



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# Characteristics of Problem Solving Success in Physics



*Marsali Wallace*

A thesis submitted in fulfilment of the requirements  
for the degree of Doctor of Philosophy  
to the  
University of Edinburgh  
May 2014

# Abstract

Skills in problem solving, including finding and applying the appropriate knowledge to a problem, are important learning outcomes from the completion of a Physics degree at University. This thesis investigates the characteristics of successful and unsuccessful novice University students solving problems in Physics in various contexts. Gaining an insight into student behaviour can clarify areas of weakness and potentially provide research based instructional strategies in these contexts.

Access to external information during problem solving, such as the Internet, is becoming an increasingly relevant research area, as students use resources for homework questions and then in employment after University. Three chapters (Chapters 3-5) investigate individual novice problem solving with and without resources, such as a textbook. Participants were from introductory years one and two of Undergraduate study at University. The results from this chapter show successful and unsuccessful approaches by students to multi-step problems. One notable result is that unsuccessful students demonstrated an inability to apply the appropriate physics concepts, with or without the availability of resources. These results have implications for the skills required in closed and open-book exams.

Three chapters of the thesis focus on the analysis of Peer Instruction (Chapters 6-8), an instructional method designed to improve conceptual understanding. Peer Instruction was used with a first year Introductory University class. Technical word use was not associated with success on Peer Instruction questions. Conversations were also analysed qualitatively. The results reflect diversity in reasoning regardless of correctness on the question. Some recommendations for the implementation of Peer Instruction are presented.

The thesis is organised as follows. A literature review was conducted in

---

relevant areas of study and is presented to set the context of the work. Three chapters report the study with novice individuals solving multi-step problems with and without resources. Three further chapters investigate successful and unsuccessful Peer Instruction discussions in Physics. The final results chapter (Chapter 9) presents a study of a group of experts solving physics problems. Overall successful and unsuccessful problem solving strategies were compared, as well as preliminary comparisons between expert and novice behaviour when solving physics problems.

# Lay Summary

Students are expected to develop strong skills in problem solving from studying Physics at University. Skills in problem solving including finding and applying the appropriate knowledge to a problem and knowing where to look up information if it cannot be remembered. These skills will also be useful in future employment. However, there is still much to learn in terms of problem solving behaviour. This can in turn inform instructional approaches at University to target student weaknesses in problem solving in different contexts.

One context, which is becoming increasingly relevant, is that of access to information during problem solving. Technology and access to information to help solve problems are widely available after University. It is therefore important to study how students use resources, such as the Internet, during problem solving. Problem solving behaviours were studied with two groups of students, one with and one without access to resources. Approaches were analysed with respect to suggested problem solving models, but also more openly from what emerged from listening to students thinking aloud whilst solving problems. Suggested problem solving models include existing instructional approaches, which try to encourage expert-like problem solving behaviour. For example, experts tend to reason qualitatively before taking a quantitative approach, such as applying an equation. Therefore some problem solving models include a describe the physics stage near the beginning of a problem solution.

Peer Instruction is an instructional techniques used to teach conceptual knowledge in order to support the describing the physics stage in a solution. Peer Instruction at University requires students to read and assimilate knowledge before their lectures. During lectures conceptual questions are posed on this material to which students discuss the answers. This thesis aimed to look at characteristics of successful and unsuccessful discussions during Peer Instruction

---

and also to feedback to the course whether questions were working and what students were finding difficult. Insight into the Peer Instruction process found diverse conversations and levels of reasoning. Some recommendations for the implementation of Peer Instruction are presented.

The final results chapter details a preliminary study with experts solving physics problems. Overall successful and unsuccessful problem solving strategies were compared, as well as preliminary comparisons between expert and novice behaviour when solving physics problems. Interesting insights into students problem solving behaviour have implications for open and closed book exams, as well as general problem solving instruction, for example with respect to applying problem solving models.

# Declaration

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The pilot study in Section 7.1 of Chapter 7 was written up for publication with a co-author R. Galloway. The publication reference is: ‘Using Smartpen Technology to Observe Student Discussions in Physics Lectures’ by M. Wallace and R. Galloway, as part of the proceedings for HEA STEM Conference 12 -13 April 2012, Imperial College, London and The Royal Geographical Society.

The candidate conducted the work for this study and the paper has been re-written for inclusion in this Chapter.

*M.B. Wallace*

May 2014

# Acknowledgements

I would like to thank my supervisor Ross Galloway for his support and conversations on problem solving. I have enjoyed the journey of research with its surprising results. I am also grateful to the past and present members of the EdPER group at Edinburgh University, including Judy Hardy and Simon Bates who have made my PhD what it is, and also my colleges Robyn Donnelly and Alison Kay for physics education and statistics conversations. Thanks also to Alison Kay and Ross Galloway for being involved in coding data, Anna Wood for reading over some of the Peer Instruction work and many thanks to Amanda Amos for a conversation on qualitative analysis at the right time.

