.4.2 ANALYSIS

The aim of this question was to examine if evidence of a relationship between anxiety and stress could be found. To begin analysis, State anxiety and Trait anxiety was investigated with respect to SCRs and and the PPG data. These comparisons were made to investigate if a relationship existed between the participant's self-reported anxiety and their stress responses to an MCQ test. As an initial step, correlation tests were used to compare relationships across all factors. Table 4.10 shows these correlations.

Table 4.10: Correlations of the factors

	State Anxiety	Trait Anxiety	SCRs	Average HR 1	Average HR2	H1 ln(RMSSD)	H2 ln(RMSSD)
State Anxiety	1						
Trait Anxiety	0.747	1					
SCRs	-0.298	-0.169	1				
Average HR 1	0.119	0.283	0.417	1			
Average HR2	0.110	0.272	0.427	0.995	1		
H1 ln(RMSSD)	-0.130	-0.228	-0.352	-0.753	-0.745	1	
H2 ln(RMSSD)	-0.065	-0.205	-0.373	-0.751	-0.744	0.985	1

Examining the correlations in Table 4.10, there appears to be weak relationships between State anxiety, Trait anxiety and all stress responses. The strong correlations between all HR and ln(RMSSD) factors are expected as they originated from the same datasets and are strongly related by nature. Given the lack of obvious relationships between anxiety and stress, multivariate clustering was used to investigate a potential relationship at a high level. Figure 4.2 is a scatter plot which depicts the cluster assignments. Table 4.11 shows the average values of the clusters.

Figure 4.2: Scatter plot showing the cluster assignments for State anxiety, Trait anxiety, SCRs and PPG data.

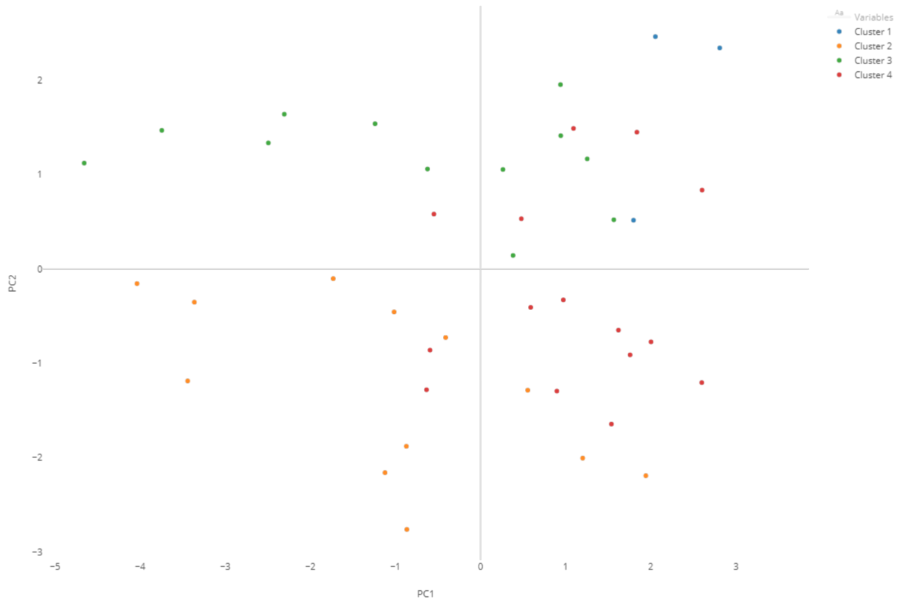


Table 4.11: Table showing the average State anxiety, Trait anxiety, SCRs and PPG data per cluster.

	State Anxiety	Trait Anxiety	SCRs	Average HR 1	Average HR2	H1 ln (RMSSD)	H2 ln (RMSSD)
Cluster 1	30	38.33	92.33	108.28	106.16	4.54	4.57
Cluster 2	63.75	69.67	7.42	74	74.61	5.33	5.39
Cluster 3	17.5	21.08	18.08	78.37	78.9	5.06	5.03
Cluster 4	39.13	69.2	16.73	99.39	99.5	4.76	4.75

To examine the relationship between the clusters, correlations between the average value of each of the clusters was examined, Table 4.12 shows the results of the correlations.

Table 4.12: Correlations of the factors

	State Anxiety	Trait Anxiety	SCRs	Avg HR1	Avg HR2	H1 ln (RMSSD)	H2 ln (RMSSD)
State Anxiety	1						
Trait Anxiety	0.862	1					
SCRs	-0.365	-0.379	1				
Avg HR 1	-0.32	-0.01	0.778	1			
Avg HR2	-0.318	0.011	0.749	0.999	1		
H1 ln(RMSSD)	0.521	0.217	-0.797	-0.975	-0.973	1	
H2 ln(RMSSD)	0.595	0.268	-0.75	-0.951	-0.951	0.994	1

Examining Table 4.12 in more detail, there are two noteworthy sets of correlations. 1) The correlations between State anxiety and ln(RMSSD) values have increased and would indicate that the lower the State anxiety the higher the ln(RMSSD) values which is a sign of lower stress.

This relationship however is not strong and will have to be validated in order to draw significant conclusions. 2) There is a strong relationship between heart rate variables and SCRs. This is extremely positive to observe as both of the variables are representative of the participants autonomic nervous system and are synchronous.

4.4.3 DISCUSSION

When the initial correlations between State anxiety, Trait anxiety, EDA (SCRs) and PPG values were investigated, a weak or extremely weak relationship was found, as shown in Table 4.10. Multivariate clustering was used to investigate if a deeper relationship could be determined by grouping students with similar profiles. No strong relationships were found between State anxiety, Trait anxiety, and SCRs. There was however one moderate correlation between State anxiety and ln(RMSSD) values. This correlation must be validated to ensure accuracy. Overall, it can be said that there was no relationship identified between self-reported anxiety and EDA and self reported anxiety and PPG values.

4.5 TE1-RQ3 - CAN A RELATIONSHIP BETWEEN STRESS (AS MEASURED BY PPG AND EDA) AND CS1 PROGRAMMING PERFORMANCE BE FOUND?

4.5.1 REQUIRED DATA

The data that is required to answer this question has been previously presented in Tables 4.3, 4.4, 4.8 and 4.9.

4.5.2 ANALYSIS

To approach this question, the SCRs and PPG data was investigated with respect to the performance data collected. These comparisons were made to investigate if a relationship existed between the participant's stress responses and behavioural responses to an MCQ test. As an initial step, correlation tests were used to compare relationship across all factors. Table 4.13 shows these correlations.

Table 4.13: Correlations of the factors

	#Correct Responses	Total Response Times	SCRs	Avg HR1	Avg HR2	H1 ln (RMSSD)	H2 ln (RMSSD)
#Correct Responses	1	0	0	0	0	0	0
Total Response Times	0.158	1	0	0	0	0	0
SCRs	0.157	0.39	1	0	0	0	0
Avg HR1	-0.159	0.138	0.418	1	0	0	0
Avg HR2	-0.176	0.132	0.427	0.996	1	0	0
H1 ln(RMSSD)	0.035	-0.009	-0.352	-0.753	-0.745	1	0
H2 ln(RMSSD)	0.031	-0.014	-0.374	-0.752	-0.744	0.985	1

In Table 4.13, examining the intersection of the performance factors (#Correct Responses and Total Response Time) and the stress factors

(Average HR and ln(RMSSD)), the range of correlations is low (-0.27 – 0.39). This suggests a weak relationship and so further investigation was carried out though the use of clustering. Given that there are 9 different factors, performing multivariate clustering on all factors returned no useful results and so is not reported here.

With the weak relationship between Total Response Time and SCRs, further investigation was conducted using bi variate clustering. Figure 4.3 is a scatter plot depicting the four clusters and average values of the clusters are shown in Table 4.14.

Figure 4.3: Scatter plot showing the cluster assignments for SCRs and Total Response Time.

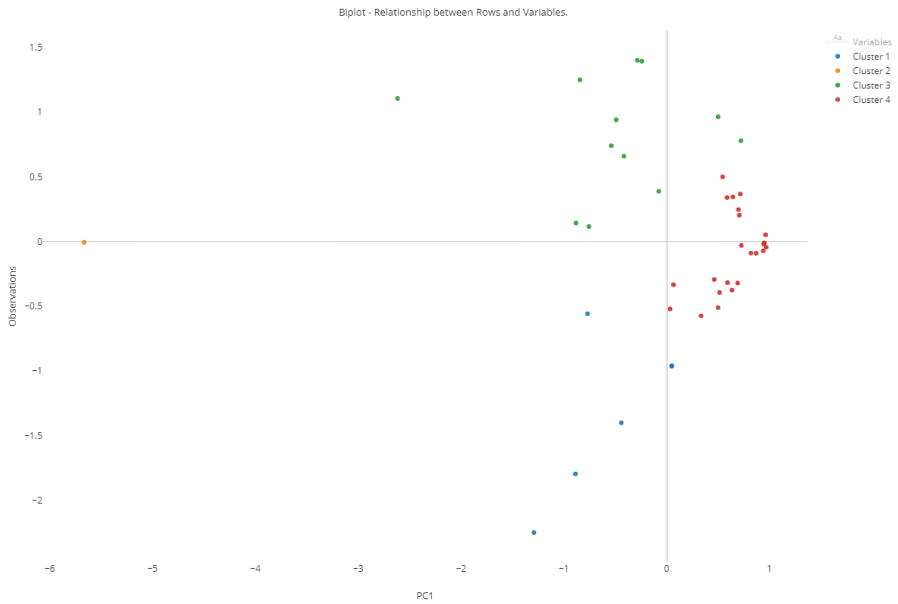


Table 4.14: Table showing the average SCRs and Total Response Time per cluster.

	Total Response Time	SCRs
Cluster 1	521.92	4.67
Cluster 2	893.41	131
Cluster 3	306.94	45.75
Cluster 4	284.12	5.48

Examining Table 4.14, there is a high level trend which suggests that as SCRs increase so does Total Response Time.

4.5.3 DISCUSSION

This research question set out to evaluate if a relationship between stress and performance could be found. It is clear that there is a relationship between Total Response Time and SCRs which suggests that the longer it took participants to respond to the MCQ test, the more stressed they became. Examining the other factors, it is apparent that there is no relationship between #Correct Responses and stress factors(SCRs and HR and ln(RMSSD)).

4.6 SUMMARY OF CHAPTER

At the start of this chapter, three research questions were defined with the hope they would inform the objectives in Chapter 1. The research questions were:

TE1-RQ1: Can a relationship between anxiety (as measured by the STAI) and CS1 programming performance be found?

Investigating if there was any relationship between anxiety and performance responses during an MCQ test returned an interesting set of findings. Focusing on Total Response Time, there appears to be a relationship with both State anxiety and Trait anxiety. As Total Response Time increases, the State anxiety and Trait anxiety levels of the participants are higher. With respect to the relationships between State anxiety, Trait anxiety and Correct Responses, no significant relationships were found. While this may appear as disappointing, it is very interesting. The data presented in this research question suggests that the relationship is more individual and follows a model such as the Individual Zone of Optimal Functioning model[90]. The IZOF model hypothesises that there is a functional relationship between anxiety and performance and that this relationship is unique for each individual. This relationship will be investigated further over the coming chapters in subsequent experiments. In conclusion to TE1-RQ1, it can be said that there **is a moderate relationship between Total Response Time and State anxiety and Trait anxiety**.

TE1-RQ2: Can a relationship between anxiety (as measured by the STAI) and stress (as measured by PPG and EDA) be found?

While investigating if there was a relationship between anxiety and stress (measured by EDA and PPG) it was shown that there was no obvious relationship between either State anxiety or Trait anxiety and the respective stress signals. Examining correlations coefficients and scatter plots, no obvious relationship was found. Using clustering techniques, it was investigated if subset profiles of participants could be used to identify a relationship. While the clustering algorithms showed clear clusters of participants, no relationship was found between the clusters that were of note. In conclusion to TE1-RQ2, **no relationship was found between anxiety and stress measures**.

TE1-RQ3: Is there a relationship between stress (as measured by PPG and EDA) and CS1 programming performance?

While investigating if there was a relationship between stress and CS1 programming performance it was shown that there is no obvious relationship between any of the factors. Multivariate analysis was carried out and clusters were formed however, the average values and ranges within the clusters showed that there was no high-level relationship between the values. In conclusion, it can be said that there is **no relationship between physiological arousal and CS1 programming performance**.

In addition to the findings within the research questions, several other key findings were made, as follows:

1. Participant State anxiety and Trait anxiety are higher than the normal values as reported in the STAI manual.
2. As the MCQ questions became harder, there was a rise in the number of SCRs observed in the participants.
3. As the MCQ questions became harder, the correct response rate reduced and the total response time increased. In addition, Question 7 appears to be an outlier with the lowest correct response rate and the highest total time taken. This is an interesting findings as Question 7 was labelled as the easiest of the hard questions.

4.7 THESIS UPDATE

The objectives of this project were to:

- To examine the relationship between anxiety and CS1 programming performance.
- To examine the relationship between stress and CS1 programming performance.
- To examine the relationship between anxiety and stress.
- To review the data obtained throughout the project, to identify analyse and identify any gender differences.

Figure 4.4 shows how Study 1 has contributed to the objectives through the identification of key findings and the identification of new factors. Chapter 5 will describe a similar study as the one one presented here with the inclusion of new factors. Given that only a measure of success was captured in Study 1, there is a clear need to obtain the participants confidence in their performance. In addition, the key findings will be validated too.

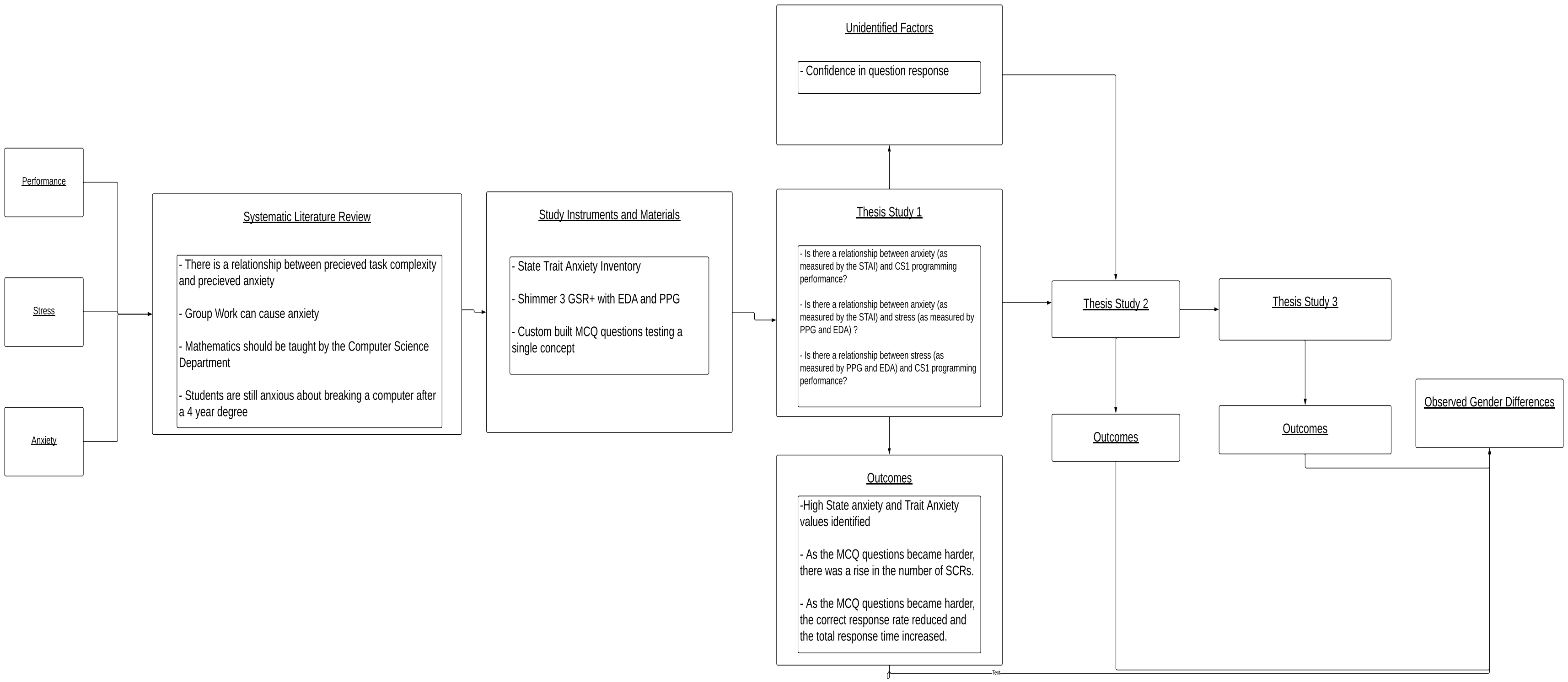


Figure 4.4: Update of the research methodology. Outcomes will be retested in the next chapter.

5 THESIS STUDY 2

The study presented in Chapter 4 attempted to find a relationship between anxiety, stress, and performance during an MCQ Programming test. The primary findings from this study can be summarised as follows:

1. Both State anxiety and Trait anxiety are higher than the normal values as reported in the STAI manual.
2. As the MCQ questions became harder, there was an increase in the number of SCRs observed in the participants.
3. As the MCQ questions became harder, the correct response rate reduced and the total response time per question increased. In addition, Question 7 appears to be an outlier with the lowest correct response rate and the highest total time taken.

The Innovation and Technology in Computer Science Education conference (ITiCSE) is a SIGCSE (Special Interest Group in Computer Science Education research) European conference held annually. In a 2015 ITiCSE Working Group led by Ihantola and Vihavainen, five "*Grand Challenges*" were defined for the CS Education community to encourage researchers and practitioners to shy away from once-off, single institutional studies and branch out into validation and re-validation studies [52]. As such, their second "*Grand Challenge*" was to "*systematically analyze and verify previous studies using data from multiple contexts to tease out tacit factors that contribute to previously observed outcomes*". Given that validation and re-validation studies are rare, an opportunity was observed in this project to facilitate the request of the working group. The study was reproduced to validate the findings that were reported in the previous chapter.

This chapter presents a follow-on study (referred to as Study 2) with a different cohort of students, which attempts to validate the findings of Chapter 4 and improve the study protocol to gain further insight. Enhancements were made during Study 2 and these will be outlined in detail. Research questions related to Study 2 will be defined in Section 5.2 and changes to the study protocol will be outlined in Section 5.4. Findings in relation to the research questions are presented in Sections 5.5,5.6,5.7, 5.8 and 5.9.

5.1 BACKGROUND

The results of Study 1 indicated that capturing EDA during an MCQ test was potentially a viable option in determining in-the-moment stress for students. In addition to this, the relationship between Total Response Time and self-reported anxiety measures (State anxiety and Trait anxiety) suggests that the longer it took a participant to respond in the study, the more anxious they were. There was however no relationship found between self-reported anxiety measures and stress. Reflecting on the study protocol and the findings of the literature review, two changes were made to the protocol used in Study 1. These changes were to gather programming self-efficacy and confidence in responses.

5.1.1 PROGRAMMING SELF-EFFICACY

In recent years, programming self-efficacy has been a topic of increased research in the area of Computer Science and in particular in relation to first-year Computer Science students. Bandura defined perceived self-efficacy as "*people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances*" [5]. Self-efficacy mediates between an individual's knowledge and their actions. Therefore, somebody may possess the knowledge or skills which are necessary to perform a particular task, however, they may not succeed due to their self-doubt or self-belief in their ability. Self-efficacy theory has emerged as an important means of understanding and predicting a person's performance.

Bergin showed that student's belief in their programming ability was the highest factor in predicting programming success [11, 13]. Quille et al., in a re-validation study found that programming self-efficacy was again found to be the main predictor of success in first-year computer science [84]. Research has shown that those with high programming self-efficacy are more likely to undertake tasks that are more challenging and demonstrate a want to learn and engage with material [96]. In addition to this, the success of a task will most likely increase the programming self-efficacy of the person involved [96]. In contrast, those with low programming self-efficacy tend to experience tasks that are easy, or at least easier, as much more difficult than they are. This leads to the person experiencing stress or anxiety [5]. This would lead to the student not succeeding in the task and further reduce their programming self-efficacy.

5.1.2 CONFIDENCE IN RESPONSES

The use of MCQ tests is perhaps the quickest and easiest way to examine students knowledge. This form of assessment, however, may allow the student to fortuitously land on the correct answer and receive marks they may not deserve. As educators, there is no way to gauge how confident a student is in their answer without asking them directly, and therefore it is difficult to know if the student is guessing, misinformed or uninformed. By considering the students' confidence in an incorrect answer, educators might be able to determine if a student is uninformed (incorrect and not confident) or misinformed (incorrect but confident) [27]. Knowing when a student is either misinformed or uninformed provides useful information on how best to support individual students.

5.2 RESEARCH QUESTIONS

By replicating Study 1 with limited changes to the study protocol, there was a unique opportunity to recreate the findings described in Chapter 4 and so, the following questions are posed:

VAL-RQ1 Are State anxiety and Trait anxiety levels higher in a CS1 population compared to the normal population?

VAL-RQ2 Can a relationship between stress and question difficulty be found?

VAL-RQ3 Can similar behavioural responses (correct/incorrect responses, Response Time) be observed in the MCQ test?

Additionally, this chapter attempts to answer the following research questions:

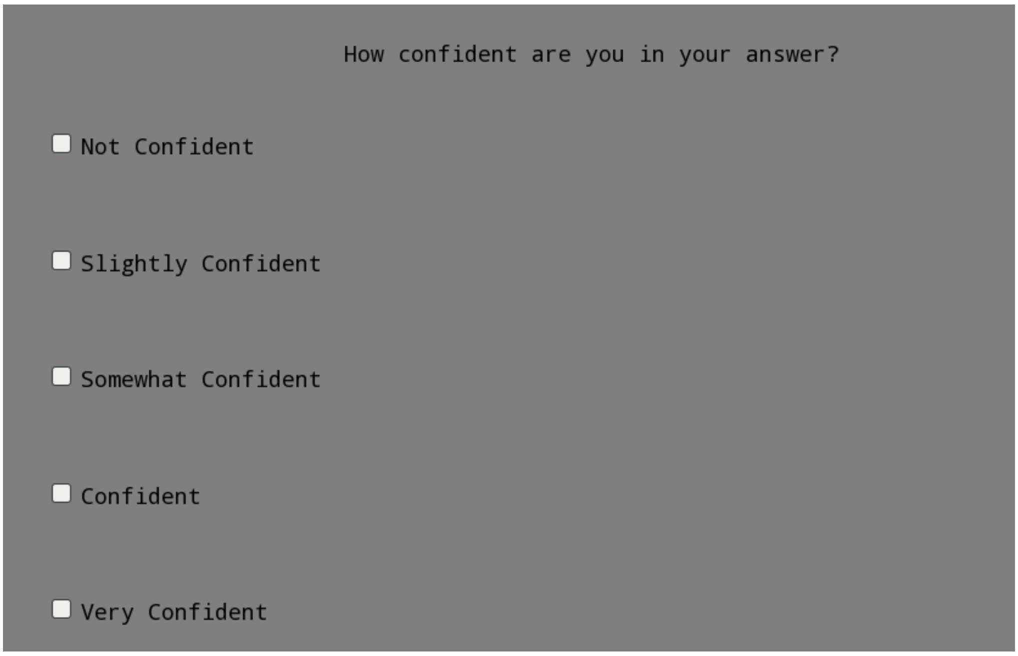
TE2-RQ1: Is there a relationship between students stress and their confidence (as measured by programming self-efficacy and confidence in response)?

TE2-RQ2: Do self-reported anxiety values align with participants confidence (as measured by programming self-efficacy and confidence in response)?

5.3 CHANGES TO STUDY 1 PROTOCOL

At the beginning of the study, programming self-efficacy was captured alongside the STAI. An alteration to the screen displaying the MCQs was made so that every programming question had one subsequent question. Participants were asked to rate their confidence in their answer by choosing one of “Not Confident", “Slightly Confident", “Somewhat Confident", “Confident" and “Very Confident". This allowed for the ability to gauge if the participant was guessing the answer as shown in Figure 5.1.

Figure 5.1: Screenshot of confidence-in-response scale.



5.4 STUDY PROTOCOL

Other than the additions described in Section 5.3, the study protocol was the same as that described in Section 4.2. This study was conducted in late 2017, a similar time when Study 1 was carried out in 2016. Ethical approval was sought and granted to carry out this research (see Appendix G, Reference Number: BSRESC-2015-017).

5.4.1 PARTICIPANT PROFILE

Forty participants (28 male, 12 female) participated in Study 2. These participants are comparable to those who undertook Study 1 with the main differences being the cohorts are a year apart but studying the same material and there were fewer participants are in the 20–22 age range in Study 2 compared to Study 1. Table 5.1 presents the age and gender breakdown of the participants Study 2.

Table 5.1: Age and gender profile of participants in Study 2.

Age	Male (N=28)	Female (N=12)
17 - 19	24 (86%)	8 (67%)
20 - 22	0 (0%)	3 (25%)
23+	4 (14%)	1 (8%)

In the following sections, the data required for each of the research questions will be presented, followed by an analysis of the data. Finally a discussion surrounding the data in relation to the research question will be provided.

5.5 VAL-RQ1 : ARE STATE ANXIETY AND TRAIT ANXIETY LEVELS HIGHER IN A CS1 POPULATION?

The following sections will outline the data required to inform the research question. Following this, a detailed analysis was carried out with a discussion of the analysis.

5.5.1 REQUIRED DATA

State Anxiety and Trait Anxiety

Table 5.2 shows the average values of the captured State anxiety and Trait anxiety of the participants of both Study 1 and Study 2. As can be seen, there is a significant increase in both State anxiety and Trait anxiety in Study 2 compared to Study 1. This is suggestive that the cohort that participated in Study 2 were more anxious than the cohort in Study 1.

Table 5.2: Average values from Study 1 Study 2 of State anxiety and Trait anxiety.

	Study 1 Average	Study 2 Average
State Anxiety	39.3	50.72
Trait Anxiety	59.3	54.67

5.5.2 DISCUSSION

In Section 4.3.2, it was noted that the Trait anxiety in the Study 1 participants appeared to be considerably higher than the reported normal college values. The STAI manual reporting normal State anxiety figures at 36 and 28 for Males and Females and Trait anxiety at 38 and 40 for Males and Females. The values stated in Table 5.2, it can be observed that both State anxiety and Trait anxiety values are considerably higher than the normal college values. Given that the findings are both a) across studies, and b) different cohorts, it is reasonable to question if Computer Science students are more anxious than the typical college student. This is explored further in Chapter 6 as a larger study was carried out exploring this as this was an unexpected finding and one that is not mirrored in the literature. Overall, it can be seen that higher State anxiety and Trait anxiety values are observed in Study 2 in comparison to normal college values and so this validates the findings in Study 1.