Special and loving thanks to my friends and family for listening to my PhD updates, reading over chapters and provided invaluable feedback; my mum, Anne Maree, for her support and discussions and my Dad, David, for his philosophical encouragement. Also thanks to Neil Masson, Kara Gibson, Duncan Wallace and Liam Humphreys for their red pen on the final version.

Finally, thanks to the Principal's Teaching Award Scheme at the University of Edinburgh for funding the smartpens in the Peer Instruction project and especially the Principal's Career Development Scholarship for my funding and giving me the opportunity to apply my research to teaching activities.



# Contents

<b>Abstract</b>	<b>i</b>
<b>Lay Summary</b>	<b>iii</b>
<b>Declaration</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vi</b>
<b>Contents</b>	<b>vii</b>
<b>List of Figures</b>	<b>xii</b>
<b>List of Tables</b>	<b>xiv</b>
<b>1 The Context: Problem Solving in Physics</b>	<b>1</b>
1.1 Definition of Problem Solving . . . . .	2
1.2 Problem Types . . . . .	3
1.3 Experts and Novices . . . . .	5
1.4 Problem Representation . . . . .	8
1.5 Knowledge Structures . . . . .	10
1.6 Problem Solving Models . . . . .	13
1.7 Working in Groups: Use of Language and Scientific Reasoning . .	19
1.8 Peer Instruction . . . . .	22
1.8.1 Theoretical foundations . . . . .	23
1.8.2 Studies of Peer Instruction . . . . .	23
1.9 Problem Solving in Summative Assessments: Open and Closed Book Exams . . . . .	28
1.10 Summary . . . . .	29
<b>2 Methodological Approaches</b>	<b>31</b>
2.1 Collecting Qualitative Data . . . . .	32
2.1.1 The Think Aloud Technique . . . . .	32
2.1.2 Retrospection . . . . .	33
2.1.3 Structured questions . . . . .	34

2.1.4	Observations and Field Notes . . . . .	35
2.1.5	The Smartpen . . . . .	35
2.2	Analysing Qualitative data . . . . .	36
2.2.1	Quantitative Methods . . . . .	36
2.2.2	Qualitative Methods . . . . .	37
2.2.3	Validating and Consolidating Coding . . . . .	38
2.3	Structure of the Thesis . . . . .	39
2.3.1	Individual Problem Solving Chapters . . . . .	39
2.3.2	Peer Instruction Chapters . . . . .	40
2.3.3	Expert Group Work Chapter . . . . .	40
2.3.4	Chosen Methods . . . . .	40
<b>3</b>	<b>Individual Problem Solving: Aims and Methodology</b>	<b>42</b>
3.1	Aims and Hypothesis . . . . .	43
3.2	Open and Closed-Book Exams . . . . .	45
3.3	Data Collection . . . . .	45
3.3.1	1B Study Design . . . . .	46
3.3.2	2B Study Design . . . . .	47
3.3.3	The Role of the Researcher . . . . .	47
3.4	The Questions . . . . .	48
3.4.1	The Warm Up Problem . . . . .	48
3.4.2	Main Question One: The Drawbridge Question . . . . .	49
3.4.3	Main Question Two: Army Cadet Question . . . . .	50
3.5	Analysis . . . . .	51
3.6	Quantitative Coding of the 1B and 2B data . . . . .	51
3.6.1	Model 1: The Minnesota Model . . . . .	52
3.6.2	Model 2: Problem Solving Behaviours . . . . .	53
3.6.3	Model 3: Representations . . . . .	55
3.6.4	Applying the Models . . . . .	55
3.6.5	Inter-rater Reliability: Validation of the Coding Scheme . . . . .	56
3.6.6	Sources of Error . . . . .	59
3.6.7	Methodological Assessment . . . . .	60
3.7	Analysis After Coding . . . . .	61
3.8	Qualitative Analysis of the 1B and 2B data . . . . .	63
3.8.1	Data Analysis 2B resources specifically . . . . .	65
3.8.2	4th and 5th year students on the Physics Skills course . . . . .	65
<b>4</b>	<b>Individual Problem Solving: Modelling Problem Solving Behaviour</b>	<b>67</b>
4.1	1B Quantitative Study Results . . . . .	67
4.1.1	Temporal maps from the drawbridge question . . . . .	69
4.1.2	Temporal maps from the army cadet question . . . . .	70
4.1.3	Transition Patterns . . . . .	71