5.6 VAL-RQ2 : IS THERE A RELATIONSHIP BETWEEN STRESS AND QUESTION DIFFICULTY?

5.6.1 REQUIRED DATA

Electrodermal Activity

The number of SCRs over the entire study was collected and broken down into the difficulty bands. Table 5.3 presents the average SCR's across the Easy, Medium and Hard questions across both Study 1 and Study 2.

Table 5.3: Average number of SCRs across each question difficulty band.

Question Difficulty	SCRs Study 1	SCRs Study 2
Easy	2.23	3.02
Medium	4.95	5.97
Hard	8.64	9.32

5.6.2 PHOTOPLETHYSMOGRAPHY

The Heart Rate of each participant was collected across the entire study. Similar to Study 1, the PPG signal from Study 2 was analysed similarly as outlined in Section 4.4.1.2. Signals were broken down into halves and results from both Study 1 and Study 2 are presented in Table 5.4.

Table 5.4: Average $ln(RMSSD)$ values and Average Heart Rate values for both Study 1 and Study 2.

	Study 1	Study 2
Half 1 ln (RMSSD)	4.99	4.12
Half 2 ln (RMSSD)	5.00	3.82
Average Heart Rate Half 1	85	93.95
Average Heart Rate Half 2	85.41	91.39

5.6.3 DISCUSSION

Electrodermal Activity

Table 5.3 presents the average SCR's across the Easy, Medium and Hard questions for both Study 1 and Study 2. As can be seen, there is a considerable increase in the number of SCRs as the difficulty level increases in both Study 1 and Study 2. This is what one would expect; given harder questions participants would become more aroused and consequently sweat more and therefore more SCRs would be observed.

Comparing the increase in SCRs Study 1 to SCRs Study 2 in Table 5.3, the findings on the increase in SCRs across the difficulty bands is compelling. In both studies, questions were evenly counterbalanced to ensure there were no confounding effects. In both studies, regardless of when the questions were displayed, the average number of SCRs increased, as did the difficulty. At a high level, this would indicate a relationship with question difficulty and EDA. This study protocol was conducted with two different cohorts in subsequent years and the same relationship was observed in both cohorts. A finding which indicates that question difficulty and EDA are related is novel in the scope of Computer Science. While there is research which indicates increases in

EDA are related to increases in workload, this project has shown that increases in EDA are related to question difficulty.

Photoplethysmography

Examining Table 5.4, it can be seen that the $ln(RMSSD)$ values in Study 2 are lower than those in Study 1. However, the Average Heart Rate values recorded in Study 2 are higher than those in Study 1. In Study 2, the population could have been more nervous about participating in the study and so have higher Average Heart Rates than the participants in Study 1. However, it is hard to draw firm conclusions on the differences observed across the studies.

Overall, the increase in SCRs as the question difficulty increased was validated in Study 2. Given the inconsistency in the PPG values, it is hard to draw a conclusion to this research question and so, further investigation is required.

5.7 VAL-RQ3 : ARE SIMILAR BEHAVIOURAL RESPONSES (CORRECT/INCORRECT RESPONSES, RESPONSE TIME) OBSERVED IN THE MCQ TEST BETWEEN PARTICIPANTS?

5.7.1 REQUIRED DATA

Question Responses

All participants in Study 2 (n=40) answered all questions. Table 5.5 shows the number of correct and incorrect responses to each question for both Study 1 and Study 2. Examining Table 5.5, it can be observed that similar question responses were recorded for all questions.

Table 5.5: The results of each of the questions for Study 1 and Study 2.

	Study 1		Study 2	
	Correct	Incorrect	Correct	Incorrect
Question 1	42	0	39	1
Question 2	40	2	39	1
Question 3	29	13	36	4
Question 4	42	0	40	0
Question 5	37	5	39	1
Question 6	23	19	31	9
Question 7	11	31	19	21
Question 8	24	18	22	18
Question 9	18	24	24	16

Response Times

Table 5.6 provides a breakdown of response times for correct and incorrect answers for both Study 1 and Study 2. Similar to the response s recorded in Study 1, the general trend appears to be that as the questions get harder, the average time taken to respond to the questions increases in Study 2.

5.7.2 DISCUSSION

MCQ Results

Examining Table 5.5, there is a clear reduction in Correct responses as the question difficulty increases. Examining the "Easy" questions (Questions 1, 2 and 3) there were only 6 incorrect responses in Study

Table 5.6: Average time taken in seconds (s) to respond correctly and incorrectly to each question for Study 2.

	Study 1		Study 2	
	Correct	Incorrect	Correct	Incorrect
Question 1	14.64	0	13.46	10.59
Question 2	13.8	12.82	13.87	31.14
Question 3	21.22	15.65	16.72	15.29
Question 4	20.37	0	20.18	0
Question 5	24.07	25	27.55	21.51
Question 6	53.11	70.18	58.88	46.81
Question 7	95.97	75.65	104.35	100.45
Question 8	39.58	32.28	36.73	40.48
Question 9	71.53	66.11	76.5	61.66

2. The same participant that got Question 1 incorrect also got Question 2 incorrect but was not one of the four participants to get Question 3 incorrect. This participant also got all of the "Hard" questions incorrect. Examining the "Medium" questions (Questions 4, 5 and 6) there were only 10 incorrect responses. All participants got Question 4 correct. This is reflected in the responses in Study 1 and can be seen in Table 5.5. The same participant that got Question 5 incorrect also got Questions 6, 7 and 9 incorrect. Examining Table 5.5, it is apparent that participants struggled with the "Hard" questions (Question 7, 8 and 9) as there were 55 incorrect answers in Study 2 (77% of overall incorrect responses). This trend of high incorrect responses was also observed in Study 1 where there were 73 incorrect answers.

There was an important finding across both Study 1 and Study 2. Initially, Question 7 (substring) was ranked as the "easiest" of the "Hard" questions when the questions were designed but in Chapter 4 it was identified as the hardest question based on participant performance. This was unusual and unexpected as there was no prior research which indicates that substring is a hard/difficult concept to understand. There was a clear need to re-validate this finding. In Study 2, the same pattern was observed with Question 7 being identified again as the "hardest" question based on participant performance. This validates the claim that substrings are potentially a stumbling concept for the students and should be given appropriate explanation and time when presenting the concept. This appears to be a novel finding and is not a concept that has been identified in the literature as a stumbling block [91].

Response Times

Examining Table 5.6, there was a significantly longer response time observed for Question 7 when compared to all other questions. This validated our earlier claim in Section 4.3.1.3 that Question 7 was harder than had been expected. This may add credence to the fact that substring is a concept that needs a lot of attention. Again, similar to the findings in Section 4.3.1.3, a similar trend is seen in the incorrect responses. As the questions become more difficult, the average response time is longer in general. With respect to Question 2, only one person got it wrong and so the response time presented in Table 5.6 should be interpreted with caution.

Overall, similar behavioural responses (correct/incorrect responses, Response Time) in Study 1 were observed in Study 2. This validation now adds supports to the claim that substrings are a stumbling point for the first-year students that participated in this study. It also validates the finding that correct response times were, on average, longer than the incorrect response times.

5.8 TE2-RQ1 : IS THERE A RELATIONSHIP BETWEEN STUDENTS STRESS AND THEIR CONFIDENCE (AS MEASURED BY PROGRAMMING SELF-EFFICACY AND CONFIDENCE IN RESPONSE)?

5.8.1 REQUIRED DATA

Data presented in Table 5.3 is used to help answer this question. The Heart Rate data would have been used to inform this research question, however, given the ambiguous data presented in Table 5.4, the heart rate data was left out of this analysis.

Programming Self-efficacy

Bergin's programming Self-efficacy questionnaire was marked in the range of 10 – 40 given that the lowest response on each question was 1. The higher the score a student gets in this questionnaire the higher the level of self-efficacy the student has in their programming ability. Consequently, the lower the score, the lower the level of self-efficacy the student has in their programming ability. Overall, the average programming self-efficacy was 29.20. Breaking this down by gender, male students have an average score of 29.74 compared to an average score of 28 for females. One male failed to complete the survey correctly and so was excluded from the analysis. This result is in line with previous research in the area with males participants reporting higher levels of self-efficacy in their programming ability compared to their female counterpart. A Shapiro-Wilks test was used to test the normality of the programming self-efficacy results to determine what comparison tests could be used. The test returned a value of $p = 0.014$ indicating that the data is not normal meaning a non-parametric correlation test will have to be used instead of Parsons correlation test.

5.8.2 CONFIDENCE IN RESPONSES

As discussed in Section 5.1, participant confidence in their answers was collected for each question. Nine confidence values were collected for each participant, one for each question. Table 5.7 shows the breakdown of their recorded confidence for each question.

Table 5.7: Confidence and correctness in answers are outlined here. (CC - Correct and Confident, CNC - Correct and Not Confident, IC - Incorrect and Confident and INC - Incorrect and Not Confident)

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
CC	39	39	36	40	39	31	16	21	20
CNC	0	0	0	0	0	0	0	0	0
IC	1	1	4	0	0	5	16	16	16
INC	0	0	0	0	1	4	5	2	0

From Table 5.7 it can be seen that when a correct answer was given, no participant indicated that they were not confident (CNC - Correct and Not Confident). While the majority of participants were Correct and Confident, there is a definite shift towards Incorrect and Confident answers as the questions get harder. Examining Questions 7, 8 and 9, it can be seen that there is a large cohort of students that were incorrect in their answers but were confident in their response which is an indication that perhaps these participants thought they were correct in their answers but were perhaps misinformed. Of the 16 participants that were Incorrect and Confident for Question 7, 8 and 9, only three participants were Incorrect and Confident.

To compare confidence in responses with other multi-value variables, a data reduction technique known as Principal Component Analysis was used to reduce the data into a single component that explains the most variability in the data. An important point to note here is that the combined confidence values are in the range of -1.617 to 5.434. While these values seem random, the algorithm used normalises the results. While this is a large spread in values, the more negative a value was, the more confident someone was in their responses overall. These values were tested for normality and the test returned a p-value of $p = 0.0003$ indicating that the data is not normal. This will be taken into account in further analysis.

5.8.3 ANALYSIS

Confidence-in-response, programming self-efficacy and Electrodermal Activity were correlated and the results are shown in Table 5.8. With initial correlations of $r = -0.145$ and $r = 0.09$, this is suggestive there may not be a relationship between Programming self-efficacy, Confidence-in-response and Electrodermal Activity.

Table 5.8: Correlations of SCRs, Programming self-efficacy and Confidence-in-response

	SCRs	Programming Self-efficacy	Confidence-in-response
SCRs	1	-	-
Programming Self-efficacy	-0.145	1	-
Confidence-in-response	0.09	-0.243	1

To investigate if there is a relationship on a deeper level, clustering was used. Figure 5.2 shows the clusters identified while Table 5.9 shows the averages of Confidence-in-response and SCRs of each of the clusters and Table 5.9 contains the average values of each of the clusters.

Figure 5.2: Clusters of SCRs Programming self-efficacy and Confidence-in-response values.

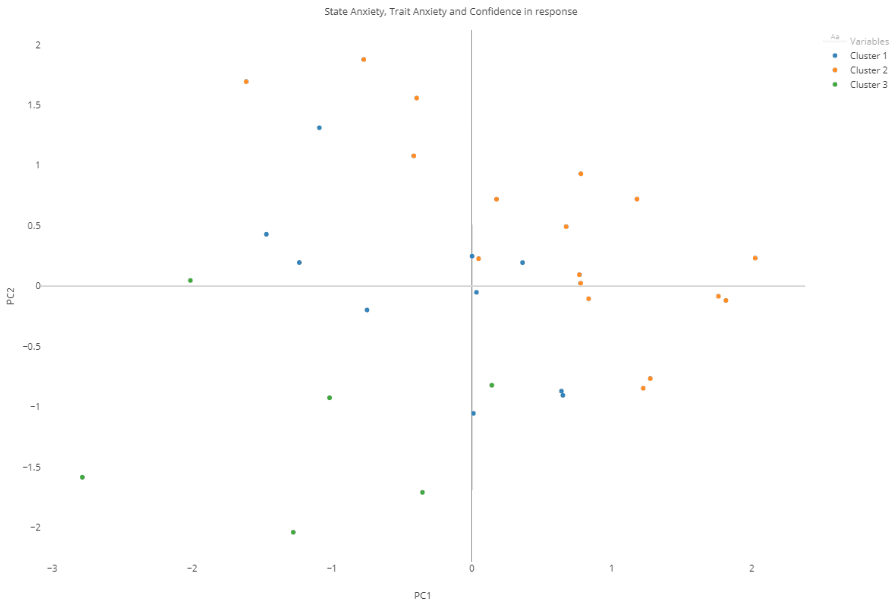


Table 5.9: Average values of clusters

	SCRs	Programming Self-efficacy	Confidence-in-response
Cluster 1	33	20.6	-0.70
Cluster 2	43.35	33.94	-0.43
Cluster 3	26.66	24	2.61

Table 5.10: Correlations of SCRs, Programming self-efficacy and Confidence-in-response clusters

	SCRs	Programming Self-efficacy	Confidence-in-response
SCRs	1		
Programming Self-efficacy	0.80	1	
Confidence-in-response	-0.74	-0.20	1

The correlations between the clusters were calculated and are shown in Table 5.10.Examining the between-cluster correlations, it is seen there is a positive relationship between SCRs and Programming self-efficacy which suggests that those who get stressed are more likely to have higher levels of Programming self-efficacy. A negative relationship between SCRs and Confidence-in-response was also observed. This would suggest that, on a deep level, the more confident in your answers you are the less you tend to sweat. While this a high-level interpretation of the results displayed here, there are other factors at play and further investigation is required to validate this claim.

5.8.4 DISCUSSION

This research question set out to investigate if there was a relationship between Confidence-in-response, programming self-efficacy and stress while completing an MCQ test. Both Heart Rate and EDA values were collected; however, Heart Rate was dropped from this analysis as there was conflicting data between Study 1 and Study 2. Initially, weak insignificant correlations were found between SCRs, confidence-in-response and programming self-efficacy. By using clustering, it allowed for an overview of how the three variables interacted with each other. Following the clustering, it appears that there are strong relationships between SCRs and both confidence-in-response and programming self-efficacy. The correlations are seen at a high level and so the results should be taken with caution but this is an indication of a relationship. There is a general trend that emerges, the lower the confidence-in-response and programming self-efficacy the lower the number of SCRs. This leads to the conclusion that the lower in confidence someone is the lower the levels of stress.

5.9 te2-rq2 : do self-reported anxiety values align with participants (as measured by programming self-efficacy and confidence in response)?

5.9.1 required data

The data required to inform this research question has been presented previously in Section 5.8.1.1, Table 5.2 and Table 5.7.

5.9.2 analysis

To begin analysis on this research question, correlations were calculated between State anxiety, Trait anxiety, Programming self-efficacy and Confidence-in-response. The correlations were calculated and are outlined in Table 5.11. It can be observed that State anxiety and Trait anxiety are closely related as would be expected. It can also be observed that there is a moderate negative relationship between State anxiety and programming self-efficacy and Trait anxiety and programming self-efficacy. This indicates that lower the participants programming self-efficacy the higher their State anxiety and Trait anxiety.

Table 5.11: Correlations within clusters of State anxiety, Trait anxiety, programming self-efficacy and confidence-in-response.

	State anxiety	Trait anxiety	Programming self-efficacy	Confidence-in-response
State anxiety	1			
Trait anxiety	0.628	1		
Programming self-efficacy	-0.47	-0.55	1	
Confidence-in-response	-0.016	0.1	-0.24	1

A multivariate clustering approach was taken to investigate if there could be a relationship between the four factors. The clustering algorithm returned three clusters and these are depicted in a scatter plot presented in Figure 5.3. The averages of the clusters were then calculated and are shown in Table 5.12.

Figure 5.3: Clusters of participants State anxiety, Trait anxiety and Confidence-in-response values displayed on a scatter plot.

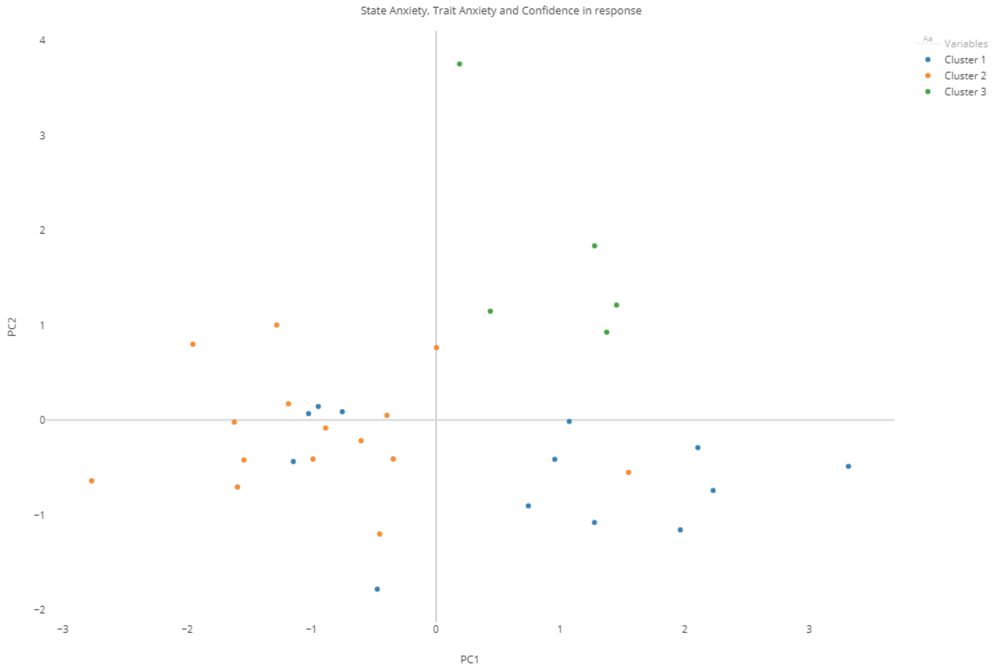


Table 5.12: Average values of clusters of State anxiety, Trait anxiety Programming self-efficacy and confidence-in-response.

	State Anxiety	Trait Anxiety	Programming self-efficacy	Confidence-in-response
Cluster 1	56.92	58.61	23.071	-0.77
Cluster 2	47.46	52.53	34.33	-0.20
Cluster 3	53.2	60	22.4	2.891

Examining Table 5.12, it seems like there is an overarching trend. As Programming self-efficacy increases so too does confidence-in-response while anxiety decreases. The correlations between the clusters were then calculated and are outlined in Table 5.13.

Table 5.13: Correlations within clusters of State anxiety, Trait anxiety, programming self-efficacy and confidence-in-response.

	State Anxiety	Trait Anxiety	Programming self-efficacy	Confidence-in-response
State Anxiety	1			
Trait Anxiety	0.83	1		
Programming self-efficacy	-0.89	-0.99	1	
Confidence-in-response	-0.02	0.52	-0.41	1

When the correlations were calculated across the averages of the clusters. In comparison to the correlations in Table 5.11, the relationships between the variables are stronger in Table 5.13. There was a clear negative relationship between State anxiety, Trait anxiety and Programming self-efficacy indicating that the lower the programming self-efficacy the higher both State anxiety and Trait anxiety. There also appears to be a moderate relationship between Confidence-in-response and Programming self-efficacy.

5.9.3 DISCUSSION

This research question explored the relationship between a student's confidence in their answers and their State anxiety and Trait anxiety. Initially, when the correlations were calculated, the results returned were weak. This is perhaps unsurprising as one could be confident in their responses to questions but still anxious due to the nature of the setting. However, given the scope of this thesis, there was a need to investigate the relationship on a deeper level. This was done by using clustering as it was possible that clustering could show a deeper relationship by limiting the variance of the variables. When the variables were clustered against each other, profiles were found with the general trend of, as both State anxiety and Trait anxiety increase, Confidence-in-responses reduces. However, no strong relationship was found between State anxiety, Trait anxiety and Confidence-in-response. Based on the results presented, there is a weak relationship between confidence-in-responses and both State anxiety and Trait anxiety. While this finding is at a high level, further research is required to investigate if there is a stronger bi-variate relationship.

5.10 summary of chapter

Given the "*Grand Challenges*" proposed by the 2015 iTICSE working group, this chapter provided the unique opportunity to recreate the findings of Study 1 outlined in Chapter 4. To this end, the following questions were proposed:

VAL-RQ1 Are State anxiety and Trait anxiety levels higher in a CS1 population?

VAL-RQ2 Is there a relationship between physiological responses and question difficulty?

VAL-RQ3 Are similar behavioural responses (correct/incorrect responses, Response Time) observed in the MCQ test between participants?

VAL-RQ1 examined the considerably higher State anxiety and Trait anxiety measures that were discovered in Study 1. By conducting the study again with a similar cohort to that in Chapter 4, it was shown that State anxiety and Trait anxiety values were again higher than the normal college student values that were reported in the STAI manual [102]. This now suggests that CS students are considerably more anxious than the average college student. This finding, coupled with the findings in the *My World Survey and My World Survey* 2 and the *Union Students of Ireland Report* are suggestive that CS students are anxious and educators should be mindful of this. To verify this, a large scale study of first-year CS was conducted at Maynooth University and will be discussed further in Chapter 6.

VAL-RQ2 set out to recreate the findings from Study 1 relating to the stress measures. In Section 4.4.1.1 it was noted that as the question difficulty rose, so too did the number of SCRs a participant experienced. When this pattern was investigated in Study 2, again it was observed that as questions became more difficult, the number of SCRs a participant experienced increased. This is a novel finding and is not mirrored in any literature regarding CS assessment. This is a significant finding and one that should lead to further research.

VAL-RQ3 investigated the similarities of the behavioural responses noted in Study 1. Initially the correct/incorrect responses were examined in Section 5.7.1.1. As observed in Study 1, as the questions became more difficult, the number of incorrect responses increased. This was observed across both studies. Interestingly, Question 7 (substring) in both studies proved to be the "hardest" question with the most incorrect responses in both studies. This was surprising and to see that it was the hardest question in terms of participant correct responses over both studies validates the claim that substrings may be a stumbling block for first-year CS students.

In addition, the response times to the questions were investigated. Again, as the questions became more difficult, the response times on the questions became longer. In general, the response times for the incorrect responses were shorter than the response times for the correct responses. This is observed across both studies.

In addition, this study set out to investigate the following research questions:

TE2-RQ1: Is there a relationship between students stress and their confidence when answering MCQ questions?

TE2-RQ2: Do self-reported anxiety values align with participants confidence throughout a set of MCQs questions?

TE2-RQ1 examined the relationship between confidence in student's answers and their physiological responses. Initially, when the correlations were tested, weak correlations were observed. The positive correlations observed here do support a weak relationship, however, caution should be taken with this. When the Confidence-in-response was clustered against Electrodermal Activity and Heart Rate, profiles of students were found and a strong relationship between Confidence-in-response and both Heart Rate and Electrodermal Activity was shown. This is important as now it can be seen that if a student is relaxed (low sweat rate) when completing a problem, they are more confident in their answers. This can be important for real-time intervention systems in online distance learning system as confidence could potentially be used.

TE2-RQ2 explored the relationship between a student's confidence in their answers and their psychological responses. Given that there are two psychological variables, State anxiety and Trait anxiety, they were tested independently. The correlation tests that were conducted showed that there was a weak relationship between both State anxiety and Trait anxiety were compared to confidence in responses. This is perhaps unsurprising as one could be confident in their responses to questions but still anxious due to the nature of the setting. When the variables were clustered against each other, profiles were found and no strong relationship was found. Based on our results, there was a weak relationship between confidence in responses and anxiety. While there is little research in this area, further research is needed to properly support this finding. A larger, longitudinal study would be ideal here where several MCQ tests could be conducted at multiple stages throughout the year and confidence in each of those questions would be recorded. This would allow for the individual tracking of confidence over a longer period and would provide multiple data points for the refinement of the clustering algorithm.

Given the objectives outlined in Chapter 1, the Research Questions addressed in this chapter inform the objectives in the following ways:

1. To gather evidence on the relationship between anxiety and performance.

   To inform this objective, **VAL-RQ1** and **EXP2-RQ2** were used. It now appears that the students who participated in CS1 and Study 1 and Study 2 tend to experience higher levels of anxiety than a normal college population. In addition, there is a weak relationship between anxiety and confidence in responses.

2. To examine the relationship between physiological arousal and performance.

   VAL-RQ2 informed this objective and now it can be said that as question difficulty becomes harder, the higher the number of SCRs there are. Unfortunately, no recommendation can be made with respect to the use of a PPG as conflicting data was obtained.

3. To examine the relationship between anxiety and physiological arousal. **TE2-RQ1** informed this objective through the use of confidence in responses. When the confidence in responses was examined in relation to physiological arousal there was a weak relationship found.

5.11 THESIS UPDATE

The original objectives of this project were to:

- To examine the relationship between anxiety and CS1 programming performance.
- To examine the relationship between stress and CS1 programming performance.
- To examine the relationship between anxiety and stress.
- To review the data obtained throughout the project, to identify analyse and identify any gender differences.

This chapter set out to 1) verify the findings from Study 1 and was successful in doing so, and, 2) identify relationships between anxiety and stress factors, confidence in responses and programming self-efficacy. The outcomes from this chapter have informed the objectives in the following manner:

- To examine the relationship between anxiety and CS1 programming performance:

  Study 2 allowed for the anxiety levels of a CS1 population to be re-tested and validated. Given that the anxiety levels are higher than the normal college population, it is reasonable to assume that the relationship between anxiety and performance found in Chapter 4 is valid here also.

- To examine the relationship between stress and CS1 programming performance:

  Over the course of Study 1 and Study 2 it was observed that as question difficulty increased so too the number of SCRs a participant exhibited and so it can be determined that there is a possible relationship between stress and performance.

Figure 5.4 shows how the project has updated with the inclusion of the findings of this chapter. Chapter 6 will present a large scale study conducted with over 180 participants.

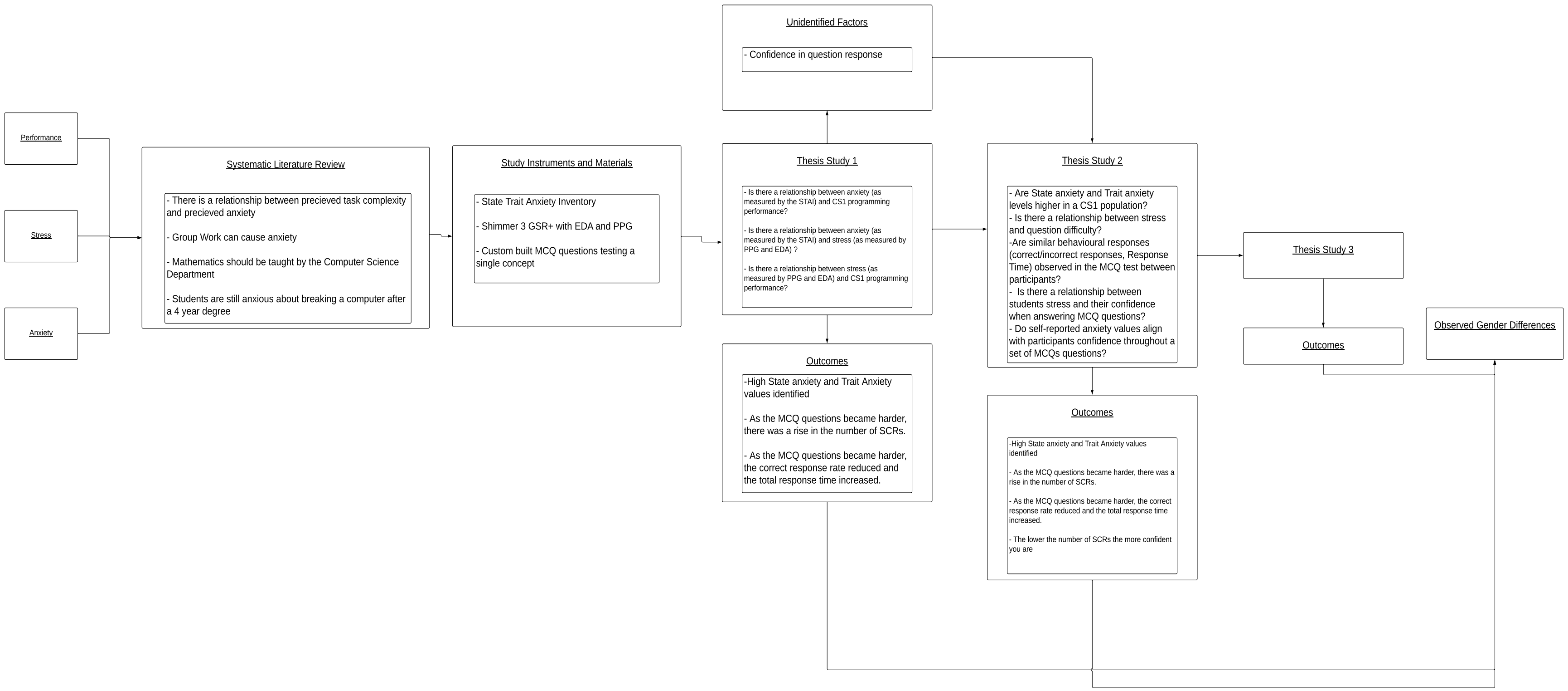


Figure 5.4: Update of the research methodology. Outcomes will be taken into account in Study 3 and will inform on gender differences.

6

LARGE SCALE STUDY OF ANXIETY IN COMPUTER SCIENCE

The studies described in Chapters 4 and 5 found that CS students were more anxious than a normal college population [102]. In addition, a relationship between State anxiety and Trait anxiety and Programming self-efficacy was uncovered. To investigate these further, a large scale study of Computer Science students was carried out to determine if the findings could be replicated. The participants who took part in this study are comparable to those who undertook Study 1 and Study 2 with the main differences being the cohorts are incremental academic years apart with more students in every age category. The first-year students had completed CS1 and were in their first week of CS2. In this chapter, the instruments and study protocols are described and the results of the survey presented are discussed.

6.1 RESEARCH QUESTIONS

Chapters 4 and 5 found higher than normal State anxiety and Trait anxiety levels were present in the Computer Science students who participated. The goal of the this large scale study (referred to as Study 3 throughout) was to test if the same findings would be replicated with a significantly larger CS student cohort. As part of this large scale study, programming self-efficacy was also collected to further investigate the relationship between it and anxiety. As such, this chapter addresses the following research questions:

SUR-RQ1: Are the students who took part in CS1 at Maynooth University more anxious than an average College population?

SUR-RQ2: Is there a relationship between a students perceived anxiety levels and their CS1 exam score?

SUR-RQ3: Is there a relationship between a students programming self-efficacy and their perceived anxiety levels?

SUR-RQ4: Is there a relationship between a students programming self-efficacy and their CS1 exam score?

SUR-RQ5: Given that students have completed CS1, how do they rate their programming self-efficacy?

6.2 STUDY PROTOCOLS

This section will describe the surveys that were used and how the data was collected. Full ethical approval was granted for this study (see Appendix G, Reference Number: SRESC-2018-114).

6.2.1 SURVEYS

As part of this study, the participant's age and gender were collected. The STAI questionnaire was used to collect participants self-reported State anxiety and Trait anxiety. It must be noted at this point that the STAI is normalised in the US and so the results presented in this Chapter should be interpreted with care. In addition to this, the participants programming self-efficacy was collected using Bergin's Programming Self-Efficacy questionnaire.

6.2.2 DATA COLLECTION

Microsoft forms were used to collect the responses to the questionnaires. The study was conducted during a participants allocated lab session for CS2 at the beginning of their first lab in the second semester. Following the completion of the questionnaires, participants allowed the researchers to collect their final exam grades from the module lecturer.

6.2.3 GENDER AND AGE PROFILE

Conducting the cohort wide study allowed for the opportunity to build on the previous small cohort results that had previously been collected in Study 1 and Study 2. For Study 3, the response rate for the survey was approximately 65% (182 of 280 students). The remaining 35% were likely absent on the day of the study (approximately 30 students), may have dropped Computer Science at that point and class lists may have not of been updated (approximately 50 students) or the students did not want to participate (approximately 18 students). The gender break down was 145 males to 37 females, equating to an approximate 80:20 male-female ratio. This gender split is similar to the cohorts described in Study 1 (n=42, 70:30 male-female) and Study 2 (n=40, 70:30 male-female). Table 6.1 shows age and gender breakdown of students within Study 3. The majority of the sample class (95%) are aged 17 – 22, which is to be expected.

Table 6.1: Age and gender breakdown of students in both the First year.

Age	Male	Female
17 – 19	122	33
20 – 22	15	4
23 – 25	1	0
26 +	7	0

6.3 sur-rq1: are the students who took part in cs1 at maynooth university more anxious than an average college population?

6.3.1 required data

Anxiety

As previously discussed, anxiety levels were measured through the use of the STAI questionnaire. Average State anxiety and Trait anxiety scores for Study 3 and normal values are presented in Table 6.2.

Table 6.2: Average State anxiety and Trait anxiety levels.

	CS Students	Normal
State anxiety	54	36 – 38
Trait anxiety	55	38 – 40

6.3.2 analysis

From the STAI manual [102], the State anxiety averages for male and female students in a normal college population are 36.47 and 38.76 respectively. The Trait anxiety averages for male and female students in a normal College population are 38.30 and 40.40 respectively. Examining Table 6.2, it is clear that the first-year undergraduate students who participated in the survey are more anxious than the normal population. Figure 6.1, illustrates State anxiety and Trait anxiety in this cohort compared with the normal averages (highlighted by horizontal lines). It should be noted that the higher anxiety levels in the students may not be caused by studying Computer Science, rather college life in general. Further study is required to identify the root cause the high anxiety levels.

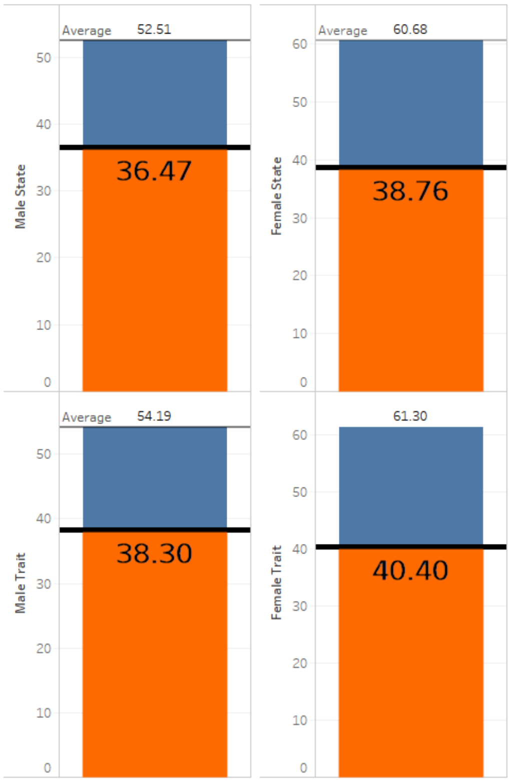


Figure 6.1: Bar chart showing the first year averages of male and female state anxiety and trait anxiety and the normal College population levels marked in horizontal lines.

6.3.3 DISCUSSION

Overall, three different cohorts [Study 1 (State anxiety = 39, Trait anxiety = 53), Study 2 (State anxiety = 51, Trait anxiety = 55), Study 3 (State anxiety = 54, Trait anxiety = 55)] were collected over three years with a total of 264 participants. Given the consistently higher scores, it is reasonable to conclude that the CS1 undergraduate population is considerably more anxious than the average College population. This significant, conclusive, novel finding is a valuable contribution of this thesis. Future work would be valuable to determine how best to work with such cohorts of students given their currently observed heightened anxiety levels.

6.4 IS THERE A RELATIONSHIP BETWEEN A STUDENTS PERCEIVED ANXIETY LEVELS AND THEIR CS1 EXAM SCORE?

6.4.1 REQUIRED DATA

Part of the data required to inform this research question has been presented in Table 6.2. In addition, the exam results that was collected have the following properties:

6.4.2 ANALYSIS

To investigate this question, several approaches were taken. Initially, the correlation between State anxiety and Exam results was examined to investigate if a relationship between State anxiety and Exam performance existed and if it did, what was the relationship. Following this, the correlation between Trait anxiety and Exam results was examined

Table 6.3: Properties of exam results.

	Value
Mean	50.91
Median	53
Mode	40
Standard Deviation	19.80
Variance	392.38

for the same reasoning. Table 6.4 shows the correlation values for each of the correlation pairs for the study. It was known previously that State anxiety and Trait anxiety were highly correlated from the STAI manual [102].

Table 6.4: Correlations between State anxiety, Trait anxiety and Exam Mark within the study.

	State	Trait	Exam Mark
State	1	-	-
Trait	0.842[1]	1	-
Exam Mark	-0.327	-0.197	1

From Table 6.4, there is a weak negative correlation between State anxiety measures and the Exam marks of $r = -0.327$ with an even weaker correlation between Trait anxiety and Exam marks of $r = -0.197$. This is suggestive of no relationship between the measures of anxiety and CS1 exam performance. A similar relationship was noted in both Study 1 and Study 2. While there is no obvious linear relationship, there

1 This correlation re-validates the relationship between State anxiety and Trait anxiety as reported in the STAI manual [102]

may be a model whereby people may perform better at different levels of anxiety, however, this requires investigation.

Subsequently, both State anxiety and Trait anxiety measures were run through a k-means clustering algorithm with Exam marks to identify any patterns in the relationship between anxiety and performance. Figure 6.2 shows the outcome of the clustering algorithm when State anxiety was plotted against Exam mark and Table 6.5 describes the clusters.

Figure 6.2: Results of clustering State anxiety and exam mark.

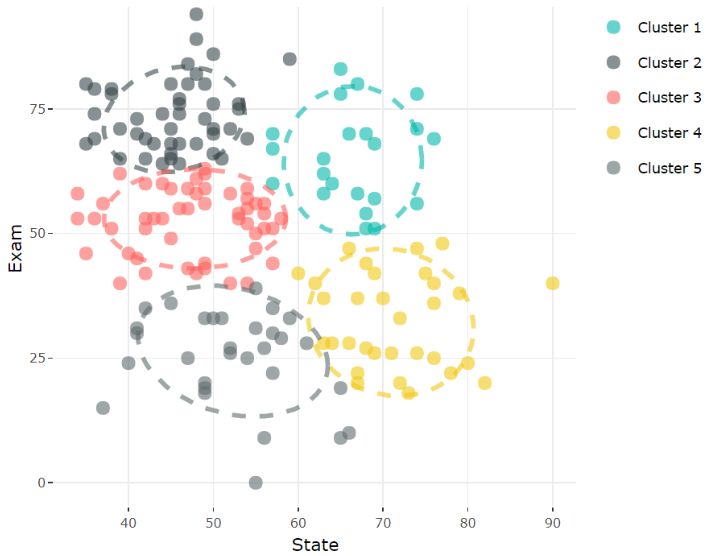


Table 6.5: Description of clusters of State anxiety and Exam Mark

Cluster	Description
Cluster 1	high State anxiety and high exam mark
Cluster 2	low State anxiety and high exam mark
Cluster 3	low State anxiety and a medium exam mark
Cluster 4	high State anxiety and low exam mark
Cluster 5	low State anxiety and low exam mark

Examining Figure 6.2, there are 5 clusters, each unique. This suggests that even inside the clusters there is no clear relationship between State anxiety and Exam performance. Examining Clusters 2, 3 and 5 in Figure 6.2, it can be observed that while State anxiety remains predominately less than 60 points, Exam performance has a range of values from 0 – 100. This suggests that while anxiety remains constant, exam performance may fluctuates. Examining Clusters 1 and 4 in Figure 6.2, there is evidence of high State anxiety with scores greater than 60 points and again, exam performance fluctuates between low performance and high performance. This strengthens the argument that there is no general relationship between State anxiety and Exam performance.

Given the strong relationship between State anxiety and Trait anxiety it would be expected that a similar relationship would be found as was observed between State anxiety and Exam performance. Figure 6.3 presents the outcome of the clustering algorithm when Trait anxiety was plotted against Exam mark and Table 6.6 describes the clusters.

Figure 6.3: Clusters of participants Trait anxiety and Exam mark.

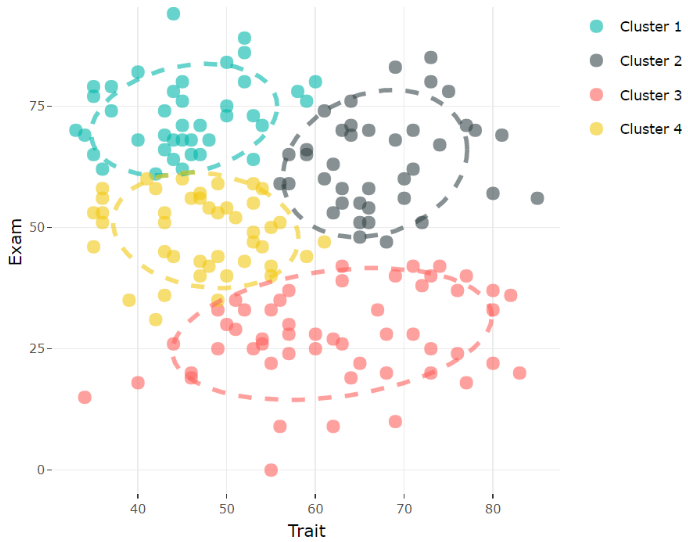


Table 6.6: Description of clusters of Trait anxiety and Exam Mark

Cluster	Description
Cluster 1	low Trait anxiety and high exam mark
Cluster 2	high Trait anxiety and high exam mark
Cluster 3	low exam mark but covers the range of Trait anxiety
Cluster 4	low Trait anxiety and a medium exam mark

Similar to State anxiety and exam performance, these clusters are unique with no overlap: each cluster represents a different combination of Trait anxiety and Exam performance. Examining Clusters 1 and 2 in Figure 6.3, it can be observed that exam performance is high (Cluster 1 average 73% and Cluster 2 average 63%) with a Trait large difference in Trait anxiety (Cluster 1 average Trait anxiety = 45 and Cluster 2 average Trait anxiety = 67).

Interestingly, unlike when State anxiety was clustered against exam mark, those with low exam marks were not grouped into two clusters but rather one cluster (Cluster 3): this cluster covered a large range of Trait anxiety values. Given this, the participants in Cluster 3 in Figure 6.3 were compared to the participants in Clusters 4 and 5 in Figure 6.2. Of the 61 participants in Clusters 3 and 4 in Figure 6.2, all but four are present in Cluster 3 in Figure 6.3. With the other clusters, only 12 participants did not map into the same clusters across State anxiety and Trait anxiety. This reinforces the strong relationship between State anxiety and Trait anxiety.

6.4.3 DISCUSSION

This research question involved the examination of a large data set from a CS1 group of students. Throughout the analysis, there was no overall relationship found between either State anxiety or Trait anxiety and performance in their CS1 exam. However, when cluster analysis was performed, an interesting finding was uncovered, irrespective of the level of anxiety presented, the performance was not affected. This suggests again that the relationship between anxiety and performance is unique and can not be generalised.

6.5 IS THERE A RELATIONSHIP BETWEEN A STUDENTS PROGRAMMING SELF-EFFICACY AND THEIR PERCEIVED ANXIETY LEVELS?

6.5.1 REQUIRED DATA

To inform this research question, data presented in Table 6.2 will be used along with the data presented in Section 6.5.1.1

Programming self-efficacy

The programming self-efficacy scale was marked in the range of 10 – 40. The higher the score, the more confidence a student is in their programming ability. An average score of 27 was found with a standard deviation of 7.75 points and median score of 26. This indicates that students are only slightly confident in their programming ability.

6.5.2 ANALYSIS

This research question set out to investigate the relationship between programming self-efficacy and anxiety with the hypothesis that as ones programming self-efficacy rises, their anxiety reduces. To this end, State anxiety was compared with programming self-efficacy and Trait anxiety was compared with programming self-efficacy. The results of these comparisons are discussed next.

State anxiety and Programming Self-efficacy

Initially when the relationship between State anxiety and programming self-efficacy for Study 3 was investigated, a moderate correlation was found ($r = -0.514$, $p < 00001$). Further inspection to determine if there was a stronger relationship at a more subtle level was carried out using clustering. Figure 6.4 shows that there were three clusters of participants and are described in Table 6.7

Figure 6.4: Chart showing Self-efficacy vs State anxiety clusters.

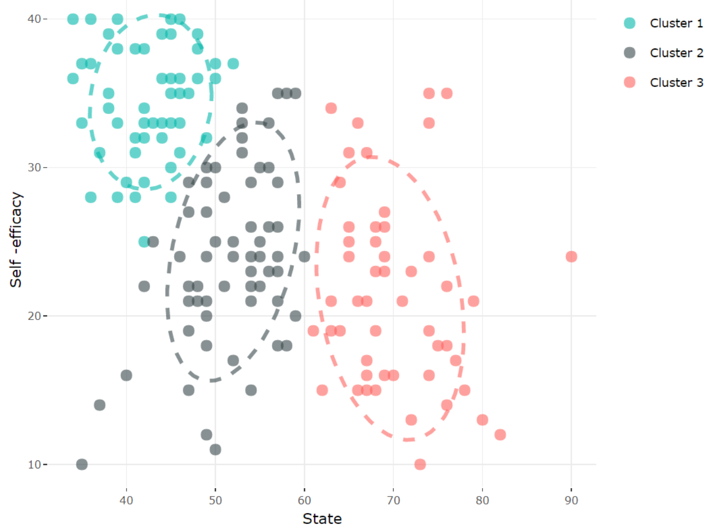


Table 6.7: Description of clusters of Self-efficacy and State anxiety

Cluster	Description
Cluster 1	medium exam mark and medium self-efficacy
Cluster 2	high exam mark and high self-efficacy
Cluster 3	low exam mark and a low self-efficacy

Looking at Figure 6.4 in conjunction with the moderate correlation ($r = -0.514$) it can be stated that there is a negative relationship between State anxiety and programming self-efficacy. This suggests that those with a high level of State anxiety tend to have low programming self-efficacy. Similarly, those with low State anxiety tend to have high programming self-efficacy. A Pearson's correlation test was performed and there was a between cluster correlation of $r = -0.85$. This suggests that there is a strong negative correlation between the profiles of State anxiety and self-efficacy identified, indicating that the lower the programming self-efficacy of a student is the higher their State anxiety, meaning that those who are more anxious have lower programming self-efficacy.

Trait Anxiety and Programming Self-efficacy

Given that there is a strong correlation between State anxiety and Trait anxiety, one would expect a similar relationship between Trait anxiety and programming self-efficacy as was observed in Figure 6.4. Figure 6.5 shows the relationship between Trait anxiety and programming self-efficacy. There is a correlation of ($r = -0.396$, $p < 0.00001$) which is considerably weaker than that found between State anxiety and programming self-efficacy ($r = -0.514$).

Using clustering techniques, the relationship between Trait anxiety and programming self-efficacy was examined to see if groups of students with similar programming self-efficacy and Trait anxiety profiles could be found. Again, three clusters were identified. Figure 6.5 shows the three clusters identified and these clusters are described in Table 6.8. A correlation test was run between the clusters identified and this returned a correlation value of $r = -0.95$ meaning that, at a high level, those with higher levels of Trait anxiety have lower levels of programming self-efficacy.

Figure 6.5: Chart showing Self-efficacy vs Trait anxiety clusters.

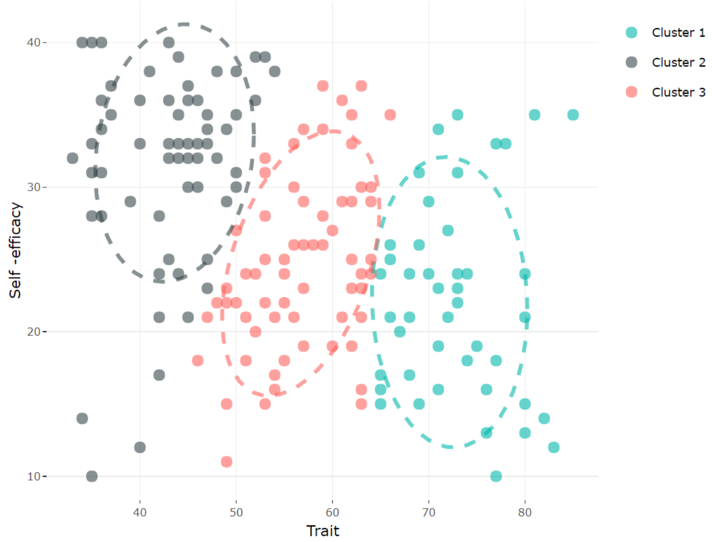


Table 6.8: Description of clusters of Self-efficacy and State anxiety

Cluster	Description
Cluster 1	high Trait anxiety with predominately low self-efficacy
Cluster 2	low Trait anxiety and high programming self-efficacy
Cluster 3	medium Trait anxiety and predominately medium programming self-efficacy

Examining the cluster assignments between State anxiety, Trait anxiety and programming self-efficacy, the participants who are in the extremities of Clusters 1 and 3 in Figure 6.4 are the same participants in the extremities in Clusters 1 and 2 in Figure 6.5.

There is a little bit of cross over between cluster assignments when examining Figure 6.4 Cluster 2 and Figure 6.5 Cluster 3. This accounts for 41 participants, however, they are at the extremities of the clusters. This crossover is to be expected given the median value for programming self-efficacy is 26 and the average is 27.

6.5.3 Discussion

Overall there appears to be a negative relationship between anxiety and programming self-efficacy; the less programming self-efficacy one has, the more State anxiety and Trait anxiety one has. This is a significant finding as future work can look at interventions to reduce the anxiety within the CS population and in turn, this will potentially increase the programming self-efficacy of the class, and in turn increase exam performance (as described in Section 6.6).

6.6 Is there a relationship between a students programming self-efficacy and their CS1 exam score?

6.6.1 Required data

The data required to inform this research question has previously been presented in Section 6.4.1 and Section 6.5.1.

6.6.2 Analysis

In previous research by Bergin et al. and Quille et al., programming self-efficacy is one of the most significant factors when it comes to predicting success [11, 84]. Given this, it is reasonable to expect a high correlation between programming self-efficacy and performance. From Study 3, Figure 6.6 shows a linear relationship between programming self-efficacy and exam mark as shown by the trend line, when these are plotted against each other. A Pearson's correlation was conducted with an r-value of $r = 0.657$.

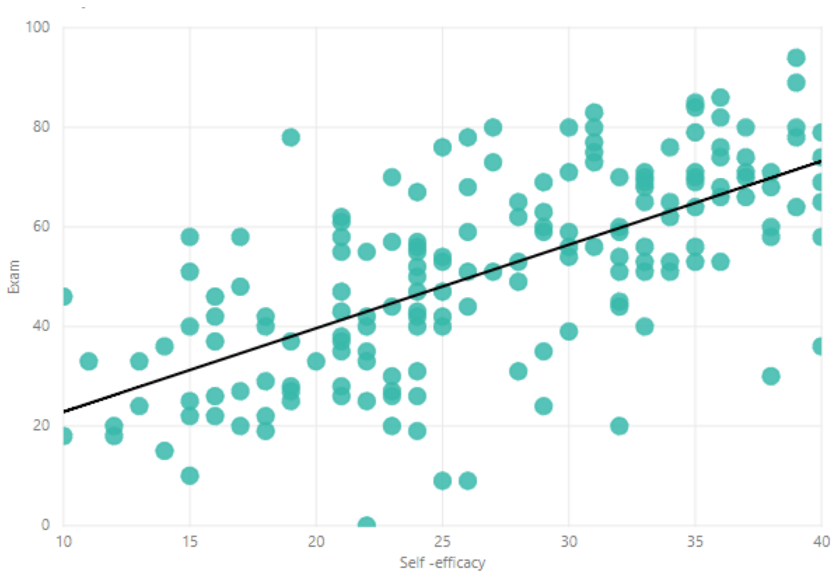


Figure 6.6: Chart showing Self-efficacy vs Exam mark and the trend line showing the linear relationship for the First Year survey.

6.6.3 Discussion

Given the strong correlation observed along with the clear linear trend, it appears that a students programming self-efficacy is directly related to their CS1 result. This is an important finding as if we, as educators can increase the programming self-efficacy of our students, we could potentially increase their exam results.

6.7 Given that students have completed a CS1 module, how do they rate their programming self-efficacy?

6.7.1 Required data

The data required to inform this research question has previously been presented in Section 6.5.1.

6.7.2 Analysis

Given that programming self-efficacy is context-specific, it can change in a short period of time [77]. In Maynooth University the CS1 module consists of 36 hours of lectures, 36 hours of mandatory programming labs, 48 hours of recommended independent course study, with additional support services in computer programming freely available in the department. For this study, the programming self-efficacy survey was completed at the beginning of the second semester meaning that all participating students were exposed to at least 12 weeks worth of Computer Science lectures. This is important as there is a steep learning curve associated with learning Computer Science and so one's confidence could change rapidly. Table 6.9 shows the number of participants broken into varying programming self-efficacy bands.