---

4.2	2B Quantitative Study Results . . . . .	75
4.2.1	Transition Patterns . . . . .	78
4.2.2	Resources Use . . . . .	81
4.3	Results From Both 1B and 2B Studies . . . . .	83
4.4	Summary . . . . .	85
<b>5</b>	<b>Individual Problem Solving: Qualitative Analysis</b>	<b>86</b>
5.1	1B Qualitative Study Results . . . . .	86
5.1.1	The Drawbridge Question . . . . .	86
5.1.2	The Army Cadet Question . . . . .	90
5.2	1B Discussion . . . . .	94
5.3	2B Qualitative Study Results . . . . .	95
5.3.1	The Drawbridge Question . . . . .	95
5.3.2	The Army Cadet Question . . . . .	100
5.4	2B Discussion . . . . .	105
5.4.1	Resource Use . . . . .	106
5.5	Discussion . . . . .	108
5.6	Further Investigation: 4th and 5th Year Students . . . . .	113
5.6.1	Student A . . . . .	113
5.6.2	Student B . . . . .	114
5.6.3	Summary . . . . .	117
5.7	Open and Closed-Book Exams . . . . .	117
5.8	The Minnesota Model . . . . .	118
5.9	Conclusions and Implications for Teaching . . . . .	120
5.10	Future research . . . . .	121
<b>6</b>	<b>Peer Instruction: Aims and Methodology</b>	<b>122</b>
6.1	Aims . . . . .	123
6.2	Cohort Background . . . . .	124
6.3	Data Collection Methodology . . . . .	125
6.3.1	Constructing the Study . . . . .	126
6.3.2	Deployment of the study and Timeline . . . . .	127
6.3.3	Running the Study and Collecting Data . . . . .	127
6.4	Risks . . . . .	127
6.5	Limitations of this Method . . . . .	128
6.6	Statistics of Smartpen Use . . . . .	128
6.7	Investigating the Data Collection Method . . . . .	129
6.8	Peer Instruction in Context . . . . .	132
6.9	Summary . . . . .	134
<b>7</b>	<b>Peer Instruction: Use of Technical Language</b>	<b>135</b>
7.1	The Pilot Study . . . . .	136
7.1.1	Pilot Study Method . . . . .	136

7.1.2	Pilot Study Results . . . . .	137
7.1.3	Conclusions of the Pilot Study . . . . .	140
7.2	Methodology . . . . .	141
7.2.1	Developing the List of Technical Words . . . . .	141
7.2.2	Criteria for Coding . . . . .	142
7.2.3	Determining percentage agreement between coders . . . . .	143
7.2.4	Counting Technical Words and other variables . . . . .	145
7.2.5	Matching to Responses for Each Question . . . . .	145
7.2.6	Data analysis . . . . .	146
7.3	Strengths and Limitations of the Study . . . . .	147
7.4	Results . . . . .	148
7.4.1	Inter-rater Agreement . . . . .	148
7.4.2	Technical Word Use . . . . .	149
7.5	Discussion . . . . .	154
7.6	Conclusions . . . . .	156
7.7	Future Research . . . . .	157
<b>8</b>	<b>Peer Instruction: Further Analysis</b>	<b>158</b>
8.1	Method . . . . .	159
8.1.1	Qualitative Analysis . . . . .	159
8.1.2	Inclusion of Technical Language . . . . .	161
8.2	Strengths and Limitations . . . . .	163
8.3	Results . . . . .	164
8.3.1	Case Study 1: Acceleration of a ball thrown vertically . . . . .	164
8.3.2	Case Study 2: Rocks Item A . . . . .	168
8.3.3	Case Study 3: Friction . . . . .	172
8.3.4	Case Study 4: Work done by a robot . . . . .	176
8.4	Summary . . . . .	182
8.4.1	Other Categories . . . . .	182
8.5	Discussion . . . . .	185
8.5.1	Intellectual position of the student . . . . .	187
8.5.2	Technical word use . . . . .	188
8.5.3	Instructional Strategies . . . . .	189
8.6	Conclusions . . . . .	190
8.6.1	Future Research . . . . .	191
<b>9</b>	<b>Expert Group Problem Solving</b>	<b>192</b>
9.1	Method . . . . .	193
9.2	The Questions . . . . .	194
9.3	Method - Question 1 . . . . .	195
9.3.1	Development of the coding scheme . . . . .	195
9.3.2	Error in coding . . . . .	198
9.4	Method - Question 2 . . . . .	198