Table 6.9: Programming Self-efficacy broken into bands with counts of participants in each band

Programming Self-efficacy	Count
10 – 15	16
16 – 20	23
21 – 25	47
26 – 30	27
31 – 35	41
36 – 40	28

6.7.3 Discussion

Given that the median value of the programming self-efficacy of the participating students was 26, a score of less than 26 is considered low programming self-efficacy and a score greater than 26 is considered high programming self-efficacy. Taking this into account, 48% of the participants have low programming self-efficacy with 21% self-assessing themselves with a score of 20 or less indicating very low levels of belief in their programming ability. The low levels of self-efficacy are particularly concerning given the strong relationship between programming performance and self-efficacy. Given the ongoing coverage of programming in a CS degree, this can have an impact on their success in many subsequent modules.

6.8 SUMMARY OF CHAPTER

This study set out to investigate the mental health in our first-year Computer Science population following the findings of previous studies as outlined in Chapters 4 and 5. In the previous studies, it was suggested that Computer Science students are more anxious than their normal college counterparts. This study set out with five research questions in mind, each question focusing on at most 2 of the following aspects: Anxiety, programming self-efficacy, and, final exam mark. All research questions inform objective 1 from Chapter 1 (**To examine the relationship between anxiety and CS1 programming performance.**)

SUR-RQ1 focused explicitly on the results of the State Trait Anxiety Inventory. It was shown that there is a significant increase in both State anxiety and Trait anxiety in out first-year population when compared to the normal college population reported in the STAI manual [102]. This finding is important and future work should focus on how we can utilise this information to better cater for this student group.

SUR-RQ2 examined the relationship between anxiety and exam marks obtained in CS1. It was observed that there is no significant correlation between the two factors. While this may seem like a non-result, it provides us with an important finding. Previous research in the field of anxiety would suggest that there is an optimal zone which is unique to everyone, where a certain amount of anxiety would lead to somebody's best performance. This theory is known as the Individual Zone of Optimal Functioning (IZOF), a sports psychology theory which posits that everyone is individual and will perform with their unique level of anxiety. Findings from this study would suggest that this theory might be multi-disciplinary and potentially applicable in the CS Education domain. This is supported through the discussions based on the observations of Figures 6.2 and 6.3. Within these figures, no relationship between anxiety and exam performance is observed. However, it appears that there are students with varying degrees of anxiety performing well. This is highly suggestive that the relationship between anxiety and performance is unique and individual.

SUR-RQ3 was concerned with the relationship between anxiety and self-efficacy and arguably produced the most interesting results. A negative relationship was observed between the two factors; that is, anxious students are less self-confident in their ability. This is seen across State anxiety and Trait anxiety. This finding is not found in any literature and is a novel and unique finding in the field of CS Education. A cautionary note should be made at this point; while a relationship was found, it does not imply that reducing anxiety improves programming self-efficacy. Educators should, however, attempt to reduce anxiety within CS lectures and labs and put in place interventions to improve programming self-efficacy.

SUR-RQ4 looked at the relationship between self-efficacy and exam mark. The relationship was extremely linear which was to be expected. Those students that are confident in their ability were the top achievers in the class, whereas, in general, those who were lower in confidence were the low achievers in the class.

SUR-RQ5 related to programming self-efficacy. Given that 48% of the first-year cohort rate their programming self-efficacy as less than 26 points, the median value is concerning (the median value of 26). Given that the first-year students have had at least 36 hours of lectures, 36 hours of labs and a suggested 48 hours of independent course study, overall self-efficacy in programming is still reasonably low. Educators should now begin to look for methods to improve the programming self-efficacy of the students which will lead to more competent programmers.

While this study provided some much-needed insight into anxiety in first-year CS students, there is still much to be done. Anxiety amongst all CS students should be recorded. If the results of this study are mirrored across all undergraduate years, then this requires attention immediately.

6.9 THESIS UPDATE

To this point of the project, three of the four objectives have been investigated. These were:

- To gather evidence on the relationship between anxiety and performance.

  It is clear from the Study 1, Study 2 and Study 3 that there is no generalisable relationship between the anxiety and the performance in CS1. These findings suggest that the relationship between anxiety and performance is far more individualised and so a model similar to the IZOF model might be more applicable in a computer science setting.

- To examine the relationship between stress and performance.

Measuring stress using the EDA and PPG sensors, the data collected from these sensors was related to performance metrics collected during an MCQ test. There appears to be a relationship between the number of SCRs recorded for a student and the difficulty of the questions being asked; that is as the question difficulty rises so too does the number of SCRs. This finding is novel as now there is a method of tracking stress in a realtime environment.

- To examine the relationship between anxiety and stress.

  With anxiety and stress being two distinct metrics, a relationship between the metrics was attempted to be drawn. While multiple different methods were attempted to draw a relationship, there was no relationship found between the metrics in any of the studies in this project.

Figure 6.7 shows an updated state of the project. Chapter 7 will discuss the gender differences that were observed throughout this project.

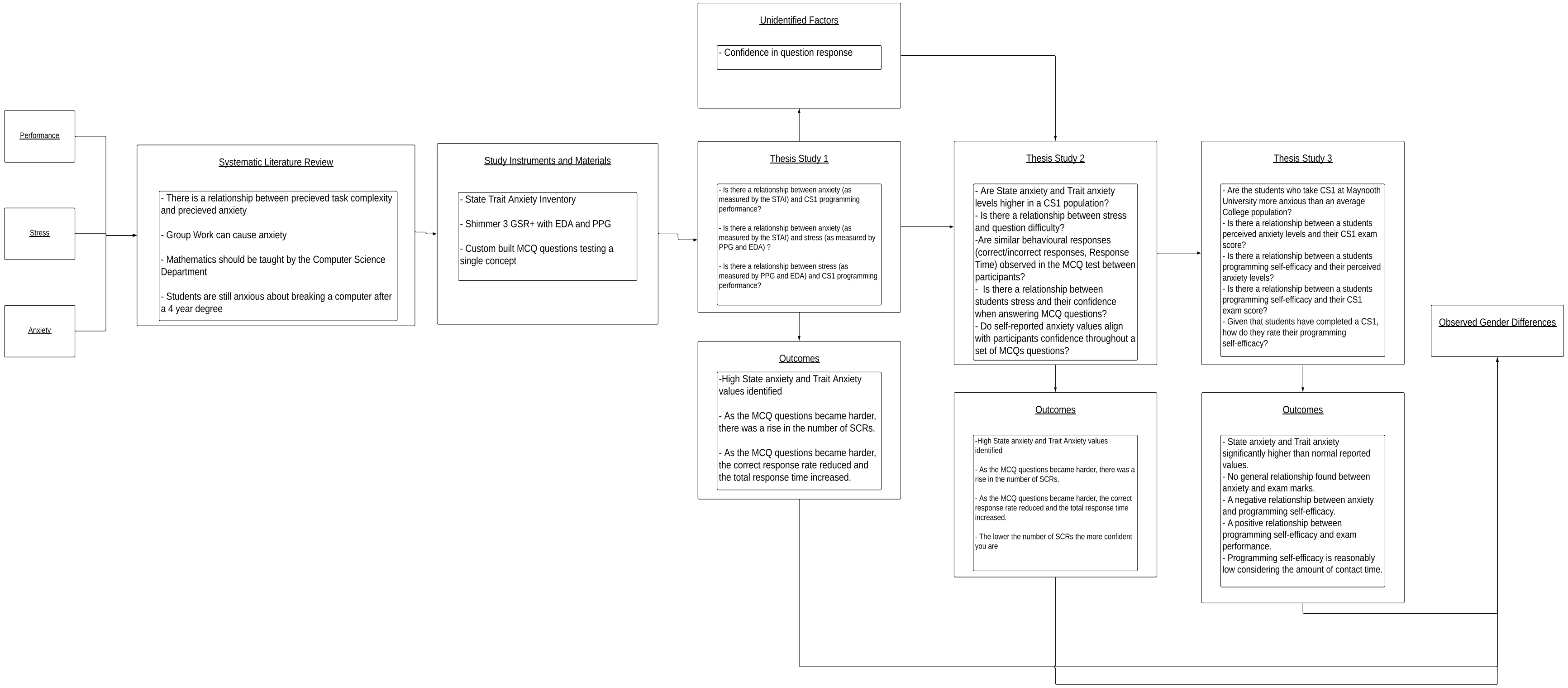


Figure 6.7: Update of the research methodology. All studies have been concluded and now gender differences will be investigated.

7

GENDER DIFFERENCES IN COMPUTER SCIENCE

In this chapter, gender differences are examined using the behavioural, psychological and stress data collected from the studies outlined in Chapter 4, Chapter 5 and Chapter 6. First, the motivation for this work is described, then the research questions are presented. The chapter concludes with results and discussion.

7.1 MOTIVATION AND BACKGROUND

Gender differences have existed in CS with a multitude of factors being identified. Gender differences are a much more systematic issue than the differences that are being described in this chapter [98]. A significantly lower number of female students choose to study CS compared to male students at third level education [14, 15, 49, 99]. In the 1980s, CS had one of the highest rates of gender balance in graduate programmes, but this has declined considerably in recent years [112]. At Maynooth University a gender imbalance is evident from first to final year. First-year CS modules and in particular CS1, tend to attract students on many different degree streams. By contrast final year CS is taken only by students who wish to graduate with a CS qualification. Final year CS is consequently male-dominated. In 2008 there was a gender split of 78% male and 22% female in final year CS at Maynooth University. However, by 2018 this split had increased to an 89% male to 11% female split indicating a clear gender imbalance. This is not unusual. The Higher Education Authority in Ireland released a report in 2016 documenting a country-wide gender difference in the fields of CS and Maths, with a split of 81% male and 19% female [49].

Female perception of STEM can be negative with views that Computer Science is a "nerdy" and male-dominated subject being common. Several reasons have been cited for this including CS majors being deficient in interpersonal skills and male tutors/educators displaying a superiority complex [15]. Several studies have interviewed women to better understand what it is like to be a woman in Computer Science. Of particular note respondents indicated that they couldn't see the point of coding and that they preferred to code alone at home and not in a lab as they did not feel like they belonged there. [79]. In another study, female respondent's indicated that they felt uncomfortable with assistance given to them by male-only tutors [94].

Furthermore female students display significantly lower confidence and programming self-efficacy in computer science when compared to males [7, 14, 64, 86, 99]. This is concerning as programming self-efficacy is significantly correlated to success in CS1 [84].

A recent large scale study, involving 690 students across 11 different institutions, examined perceived self-efficacy and test anxiety during a programming exam [84]. Findings from this study indicated significant differences in the self-efficacy and test anxiety of genders within CS1. An interesting finding from the study was that males tended to outperform females at the early stage of CS1. However, at the later stages of CS1 females tended to outperform males. Quille et al. [84] suggested that this difference may be caused by females having lower programming self-efficacy than males. The study also found that females have greater test anxiety and that this may affect performance [84].

Further evidence from around the world shows similar trends. In the USA, the percentage of females pursuing a degree in CS has gone from 40% in 2000-2001 to 26% in 2008-2009. The percentage of women receiving degrees in Computer Science is even lower, with a 28% completion rate in 2000 compared to 17.7% in 2008 [94]. This number further decreased in 2011 with the release of the Computer Research Association report stating that less than 12% of CS degrees were awarded to women [86]. The U.S. Department of Commerce, Economics and Statistics Administration released a report outlining that only 27% of the workforce in CS and math were women [8]. More recently, in 2017, the U.S. Department of Labour reported that only 25.5% of people working in the Computer Science and IT field were women [59]. Similar trends were noted in the UK with the WISE Campaign citing only 14% of the ICT workforce are women [113].

Over the years many initiatives have tried to address the downward trend in female participation in STEM. Of note, the European Union (EU) funded one of the largest recent initiatives which aimed to encourage females to participate in STEM subjects through various activities across many EU countries. A website entitled: "Science: It's a girls thing!" aimed at teenage girls aged 13-18 was developed and included information on careers within the STEM fields and a quiz to discover their "inner researcher". Accompanying the website was a video which depicted women "scientists" conducting work in stiletto heels. This video was referred to as offensive and after criticism was removed [99]. Other initiatives such as *Girls Who Code*, *SciGirls* and *GirlsInc* have been created. Indicative reports suggest that they are improving the uptake of females in Computer Science, however, no formal studies have been conducted.

Chapter 2 of this thesis outlined an extensive systematic literature review of the role of anxiety in CS with an emphasis on learning to program [70]. Findings from this review suggest that students are anxious when learning to program resulting from a multitude of factors, including, task complexity, the modality of programming assessments and general anxiety of using a computer. Of particular note from the systematic review was that very few studies have examined gender differences related to identified anxiety. New studies that attempt to understand and address gender imbalances in CS are crucial. As a final contribution to this thesis, a study focused on gender was carried out using the behavioural, stress and psychological data captured in the earlier chapters and is documented in the following sections.

7.2 RESEARCH QUESTIONS

Given the observations that were made throughout Chapters 5, 6 and 7, the following research questions were proposed:

GD-RQ1 : Are there differences in the stress signals (Electrodermal Activity and Heart Rate Variability) between male and female students during an MCQ test in a controlled lab setting?

GD-RQ2 : Are there differences in State and Trait anxiety between male and female students in CS1?

GD-RQ3 : Are there differences in the behavioural activity (correct responses & response time) between male and female students during an MCQ test in a controlled lab setting?

GD-RQ4 : Are there differences in programming self-efficacy between male and female students in CS1?

7.3 ARE THERE DIFFERENCES IN THE STRESS SIGNALS (ELECTRO-DERMAL ACTIVITY, HEART RATE VARIABILITY) BETWEEN MALE AND FEMALE STUDENTS DURING AN MCQ IN A CONTROLLED LAB SETTING?

This section presents the gender differences found in both the Electrodermal Activity and the PPG data during Study 1 and Study 2.

7.3.1 ANALYSIS

Electrodermal Activity

Examining the gender differences in Electrodermal Activity during an MCQ test, the first step was to compare the number of SCRs across genders for both Study 1 and Study 2. Table 7.1 shows the SCRs broken down by gender for both studies.

Table 7.1: Average number of SCRs across the studies and each question difficulty band categorised by gender

	Study 1			Study 2		
	M (N=28)	F (N=11)	*p*	M (N=24)	F (N=10)	*p*
Total	16.64	8.72	0.056	40.375	26	0.068
Easy	2.6	1.16	**0.044**	3.875	1	**0.0015**
Medium	6.03	2.25	**0.018**	6.95	3.6	**0.0331**
Hard	9.96	5.33	0.11	10.2	7.2	0.147

As can be seen in Table 7.1 there is a significant gender difference in the number of SCRs over both Study 1 and Study 2. Male participants exhibited almost double the number of SCRs throughout the studies compared to female participants. This suggests that males are prone to sweat more in MCQ tests than female students. This is due to the levels of arousal which are experienced. The more stressed one is the more sweat is produced. This is a novel finding and one that has not been explored before in the area of stress signals in a Computer Science exam.

A Welch's t-test was used to compare the average of the SCRs across the genders. When examining the comparative levels (Easy, Medium and Hard) within the genders across the studies, SCR levels raised nearly 3 times from Study 1 to Study 2. In both studies, there is a strong numerical difference between the genders and the difference is close to significant in both studies with p-values of $p = 0.056$ (Study 1) and $p = 0.068$ (Study 2). This suggests that the differences observed here are not by chance. The numerical difference between the studies (Study 1 - 16.64 (male) & 8.72 (female), Study 2 - 40.375 (male) & 26 (female) here is interesting and could be due to the Study 2 cohort being more stressed than the Study 1 cohort.

To further examine the gender difference in EDA, the number of SCRs over the difficulty bands was reviewed. Table 7.1 provides the number of SCRs across question difficulty bands. Similar to SCRs int the "Total Row", male participants tended to be significantly more stressed throughout all difficulty bands of the programming comprehension questions suggesting that males may be becoming more stressed than female students. Further to this, observing the increase in SCRs across the difficulty bands, the number of female SCRs increased the most between Medium and Hard (approximately 100% increase) whereas male SCRs increased the most between Easy and Medium (approximately 100% increase) indicating that the female participants were potentially more stressed by the harder questions whereas the males were potentially more stressed by the medium questions. These novel differences are important with potentially significant impacts. Should the use of wearable technology become more prevalent in classroom situations, knowing that males exhibit higher levels of SCRs when compared to females is important as interventions can be deployed in a somewhat targeted sense.

To examine if the differences between the genders were significant and to avoid Type 1 error[1], the effect size between the genders was investigated. Table 7.2 shows the Cohen d values.

Table 7.2: Cohen d values of the differences between the genders between EDA values

	Study 1	Study 2
Easy	0.5	0.97
Medium	0.59	0.6
Hard	0.35	0.36

Examining Table 7.2, the effect size of the gender differences of the Easy questions in Study 2 is considered large. This means that the difference between the genders is large and significant. All other differences have a medium to small effect. The larger the Cohen d value, the larger percentage of non-overlap between the groups meaning that the difference between the groups is more significant.

Photoplethysmography

Examining differences in PPG data during an MCQ test, the first step was to compare the Heart Rate of the two halves and the RMSSD of the two halves of both Study 1 and Study 2. The decision was made to analyse over halves to allow for reliable HRV measures to be captured. Table 7.3 presents the gender difference in PPG data on average over the course of the studies.

1 A Type 1 error is an error which leads you to reject a true null hypothesis

Table 7.3: Average ln(RMSSD) and Heart Rate values for each study half categorised by gender

	Study 1			Study 2		
	M (28)	F(11)	p	M(24)	F(10)	p
Half 1 ln (RMSSD)	4.892	5.252	0.089	4.196	3.872	0.06
Half 2 ln (RMSSD)	4.913	5.221	0.118	3.896	3.641	0.144
Half 1 Heart Rate (BPM)	89.1	80.90	0.1	91.87	96.03	**0.044**
Half 2 Heart Rate (BPM)	89.05	81.77	0.12	88.08	94.70	**0.013**

Examining Table 7.3, a gender difference can be seen in both Study 1 and Study 2 across all variables. Examining both the Half 1 and Half 2 of ln(RMSSD), males have lower values compared to females. The opposite can be seen in Study 2, with females showing lower ln(RMSSD) scores when compared to males. Table 7.3 also shows the significance values when compared across the genders over both studies with values which indicate a close to significant difference ($p = 0.089$ and $p = 0.06$) in Half 1 ln (RMSSD) values.Table 7.3 also shows the significance values when compared across the genders over both studies with values which indicate a close to significant difference ($p = 0.089$ and $p = 0.06$) in Half 1 ln (RMSSD) values.

Examining the Heart Rate in Study 1, in both halves, male participants showed higher heart rates, however, neither halves showed any significant differences between the genders as indicated by the non-significant p-values ($p = 0.1$ and $p = 0.12$).This suggests that there is no gender difference in heart rates. There is a significant observation to be made with regards to the PPG data in terms of gender. The PPG values in Study 2 are not consistent with Study 1 and therefore no conclusion can be drawn with respect to gender differences in PPG values.

7.3.2 DISCUSSION

While no studies have investigated the difference in stress signals between genders in an MCQ test, other studies in different disciplines (public speaking, psychological tests and responses to music) have found that females have higher SCRs [19, 88] or have found no discernible difference [51]. Although the settings/focus are very different this is interesting to note, and further study is justified. Overall it appears that male students are more stressed when answering MCQ type questions when compared to female students. In Section 7.5.1.2, evidence that females are faster and more accurate when responding to the MCQ questions will be provided. This could be an indication that, in general, for male participants, the situation was too stressful for them to perform optimally.

In Study 1, male participants have a lower PPG values than their female counterparts and this is seen in Table 7.3. However, examining Study 2, the opposite is found. This finding is not common and somewhat goes against what is reported in the literature where males tend to have lower PPG scores than females [110]. This inconsistency could be due to poor data reliability from the PPG. This could be caused by the placement of the sensor at the extremity of the finger and with the likelihood of movement, the data could become unreliable. Also, when comparing normal figures found in `HRVCourse.com`, PPG levels for this age group are typically significantly lower, approximately 24% lower than what was observed in the study. The observed increase in PPG scores could be attributed to test-anxiety. This, for now, is a working theory. This project could not contribute a meaning explanation for the difference in PPG scores.

Examining the EDA signals on both studies, male participants exhibited higher EDA spikes than female participants. A large-scale recent study found that female students outperformed male students at the latter stages of CS1 [84]. At the time this study was conducted, students had completed between 8–10 weeks of a 12 week CS1 module. Given that female students perform better in the latter stages of CS1, and that they were more accurate and faster when responding to the MCQ questions (discussed in Section 7.5.1.2), females could ultimately be more confident in their ability and may not be getting stressed or aroused when presented with the programming questions and thus produce fewer SCRs than male students. Based on the analysis, EDA provides a clearer indication that the participant is emotionally aroused and so is potentially more stressed in general studies.

7.4 ARE THERE DIFFERENCES IN STATE ANXIETY AND TRAIT ANXIETY BETWEEN MALE AND FEMALE STUDENTS IN CS1?

7.4.1 ANALYSIS

The following sections will present the analysis of the data required to inform the research question.

State Anxiety and Trait Anxiety

Table 7.4 presents the State anxiety and Trait anxiety data gathered from Study 1 (Chapter 5), Study 2 (Chapter 6) and Study 3 (Chapter 7). Examining the average State anxiety and Trait anxiety values, it can be seen there is no significant difference between male and female participants within either Study 1 and Study 2. Both male and female State anxiety and Trait anxiety averages appear similar (within 4 points) within both Studies 1 and 2. This finding is consistent with the normative averages in the STAI averages for college students [102]. Examining State anxiety and Trait anxiety values across the three studies, differences in values were noted. There was a rise in both State anxiety and Trait anxiety in the male participants across Study 1, Study 2 and the First Year survey. This would suggest that the population participating in the studies, year on year were more anxious. In general, the female cohorts, while experiencing a drop in both State anxiety and Trait anxiety in Study 2, display the same trend and appear more anxious year on year.

Table 7.4: Average State anxiety and Trait anxiety values for male and female participants and associated p-values. N values in brackets ()

	State Anxiety			Trait Anxiety		
	Male	Female	p	Male	Female	p
NV[2]	*36.47*	*38.76*	-	*38.30*	*40.40*	-
Study 1 [3]	38.96(30)	40.25(12)	0.44	52(30)	56.83(12)	0.32
Study 2 [4]	49.82(28)	52.83(12)	0.155	54.32 (28)	55.5 (12)	0.38
Study 3 [5]	52.51 (145)	60.67 (37)	**0.0009**	54.18 (145)	61.29 (37)	**0.003**

Interestingly in Table 7.4, the State anxiety and Trait anxiety data from Study 3 contains a significant gender difference. Within the State anxiety and Trait anxiety data from Study 3, females exhibited significantly higher State anxiety and Trait anxiety with p-values of $p = 0.0009$ and $p = 0.003$. Examining further, the effect size of the differences were calculated with Table 7.5 showing the Cohen d values. The results from Study 3 show that there is a medium to large effect with d-values of $d = 0.65$ and $d = 0.54$ which supports that the differences found in Table 7.4 are not by chance.

2 Normative Values
3 Chapter 4
4 Chapter 5
5 Chapter 6

Table 7.5: Cohen d values for State and Trait values

	State	Trait
Study 1	0.051	0.160
Study 2	0.353	0.100
Study 3	0.650	0.546

This finding was unexpected however as there was no indication of a strong effect in Study 1 and Study 2 considering Study 3 was conducted with a similar cohort. This could be due to the lower numbers in both Study 1 and Study 2. However, in Study 1 and Study 2 both State and Trait values for male and females are higher than the normal college values for State anxiety and Trait anxiety. Previously, data captured from a large multi-institutional international study found that females have greater test anxiety than males [84]. In all three studies, females had higher State anxiety and Trait anxiety when compared to males, however, is normal [102].

7.4.2 DISCUSSION

After examining the data presented in Table 7.4 and Table 7.5, there is a strong gender difference observed in both State anxiety and Trait anxiety in Study 3. This study was conducted with 182 participants. Examining Study 1 and Study 2, while the difference may not have a strong effect size, there was an obvious gender difference in State anxiety and Trait anxiety. This finding can now be utilised to ensure appropriate interventions are gender specific rather than a general intervention.

7.5 ARE THERE DIFFERENCES IN THE BEHAVIOURAL ACTIVITY (CORRECT RESPONSES & RESPONSE TIME) BETWEEN MALE AND FEMALE STUDENTS DURING AN MCQ TEST IN A CONTROLLED LAB SETTING?

7.5.1 ANALYSIS

Correct Responses

The gender differences in the correct and incorrect responses from Study 1 and Study 2 were examined. When analysing this data, the responses were binary i.e. either correct or incorrect. Table 7.6 illustrates the percentage of correct and incorrect answers with the number of correct responses per question broken down by gender for both Study 1 and Study 2.

Examining Table 7.6 there is little differences between the percentages of correct answers between the genders in Study 1. Interestingly, there was a difference in Question 5 where all females were correct compared to only 83% males were correct. Examining Study 2, differences were observed. Question 6 shows a sizable difference between the genders with males outperforming females. However, on Question 9, female students outperformed male students. While there are gender differences in the responses between male and female students, none are significant. However, if this study was conducted with a larger cohort, perhaps significant gender differences might be uncovered.

Response Times

The response times of the correct responses were examined next. Table 7.7 documents response times along with significant values per question for both Study 1 and Study 2.

Table 7.6: Percentage of correct answers, average time taken in seconds (s) to respond correctly to each question, grouped by gender

	Study 1		Study 2	
	Male (30)	Female (12)	Male (28)	Female (12)
Question 1	100% (30)	100% (12)	96.43% (27)	100% (12)
Question 2	96.7% (29)	91.7% (11)	96.43% (27)	100% (12)
Question 3	70% (21)	66.7% (8)	96.43% (27)	75% (9)
Question 4	100% (30)	100% (12)	100% (28)	100% (12)
Question 5	83.3% (25)	100% (12)	100% (28)	91.67% (11)
Question 6	56.7% (17)	50% (6)	85.71% (24)	58.33% (7)
Question 7	23.3% (7)	33.3% (4)	53.57% (15)	33.33% (4)
Question 8	56.7% (17)	58.3% (7)	53.57% (15)	58.33% (7)
Question 9	46.7% (14)	33.3% (4)	53.57% (15)	75% (9)

A Welch's t-test was used to compare the data at a 95% significance level. Breaking the response times down on a per question level, females, on average, responded faster than males and in some cases were significantly faster in both studies. All participants had to respond to each question. There was no time limit enforced on answering the questions. Although a significant difference was found on five out of the nine questions in Study 1 and two of the nine were significantly different in Study 2, only Question 5 showed a significant difference in both studies. This is likely a random finding as there is no evidence to support that females understand nested concepts (nested if statements or nested while loops) more than males.

In general females were significantly faster and more accurate than males when completing the programming comprehension questions at this stage of the module; this is in line with previous research [84].

Table 7.7: Percentage of correct answers, average time taken in seconds (s) to respond correctly to each question, grouped by gender

	Study 1					Study 2				
	% Correct		Response Time (s)			% Correct		Response Time (s)		
	Male(30)	Female(12)	Male(30)	Female(12)	p-value	Male(28)	Female (12)	Male(28)	Female(12)	p-value
Q1	100%	100%	14.98	13.79	0.307114	96.43%	100%	14.03	12.18	0.2199
Q2	96.67%	91.67%	14.78	11.63	**0.0265**	96.43%	91.67%	14.06	13.45	0.3638
Q3	70%	66.67%	23.39	15.49	**0.0398**	96.43%	66.67%	16.59	17.1	0.4292
Q4	100%	100%	22.05	16.15	**0.01735**	100%	100%	20.8	18.72	0.2113
Q5	83.33%	100%	26.65	18.68	**0.002928**	100%	100%	31.08	18.59	**0.00005**
Q6	56.67%	50%	50.85	59.49	0.273882	85.71%	50%	61.8	48.87	0.1032
Q7	23.33%	33.33%	110.81	69.97	0.1584	53.57%	33.33%	110.1	82.81	0.1874
Q8	56.67%	58.33%	41.03	36.05	0.28744	53.57%	58.33%	35.55	39.26	0.2569
Q9	46.67%	33.33%	79.1	45.01	0.0549	53.57%	33.33%	85.81	61	**0.0382**

7.5.2 Discussion

Analysis of both correct responses and response times indicated the presence of gender differences. Examining Table 7.6 there are slight differences in the number of correct responses between the genders in Study 1. This difference was not observed in Study 2 however. While there are differences, these require further investigation. This investigation could involve a longitudinal study involving a larger sample size with a stronger gender balance. Given that the cohorts were similar across the studies, it can be concluded that there is no significant gender difference in the number of correct responses.

Examining the correct response times presented in Table 7.7 females were, in general, faster and more accurate than their male counterparts. This is a unique finding and one that is not found in the literature. While there appear to be gender differences, the sample size of each study is too small to draw strong conclusions.

While gender differences were observed they should be interpreted with caution as there were comparatively small numbers of male and female students. Further testing on a larger cohort with equal gender balances would allow for a more equal comparison and stronger conclusions.

7.6 Are there differences in programming self-efficacy between male and female students in CS1?

7.6.1 Analysis

To investigate gender differences in programming self-efficacy, two data sets were used; the study data from Study 2 and Study 3. Table 7.8

shows the average programming self-efficacy values by gender for each data set.

Table 7.8: Programming Self-Efficacy values (10-40 points) for male and female participants and associated p-values

	Male	Female	p-Value
Study 2	29.74	28	0.255
Study 3	27.4	24.54	**0.03**

7.6.2 discussion

In Table 7.8, examining the results from Study 2 it can be seen there is a numerical difference between Male and Female students in programming self-efficacy, however, there is no statistical difference. It is known that there is a gender difference in programming self-efficacy and this is seen in this study. From the results in Study 3, there is a clear gender difference with males more confident in their programming abilities than their female counterparts. This difference is statistically significant as well. In general, there are gender differences in programming self-efficacy in Computer Science. Given that two independent cohorts are exhibiting similar results, these results are generalisable.

7.7 summary of chapter

Throughout this thesis, gender differences have been noted. Examining the stress signals, EDA and PPG, gender differences were found across both Study 1 and Study 2. While a small number of studies have noted gender differences, it appears to be the first time these differences have been noted in an MCQ test for CS students.

In addition to the stress signals, behavioural differences were also noted. It was shown that female students were, in general, faster and more accurate when responding to multiple-choice programming questions. This appears to be another novel finding and is worthy of further investigation. Perhaps repeated longitudinal testing would allow for a deeper analysis of these differences. The study would use standard tasks with a high test-retest reliability with a large gender-balanced cohort and PPG and ECG data would be collected. This would provide an opportunity to a) validate the PPG data off an ECG machine and b) investigate the gender difference further.

While it may have been previously known that there is a gender difference in both State anxiety and Trait anxiety, these differences were noted across Study 1, Study 2 and Study 3. In addition, significant gender differences were noted in programming self-efficacy.

This project uncovered findings on gender differences in relation to stress signals (EDA of note), and State and Trait anxiety in Computer Science. With respect to programming self-efficacy, differences have previously been found between the genders and further evidence was found in this project in support of the findings.

7.8 thesis update

This chapter set out to inform Objective 4 in Section 1.2: **To review the data obtained throughout the project, to identify analyse and identify any gender differences.**. While there have been many instances shown that a gender difference exists across the broad area of Computer Science and even broader the area of IT, no such studies have investigated the gender differences within Computer Science programming. This project has shown that there are fine-grained differences such as the differences in responses to programming questions and the differences in stress signals during an MCQ test. These differences can now be taken into account when developing systems for interventions for students. Figure 7.1 now shows an updated logical flow of the project.

Chapter 8 discusses the conclusions of this project.

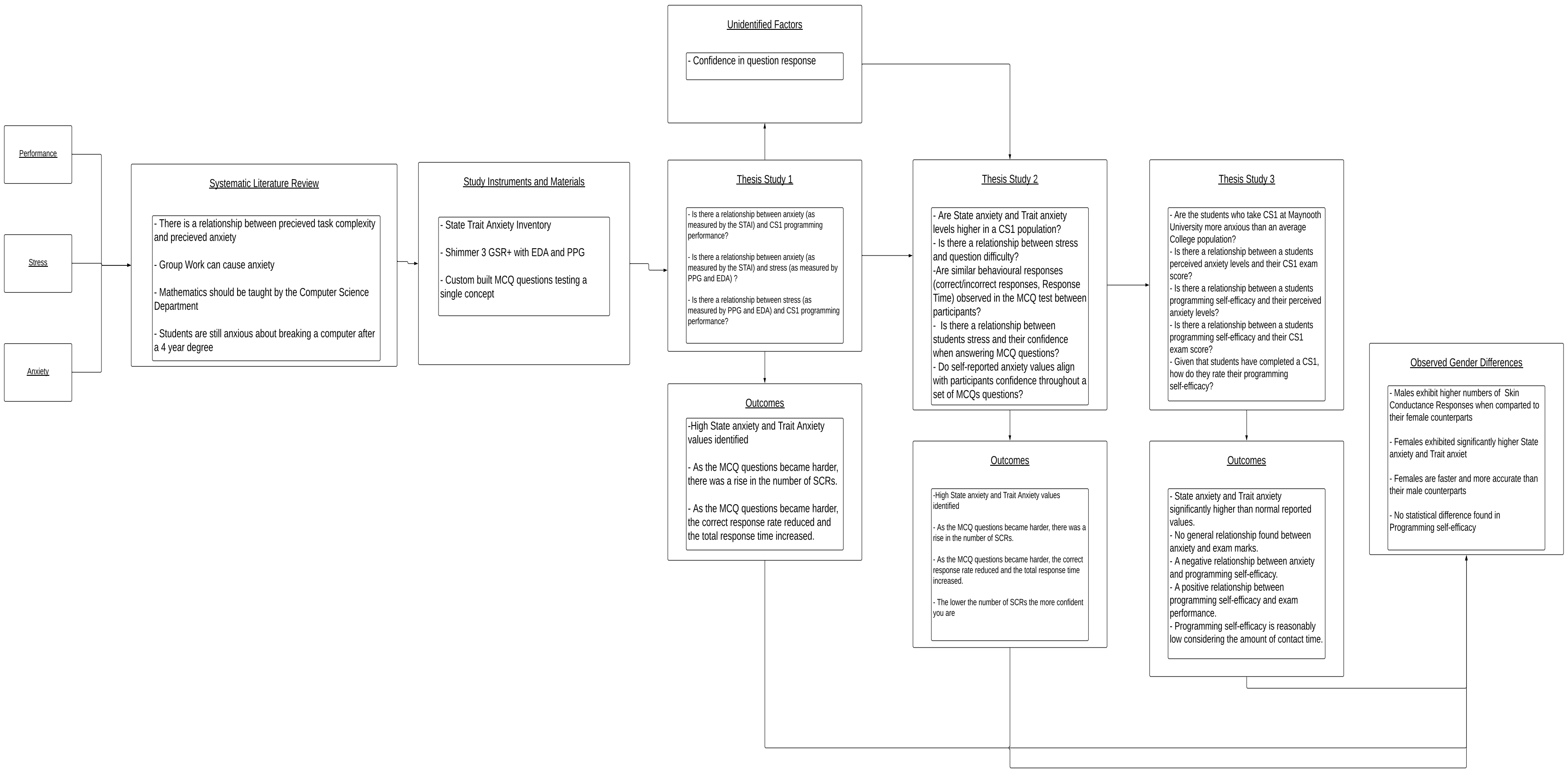


Figure 7.1: All objectives have been informed and are outlined.

8 CONCLUSIONS AND FUTURE WORK

8.1 CONCLUSIONS

The mental health of third level students is potentially at an all-time low. Reports such as the *My World Survey*, the *My World Survey 2* and the *Union of students of Ireland Report* indicate that a large amount of third level students in Ireland are suffering from mental health issues. For students, mental well-being is associated with affective learning, and their ability to navigate through third level education, coping with the challenges and stresses of student life. As such, this project attempted to investigate the effects that mental health factors such as stress and anxiety have on programming performance within a first-year CS population by designing experimental studies to inform the following objectives:

- To gather evidence on the relationship between anxiety and performance.
- To examine the relationship between stress and performance.
- To examine the relationship between anxiety and stress.
- To review the data obtained throughout the project, to identify analyse and identify any gender differences.

As this project progressed, the research model was updated at the end of each chapter to allow the reader to understand how the chapters interact with each other. Each chapter built upon the previous chapters and all results from the studies and chapters inform the objectives laid out in Chapter 1. Figure 7.1 shows the complete flow of the project.

The following sections will focus on the findings of each of the project objectives.

8.1.1 To gather evidence on the relationship between anxiety and performance.

From the outset, an attempt was made to relate anxiety and performance. No generalised relationship could be found between measures of anxiety and performance (TE1-RQ1, TE2-RQ2, Chapter 7 Research Questions) and as such, models such as the Individual Zone of Optimum Functioning might be able to explain this individualised relationship. This will require further investigation as will be outlined in Section 8.2.

8.1.2 To examine the relationship between stress and performance.

Through using EDA and PPG, this project was capable of determining stress levels while completing an MCQ test (TE1-RQ3, VAL-RQ2 and TE2-RQ1). There were three levels of difficulty within the test, Easy, Medium and Hard which allowed for the comparison of stress across the difficulty bands. Across both Study 1 and Study 2, it was observed that as the question difficulty increased so too did the number of SCRs. While PPG may be a reasonable measure of stress, it was noted that conflicting data was gathered over the course of the studies and so no reasonable conclusion can be drawn.

8.1.3 to examine the relationship between anxiety and stress.

Given that in Study 1 and Study 2 (Chapter 4, Chapter 5) incorporated the use of the State Trait Anxiety Inventory and stress (as measured by EDA and PPG), an attempt was made to draw a relationship between the factors (TE1-RQ2). Both bivariate and multivariate analysis were employed to aide the investigation. While there was no strong relationship found, there was a weak relationship between State anxiety, Trait anxiety and Electrodermal Activity. A relationship could not be found between State anxiety, Trait anxiety and PPG signals. This is due to the conflicting data observed between Study 1 and Study 2. Overall, the use of Electrodermal Activity to determine anxiety is a possibility while further research is required with the use of Photoplethysmogram.

8.1.4 to review the data obtained throughout the project, to identify analyse and identify any gender differences.

Throughout this project there has been an observation of a strong gender difference across all stress, psychological and behavioural aspects relating to an MCQ test. Examining the stress signals, EDA and PPG, gender differences were found across both Study 1 and Study 2 and in the First Year Study. While there were known gender differences in stress signals, this project has noted that there are differences in a CS population during an MCQ test. Also, it was shown that female students were, in general, faster and more accurate when responding to MCQ tests than their male counterparts.

Examining the stress signals, and State anxiety and Trait anxiety, the gender differences that are already well established and this project has succeed in supporting the literature to this extent. Similarly, differences in Programming self-efficacy were found and validated and align with differences observed in the literature.

8.1.5 FURTHER CONTRIBUTIONS

As a starting point to this project, a detailed systematic review of the existing literature was carried out. This systematic review compiled all of the literature relating to anxiety when learning to program and found that students who take Computer Science as part of their degree experience anxiety. Throughout this review, longitudinal studies were identified which investigated students anxiety levels when receiving programming errors and found that anxiety levels did increase. Other areas of programming were identified as causes of anxiety such as debugging and difficult programming tasks. This review provided much-needed direction in the area of anxiety when learning to program. No such review had previously been conducted.

Given that each study described (Chapter 4, Chapter 5 and Chapter 6) incorporated the use of the State Trait Anxiety Inventory, there was a unique opportunity to validate all of the findings relating to the STAI. Initially, in Study 1, higher Trait anxiety levels were noted in female participants compared to the normal population (Section 4.3.1.1). In Study 2, higher State anxiety levels and higher Trait anxiety levels were noted in both male and female participants compared to normal college students. This validated our finding from Study 1, but posed a question, *"Are CS students more anxious that the normal college population?"*. A large scale study was conducted with the first-year CS cohort and the finding observed in Study 1 and Study 2 was re-validated, with significantly higher State anxiety and Trait anxiety levels noted. This is a novel contribution of this project and this finding provides the foundations for future research into anxiety in Computer Science.

The relationship between anxiety and programming self-efficacy (detailed in Section 6.5) is a novel finding. While it may seem intuitive, this finding is important. If, as educators, a way can be found to reduce anxiety among student, this could potentially increase the levels of programming self-efficacy within the CS population.

8.2 future work

This project has uncovered many possible avenues for future research. In this section, future possible directions are discussed.

8.2.1 anxiety and performance in computer science

Looking into the area of sports psychology, the IZOF model, a model which relates anxiety and performance on an individual level, has been successful with high-performance athletes improving their success in their fields. Examining the relationships, or the lack of relationship, between anxiety and performance it is clear that a model such as the IZOF might be optimal. An attempt should be made to map the model into the Computer Science space. This would involve a longitudinal study whereby students would rate their anxiety levels before multiple tests/challenges and reflect after on how they felt during the test/challenge. This would allow for an individualised relationship between anxiety and performance to be formed and so could be used to allow the student to perform optimally.

8.2.2 MENTAL HEALTH OF CS STUDENTS

As a follow on to this work, a large scale study across all CS students should be carried out. The finding that first-year CS students are more anxious is novel. The question has to be asked if this pattern of higher anxiety exists across all CS students. By doing a large scale study like this, two contributions are foreseen:

1. A re-validation of the findings presented in this project would be valuable. This sits in line with the *"Grand Challenges"* defined by the 2015 ITiCSE working group [52].

2. A validation of the STAI within the scope of CS. This would allow for other anxiety surveys which attempt to measure anxiety to be validated. This would be a valuable study and one which would serve a community far wider than just the Computer Science community.

8.2.3 USE OF PHYSIOLOGICAL SENSORS

Given that the relationship be EDA and question difficulty was so strong, the use of EDA in MCQ test situations should be explored further. Creating a platform whereby stress can be measured within the platform would allow for real-time interventions at a point where students begin to become emotionally aroused and likely stressed. By utilising the signals and employing the interventions, students could potentially become less stressed and perform better.

BIBLIOGRAPHY

[1] Amir Ahmad and Lipika Dey. "A k-mean clustering algorithm for mixed numeric and categorical data". In: *Data Knowledge Engineering* 63.2 (2007), pp. 503 –527. ISSN: 0169-023X. DOI: https://doi.org/10.1016/j.datak.2007.03.016. URL: http://www.sciencedirect.com/science/article/pii/S0169023X0700050X.

[2] John J B Allen and John P. Kline. "Frontal EEG asymmetry, emotion, and psychopathology: The first, and the next 25 years". In: *Biological Psychology* 67.1-2 (2004), pp. 1–5. ISSN: 03010511. DOI: 10.1016/j.biopsycho.2004.03.001.

[3] Dominik R Bach, Karl J Friston, and Raymond J Dolan. "Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations". In: *International Journal of Psychophysiology* 76.1 (2010), pp. 52–55.

[4] Mustafa Baloglu and Recep Kocak. "A multivariate investigation of the differences in mathematics anxiety". In: *Personality and Individual Differences* 40.7 (2006), pp. 1325–1335. ISSN: 01918869. DOI: 10.1016/j.paid.2005.10.009.

[5] Albert Bandura. "Social foundations of thought and action". In: *Englewood Cliffs, NJ* 1986 (1986).

[6] Yair Bar-Haim, Dominique Lamy, Lee Pergamin, Marian J Bakermans-Kranenburg, and Marinus H van IJzendoorn. "Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study." In: *Psychological bulletin* 133.1 (2007), pp. 1–24. ISSN: 0033-2909. DOI: 10.1037/0033-2909.133.1.1.

[7] Mustafa Baser. "Attitude, Gender and Achievement in Computer Programming." In: *Middle-Ease Journal of Scientific Research* 14.2 (2013), pp. 248–255.

[8] David N Beede, Tiffany A Julian, David Langdon, George McKittrick, Beethika Khan, and Mark E Doms. *Women in STEM: A gender gap to innovation*. 2011. URL: https://files.eric.ed.gov/fulltext/ED523766.pdf.

[9] Andrew Begel and Beth Simon. "Novice software developers, all over again". In: *Proceeding of the fourth international workshop on Computing education research*. Vol. 1. 425. New York, New York, USA: ACM Press, 2008, pp. 3–14. ISBN: 9781605582160. DOI: 10.1145/1404520.1404522. URL: http://portal.acm.org/citation.cfm?doid=1404520.1404522.

[10] Susan Bergin. "Statistical and Machine Learning Models to Predict Programming Performance". PhD thesis. 2006.

[11] Susan Bergin and Ronan Reilly. "Programming: Factors That Influence Success". In: *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education*. SIGCSE '05. New York, NY, USA: ACM, 2005, pp. 411–415. ISBN: 1-58113-997-7. DOI: 10.1145/1047344.1047480. URL: http://doi.acm.org/10.1145/1047344.1047480.

[12] Susan Bergin and Ronan Reilly. "Predicting introductory programming performance: A multi-institutional multivariate study". In: *Computer Science Education* 16.4 (2006), pp. 303–323. ISSN: 0899-3408. DOI: 10.1080/08993400600997096. URL: http://www.tandfonline.com/doi/abs/10.1080/08993400600997096.

[13] Susan Bergin and Ronan Reilly. "Predicting introductory programming performance: A multi-institutional multivariate study". In: *Computer Science Education* 16.4 (2006), pp. 303–323. DOI: 10.

1080/08993400600997096. eprint: doi.org/10.1080/08993400600997096. URL: doi.org/10.1080/08993400600997096.

[14] Sylvia Beyer, Michelle DeKeuster, Kathleen Walter, Michelle Colar, and Christina Holcomb. "Changes in CS Students' Sttitudes Towards CS over Time: An Examination of Gender Differences". In: *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education*. SIGCSE '05. New York, NY, USA: ACM, 2005, pp. 392–396. ISBN: 1-58113-997-7. DOI: 10.1145/1047344.1047475. URL: http://doi.acm.org/10.1145/1047344.1047475.

[15] Sylvia Beyer, Kristina Rynes, Julie Perrault, Kelly Hay, and Susan Haller. "Gender Differences in Computer Science Students". In: *SIGCSE Bull.* 35.1 ().

[16] Collin A. Blanchard, Holly A. Buff, Travis D. Cook, Raquel E. Dottle, Gideon B. Luck, Alani L. Peters, Virginia L. Pettit, Isaak Matthew Ramirez, and Jessica E. Wininger. "Print ("Good Luck!"): Measuring the Effect of Autogenerated Social Encouragement on Student Anxiety". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA '18. New York, NY, USA: ACM, 2018, SRC12:1–SRC12:6. ISBN: 978-1-4503-5621-3. DOI: 10.1145/3170427.3180287. URL: http://doi.acm.org/10.1145/3170427.3180287.

[17] John T Cacioppo, Louis G Tassinary, and Gary Berntson. *Handbook of psychophysiology*. Cambridge University Press, 2007.

[18] Yekta Said Can, Niaz Chalabianloo, Deniz Ekiz, and Cem Ersoy. "Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study". In: *Sensors* 19.8 (2019), p. 1849.

[19] Eduvigis Carrillo, Luis Moya-Albiol, Esperanza Gonzalez-Bono, Alicia Salvador, Jorge Ricarte, and Jesus Gomez-Amor. "Gen-

der differences in cardiovascular and electrodermal responses to public speaking task: the role of anxiety and mood states". In: *International Journal of Psychophysiology* 42.3 (2001), pp. 253 –264. ISSN: 0167-8760. DOI: https://doi.org/10.1016/S0167-8760(01)00147-7. URL: http://www.sciencedirect.com/science/article/pii/S0167876001001477.

[20] Sheeson E. Chang. "Computer anxiety and perception of task complexity in learning programming-related skills". In: *Computers in Human Behavior* 21.5 (2005), pp. 713–728. ISSN: 07475632. DOI: 10.1016/j.chb.2004.02.021. URL: http://linkinghub.elsevier.com/retrieve/pii/S074756320400038X.

[21] Kwang-Ho Choi, Junbeom Kim, O. Sang Kwon, Min Ji Kim, Yeon Hee Ryu, and Ji-Eun Park. "Is heart rate variability (HRV) an adequate tool for evaluating human emotions? - A focus on the use of the International Affective Picture System (IAPS)". In: *Psychiatry Research* 251 (2017), pp. 192 –196. ISSN: 0165-1781. DOI: https://doi.org/10.1016/j.psychres.2017.02.025. URL: http://www.sciencedirect.com/science/article/pii/S0165178116312550.

[22] Siew Lian Chua, Der-Thanq Chen, and Angela F.L. Wong. "Computer anxiety and its correlates: a meta-analysis". In: *Computers in Human Behavior* 15.5 (1999), pp. 609–623. ISSN: 07475632. DOI: 10.1016/S0747-5632(99)00039-4. URL: http://linkinghub.elsevier.com/retrieve/pii/S0747563299000394.

[23] J A Coan, J J Allen, and E Harmon-Jones. "Voluntary facial expression and hemispheric asymmetry over the frontal cortex." In: *Psychophysiology* 38.6 (2001), pp. 912–925. ISSN: 0048-5772. DOI: 10.1111/1469-8986.3860912.

[24] James A. Coan and John J B Allen. "Frontal EEG asymmetry as a moderator and mediator of emotion". In: *Biological Psychol-*

ogy 67.1-2 (2004), pp. 7–49. ISSN: 03010511. DOI: 10.1016/j.biopsycho.2004.03.002.

[25] Cornelia Connolly, Eamonn Murphy, and Sarah Moore. "Programming Anxiety Amongst Computing Students - A Key in the Retention Debate?" In: *IEEE Transactions on Education* 52.1 (2009), pp. 52–56. ISSN: 0018-9359. DOI: 10.1109/TE.2008.917193. URL: http://ieeexplore.ieee.org/document/4569871/.

[26] The Joint Task Force on Computing Curricula, Association for Computing Machinery, and IEEE-Computer Society. *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*. Ed. by ACM Computing Curricula Task Force. ACM, Inc, 2013. ISBN: 9781450323093. DOI: 10.1145/2534860. URL: http://dl.acm.org/citation.cfm?id=2534860.

[27] Donald A Curtis, Samuel L Lind, Christy K Boscardin, and Mark Dellinges. "Does student confidence on multiple-choice question assessments provide useful information?" In: *Medical Education* 47.6 (2013), pp. 578–584. DOI: 10.1111/medu.12147. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/medu.12147. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/medu.12147.

[28] Richard J Davidson, Paul Ekman, Clifford D Saron, Joseph A Senulis, and Wallace V Friesen. "Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology: I." In: *Journal of personality and social psychology* 58.2 (1990), p. 330.

[29] Robert Deloatch, Brian P Bailey, and Alex Kirlik. "Measuring Effects of Modality on Perceived Test Anxiety for Computer Programming Exams". In: *SIGCSE '16 Proceedings of the 47th ACM Technical Symposium on Computing Science Education* (2016), pp. 291–296. DOI: 10.1145/2839509.2844604.

[30] Robert Deloatch, Brian P. Bailey, Alex Kirlik, and Craig Zilles. "I Need Your Encouragement!: Requesting Supportive Comments on Social Media Reduces Test Anxiety". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. New York, NY, USA: ACM, 2017, pp. 736–747. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025709. URL: http://doi.acm.org/10.1145/3025453.3025709.

[31] Barbara A Dooley and Amanda Fitzgerald. *My world survey: National study of youth mental health in Ireland*. Headstrong and UCD School of Psychology, 2012.

[32] Barbara Dooley, cliodhna o'connor, Amanda Fitzgerald, and Aileen Oreilly. *My World Survey 2: The National Study of Youth Mental Health in Ireland*. Nov. 2019. ISBN: 978-1-5272-4922-6.

[33] E. Doyle, I. Stamouli, and M. Huggard. "Computer anxiety, self-efficacy, computer experience: an investigation throughout a computer science degree". In: *Proceedings Frontiers in Education 35th Annual Conference* (2005), pp. 3–7. ISSN: 15394565. DOI: 10.1109/FIE.2005.1612246.

[34] Paul Ekman, Maureen O Sullivan, Mark G Frank, Source Psychological Science, No May, Paul Ekman, Maureen O Sullivan, and Mark G Frank. "Association for Psychological Science, Sage Publications, Inc." In: 10.3 (2016), pp. 263–266.

[35] Katrina Falkner, Nickolas J.G. Falkner, and Rebecca Vivian. "Collaborative learning and anxiety". In: *Proceeding of the 44th ACM technical symposium on Computer science education*. New York, New York, USA: ACM Press, 2013, p. 227. ISBN: 9781450318686. DOI: 10.1145/2445196.2445268. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-84876216299{\&}partnerID=40{\&}md5=a9a02411aa004dcb93de8b92e6b472behttp://dl.acm.org/citation.cfm?doid=2445196.2445268.