9.5	Strengths and Limitations . . . . .	199
9.6	Results . . . . .	200
9.6.1	Macrobehaviours - Experts, Question 1 . . . . .	200
9.6.2	Macrobehaviours - Post Graduates, Question 1 . . . . .	203
9.6.3	Resources Use - Experts, Question 2 . . . . .	205
9.7	Discussion . . . . .	206
9.7.1	Coding and the Minnesota Model . . . . .	208
9.7.2	Recording Technologies in Group Work . . . . .	209
9.8	Conclusions . . . . .	209
9.9	Future Research . . . . .	210
<b>10</b>	<b>Conclusions</b>	<b>211</b>
10.1	Summary . . . . .	211
10.2	Comparing Expert and Novice Behaviour . . . . .	213
10.3	A New Model . . . . .	215
10.4	Authentic Methods . . . . .	217
10.5	Instructional Recommendations . . . . .	218
10.6	Future Research . . . . .	219
	<b>Appendices</b>	<b>222</b>
<b>A</b>	<b>Individual Problem Solving and Resources Use</b>	<b>222</b>
A.1	Physics Questions . . . . .	222
A.2	Answers to Physics Questions . . . . .	224
A.3	Allocation of Marks . . . . .	227
A.4	Master Sheet for Researcher - 1B Students . . . . .	229
A.4.1	Overall Plan . . . . .	229
A.4.2	Welcome and warm up question . . . . .	230
A.4.3	First Main Diagram Question . . . . .	231
A.4.4	Second Main Question . . . . .	232
A.4.5	Retrospection . . . . .	233
A.4.6	Validity of think aloud . . . . .	233
A.5	Consent Form . . . . .	235
A.6	Temporal Maps . . . . .	237
A.7	Transition Maps . . . . .	252
<b>B</b>	<b>Smartpen Use: Code of Conduct</b>	<b>267</b>
<b>C</b>	<b>The List of Technical Words</b>	<b>270</b>
<b>D</b>	<b>Expert Group Problem Solving</b>	<b>273</b>
D.1	Question and Answers from Expert Group Problem Solving . . . . .	274
	<b>Bibliography</b>	<b>277</b>

**Publications**

**286**

# List of Figures

3.1	Drawbridge question diagram . . . . .	49
4.1	1B Student 5 Drawbridge question temporal map . . . . .	69
4.2	1B Student 4 Drawbridge question temporal map . . . . .	69
4.3	1B Student 6 Army cadet question temporal map . . . . .	70
4.4	1B Student 3 Army cadet question temporal map . . . . .	70
4.5	1B Transitions Student 5 Drawbridge question . . . . .	72
4.6	1B Transitions Student 6 Army Cadet question . . . . .	72
4.7	1B Transitions Student 4 Drawbridge question . . . . .	73
4.8	1B Transitions Student 3 Army Cadet question . . . . .	73
4.9	2B Student 4 Drawbridge question temporal map . . . . .	76
4.10	2B Student 2 Drawbridge question temporal map . . . . .	76
4.11	2B Student 6 Army cadet question temporal map . . . . .	77
4.12	2B Student 7 Army cadet question temporal map . . . . .	77
4.13	2B Transitions Student 4 Drawbridge question . . . . .	78
4.14	2B Transitions Student 6 Army cadet question . . . . .	79
4.15	2B Transitions Student 2 Drawbridge question . . . . .	79
4.16	2B Transitions Student 7 Army cadet question . . . . .	80
4.17	Resources use on the drawbridge question . . . . .	82
4.18	Resources use on the army cadet question . . . . .	83
4.19	Complex Score 1 and exam score . . . . .	84
4.20	Normalised complexity score 2 and exam score . . . . .	84
4.21	Normalised number of transitions and exam score . . . . .	85
5.1	Gerace's knowledge structure model . . . . .	111
5.2	Questions A6 and A7, 2010 paper . . . . .	113
5.3	B4 question, 2008 paper . . . . .	115
5.4	B5 question, 2008 paper . . . . .	116
5.5	Re-ordered Minnesota Model problem solving strategies . . . . .	119
6.1	Scores of students who volunteered to use a smartpen compared to the whole year group . . . . .	130
6.2	PI marks compared to Post FCI score . . . . .	133

7.1	Mean Number of Technical words used for each clicker response type. . . . .	150
7.2	Mean $h_{index}$ value for each clicker response type. . . . .	151
7.3	Mean number of different technical words used for each clicker response type . . . . .	152
7.4	Mean number of technical words used in each clicker question and normalised class gain . . . . .	153
7.5	Average number of technical words and class gain . . . . .	154
8.1	Stages of Qualitative Analysis . . . . .	160
8.2	Case Study 1: Clicker question . . . . .	164
8.3	Case Study 2: Clicker question . . . . .	169
8.4	Case Study 3: Clicker question . . . . .	173
8.5	Case Study 4: Clicker question . . . . .	176
9.1	Macrobehaviours of the Experts . . . . .	201
9.2	Macrobehaviours of the Post Graduates . . . . .	204
10.1	Summary of Expert and Novice Behaviour . . . . .	214
10.2	A New Model of Problem Solving Behaviour . . . . .	216
A.1	Drawbridge question . . . . .	223
A.2	Diagram of the drawbridge question . . . . .	225
A.3	Diagram of the army cadet question . . . . .	226
D.1	Diagram for one of the expert problems . . . . .	274