[36] Ernst Fehr and KM Schmidt. "A Theory of Fairness,Competition and Cooperation". In: *Quarterly journal of Economics* 114.August (1999), pp. 817–868. ISSN: 0033-5533. DOI: 10.1162/003355399556151. arXiv: arXiv:1011.1669v3. URL: http://www.jstor.org/stable/2586885.

[37] James B. Fenwick, Cindy Norris, Andrew R. Dalton, and William Kreahling. "24/7 lectures as an exam review technique". In: *Proceedings of the 41st ACM technical symposium on Computer science education*. New York, New York, USA: ACM Press, 2010, p. 455. ISBN: 9781450300063. DOI: 10.1145/1734263.1734416. URL: http://dl.acm.org/citation.cfm?id=1734416http://portal.acm.org/citation.cfm?doid=1734263.1734416.

[38] George Fink. "Stress: definition and history". In: *Stress science: neuroendocrinology* (2010), pp. 3–9.

[39] William Fone. "Using a familiar package to demonstrate a difficult concept". In: *ACM SIGCSE Bulletin* 33.3 (2001), pp. 165–168. ISSN: 00978418. DOI: 10.1145/507758.377675. URL: http://portal.acm.org/citation.cfm?doid=507758.377675.

[40] Charlotte Gerritsen, Jeroen de Man, and Justin van der Meij. "Physiological and Subjective Response to Injustice: The Effects of Unjust Evaluations on Physiological Responses and Subjective Experiences". In: *International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. Vol. 3. IEEE, 2013, pp. 179–182. ISBN: 978-0-7695-5145-6. DOI: 10.1109/WI-IAT.2013.178. URL: http://ieeexplore.ieee.org/document/6690723/.

[41] Yori Gidron. "Trait Anxiety". In: *Encyclopedia of Behavioral Medicine*. Ed. by Marc D. Gellman and J. Rick Turner. New York, NY: Springer New York, 2013, pp. 1989–1989. ISBN: 978-1-4419-1005-9. DOI: 10.1007/978-1-4419-1005-9_1539. URL: https://doi.org/10.1007/978-1-4419-1005-9_1539.

[42] Margaret Grigg and Shekhar Saxena. "Promoting mental health nursing research in low and middle income countries". In: *International Nursing Review* 51.4 (2004), pp. 194–195. ISSN: 0020-8132. DOI: 10.1111/j.1466-7657.2004.00268.x. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2408399{\&}tool=pmcentrez{\&}rendertype=abstracthttp://doi.wiley.com/10.1111/j.1466-7657.2004.00268.x.

[43] Gina M Grimshaw, Joshua J Foster, and Paul M Corballis. "Frontal and parietal EEG asymmetries interact to predict attentional bias to threat". In: *BRAIN AND COGNITION* 90 (2014). DOI: 10.1016/j.bandc.2014.06.008.

[44] Jan L. Guynes. "Impact of system response time on state anxiety". In: *Communications of the ACM* 31.3 (1988), pp. 342–347. ISSN: 00010782. DOI: 10.1145/42392.42402. URL: http://dl.acm.org/citation.cfm?id=42402http://portal.acm.org/citation.cfm?doid=42392.42402.

[45] Dirk Hagemann, Ewald Naumann, Julian F Thayer, and Dieter Bartussek. "Does resting electroencephalograph asymmetry reflect a trait? an application of latent state-trait theory." In: *Journal of personality and social psychology* 82.4 (2002), pp. 619–641. ISSN: 0022-3514. DOI: 10.1037/0022-3514.82.4.619.

[46] John Hamer, Helen C. Purchase, Paul Denny, and Andrew Luxton-Reilly. "Quality of peer assessment in CS1". In: *Proceedings of the fifth international workshop on Computing education research workshop*. New York, New York, USA: ACM Press, 2009, p. 27. ISBN: 9781605586151. DOI: 10.1145/1584322.1584327. URL: http://portal.acm.org/citation.cfm?doid=1584322.1584327.

[47] Eddie Harmon-Jones. "Clarifying the emotive functions of asymmetrical frontal cortical activity". In: *Psychophysiology* 40.6 (2003), pp. 838–848. ISSN: 00485772. DOI: 10.1111/1469-8986.00121.

[48] Eddie Harmon-Jones, Philip A. Gable, and Carly K. Peterson. "The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update". In: *Biological Psychology* 84.3 (2010), pp. 451–462. ISSN: 03010511. DOI: 10.1016/j.biopsycho.2009.08.010. URL: http://dx.doi.org/10.1016/j.biopsycho.2009.08.010.

[49] David Harmon and Stephen Erskine. *Eurostudent Survey VI*.

[50] Robert K. Heinssen, Carol R. Glass, and Luanne A. Knight. "Assessing computer anxiety: Development and validation of the Computer Anxiety Rating Scale". In: *Computers in Human Behavior* 3.1 (1987), pp. 49 –59. ISSN: 0747-5632. DOI: https://doi.org/10.1016/0747-5632(87)90010-0. URL: http://www.sciencedirect.com/science/article/pii/0747563287900100.

[51] W. Huang and R. Benjamin Knapp. "An exploratory study of population differences based on massive database of physiological responses to music". In: *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. 2017, pp. 524–530. DOI: 10.1109/ACII.2017.8273649.

[52] Petri Ihantola et al. "Educational Data Mining and Learning Analytics in Programming: Literature Review and Case Studies". In: *Proceedings of the 2015 ITiCSE on Working Group Reports*. ITICSE-WGR '15. Vilnius, Lithuania: ACM, 2015, pp. 41–63. ISBN: 978-1-4503-4146-2. DOI: 10.1145/2858796.2858798. URL: http://doi.acm.org/10.1145/2858796.2858798.

[53] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian. "Comprehensive database for facial expression analysis". In: *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. IEEE. 2000, pp. 46–53.

[54] Emily S Kappenman, Jaclyn L Farrens, Steven J Luck, Greg Hajcak Proudfit, Daniela M Pfabigan, and Anna Grubert. "Behavioral and ERP measures of attentional bias to threat in the dot-probe task: poor reliability and lack of correlation with anxiety". In: (2014). DOI: 10.3389/fpsyg.2014.01368.

[55] Onder Kavakci, Murat Semiz, Aysegul Kartal, Ayfer Dikici, and Nesim Kugu. "Test anxiety prevalance and related variables in the students who are going to take the university entrance examination". In: *Dusunen Adam: The Journal of Psychiatry and Neurological Sciences* 27.4 (2014), pp. 301–307. ISSN: 10188681. DOI: 10.5350/DAJPN2014270403. URL: http://www.dusunenadamdergisi.org/tr/MakaleDetay.aspx?MkID=1026.

[56] Theodore D. Kemper and Richard S. Lazarus. "Emotion and Adaptation." In: *Contemporary Sociology* 21.4 (1992), p. 522. DOI: 10.2307/2075902.

[57] Barbara Kitchenham and Stuart Charters. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. 2007.

[58] Ernst H W Koster, Geert Crombez, Bruno Verschuere, and Jan De Houwer. "Selective attention to threat in the dot probe paradigm: Differentiating vigilance and difficulty to disengage". In: *Behaviour Research and Therapy* 42.10 (2004), pp. 1183–1192. ISSN: 00057967. DOI: 10.1016/j.brat.2003.08.001.

[59] Bureau of Labor Statistics. *Women in the labor force: a databook*. 20117. URL: https://www.bls.gov/opub/reports/womens-databook/2017/home.htm.

[60] Richard D. Lane, Kateri McRae, Eric M. Reiman, Kewei Chen, Geoffrey L. Ahern, and Julian F. Thayer. "Neural correlates of heart rate variability during emotion". In: *NeuroImage* 44.1 (2009), pp. 213 –222. ISSN: 1053-8119. DOI: https://doi.org/10.1016/

j.neuroimage.2008.07.056. URL: http://www.sciencedirect.com/science/article/pii/S1053811908008926.

[61] Peter J Lang, Margaret M Bradley, and Bruce N Cuthbert. "International affective picture system (IAPS): Technical manual and affective ratings". In: *NIMH Center for the Study of Emotion and Attention* 1 (1997), pp. 39–58.

[62] Richard S Lazarus. "From psychological stress to the emotions: A history of changing outlooks". In: *Annual review of psychology* 44.1 (1993), pp. 1–22.

[63] David Lester, Bijou Lester, and Simon James. "A short computer anxiety scale". In: *Perceptual and motor skills* 100 (July 2005), pp. 964–8. DOI: 10.2466/PMS.100.3.964-968.

[64] Alex Lishinski, Aman Yadav, Jon Good, and Richard Enbody. "Learning to Program: Gender Differences and Interactive Effects of Students' Motivation, Goals, and Self-Efficacy on Performance". In: *Proceedings of the 2016 ACM Conference on International Computing Education Research*. ICER '16. New York, NY, USA: ACM, 2016, pp. 211–220. ISBN: 978-1-4503-4449-4. DOI: 10.1145/2960310.2960329. URL: http://doi.acm.org/10.1145/2960310.2960329.

[65] Daniel Macher, Manuela Paechter, Ilona Papousek, and Kai Ruggeri. "Statistics anxiety, trait anxiety, learning behavior, and academic performance". In: *European Journal of Psychology of Education* 27.4 (2012), pp. 483–498. ISSN: 02562928. DOI: 10.1007/s10212-011-0090-5.

[66] Matthew M. Maurer. "Computer anxiety correlates and what they tell us: A literature review". In: *Computers in Human Behavior* 10.3 (1994), pp. 369–376. ISSN: 07475632. DOI: 10.1016/0747-

5632(94)90062-0. URL: http://linkinghub.elsevier.com/retrieve/pii/0747563294900620.

[67] Ulf Melin and Stefan Cronholm. "Project oriented student work". In: *ACM SIGCSE Bulletin* 36.3 (2004), p. 87. ISSN: 00978418. DOI: 10.1145/1026487.1008021. URL: http://portal.acm.org/citation.cfm?doid=1026487.1008021.

[68] Harlan D Mills. "How to write correct programs and know it". In: *ACM SIGPLAN Notices* 10.6 (1975), pp. 363–370. ISSN: 03621340. DOI: 10.1145/390016.808459. URL: http://portal.acm.org/citation.cfm?doid=390016.808459.

[69] Grace Ngai, Winnie W.Y. Lau, Stephen C.F. Chan, and Hong-va Leong. "On the implementation of self-assessment in an introductory programming course". In: *ACM SIGCSE Bulletin* 41.4 (2010), p. 85. ISSN: 00978418. DOI: 10.1145/1709424.1709453. URL: http://portal.acm.org/citation.cfm?doid=1709424.1709453.

[70] Keith Nolan and Susan Bergin. "The Role of Anxiety when Learning to Program: A Systematic Review of the Literature". In: *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*. Koli Calling '16. New York, NY, USA: ACM, 2016, pp. 61–70. ISBN: 978-1-4503-4770-9. DOI: 10.1145/2999541.2999557. URL: http://doi.acm.org/10.1145/2999541.2999557.

[71] Keith Nolan, Susan Bergin, and Aidan Mooney. "An Insight Into the Relationship Between Confidence, Self-efficacy, Anxiety and Physiological Responses in a CS1 Exam-like Scenario". In: *Proceedings of the 1st UK & Ireland Computing Education Research Conference*. UKICER. New York, NY, USA: ACM, 2019, 8:1–8:7. ISBN: 978-1-4503-7257-2. DOI: 10.1145/3351287.3351296. URL: http://doi.acm.org/10.1145/3351287.3351296.

[72] Keith Nolan, Aidan Mooney, and Susan Bergin. "Examining the Role of Cognitive Load when Learning to Program". In: *The 3rd International Workshop on Eye Movements in Programming*. 2015.

[73] Keith Nolan, Aidan Mooney, and Susan Bergin. "Facilitating student learning in Computer Science: large class sizes and interventions". In: *International Conference on Engaging Pedagogy* (2015).

[74] Keith Nolan, Aidan Mooney, and Susan Bergin. "A Picture of Mental Health in First Year Computer Science". In: *Proceedings of the 10th International Conference on Computer Science Education: Innovation and Technology*. CSEIT'19. Bangkok, Thailand: Global Science and Technology Fourm, 2019. ISBN: 2251-2195.

[75] Keith Nolan, Aidan Mooney, and Susan Bergin. "An Investigation of Gender Differences in Computer Science Using Physiological, Psychological and Behavioural Metrics". In: *Proceedings of the Twenty-First Australasian Computing Education Conference*. ACE '19. New York, NY, USA: ACM, 2019, pp. 47–55. ISBN: 978-1-4503-6622-9. DOI: 10.1145/3286960.3286966. URL: http://doi.acm.org/10.1145/3286960.3286966.

[76] Mark Noone, Keith Nolan, and Aidan Mooney. "Hybrid Java Programming: A Visual-Textual Programming Language Workshop". In: *UK & Ireland Computing Education Research Conference*. *Workshop in UK & Ireland Computing Education Research Conference*. 2019.

[77] Else Ouweneel, Wilmar B Schaufeli, and Pascale M Le Blanc. "Believe, and you will achieve: Changes over time in self-efficacy, engagement, and performance". In: *Applied Psychology: Health and Well-Being* 5.2 (2013), pp. 225–247.

[78] J Owolabi, P Olanipekun, and J Iwerima. "Mathematics Ability and Anxiety, Computer and Programming Anxieties, Age and

Gender as Determinants of Achievement in Basic Programming". In: *GSTF Journal on Computing (JoC)* 3.4 (2014), p. 47. ISSN: 2010-2283. DOI: 10.7603/s40601-013-0047-4. URL: http://www.globalsciencejournals.com/article/10.7603/s40601-013-0047-4.

[79] Reena Pau, Wendy Hall, Marcus Grace, and John Woollard. "Female Students' Experiences of Programming: It's Not All Bad!" In: *Proceedings of the 16th Annual Joint Conference on Innovation and Technology in Computer Science Education*. ITiCSE '11. New York, NY, USA: ACM, 2011, pp. 323–327. ISBN: 978-1-4503-0697-3. DOI: 10.1145/1999747.1999837. URL: http://doi.acm.org/10.1145/1999747.1999837.

[80] Kalina Petrova, Dirk Wentura, and Christina Bermeitinger. "What Happens during the Stimulus Onset Asynchrony in the Dot-Probe Task? Exploring the Role of Eye Movements in the Assessment of Attentional Biases". In: (2013). DOI: 10.1371/journal.pone.0076335.

[81] A. Price and HA Smith. *USI National Report on Student Mental Health in Third Level Education*.

[82] C. W E M Quaedflieg, F. T Y Smulders, T. Meyer, F. Peeters, H. Merckelbach, and T. Smeets. "The validity of individual frontal alpha asymmetry EEG neurofeedback". In: *Social Cognitive and Affective Neuroscience* 11.1 (2015), pp. 33–43. ISSN: 17495024. DOI: 10.1093/scan/nsv090.

[83] Keith Quille, Susan Bergin, and Aidan Mooney. "PreSS#, A Web-Based Educational System to Predict Programming Performance". In: *International Journal of Computer Science and Software Engineering (IJCSSE)* 4.7 (2015). This work is licensed under a Creative Commons Attribution 3.0 Unported License., pp. 178–189. URL: http://eprints.maynoothuniversity.ie/6503/.

[84] Keith Quille, Natalie Culligan, and Susan Bergin. "Insights on Gender Differences in CS1: A Multi-institutional, Multi-variate Study." In: *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*. ITiCSE '17. New York, NY, USA: ACM, 2017, pp. 263–268. ISBN: 978-1-4503-4704-4. DOI: 10.1145/3059009.3059048. URL: http://doi.acm.org/10.1145/3059009.3059048.

[85] Michael de Raadt. "Student Created Cheat-sheets in Examinations: Impact on Student Outcomes". In: *Proceedings of the Fourteenth Australasian Computing Education Conference*. ACE '12. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2012, pp. 71–76. ISBN: 978-1-921770-04-3. URL: http://dl.acm.org/citation.cfm?id=2483716.2483725.

[86] Katie Redmond, Sarah Evans, and Mehran Sahami. "A Large-scale Quantitative Study of Women in Computer Science at Stanford University". In: *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*. SIGCSE '13. New York, NY, USA: ACM, 2013, pp. 439–444. ISBN: 978-1-4503-1868-6. DOI: 10.1145/2445196.2445326. URL: http://doi.acm.org/10.1145/2445196.2445326.

[87] Raoul Rickenberg and Byron Reeves. "The effects of animated characters on anxiety, task performance, and evaluations of user interfaces". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. Vol. 2. 1. New York, New York, USA: ACM Press, 2000, pp. 49–56. ISBN: 1581132166. DOI: 10.1145/332040.332406. URL: http://portal.acm.org/citation.cfm?doid=332040.332406.

[88] Sonja Rohrmann, Henrik Hopp, and Markus Quirin. "Gender Differences in Psychophysiological Responses to Disgust". In: *Journal of Psychophysiology* 22.2 (2008), pp. 65–75. DOI: 10.1027/

0269-8803.22.2.65. eprint: https://doi.org/10.1027/0269-8803.22.2.65. URL: https://doi.org/10.1027/0269-8803.22.2.65.

[89] Morris Rosenberg. "Rosenberg self-esteem scale (RSE)". In: *Acceptance and commitment therapy. Measures package* 61 (1965), p. 52.

[90] Montse C. Ruiz, John S. Raglin, and Yuri L. Hanin. "The individual zones of optimal functioning (IZOF) model (1978–2014): Historical overview of its development and use". In: *International Journal of Sport and Exercise Psychology* 15.1 (2017), pp. 41–63. DOI: 10.1080/1612197X.2015.1041545. eprint: https://doi.org/10.1080/1612197X.2015.1041545. URL: https://doi.org/10.1080/1612197X.2015.1041545.

[91] Kate Sanders and Robert McCartney. "Threshold Concepts in Computing: Past, Present, and Future". In: *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*. Koli Calling '16. New York, NY, USA: ACM, 2016, pp. 91–100. ISBN: 978-1-4503-4770-9. DOI: 10.1145/2999541.2999546. URL: http://doi.acm.org/10.1145/2999541.2999546.

[92] Carmen Sandi. "Stress and cognition". In: *Wiley Interdisciplinary Reviews: Cognitive Science* 4.3 (2013), 245–261. DOI: 10.1002/wcs.1222. URL: http://dx.doi.org/10.1002/wcs.1222.

[93] Thiago Donizetti dos Santos and Vagner Figueredo de Santana. "Computer Anxiety and Interaction: A Systematic Review". In: *Proceedings of the Internet of Accessible Things*. W4A '18. New York, NY, USA: ACM, 2018, 18:1–18:10. ISBN: 978-1-4503-5651-0. DOI: 10.1145/3192714.3192825. URL: http://doi.acm.org/10.1145/3192714.3192825.

[94] Donna Saulsberry. "Dwindling Number of Female Students: What Are We Missing?" In: *Proceedings of the 13th Annual Conference on*

Information Technology Education. SIGITE '12. New York, NY, USA: ACM, 2012, pp. 221–226. ISBN: 978-1-4503-1464-0. DOI: 10.1145/2380552.2380616. URL: http://doi.acm.org/10.1145/2380552.2380616.

[95] Stefan C Schmukle. "Unreliability of the Dot Probe Task". In: *European Journal of Personality Eur. J. Pers* 19 (2005), pp. 595–605. DOI: 10.1002/per.554.

[96] Dale H Schunk. "Self-efficacy and achievement behaviors". In: *Educational psychology review* 1.3 (1989), pp. 173–208.

[97] Michael James Scott and Gheorghita Ghinea. "Measuring enrichment". In: *Proceedings of the tenth annual conference on International computing education research*. New York, New York, USA: ACM Press, 2014, pp. 123–130. ISBN: 9781450327558. DOI: 10.1145/2632320.2632350. URL: http://dl.acm.org/citation.cfm?id=2632350http://dl.acm.org/citation.cfm?doid=2632320.2632350.

[98] Lily Shashaani. "Gender differences in computer attitudes and use among college students". In: *Journal of educational computing research* 16.1 (1997), pp. 37–51.

[99] Jane Sinclair and Sara Kalvala. "Exploring Societal Factors Affecting the Experience and Engagement of First Year Female Computer Science Undergraduates". In: *Proceedings of the 15th Koli Calling Conference on Computing Education Research*. Koli Calling '15. New York, NY, USA: ACM, 2015, pp. 107–116. ISBN: 978-1-4503-4020-5. DOI: 10.1145/2828959.2828979. URL: http://doi.acm.org/10.1145/2828959.2828979.

[100] David A Sousa. *Mind, brain, & education: Neuroscience implications for the classroom*. Solution Tree Press, 2014.

[101] Charles D Spielberger. *State-Trait Anxiety Inventory: Bibliography (2nd ed.)* CA: Consulting Psychologists Press, 1989.

[102] Charles D Spielberger, R. Gorsuch R. L.and Lushene, P. R. Vagg, and G. A Jacobs. *Manual for the State-Trait Anxiety Inventory*. CA: Consulting Psychologists Press, 1983.

[103] Søren Risløv Staugaard. "Reliability of two versions of the dot-probe task using photographic faces". In: *Psychology Science Quarterly* 51.3 (2009), pp. 339–350.

[104] J. R. Stroop. *Studies of Interference in Serial Verbal Reactions*. 1992.

[105] Francis Suraweera. "Using a Familiar Package to Demonstrate a Difficult Concept: Using an Excel Spreadsheet Model to Explain the Concepts of Neural Networks to Undergraduates." In: *ACM SIGCSE Bulletin* 34.4 (2002), p. 117. ISSN: 00978418. DOI: 10.1145/820127.820181. URL: http://portal.acm.org/citation.cfm?doid=820127.820181.

[106] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard. "Automatic identification of artifacts in electrodermal activity data". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2015, pp. 1934–1937. DOI: 10.1109/EMBC.2015.7318762.

[107] John Todman and Kenneth Day. "Computer anxiety: the role of psychological gender". In: *Computers in Human Behavior* 22.5 (2006), pp. 856–869. ISSN: 07475632. DOI: 10.1016/j.chb.2004.03.009. URL: http://linkinghub.elsevier.com/retrieve/pii/S0747563204000548.

[108] Prima Vitasari, Tutut Herawan, Muhamad Nubli Abdul Wahab, Ahmad Othman, and Suriya Kumar Sinnadurai. "Exploring Mathematics Anxiety among Engineering students". In: *Procedia - Social and Behavioral Sciences* 8.5 (2010), pp. 482–489. ISSN: 18770428.

DOI: 10.1016/j.sbspro.2010.12.066. URL: http://dx.doi.org/10.1016/j.sbspro.2010.12.066http://linkinghub.elsevier.com/retrieve/pii/S1877042810021713.

[109] Prima Vitasari, Muhammad Nubli Abdul Wahab, Ahmad Othman, Tutut Herawan, and Suriya Kumar Sinnadurai. "The relationship between study anxiety and academic performance among engineering students". In: *Procedia - Social and Behavioral Sciences* 8.5 (2010), pp. 490–497. ISSN: 18770428. DOI: 10.1016/j.sbspro.2010.12.067. URL: http://dx.doi.org/10.1016/j.sbspro.2010.12.067.

[110] Andreas Voss, Rico Schroeder, Andreas Heitmann, Annette Peters, and Siegfried Perz. "Short-Term Heart Rate Variability-Influence of Gender and Age in Healthy Subjects". In: *PloS one*. 2015.

[111] J M Williams, Andrew Mathews, and C MacLeod. "The emotional Stroop task and psychopathology." In: *Psychological bulletin* 120.1 (1996), pp. 3–24. ISSN: 0033-2909. DOI: Doi10.1037/0033-2909.120.1.3.

[112] Brenda Wilson. "Improving Comfort Level of Females in the First Computer Programming Course: Suggestions for CS Faculty". In: *J. Comput. Sci. Coll.* 23.4 ().

[113] *Women in STEM workforce 2017*. URL: https://www.wisecampaign.org.uk/statistics/women-in-stem-workforce-2017/.

[114] Habib Yaribeygi, Yunes Panahi, Hedayat Sahraei, Thomas P. Johnston, and Amirhossein Sahebkar. "The impact of stress on body function: a review". In: *EXCLI Journal; 16:Doc1057; ISSN 1611-2156* (2017). DOI: 10.17179/excli2017-480. URL: http://www.excli.de/vol16/Sahebkar_Panahi_21072017_proof.pdf.

A

PUBLICATIONS

The work described in this thesis has appeared in the following publications:

[70] Keith Nolan and Susan Bergin. "The Role of Anxiety when Learning to Program: A Systematic Review of the Literature". In: *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*. Koli Calling '16. New York, NY, USA: ACM, 2016, pp. 61–70. ISBN: 978-1-4503-4770-9. DOI: 10.1145/2999541.2999557. URL: http://doi.acm.org/10.1145/2999541.2999557

[75] Keith Nolan, Aidan Mooney, and Susan Bergin. "An Investigation of Gender Differences in Computer Science Using Physiological, Psychological and Behavioural Metrics". In: *Proceedings of the Twenty-First Australasian Computing Education Conference*. ACE '19. New York, NY, USA: ACM, 2019, pp. 47–55. ISBN: 978-1-4503-6622-9. DOI: 10.1145/3286960.3286966. URL: http://doi.acm.org/10.1145/3286960.3286966

[74] Keith Nolan, Aidan Mooney, and Susan Bergin. "A Picture of Mental Health in First Year Computer Science". In: *Proceedings of the 10th International Conference on Computer Science Education: Innovation and Technology*. CSEIT'19. Bangkok, Thailand: Global Science and Technology Fourm, 2019. ISBN: 2251-2195

[71] Keith Nolan, Susan Bergin, and Aidan Mooney. "An Insight Into the Relationship Between Confidence, Self-efficacy, Anxiety and Physiological Responses in a CS1 Exam-like Scenario". In: *Pro-*

ceedings of the 1st UK & Ireland Computing Education Research Conference. UKICER. New York, NY, USA: ACM, 2019, 8:1–8:7. ISBN: 978-1-4503-7257-2. DOI: 10.1145/3351287.3351296. URL: http://doi.acm.org/10.1145/3351287.3351296

Additional Contributions

In addition to the list of contributing publications, the following CS Education publications were also completed:

[73] Keith Nolan, Aidan Mooney, and Susan Bergin. "Facilitating student learning in Computer Science: large class sizes and interventions". In: *International Conference on Engaging Pedagogy* (2015)

[72] Keith Nolan, Aidan Mooney, and Susan Bergin. "Examining the Role of Cognitive Load when Learning to Program". In: *The 3rd International Workshop on Eye Movements in Programming*. 2015

[76] Mark Noone, Keith Nolan, and Aidan Mooney. "Hybrid Java Programming: A Visual-Textual Programming Language Workshop". In: *UK & Ireland Computing Education Research Conference*. *Workshop in UK & Ireland Computing Education Research Conference*. 2019

B

STATISTICAL TECHNIQUES

This chapter presents the relevant statistical definitions that are used throughout this thesis.

- **Correlation Coefficient**

  A numerical index that reflects the relationship between two variables usually denoted by r. The correlation coefficient takes a value between -1 and 1 where -1 would indicate a perfect negative correlation, 0 would indicate no correlation and 1 would indicate a perfect positive correlation.

- **Pearson's Correlation Coefficient**

  The Pearson's Correlation Coefficient examines the relationship between two variables, but both of those variables are continuous in nature. A Pearson's Correlation is based on the assumption that both variables are normal. The formula for calculating it is outlined in Equation B.1.

  $$r_{XY} = \frac{N\sum XY - \sum X \sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}} \tag{B.1}$$

- **Shapiro-Wilks test of normality**

  The Shapiro-Wilks test of normality tests the null hypothesis that a sample comes from a normally distributed population where:

  $$W = \frac{(\sum_{i=1}^{n})a_i x_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{B.2}$$

  n is the number of observations, x_i is the values of the ordered sample and a_i is the tabulated coefficients. If $W = 1$ the given data is normally distributed. When W is significantly smaller than 1, the assumption of normality is not met.

- **Welch's t-test**

  A Welch's t-test compares the mean scores of two groups on a given variable. It is based on the assumptions that the dependent variable is normally distributed and that the two groups have approximately equal variance on the dependent variable. Once satisfied the t-test can be calculated using:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{B.3}$$

  where $\overline{X_i}$, s_i^2 and N_i are the samples mean, variance and size respectively where $i \in 1, 2$.

- **Alpha Significance**

  Alpha Significance, α, is the probability of rejecting the null hypothesis when it is true. For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference. This is used when testing the significance of a relationship either in a correlation test or a t-test.

- **Spearman's rank-order correlation**

  The Spearman's rank-order correlation is a non-parametric test of the Pearson's Correlation Coefficient. The Spearman correlation can be used when the assumptions of the Pearson correlation are violated. The following formula, Equation B.4, is used to calculate the Spearman rank correlation: Equation B.4.

$$r_s = \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{B.4}$$

- **The Elbow Method**

The Elbow method was chosen for its simplicity and still allows for the human element of seeing the clusters. The Elbow method chooses the number of clusters by attempting to minimise the variance of the data points inside the clusters while maximising the variance between the clusters. The number of clusters is then picked by examining the point where adding an extra cluster does not give a better model of the data than choosing one cluster less. The number of clusters was automatically picked by running the elbow method multiple times determining the point where most of the variance was explained without overloading the number of clusters. As an example, we present the clusters that exist between State and Trait anxiety.

Figure B.1: Elbow method showing the number of clusters

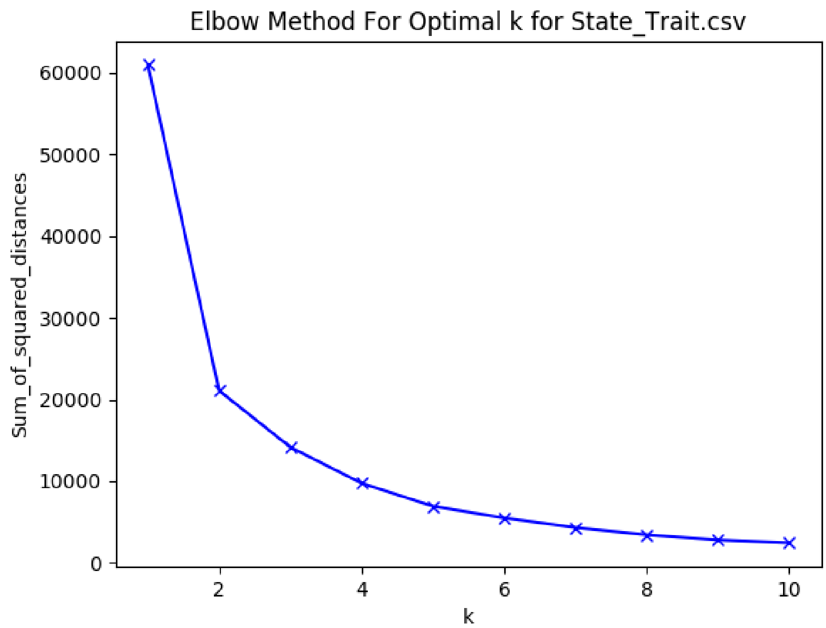
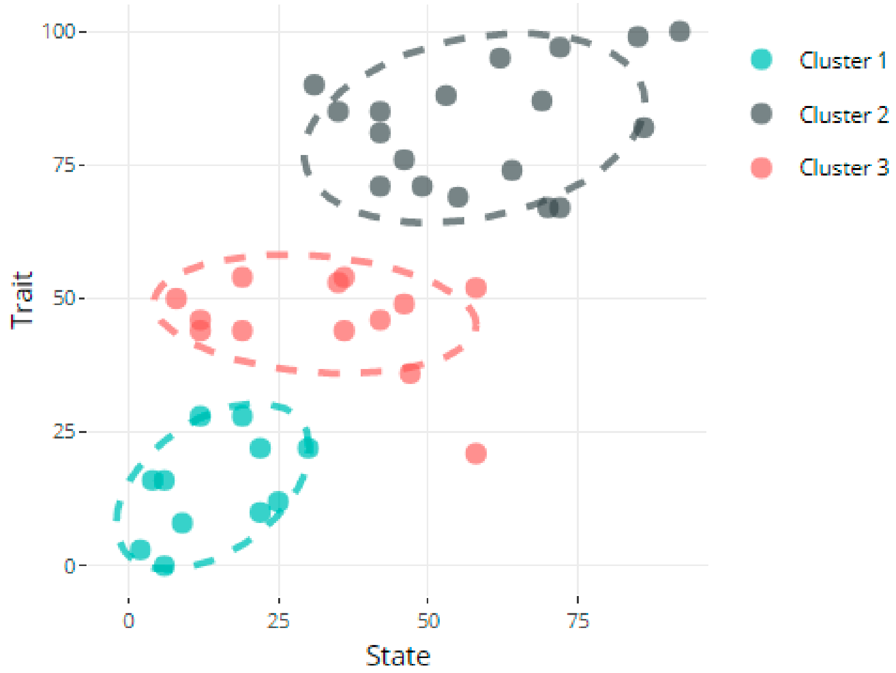


Figure B.1 illustrates the various clusters found using the elbow method, with three clusters found to be optimal. These three clusters are then presented in Figure B.2. Cluster 1 represents those with low State and low Trait anxiety. Cluster 2 represents those

Figure B.2: Clusters of participants by State and Trait anxiety.



with low – medium State and medium Trait. Cluster 3 represents those with medium – high State and high Trait. As can be seen, there are outliers in the clusters as denoted by the points outside the ellipses drawn around the clusters. While these outliers exist, a positive linear relationship can be observed between State and Trait anxiety.

- **Cohen-d**

The Cohen's d effect size is a quantitative measure of the magnitude of the experimenter effect. The larger the effect size the stronger the relationship between two variables. It is calculated by using the following equation:

$$Cohen - d = \frac{M_2 - M_1}{SD_{pooled}} \tag{B.5}$$

$$SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}} \tag{B.6}$$

A d-value of 1 indicates the two groups differ by 1 standard deviation, a d-value of 2 indicates they differ by 2 standard deviations. Cohen suggested that d=0.2 be considered a "small" effect size, 0.5 represents a "medium" effect size and 0.8 a "large" effect size. This means that if two groups' means don't differ by 0.2 standard deviations or more, the difference is trivial, even if it is statistically significant.

C

ANALYSIS TECHNIQUES

C.1 ELECTRODERMAL ACTIVITY

The analysis of the Electrodermal Activity for this project was conducted using the MIT EDA Explorer tool. This tool was chosen due to the high citation count in other research projects and the flexibility of the tool to analyse data from different devices.

The EDA Explorer is an automatic artifact detection tool which uses a machine learning algorithm to clean the data and report meaning results. The team collected 1560 samples of EDA portions. Experts were asked to label the data samples as either "Clean" or "artifact". Both of the experts agreed that an artifact was:

- A peak which does not show exponential decay, depending on the context (e.g. if two SCRs occur close together in time, the first response may not decay before the second begins, yet this is not considered an artifact)
- Quantization error with >= 5%of signal amplitude
- A sudden change in EDA correlated with motion
- A SCL<= 0

Prior to the experts labelling the samples, the samples were filtered using a 1Hz low pass filter. The experts agreed on 80.71% of the samples and so a ground truth was established.

Prior to the training of the machine learning algorithm, several statistical features were calculated from the samples. In addition, several wavelet coefficients. The wavelet coefficients were calculated using a Discrete Haar Wavelet Transformation. To ensure the models are overfitted, a Wrapper feature selection was used which is robust to overfitting.

Several machine learning techniques were then tested (neural networks, random forests, naive bayes, nearest neighbour, logistic regression and support vector machines). Of the various algorithms, a support vector machine was the best preforming algorithm.

As an output of the tool, the number of SCRs evident in the dataset is returned which can allow other researchers an insight into participants arousal levels.

C.2 PHOTOPLETHYSOMGRAM

The beat-to-beat data was cleaned using a low pass filter to reduce noise and sharpen the beat-to-beat peaks. Following this, a peak detection algorithm was used to create interbeat intervals (IBI) which is the time between successive beats of the heart. Finally, customised software was developed to analyse the IBI file and determine the RMSSD. Equation C.1 depicts the formula to calculate $RMSSD$, where N is the number of beat-to-beat intervals and $(R-R)$ is the difference in time in subsequent beats.

$$RMSSD = \sqrt{\frac{1}{N-1}(\sum_{i-1}^{N-1}((R-R)_{i+1}-(R-R)_i)^2)} \tag{C.1}$$

The PPG signals were analysed over the first and second halves of the experiment rather than individual questions or question bands as the time frame for reliable HRV measures were too short at the per question level. Halves were chosen by taking the total run time of the experiment and dividing by 2. All data was corrected into the same logical flow: Easy Questions, followed by Medium questions followed by Hard questions. It is important to note that the data is not evenly split in terms of what questions were included in each halves response time. One set of participant data had to be removed from the PPG data as sections of the data were lost during recording.

D

SYSTEMATIC LITERATURE REVIEW TABLE

D.1 SURVEYS USED

Table D.1: Table of surveys used among studies.

Instrument Full Name	Short name
Revised Mathematics Anxiety Scale	RMAS
Computer Attitude Scale	CAS
Computer Anxiety Rating Scale	CARS
Computer Programming Anxiety Questionnaire	CPAQ
Test Anxiety Inventory	TAI
Beck's Depression Inventory	BDI
State Trait Anxiety Inventory	STAI
Liebowitz Social Anxiety Scale	LSAS
Wender Utah Rating Scale	WURS
Adult ADHD Self-Reported Scale	A-ADHD-SRS
Statistics Anxiety Rating Scale	STARS
Computer Programming Anxiety Rating Scale	CPARS
Computer Programming Achievement	CPA

D.2 SYSTEMATIC LITERATURE REVIEW TABLE

Table D.2: Accepted studies

Authors	Citation	Publication Source	Type	Type of Study	Survey Used[1]	# participants
Baloglu et al.	[4]	Personality and Individual Differences	Journal	Questionnaire	RMAS	759
Chang	[20]	Computers in Human Behaviour	Journal	Questionnaire	CAS	307
Maurer	[66]	Computers in Human Behaviour	Journal	Literature Review	n/a	n/a
Deloatch et al.	[29]	SIGCSE	Conference	Questionnaire	Own Measure	391

Doyle et al.	[33]	Frontiers in Education	Conference	Questionnaire	CARS	163
Connolly et al.	[25]	Transactions in Education	Conference	Questionnaire	CPAQ	86
Falkner et al.	[35]	SIGCSE	Conference	Interview	n/a	10
Kavakci et al.	[55]	Dusunen Adam	Journal	Questionnaire	TAI, BDI, STAI, LSAS, WURS, A-ADHD-SRS	436
Macher et al.	[65]	Eur J Psychol Educ	Journal	Questionnaire	STARS	147
Chua et al.	[22]	Computers in Human Behaviour	Journal	Literature Review	n/a	n/a
Scott et al.	[97]	ICER	Conference	Questionnaire	Subscale of a scale	239

Todman et al.	[107]	Computers in Human Behaviour	Journal	Questionnaire	Short form of CAS	138
DeRaadt	[85]	ACE	Conference	Experimental	n/a	89
Fone	[39]	Neural Networks	Conference	Experimental	n/a	21
Gerritsen et al.	[40]	WI-IAT	Conference	Questionnaire	Not Reported	21
Guynes	[44]	Communications of the ACM	Journal	Experimental	State of STAI	93
Hamer et al.	[46]	ICER	Conference	Questionnaire	Did Not Report	1500
Fenwick et al.	[37]	SIGCSE	Conference	Questionnaire	Did Not Report	100
Melin et al.	[67]	ITICSE	Conference	Questionnaire	Did Not Report	60
Mills	[68]	ACM	Journal	Literature Review	n/a	n/a
Ngai et al.	[69]	SIGCSE	Journal	Questionnaire	Did Not Report	13
Suraweera	[105]	ACM	Journal	Literature Review	n/a	n/a

Owolabi et al.	[78]	GSTF Journal on Computing	Journal	Experimental	CARS, CPARS, CPA	160
Vitasari et al.	[108]	Procedia - Social and Behavioural Sciences	Journal	Questionnaire	Own Measure	205
Blanchard et al.	[16]	CHI	Conference	Experimental	STAI	27
Deloatch et al.	[30]	CHI	Conference	Experimental	Trait Form of STAI	1235
Dos Santos et al.	[93]	W4A	Conference	Literature Review	CARS	n/a

E

PROGRAMMING QUESTIONS

E.1 PROGRAMMING QUESTIONS

```
// Question1 difficulty : easy
public class q{
    public static void main(String [] args){
        System.out.println("Hello World!");
    }
}
// Answer A : hello world!
// Answer B : Hello world!
// Answer C : Hello World! **Correct Answer**
// Answer D : None of These
```

```
// Question2 difficulty : easy
public class q{
    public static void main(String [] args){
        String s1 = new String("How are you?");
        String s2 = new String("Hello");
        System.out.println(s2);
    }
}
// Answer A : How are you?
// Answer B : Hello **Correct Answer**
// Answer C : HelloHow are you?
// Answer D : None of These
```

```
// Question3 difficulty : easy
public class q{
    public static void main(String [] args){
        String s1 = new String("1");
        String s2 = new String("6");
        System.out.println(s1+s2+s1);
    }
}
// Answer A : 161 **Correct Answer**
// Answer B : 8
// Answer C : 116
// Answer D : None of These
```

```java
// Question4 difficulty : medium
public class q{
    public static void main(String [] args){
        int x = 5;
        if (x<5){
            System.out.println("<");
        }
        else if (x>5){
            System.out.println(">");
        }
        else{
            System.out.println("=");
        }
    }
}
// Answer A : <
// Answer B : >
// Answer C : = **Correct Answer**
// Answer D : None of These
```

```java
// Question5 difficulty : medium
public class q{
    public static void main(String [] args){
        int income = 15;
        if (income > 10){
            System.out.print("A");
            if (income < 20){
                System.out.print("B");
            }
        }
        else{
            System.out.print("C");
        }
    }
}
// Answer A : AC
// Answer B : AB **Correct Answer**
// Answer C : BC
// Answer D : None of These
```

```java
// Question6 difficulty : medium
public class q{
    public static void main(String [] args){
        int count = 0;
        int result = 0;
        while (count< 5){
            result = result + count;
            count++;
        }
        System.out.println(result);
    }
}
// Answer A : 9
// Answer B : 10 **Correct Answer**
// Answer C : 11
// Answer D : None of These
```

```
// Question7 difficulty : hard
public class q{
    public static void main(String [] args){
        int count = 1;
        String x = new String("Hello World!");
        String newX = new String("");
        while (count< 12){
            newX = newX + x.substring(count, count+1);
            count+=2;
        }
        System.out.println(newX);
    }
}
// Answer A : el ol! **Correct Answer**
// Answer B : Hello World!
// Answer C : HloWrd
// Answer D : None of These
```

```
// Question8 difficulty : hard
public class q{
    public static void main(String [] args){
        String x = new String("Hello World!");
        if(5>6){
            System.out.println(x.substring(0,5));
        }
        else{
            System.out.println(x.substring(6));
        }
    }
}
// Answer A : Hello World!
// Answer B : Hello
// Answer C : World! **Correct Answer**
// Answer D : None of These
```

```
// Question9 difficulty : hard
public class q{
    public static void main(String [] args){
        int count1 = 0;
        String star = new String("*");
        while (count1< 5){
            int count2=0;
            while (count2<= count1){
                System.out.print(star);
                count2++;
            }
            System.out.println();
            count1++;
        }
    }
}
// Answer A : *
//            * *
//            * * *
//            * * * *
// Answer B : *          **Correct Answer**
//            * *
//            * * *
//            * * * *
//            * * * * *
// Answer C : *
//            * *
//            * * *
// Answer D : None of These
```

F

INTERNATIONAL COLLABORATION

This appendix presents an international collaboration undertaken as part of this research project. It begins with an outline of the motivation and aims of the collaboration. The experiments conducted and the experimental paradigms used are discussed with the results of the experiments presented.

The appendix is laid out as follows. Section F.1 discusses the motivations for the international collaboration. Section F.2 outlines the materials used throughout this appendix. Section F.3 breaks down the different experimental paradigms decided on during the collaboration. Section F.4 describes an initial behavioural study and the outcomes of the study which form the foundation for the experimental protocols used in later research experiments.

F.1 MOTIVATION FOR INTERNATIONAL COLLABORATION

At the beginning of the project, there were plans to monitor physiological and cognitive signals to potentially understand how anxiety affects learning. This presented two challenges, 1) there was a need to learn how to capture physiological signals, and, 2) there was a need to learn how to capture cognitive signals. Members of staff within Maynooth University were skilled in collecting physiological signals but unfortunately, there was not the same expertise readily available to support the capture of cognitive signals. A close collaboration between the Department of Psychology at the University of Wuppertal and the Department of Computer Science at Maynooth University was already established and was drawn on here to help:

- Learn and develop the necessary skills to design and conduct a research experiment.
- Learn the Standard Operating Procedures of an EEG machine.
- Learn how to analyse the data gathered by the EEG.
- Design and conduct a research experiment to validate a 4–channel EEG Headband against a 64–channel EEG.

There was a need to develop the experimental skills required to conduct experiments rigorously. This was needed as the researcher had no previous experience in experimental design. In addition to learning the skills necessary to design an experiment, there was a need to understand the Standard Operating Procedures for the associated hardware.

Initially, the project had planned to utilise a four-channel EEG headband to detect signals indicative of anxiety. This introduced the need to understand and become familiar with the methods of analysing the raw EEG data produced by this equipment, taking the data and outputting meaningful results. Having the use of a four-channel EEG introduces the need to validate the signals produced by said EEG. This meant that the signals would have to be validated against a 64–channel EEG. Given that the Department of Psychology at the University of Wuppertal has extensive experience in EEG studies, the collaboration provided a quick and effective method of obtaining the relevant skills.

The international collaboration visit to the University of Wuppertal lasted for four weeks over a period from November 2015 – December 2015. During this time, the possible types of studies that could be conducted were discussed which would allow for the validation of the four-channel EEG headband.

Figure F.1: Muse headband with sensors on forehead and resting on ears.

F.2 MATERIALS

As part of this research project, a four-channel EEG headband would be utilised to attempt to capture signals indicative of anxiety. The Muse Headband, shown in Figure F.1, is a powerful, compact four-channel EEG system. The headset has four dry sensors (two mastoid and two forehead sensors) and fits over the ears and extends at an angle over the middle of the forehead when properly fitted. As the Muse only has four channels and covers the forehead and mastoid areas, a frontal asymmetry study was planned.

F.3 EXPERIMENTAL PARADIGMS

This section outlines the psychological paradigms employed in the international collaboration.

F.3.1 FRONTAL ALPHA ASYMMETRY

Frontal asymmetry is the average differences between brain activity in the frontal areas of the brain. In a study by Davidson et al. disgust was found to be associated with right-sided activation in the frontal and anterior temporal regions of the brain and happiness was accompanied by left-sided activation in the anterior temporal region [28, 82]. Frontal Alpha Asymmetry is an area that has been studied extensively as part of research on emotional and motivational processes, specifically, right and left sides brain differences in alpha power. Power bands represent a frequency range that the brain waves that can be measured using an EEG [48].

When investigating Frontal Alpha Asymmetry, the alpha frequencies (8-12Hz) is of interest. Alpha waves aid overall mental coordination, calmness, alertness, mind/body integration and learning. Frontal Alpha Asymmetry was initially detected by Davidson et al and validated by Hagemann et al. when investigating different biomarkers of personality [28, 45]. They discovered that people with increased left-frontal alpha power were found to process information positively compared to people processing the information on the right-hand side of the brain where a more negative processing mode was observed [28, 45].

F.3.2 RELATIONSHIP TO ATTENTION TO THREAT

Feelings of withdrawal have been linked to right frontal EEG activity when the person is resting and also in the face of new emotionally threatening situation [24, 48]. This bias is evident in healthy children and adults [23], individuals at increased temperament, given their high negative emotional state- or individuals with anxiety and depression [36], and individuals with a current or past history of mood disorder

[2]. In contrast, greater left frontal EEG activity has been linked to approach tendencies, involving both positive emotions, such as joy [34], and negative emotions, such as anger [47]. This would coincide with the withdrawal system of the Flight-Fright System [47]. Given these frontal lobe difference, the Emotional Stroop Paradigm (discussed in Section F.3.3) and the Dot-Probe Paradigm (discussed in Section F.3.4) were chosen as experimental paradigms as they exploit frontal lobe differences.

F.3.3 emotional stroop paradigm

The Emotional Stroop test is an adaptation of the classic Stroop test designed by Williams et al [104, 111]. Before explaining the Emotional Stroop paradigm, the Classic Stroop paradigm must first be explained. The Classic Stroop demonstrates the power of Cognitive Interference[1]. It exploits the mismatch between the name of a colour and the colour in which the word is written in. The interference here is the time delay between recognising the colour that the word is written in and responding to what colour it is written in, as shown in Figure F.2. The paradigm aims to respond with the colour that the word is written in rather than what colour the word says. It is known that it is quicker to respond with "Blue" when the word says Blue rather than if the word said Red or Green.

Following on from this, the Emotional Stroop paradigm aims to measure the attentional bias towards emotional words. In the Emotional Stroop test, emotional words (Negative Words: "Hate", "Depressed", Positive Words: "Happy", "Glad" and Neutral Words: "Indifferent", "Weary") are flashed on the screen in different colours as shown in Fig-

1 Cognitive interference refers to the unwanted and often disturbing thoughts that intrude on a person's life.

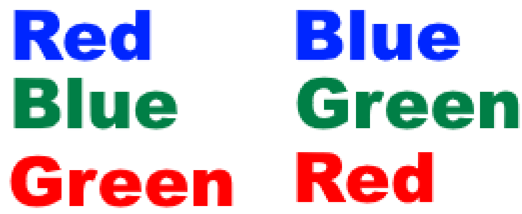


Figure F.2: Examples of the presentation of words in the Classic Stroop.

ure F.3. The objective of the test is to respond with the colour that the word is presented in and ignore the actual word. The emotional Stroop effect emphasises the conflict between the emotional relevance to the individual and the word; whereas, the classic Stroop effect examines the conflict between the mismatched colour and the word.a

Figure F.3: Examples of the presentation of negative words in the Emotional Stroop.

During the analysis of the Emotional Stroop, a difference between the response times between the positive, neutral and negative words are examined. The effect that is desired is longer response times to the negative words when compared to the positive and neutral words. The Emotional Stroop effect is difficult to detect in a normal population. By a normal population, it is meant that a population that is un-diagnosed with any mental issues such as chronic stress or anxiety. However, a slight effect can be observed whereby the cognitive interference builds up over subsequent trials and towards the end of the trial sequence, the effect is observed. The next section will discuss the Dot-Probe paradigm.

F.3.4 dot probe paradigm

The Dot Probe paradigm was designed by MacLeod et al. (1986) to overcome the lack of a result observed in a normal population seen in the Emotional Stroop test. The Dot Probe paradigm follows the following flow:

1. A focus point, usually a dot or a cross is placed in the centre of the screen. This is used to focus a participants attention. This is shown in Figure F.4 as the cross in the first screen.

2. Two stimuli then appear on the screen, one on either side of the centre point. The stimuli, in this case, are usually emotive faces (angry or happy) paired with a neutral face (no expression). The faces will stay on the screen for a set amount of time, between 800ms and 1200ms. This is shown in Figure F.4 as the two faces on the second screen. Here it is hoped that the participant's attention will be switched and fixed on the emotive face rather than the neutral face. In particular, it is hoped that participants will take longer to disengage (turn attention from the emotional image to the probe) from the angry/negative face when compared to the happy/positive face. This disengagement time is measured from the time that the emotional image disappears (Step 4) to the time the participant responds to the probe (Step 5).

3. The faces are removed from the screen.

4. A second single stimulus (usually a dot) will then appear from behind where either of the faces was. This is shown in Figure F.4 as the dot appearing in place of the happy face in the third screen.

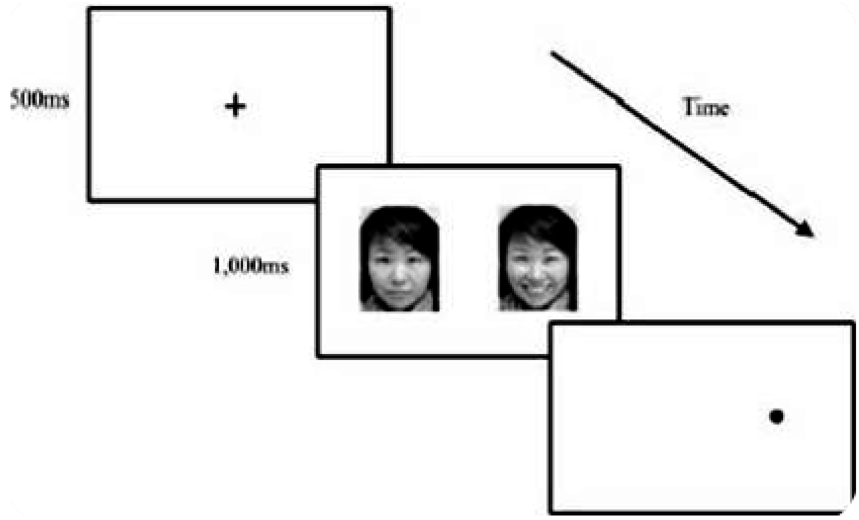


Figure F.4: Flow of a typical trial in a Dot-Probe Experiment.