# List of Tables

1.1	Well-structured and ill-structured problems . . . . .	4
3.1	Summary of first and second year courses in Physics at Edinburgh University . . . . .	43
3.2	Students participating in the 1B and 2B studies . . . . .	46
3.3	Summary of the individual problem solving research . . . . .	51
3.4	1B inter-rater agreement table . . . . .	57
3.5	1B Average initial inter-rater agreement . . . . .	57
3.6	2B inter-rater agreement table . . . . .	59
3.7	2B Average inter-rater agreement . . . . .	59
4.1	1B Marks . . . . .	68
4.2	1B Number of transition types . . . . .	74
4.3	2B Marks and resources use . . . . .	75
4.4	2B Number of transition types . . . . .	81
4.5	Spearman correlation values . . . . .	83
6.1	Smartpen holders' FCI Scores compared to the whole class distri- bution . . . . .	131
6.2	Smartpen holders' course marks . . . . .	132
6.3	Summary of the Peer Instruction research . . . . .	134
7.1	Pilot study: Case study 1 . . . . .	138
7.2	Pilot study: Case study 2 . . . . .	139
7.3	Inter-rater agreement method . . . . .	144
7.4	Overlap and All words example . . . . .	145
7.5	Inter-rater agreement and data discarded . . . . .	149
8.1	Case Study 1: Technical word use and qualitative analysis . . . . .	165
8.2	Case Study 2: Technical word use and qualitative analysis . . . . .	169
8.3	Case Study 3: Technical word use and qualitative analysis . . . . .	173
8.4	Case Study 4: Technical word use and qualitative analysis . . . . .	177
8.5	Summary of Case Studies . . . . .	182
9.1	Categories for coding expert behaviour . . . . .	197

# Chapter 1

## The Context: Problem Solving in Physics

Good problem solving skills are considered one of the key outcomes from the completion of a physics degree at university [1]. A main focus of Physics Education Research (PER) is an exploration of problem solving, including how students solve problems and techniques for teaching problem solving. Graduates from physics degrees are expected to be proficient in solving a range of problems, from real-world problems to conducting their own laboratory experiments. High quality problem solving skills also benefit students in employment post University. There have been several exciting developments in the understanding of novice and expert behaviour, such as the use of diagrams and other representations [2], modelling approaches [3, 4] and looking at non-discipline specific problem solving skills [5]. To investigate problem solving processes qualitative techniques, such as the think aloud protocol [6], are being used increasingly by the PER community.

The broad aim of this thesis is to develop a better understanding of student and expert behaviour by investigating problem solving in different guises. This can inform instruction in teaching for both individual and peer problem solving. Potential areas of research to further our understanding of student problem solving include the characteristics exhibited in successful and unsuccessful problem solving solutions, the strategies used by students, and the ways in which students talk to each other to solve physics problems.

The following review aims to provide an overview of the main themes in the extensive problem solving literature. The general landscape of problem

solving behaviour is constructed by a review of problem types, expert and novice behaviour and problem solving models. The review will also focus on areas relating directly to areas covered in this thesis, particularly an analysis of group work, peer instruction and open and closed book exams. Areas where the literature reviewed could be extended are discussed in Section 1.10. Chapter 2 contains an overview of methodological approaches for studying problem solving, a description of the structure of the research presented in this thesis (Section 2.3) and the methods chosen to examine these topics. Chapters 3 and 6 cover aims and methodology, with results and discussion in Chapters 4, 5, 7 and 8. Chapter 9 is self-contained and provides the aims, methodology, results and discussion of an additional area of research, as introduced in more detail in Section 2.3. Chapter 10 brings key areas of the thesis together and provides distilled instructional recommendations and overall conclusions.

## 1.1 Definition of Problem Solving

In order to explain the foundations for research and to clarify the terminology used throughout this work, it is important to define the concept of problem solving. Robertson [7] states that a problem signifies a goal where it is not immediately obvious how this goal is to be reached. Furthermore, the process of going from some initial state to the desired goal state may require passing through intermediate problem states [8, 9, 10]. The definition of problem solving is particularly important when discussing expert and novice problem solving behaviour (see Section 1.3); what is a problem for a novice may not be a real problem for an expert in that they can immediately see how to solve it. Ensuring questions are true problems for each participant is difficult. This may mean using different, more difficult types of questions for experts.