5. The participant will have to respond by indicating what side of the screen the dot was on.

The Dot Probe paradigm offers advantages when compared to other paradigms such as Emotional Stroop. Firstly, the Dot Probe allows for no concern that delayed latencies may result from response bias or general arousal. Secondly, it allows for the ability to manipulate the Stimulus Onset Asynchrony (SOA) which is the time interval between presentation of the critical stimuli and presentation of the probe.

When examining the Dot Probe, there are two trials, Congruent and Not Congruent. A Congruent trial is one where the probe appears behind the emotional image. A Not Congruent trial is one where the probe appears behind the neutral image. The desired effect is a slower Not Congruent trial when compared to a Congruent trial.

F.4 CONCEPT STUDIES

Initially, a short pilot experiment using the Dot Probe paradigm was conducted. As part of the setup, four sets of faces from the Cohn-Kanade (CK and CK+) database were chosen for the experiment [53].

To ensure that the experiment was fair and unbiased, it was important that every face, irrespective of emotion, appeared on both sides of the screen once. For every face, there are three individual faces, a happy face, an angry face and a neutral face. Each of the neutral faces was always paired with an emotive face. Every emotive face was placed on the left and right side of the screen once. Given these conditions, there were a total of thirty-2 trials. The participants were told to read the instructions presented on the screen before the beginning of the experiment. Following the instructions, the trials of the Dot probe began.

This initial experiment was conducted with seven participants of which, four of them showed a slight effect, that is they responded slower when the probe was on the opposite side of the negative emotive face. As this was in the planning stages of the experiment, no deep analysis was conducted. The effect that was observed while slight, was within the bounds of standard error meaning that the difference in times was close to negligible.

Following this, it was suggested by the team in the University of Wuppertal that participants should be negatively primed[2] by displaying negative pictures from International Affective Picture System (IAPS) [61]. The IAPS is a standardised database of specifically designed images used when studying emotion or attention. When using the IAPS images it is essential that ethical approval is sought from the relevant authority (See Appendix G for a copy of the ethical approval granted). Using the IAPS, the experiment was modified in the following way:

- The pictures with the faces were cropped to only contain the head and no background colour.
- The visual angle was corrected such that the pictures appeared to be two meters away from the participant.

2 Negative priming is an implicit memory effect in which prior exposure to a stimulus unfavourably influences the response to the same stimulus.

The modified experiment was run in the University of Wuppertal with 5 participants. An effect was observed with 3 of those 5 participants however the effect was marginal and within the bounds of error. While there was an effect observed, the protocol was changed to attempt to increase the observed effect. Two separate experiments were designed to attempt to show this increase. The experiments aimed to examine the following questions:

1. Will a larger threat-related bias be found if the probe is harder to detect?

2. Will a larger threat-related bias be found using the images with cropped faces (that only contain the head and no background colour)?

As such, two experiments were developed.

Experiment 1: Four participants took the experiment with the new protocol. Of those four participants, one showed an effect. Due to only one participant showing an effect, no more changes were made to this protocol and it was not continued which meant this protocol would not be capable of validating the Muse.

Experiment 2: Five participants took part in this protocol. It was found that four out of the five participants were slower when reacting to the probe when it was not under the negative face. There was also a low error rate associated with this protocol. This now appeared to be promising and so a larger, more in-depth experiment was planned.

Given that four out of five participants showed an effect in Experiment 2 it was decided to increase the number of participants in the experiment. The experiment was run again with an additional five people (bringing the total number of participants to ten). Based on all ten participants, five of them were slower when reacting to the probe when it was not under the negative face.

Compared to the initial findings when there were only five participants, these findings were disappointing. What had originally appeared to be an 80% effect dropped to a 50% effect. Upon closer investigation, a variation in participant profiles was noted. Of the ten participants, two profiles of participants were noted.

- Profile 1 : Student and aged <30, n=5
- Profile 2 : Staff and aged >30, n=5

Further analysis of the data showed a difference in results between the two profiles.

- Profile 1 - Student and aged < 30:
  - Four out of the five students showed an overall effect.
- Profile 2 - Staff and aged > 30:
  - One out of five staff members showed an effect.

From an analysis of literature reviews and other studies, the average age of the participants who participate in a Dot-Probe experiment was recorded in the range of 19 – 28 [6, 43, 58, 80, 95, 103]. Given the older age range of Profile 2, they may not have had an effect similar to the effect outlined in the literature.

Up to this point, all studies have been conducted as a pilot phase. This phase was to investigate the paradigms presented and their reliability in an undiagnosed population. The iterations of experimental protocols led to a solid experimental protocol that could potentially be used.

F.4.1 PARADIGM DISCUSSION

From the literature, the paradigms used typically show an effect on participants with clinical or social anxiety but the evidence for normal participants is conflicting. Some participants in this study may have an anxiety disorder but this was not recorded. Based on research by Bar-Haim et al., in a control population, consisting of individuals who are not clinically anxious or have high self-reported anxiety, a non-significant attentional bias is seen only in the Emotional Stroop paradigm. Over 30 studies used a block design[3] Emotional Stroop paradigm and saw that in the control group exhibited a significant effect (n = 716, d= 0.56. $p < 0.001$, CI = 0.42,0.70) [6].

The number of participants in the experiment is very small when compared to an average of 30+ in the literature [43, 54, 58, 80, 95, 103] and to see a reliable effect another pilot study with a larger sample size seemed appropriate. In the follow-up pilot study participants could be screened for social or clinical anxiety by using standard tests such as the state form of the State Trait Anxiety Inventory and the Social Interaction Anxiety Scale. Individuals who do not suffer from any class of anxiety could be used as a control and individuals who suffer from depression should be rejected from the study.

However, although this larger study seems appropriate, concerns remain about the Dot Probe paradigm, in particular concerning test-retest reliability and internal reliability [54, 80, 95, 103]. Even though these problems exist, the paradigm remains to be the "gold standard" when it comes to investigating attentional bias to threat [54]. Bar-Haim et al. compared the Dot Probe paradigm to the Emotional Stroop paradigm. They found that both paradigms are equally as effective in detecting a frontal alpha asymmetry difference. However, they did criticise the Emotional Stroop paradigm by proposing that delayed response latencies with threat-related stimuli may result from late processes that are unrelated to attention. While the Emotional Stroop is criticised, it has

3 In both the Dot Probe and Emotional Stroop three affective states were used: positive, negative and neutral. In a block design experiment stimuli in groups of the same affective states are presented.

been shown that an effect can be observed when stimuli are displayed in a block design.

F.5 BEHAVIOURAL STUDY

Following on from the findings of the initial studies, a larger study was conducted at Maynooth, as the participants that were the focus of the thesis were based in Maynooth University. Participants were studying CS1 and were gathered voluntarily. Ethical approval was sought and granted to carry out the research (letter of approval can be seen in Appendix G).

F.5.1 EXPERIMENTAL PROTOCOL

The experiment took place in the Department of Computer Science. The room chosen was selected to minimise any discomfort the participant might experience. In this experiment, the researcher and a single participant were present in the test room. The researcher was out of view from the participant throughout the experiment and stayed in the room solely to ensure the experiment ran smoothly. Participants sat at a desk in front of a monitor with a keyboard and a mouse on the desk. Participants were given an information sheet and a consent form. Participants were instructed to read the information sheet describing the experiment before commencement. Upon completion, if they had any issues or questions they were encouraged to ask for clarification. Thereafter, they were asked to sign a consent form. Following completion, the participant was instructed to begin the experiment. It is important to note at this point that the Muse headband had not been placed on the participant's heads. The objective of this experiment was to determine if a Frontal Alpha difference could be obtained through the behavioural results before the use of the Muse.

Before the beginning of the Emotional Stoop and Dot Probe experiments, images from the International Affective Picture System (IAPS)[4] were used to negatively prime the participants. This was done to induce negative bias in the would hope that this would increase a difference in the frontal alpha activity of the brain. Following this, the Emotional Stroop and the Dot Probe experiments were run with participants in a counter-balanced format, half of the participants saw the Dot Probe first and the other half saw the Emotional Stroop first.

Each participant was given the following instructions on how to respond to each trial of both the Emotional Stoop and Dot Probe as follows:

- Emotional Stroop

  -The participant was asked to respond with the colour of the word by using the directional arrows (Red == left, Green == down and Blue == right) on the keyboard.

- Dot probe

  -The participant was asked to respond by indicating which side of the screen the probe appeared on (Left Side == left and Right Side == right).

4 The International Affective Picture System (IAPS) is a database of pictures designed to provide a standardized set of pictures for studying emotion and attention. IAPS images have been widely used in psychological research. They contain images from everyday images to emotive distressing images such as mutilated human bodies.

F.6 RESULTS

This section outlines the results observed from both the Dot Probe and Emotional Stroop tests. Following this, a deeper analysis is undertaken and the results presented.

F.6.1 PARTICIPANT PROFILE

Forty-one participants (29 male, 12 female) participated in this study. Table F.1 presents the age and gender profiles of the participants.

Table F.1: Age and gender profile of participants

Age	Male (N=29)	Female (N=12)
17–19	21 (74%)	10 (84%)
20–22	4 (13%)	1 (8%)
23+	4 (13%)	1 (8%)

F.6.2 DOT PROBE EXPERIMENT

To analyse the Dot-Probe experiment, the reaction times between Congruent and Not Congruent trials for both the Happy and Angry faces were investigated. In total, each participant completed 256 trials in a random sequence. The 256 trials are derived from:

- 2 different on set times (800ms and 1200ms).
- 2 sides the probe can appear on (left and right).
- 2 different emotions (Happy or Angry).

4 Full name in Table D.1

- 2 sides of the screen images can appear on (left and right).
- 2 types of Congruence's (Congruent and Not Congruent).
- 4 types of people.

Participants were always presented with an emotive face (Happy or Angry) and a neutral face. The probe would always be displayed on one of the sides. Again, the position of the probe was counterbalanced too. Having these conditions lead to the following four categories of trials:

- Angry Congruent (Probe appears the same side of the Angry face).
- Angry Not Congruent (Probe appears the same side of the neutral face).
- Happy Congruent (Probe appears the same side of the Happy face).
- Happy Not Congruent (Probe appears the same side of the neutral face).

The amount of time taken to respond indicating which side of the screen the participant felt the probe was on was recorded. Once all of the data was collected, the data was cleaned. This meant first removing any incorrect answers, and secondly, the average response time of all trials was taken per participant. Following this, any trial response time that was greater or less than two standard deviations away from the average time was removed. This is standard practice for this experimental paradigm and is seen in the literature [43, 54, 58, 80, 95, 103]. The average times of each of the four categories over all the trials are shown in Table F.2. Each participant had 256 trials.

As is shown in Table F.2 each set of trials has the same number of correct responses and all trials have response times within a 15 millisecond response time. Interestingly, in Table F.2, when Angry Congruent

Table F.2: Average number of correct responses with their associated average response times for the Dop-Probe.

	Average number of correct trials (rounded)	Average Response time (milliseconds)
Angry Congruent	59	459.3
Angry Not Congruent	59	469.75
Happy Congruent	59	469.67
Happy Not Congruent	59	456.25

and Angry Not Congruent are compared, there is a difference of 10.45 milliseconds with the Angry Not Congruent being slower. Those 10.45 milliseconds could be attributed to threat-related bias and the participant finding it hard to break the concentration from the negative image. This theory breaks down however when examining the Happy Congruent and Happy Not Congruent. One would expect that the Happy Congruent response times would be faster than the Happy Not Congruent given the fact that there is no need to disengage from threatening material. However, as is seen in Table F.2, this is not the case as there are over 12 milliseconds in the difference between Happy Congruent and Happy Not Congruent. The 10–12 milliseconds of a difference between Happy Congruent and Happy Not Congruent is negligible and falls well within the range of a chance and so it appears there was no threat-related bias detected in our population using the Dot Probe paradigm.

It was thought that perhaps there was a confounding effect[5] and that the trials should be examined in sequential sets rather than as a whole. The Dot Probe experiment was broken down into four sections:

- Section one: first quarter of trials.

5 A confounding effect is one where the effect may not be observable at the beginning of the experiment but may be observed towards the end of the experiment

- Section two: second quarter of trials.
- Section three: third quarter of trials.
- Section four: fourth quarter of trials.

Breaking the data into these sections allowed a block-level examination with more of a focus on the later blocks. Table F.3 shows the results of the trials when broken into the four sections. Each section contained 64 trials. When the trials were broken into four sections, examining the difference in the Angry Congruent and Angry Not Congruent trials in each of the Sections, the Angry Not Congruent trials were slower. This demonstrates a possible difficulty for participants to disengage after seeing the negative faces. However, similar to the other sets of analysis these differences again were not significant with the average difference being 8 milliseconds.

When examining Table F.3 the objective is to compare the congruence's within each emotion. It is expected that the Not Congruent block would be slower than the Congruent block given that it is expected that it is harder to disengage with the emotive image and shift attention to the probe. Examining Block 1 and the Angry Congruent vs Angry Not Congruent, there is a difference of 19 milliseconds between them. This difference is not replicated through Blocks 2,3 and 4. This initial difference could be due to the participants initially getting used to the Dot-Probe paradigm.

F.6.3 EMOTIONAL STROOP

To analyse the Emotional Stroop of the experiment, the reaction time of participants when responding with the colour that the word was written in was recorded. In total, each participant completed 225 trials in a random sequence. These 225 trials were comprised of 25 Happy, Neutral

Table F.3: Trials broken into 4 sets of 64 trials considered for Dot-Probe analysis

	Red	Average Correct Trials	Average Response Time
Block 1	Angry Congruent	14.48	497.8
	Angry Not Congruent	15.03	516.19
	Happy Congruent	14.53	487.21
	Happy Not Congruent	14.43	485.56
Block 2	Angry Congruent	14.63	459.53
	Angry Not Congruent	14	459.91
	Happy Congruent	15.3	474.25
	Happy Not Congruent	15.28	456.38
Block 3	Angry Congruent	14.7	448.54
	Angry Not Congruent	15.35	456.7
	Happy Congruent	14.98	466.64
	Happy Not Congruent	14.55	431.78
Block 4	Angry Congruent	15.35	425.8
	Angry Not Congruent	14.45	432.71
	Happy Congruent	14.5	440.73
	Happy Not Congruent	14.73	436.06

and Negative words presented in three different colours. Participants were presented with all combinations.

Once all of the data was collected, the data was cleaned. This involved first removing any incorrect answers from the data. Secondly, the average response time of all trials was taken per participant. This was done to so that any trial response time that was greater or less than two standard deviations away from the average time was removed. This is done to ensure the data that is presented is not skewed. The average times of each of the three emotions with the three colours are shown in Table F.4.

Table F.4: Average number of correct responses with their associated average response times for the Emotional Stroop.

	Red	Green	Blue
Negative Words	788.48	744.93	773.66
Neutral Words	824.99	763.91	750.06
Positive Words	758.1	740.35	752.83

Table F.4 shows the average times of each emotion and colour. The average response time of all trials was 759 milliseconds with a standard deviation of 15 milliseconds. These response times are extremely close in time however, the neutral red words are a complete outlier. This is extremely unusual as it suggests that the neutral words were harder to disengage with than the red negative words. It was expected that the red negative words would have the longest reaction times with positive blue words having the least. This is not the case here. Removing the outlier in the data, the red neutral words, the red negative words has the longest response time, however, the 15 milliseconds of a standard deviation, is too little of a difference to be a significant difference.

F.7 SUMMARY

The main aim of the international collaboration was to find an experimental paradigm that could be used to validate the Muse four-channel EEG headband against a EEG. The need to validate the Muse stemmed from wanting to measure cognitive signals of anxiety in students learning to programme. It was decided that a frontal alpha asymmetry study would be conducted and so the Dot Probe and the Emotional Stroop were chosen.

The paradigms were initially tested on a small scale. They appeared to be promising through the observed small effects. A behavioural study was undertaken at Maynooth University. Following the initial analysis, there were no obvious threat-related biases in either the Dot Probe and the Emotional Stroop. This was disappointing as the participants had been negatively primed using the IAPS images.

A deeper analysis of the Dot Probe was carried out to investigate if there was a presence of a confounding effect. Closer analysis revealed this was not the case. Given the lack of results, and the literature stating that it is difficult to find the desired effects in a normal population, neither experiments were investigated further in the project. As the Muse headband could not be validated in a normal population it was not utilised in this project. This meant that anxiety could not be measured using a wireless EEG headband.

While the experiments were not a success, the international collaboration provided invaluable guidance of conducting experiments.

G

ETHICAL APPROVAL

G.1 ETHICAL APPROVAL

MAYNOOTH UNIVERSITY RESEARCH ETHICS COMMITTEE
MAYNOOTH UNIVERSITY,
MAYNOOTH, CO. KILDARE, IRELAND

Maynooth University
National University of Ireland Maynooth

Dr Carol Barrett
Secretary to Maynooth University Research Ethics Committee

15 October 2015

Dr Susan Bergin/ Mr Keith Nolan
Department of Computer Science
Maynooth University

Application for ethical approval for the following:
Application Reference Number: BSRESC-2015-017
Ethical Review Level: Tier 2, Criteria 1
Project title: An Investigation on the Role of Anxiety when Learning to Program

Dear Susan and Keith,

The above application was reviewed under Tier 2: Expedited Review and you have been granted ethical approval to carry out the project described in your application, reference number BSRESC-2015-017.

Any deviations from the project details submitted to the ethics committee will require further evaluation. This ethical approval will expire on 31 October 2020.

Thank you for participating in this process.

Kind Regards,

Dr Carol Barrett
Secretary, Maynooth University Research Ethics Committee

BSRESC-2015-017

MAYNOOTH UNIVERSITY RESEARCH ETHICS COMMITTEE
MAYNOOTH UNIVERSITY,
MAYNOOTH, CO. KILDARE, IRELAND

Maynooth University
National University of Ireland Maynooth

Dr Carol Barrett
Secretary to Maynooth University Research Ethics Committee

29 November 2017

Keith Nolan
Department of Computer Science
Maynooth University

Project entitled: An Investigation on the Role of Anxiety when Learning to Program

Dear Keith,

The above project has been evaluated under Tier 2 process: expedited review and we would like to inform you that ethical approval has been granted.

Any deviations from the project details submitted to the ethics committee will require further evaluation. This ethical approval will expire on 30 November 2019.

Kind Regards,

Dr Carol Barrett
Secretary,
Maynooth University Research Ethics Committee

C.c. Dr Susan Bergin, Department of Computer Science
Dr Aidan Mooney, Department of Computer Science

Reference Number
BSRESC-2017-021

MAYNOOTH UNIVERSITY RESEARCH ETHICS COMMITTEE
MAYNOOTH UNIVERSITY,
MAYNOOTH, CO. KILDARE, IRELAND

Maynooth University
National University of Ireland Maynooth

Dr Carol Barrett
Secretary to Maynooth University Research Ethics Committee

13 December 2018

Keith Nolan
Department of Computer Science
Maynooth University

Re: Project entitled: An Investigation on the Role of Anxiety when Learning to Program

Dear Keith,

The above project has been evaluated under Tier 2 process: expedited review and we would like to inform you that ethical approval has been granted.

Any deviations from the project details submitted to the ethics committee will require further evaluation. This ethical approval will expire on 31 December 2019.

Kind Regards,

Dr Carol Barrett
Secretary,
Maynooth University Research Ethics Committee

C.c. Dr Aidan Mooney, Department of Computer Science
Dr Susan Bergin, Department of Computer Science

Reference Number
SRESC-2018-114

H

CONSENT AND INFORMATION SHEET

H.1 SAMPLE CONSENT AND INFORMATION SHEET

Information Sheet

Purpose of the Study. I am Keith Nolan, a doctoral student, in the Department of Computer Science, Maynooth University. As part of the requirements for PhD, I am undertaking a research study` under the supervision of Dr Susan Bergin and Dr Aidan Mooney.

The research study is a survey which examines the state of anxiety within the Computer Science community. Mr Keith Nolan has conducted research in the area and has found that Computer Science students do experience anxiety due to the solitary nature of Computer Science, the use of computers and other factors (should you wish to read more, please see https://dl.acm.org/citation.cfm?id=2999557). Given that we know Computer Science students experience anxiety, we would like to determine how anxious they are when compared to the average college student. By doing this we can justify further research into the area which may inform educators on how they could change their modes of teaching and assessment which might help Computer Science students in the future.

What will the study involve? The study will require you to complete standard questionnaires. The questionnaires ask simple questions regarding the emotions that you experience and how you usually respond to them and also gather background biographical and academic information. Your answers on these questionnaire will not allow us to make psychological assessments of you.

Who has approved this study? This study has been reviewed and received ethical approval from Maynooth University Research Ethics committee. You may have a copy of this approval if you request it.

Why have you been asked to take part? You have been asked because you are a student taking Computer Science as part of your degree programme. Should you wish to take part in this study there will a 24 hour cooling off period should you wish to withdraw from the study before completing it.

Do you have to take part?

No, you are under no obligation whatsoever to take part in this research. However, we hope that you will agree to take part and give us some of your time to complete a short questionnaire. It is entirely up to you to decide whether or not you would like to take part. If you decide to do so, you will be asked to sign a consent form and given a copy and the information sheet for your own records**. If you decide to take part, you are still free to withdraw at any time without giving a reason and/or to withdraw your information up until such time as the research findings are anonymised (March 1st 2019). A decision to withdraw at any time, or a decision not to take part, will not affect your relationships with the Department of Computer Science or any of its staff members.**

Should you exclude yourself from the study?

While we envisage no adverse effects, we ask if you are under the age of 18 to self-exclude yourself. In addition to this, if you suffer from an anxiety-related condition, diagnosed or otherwise, we ask you self-exclude yourself.

What information will be collected? Information regarding the emotions that you experience and how you usually respond to them and background biographical and academic information will be collected. Your student number will be collected which will allow us to contact you in if we feel a full debrief is required. However all data will be anonymously by March 1st by Mr Keith Nolan.

What will happen after the study? Following the study, the data will be collected and analysied. Should the researchers determine that a follow-up debrief in necessary we will use your student number that was collected in the study to contact you.

Will your participation in the study be kept confidential? Yes, all information that is collected about you during the course of the research will be kept confidential. No names will be used at any time. All hard copy information will be held in a locked cabinet at the researchers' place of work, electronic information will be encrypted and held securely on MU PC or servers and will be accessed only by Mr Keith Nolan, Dr Susan Bergin and Dr Aidan Mooney.

No information will be distributed to any other unauthorised individual or third party. If you so wish, the data that you provide can also be made available to you at your own discretion.

'It must be recognised that, in some circumstances, confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of investigation by lawful authority. In such circumstances the University will take all reasonable steps within law to ensure that confidentiality is maintained to the greatest possible extent.'

What will happen to the information which you give? All the information you provide will be kept at Maynooth University in such a way that it will not be possible to identify you. On completion of the research, the data will be retained on the MU server. After ten years, all data will be destroyed (by the PI). Manual data will be shredded confidentially and electronic data will be reformatted or overwritten by the PI in Maynooth University.

What will happen to the results? The research will be written up and presented as a publication at national and international conferences and may be published in scientific journals and in the PhD thesis of Mr Keith Nolan. A copy of the research findings will be made available to you upon request.

What are the possible disadvantages of taking part? I don't envisage any negative consequences for you in taking part. However given the nature of some of the questions, should you experience any negative feelings during the survey, we encourage you to stop the survey. Should you need to talk you can speak to us directly or go to the Maynooth University Counselling services.

What if there is a problem? At the end of the questionnaire, if you experience any distress following the questionnaire, contact my supervisors Dr Susan Bergin or Dr Aidan Mooney (susan.bergin@mu.ie or aidan.mooney@mu.ie) if you feel the research has not been carried out as described above. Following this, you may contact Maynooth University Counselling Service office hours at (01) 7083554 should you need to.

Any further queries? If you need any further information, you can contact me at keith.nolan@mu.ie

If you agree to take part in the study, please complete and sign the consent form overleaf.

Thank you for taking the time to read this

Consent Form

I……………………………………agree to participate in Keith Nolans's research study titled An investigation of the role of anxiety when learning to program.

Please tick each statement below if you agree :

The purpose and nature of the study has been explained to me verbally & in writing. I've been able to ask questions, which were answered satisfactorily. ☐

I am participating voluntarily. ☐

I understand that I can withdraw from the study, without repercussions, at any time, whether that is before it starts or while I am participating. ☐

I understand that I can withdraw permission to use the data until it is irreversibly anonymized – March 1st 2019. ☐

It has been explained to me how my data will be managed. ☐

I understand the limits of confidentiality as described in the information sheet ☐

I understand that my data, in an anonymous format, may be used in further research projects and any subsequent publications if I give permission below: ☐

I agree for my data to be used for further research projects ☐
I do not agree for my data to be used for further research projects ☐

Signed……………………………………. Date………………..

Participant Name in block capitals ……………………………………………..

I the undersigned have taken the time to fully explain to the above participant the nature and purpose of this study in a manner that they could understand. I have explained the risks involved as well as the possible benefits. I have invited them to ask questions on any aspect of the study that concerned them.

Signed……………………………………. Date………………..

Researcher Name in block capitals ……………………………………………..

If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the Maynooth University Ethics Committee at research.ethics@mu.ie or +353 (0)1 708 6019. Please be assured that your concerns will be dealt with in a sensitive manner.

For your information the Data Controller for this research project is Maynooth University, Maynooth, Co. Kildare. Maynooth University Data Protection officer is Ann McKeon in Humanity house, room 17, who can be contacted at ann.mckeon@mu.ie. Maynooth University Data Privacy policies can be found at https://www.maynoothuniversity.ie/data-protection.

Two copies to be made: 1 for participant, 1 for PI

I

PARTICIPANT QUESTIONNAIRE

I.1 PARTICIPANT QUESTIONNAIRE

Maynooth University
National University of Ireland Maynooth

Participant Questionnaire

Subject's Identifier: _________________

Date: _____________ Starting Time: _________

Subject's Gender: M ☐ F ☐

What age are you? 17-19 ☐ 20-22 ☐ 23+ ☐

Does the subject wear glasses / contact lens: Yes / No

If yes, do they have corrected to normal vision? Yes / No

Which is the subject's dominant hand? Left hand / Right hand

Does the subject classify themselves as a:

Morning person ☐

Evening person ☐

Is the subject a native English speaker? Yes / No

If not, what level of English have they achieved? ________________________

Observers Signature: _____________________________________