Problem solving can also be discussed in terms of an **answer**, **solution** and **solution process** [8]. Reif defined the solution of a problem as being generated by the solution process [8] and the answer as the product of the solution. In other words, the solution process provides a solution which can lead to a final answer. The solution alone can only provide limited information on the solution process, just as the answer can only provide limited information on the solution. Students are usually asked to show their working in homework exercises and exams to

show their solution and not just the final answer. Solutions can be assessed by traditional exam-style marking schemes or by marking rubrics designed to assess problem solving ability and strategic thinking, as well as correctness [11]. The solution process is more difficult to examine and ways of studying this, such as the think aloud protocol, are discussed in Chapter 2.

## 1.2 Problem Types

Students are expected to develop a range of skills through Physics and one way to do this is the practise of different problem types [12, 13]. The amount of information given on the initial state, process and goal state depends on the problem type. Classifying problems on a spectrum from well-structured to ill-structured is a common theme in defining problem types [12, 13, 14, 15, 16, 17]. Generally a problem is classed as well-structured if the data and outcomes are well-defined in the problem statement and the methods are familiar to the student [12, 13]. Ill-structured problems are those where there is some ambiguity as to the methods to use and there are ill-defined goals [16].

In his well-cited study considering the typology of problems [12] Jonassen gathered hundreds of problems and split them into 11 problem types, ranging from well-structured to ill-structured. Well-structured problems include logical problems and algorithmic problems, such as tasks and exercises, whereas case analysis and design problems are examples of ill-structured problems. Ill-structured problems may need problem structuring and can be real life problems with no correct solution. A full table of these is given in Table 1.1. However, these classifications do not account for other factors such as an indication of the prior knowledge required to solve the problem, whether it involves domain specific or domain general knowledge [7] or how much knowledge is required to solve the problem (classified by Robertson [7] as knowledge rich and knowledge lean problems).

Well-structured and ill-structured problems require different cognitive processes. Ill-structured problems are said to use higher order functions and are considered more difficult [16]. Jonassen [12, 13] and Fortus [17] claim that certain well-structured problem solving skills are needed as sub-skills for being able to do ill-structured problems and that multiple problem types require the use of

Structure	Type
Well-structured	Logical Problems Algorithmic Problems Story Problems Rule Using Problems Decision making problems Trouble Shooting problems Diagnostic Solution problems Strategic Performance Problems Case Analysis Problems Design Problems
Ill-structured	Dilemmas

Table 1.1: Well-structured and ill-structured problems adapted from Jonassen [12]. This table shows a spectrum of problems types from well-structured problems at the top to ill-structured problems at the bottom of the table.

a variety of skills. Jonassen argues that instruction should be designed to suit different problem types, which require a range of skills. There is some debate over whether ill-structured problems are used enough in education to develop the skills they require [5]. Labs and student investigatory project work would be included in this category. However, workshops - where students solve problems in groups each week to complement lectures - will also have a certain number of well-structured problems and exercises in order for the students to be able to master the necessary techniques.

There is also a question about which problem solving skills students will need after University, the inference being that employers want a graduate who has the ability to develop expertise in new skill areas and solve problems with ill-defined goals [18]. If the goal of instruction at University is to develop students skilled in answering the spectrum of well to ill-structured questions then presenting a mix of problem types seems appropriate. Many ways to examine higher-order cognitive processes already exist, including group projects, individual research projects, more difficult problems and open-book exams. The impact of open-book exams on students' problem solving behaviour and the potential skills needed to solve problems in an open and closed book format are examined in Section 1.9 of this review and in Chapters 3, 4 and 5.

## 1.3 Experts and Novices

There has been extensive research comparing expert problem solving behaviour to that of novices. Researchers can use either absolute or relative methods to study expertise [16]. Absolute methods focus on an in-depth analysis of a high-level expert, whereas relative methods involve the comparison of an expert and novice when completing the same task. Hardiman et al. [19] questioned whether the criteria for qualifying as an expert in studies was strict enough. According to Nokes et al. [16] this is of less importance in relative methods. Both methods are used in this thesis; relative methods to compare successful and unsuccessful students, and a study of experts in Chapter 9 to provide an absolute measure of expert behaviour.

Expert and novice differences have been investigated with respect to the specific skill of categorising problems. A seminal paper by Chi et al. in 1981 [20] reported work in which experts and novices categorised physics problems without solving them. It was found that experts grouped problems by physics principles and laws whereas novices categorised problems by surface features of the problem statement. The skill of being able to categorise problems was also studied by Leonard et al. [21]. According to Leonard et al., students who were taught categorisation skills improved in their ability to categorise problems according to principles and indicated improved recall of the principles they learnt, compared to a class not taught categorisation skills. However, neither Chi et al. [20] nor Leonard et al. [21] demonstrated the link between skill of categorising problems and problem solving ability overall. Instead Leonard et al. referred to a previous study by Hardiman et al. [19]. This previous study [19] claims to show a link between problem categorisation and problem solving ability by showing that higher scoring problem solvers use principles in reasoning more than lower scoring problem solvers. However this does not show that teaching categorisation to novices improves problem solving abilities in general. They acknowledge that the use of principles and surface features by novices and experts is not clear cut:

*..the conclusion that novices focus almost exclusively on surface-feature similarity is unwarranted. p.633 [19]*

There is therefore debate over the amount of emphasis placed on the results from the Chi et al. [20] study, a point made by Adams [5]. Adams noted that while the

way experts and novices categorise problems undoubtedly provides an insight into the problem solving process, it seems that improving categorisation skills is only a step towards improved problem solving. In her thesis, Adams found that many other skills were needed for a successful solution, such as beliefs and attitudes [5]. Furthermore, Kohl and Finkelstein [2] did not observe any use of surface features by novices in grouping dynamics and electrostatics problems. They in fact observed them using multiple representations - a technique classed as an expert trait (see Section 1.4 for further discussion on representation use). It is evident, therefore, that this area is not straightforward. Teaching categorisation alone has not been shown to explicitly improve problem solving skills when assessed by a task other than categorisation, but there are clear links between use of principles and problem solving ability [19, 20].

One example of an expert approach is to reason qualitatively at the beginning of the problem in order to define the relevant physics concepts. Van Heuvelen [22] proposes that experts use qualitative representations to understand the problem and to help build the appropriate mathematical relations. Qualitative reasoning is considered an expert-like trait by many authors [21, 22, 23] including those who researched and designed problem solving models. They ordered the behaviours ‘describing the physics’ or ‘analysis’ before ‘implementation’ [3, 24], as discussed in Section 1.6.

Van Heuvelen devised an instructional technique to encourage students to focus on qualitative reasoning at the start of physics problems before being exposed to the more quantitative aspects [22]. Throughout, students were encouraged to use multiple representations such as diagrams and equations, and to evaluate their final solution. The course also included complex case study problems incorporating knowledge from different domains of physics. Van Heuvelen reported improved scores on both qualitative and quantitative questions for the intervention group compared to the traditionally taught group. However as the intention was to reform the whole course it is hard to disentangle the effects of other instructional changes - such as the use of multiple representations, which is generally thought to be an expert skill [22, 25] - from the benefits of reasoning qualitatively.

There is more to expert behaviour than categorisation and qualitative reasoning. These are just the first of seven problem solving stages defined by

Nokes et al. [16], discussed further in Section 1.6. On the whole, Jonassen argues that a good mental problem representation is the key to problem solving [26], including mapping the statement onto prior knowledge, setting the problem space, establishing clear goals and structure, and understanding the information provided and the strategies to use. These approaches will assist with the rest of the solution.

Familiarity with relevant physics problems potentially also helps experts. Jonassen argues that experts are better problem solvers because they understand what to do to solve the problem by recognising the solution [26]. This is related to cueing of appropriate knowledge as discussed by Redish and Sabella [27] and discussed in further detail in section 1.5. Being able to refer to worked examples after years of experience is helpful for a number of reasons, such as recognising which principles apply. This can mean, however, that what is a problem to a novice may be an exercise to an expert. Experts' wealth of past experience makes setting them true problems [28] quite difficult, especially in the area of introductory mechanics, which is often covered in introductory physics courses. Therefore comparing expert to novice behaviour is more complicated than it first seems. It is difficult to know whether this affected tasks, such as the categorisation task by Chi et al. [20], as experts, unlike novices, may have been able to see how to solve the problem whilst categorising it.

To look at this directly, expert behaviour on counter-intuitive, unfamiliar problems was examined by Singh [29]. Twenty physics professors (experts) were given a dynamics problem which they had not come across before, meaning they had no useful intuition, experience or familiarity to aid their solution process. Sixty seven calculus based physics students were also given the question and the problem solving traits of the two groups were compared. All of the professors used memory, past experiences (learning) or real-life experiences at some point in the process. Singh suggested that experience and familiarity of a problem are important factors in experts' problem solving.

*...they(experts) often started by visualizing and analyzing the problem qualitatively and searching for useful conservation principles before resorting to other routes. They were much more likely to draw analogies and map the unfamiliar problem onto a familiar one. They often examined limiting cases; a strategy rarely employed by students.*



p.1106 [29]

However, professors still faced similar difficulties to students when put in an unfamiliar situation, often only considering one out of the two important parameters of the problem. None of the experts solved the problem within the time set, but they demonstrated superior problem solving skills compared to novices. Singh believed that given enough time experts would have arrived at a solution, as their ability to know what to try and when (i.e. their systematic approach) and their problem solving traits would eventually have led them to the answer. This study demonstrates the importance of practice on similar questions. It also indicates that general problem solving skills unrelated to experience on similar problems can be applied, such as examining limiting cases. This study confirms that it is important to present experts with questions that are real problems in order to give them a similar experience to students, who usually work on new or relatively unfamiliar material.

In summary, expertise is based on intuition, knowledge and experience. Categorisation and qualitative analysis skills are important, but only a subset of the required skills to achieve a successful solution. Effective problem solving skills were demonstrated by experts even when they were unable to successfully use their intuition or solve the problem. Nevertheless, it appears that being able to use worked examples and match the problem with existing knowledge is a key expert trait.

## 1.4 Problem Representation

Problem representations in this context are of a different nature to mental representations discussed above. Representations in this sense are the externalisation of working during a problem, three examples of which are diagrams, equations and words. It is generally thought that the use of multiple representations is an expert-like behaviour and that it can improve performance. Many courses have therefore tried to focus on promoting use of multiple representations in instruction [22, 25].

Kohl and Finkelstein have been among the main, more recent, contributors to research into external representation use in physics [2, 30, 31, 32]. Kohl et al. [30] coded students' solutions from equivalent classes in two different US universities.

Students completed four basic problems and one more difficult problem [30]. The majority of students drew both pictures and a force body diagram (FBD) in the harder challenge problem. In the challenge question students who drew a correct and complete FBD outperformed those who had not drawn an FBD or who had drawn an incorrect FBD. This trend was not as clear with the basic problems. The authors speculate that students in the more difficult problem benefited more from using an FBD than in the basic problems. Kohl et al.[30] considered possible sources of systematic error. Exam score or SAT entrance scores were relatively similar between the populations of the two universities, but the authors acknowledge that there may have been additional differences between these cohorts. Their study is interesting, but considers representational use in a very limited context of problems involving force body diagrams.

Although multiple representation use is considered an expert-trait [30], Kohl et al. [30] found that while multiple representations were used by both classes of students it did not mean that they were successful in solving the problem. Multiple representation use by both experts and novices was also confirmed by Kohl and Finkelstein [2]. They speculate that perhaps novices are now using multiple representations due to instruction through reformed PER courses which encourage this. Reformed courses base instruction on PER findings, such as active learning techniques (Section 1.8), and aim to continually assess pedagogy through student understanding. Traditional methods usually involve didactic teaching methods, whereas reformed courses aim to keep students engaged and active throughout the activities in the course.

In a later study by Kohl and Finkelstein [2], students and experts were interviewed solving physics problems. Representational use was coded with respect to problem solving behaviours, according to an adapted version of the Schoenfeld model, discussed in Section 1.6. They found that most students actually used multiple representations and drew a diagram, but it was how they used them that was important [2]. They concluded that while students may be taught to use multiple representations, they would also require meta-level skills to understand and make best use of them. They designed the study to mitigate against possible errors. A harder question was included to ensure it was not just an exercise for the expert group, and the study included a novel type of situation where students had to group similar representations, to reduce reliance

on the expected solution routine practiced in class. The coding scheme is perhaps unclear and overlapping in places (for example between analysis, exploration and implementation), but this may be due to the limited space to describe each code. Further study is still necessary to keep up with pedagogical developments in PER and to determine whether students still have a formula-centred approach even if they are drawing a diagram.

Resources can in fact be used in many different ways. Suwa and Tversky [33] have found many reasons as to why representations are useful, including: reducing the load on the working memory; cueing memory that may not have been retrieved otherwise; and generating new ideas. It would therefore be interesting to consider further how representations are applied in physics beyond the problem solving behaviours discussed by Kohl and Finkelstein [2], such as the use of representations as tools, visualisations and ways to generate new ideas [33]. Information on this may help inform students not just to draw a diagram, but also how to use the diagram and what to include on it.

## 1.5 Knowledge Structures

Knowledge structure is a theoretical construct. It is the organisation of mental resources and the strength of links between these resources. The terms ‘mental resources’ and ‘resources’ should not be confused: in this thesis ‘resources’ refers to external information, such as a textbook or the internet, while ‘mental resources’ refers to conceptual and procedural knowledge of a subject area [34, 35]. Nokes et al. [16] claim not only that experts have a greater amount of procedural and conceptual knowledge, but that it is also more organised and accessible to them.

Redish and Sabella [27] reason that particular knowledge structures are activated depending on certain cues and the interpretation of these cues. Beatty and Gerace studied words cued by specific physics terms to indicate the strength of the association between nodes of knowledge and to investigate cognitive knowledge structures [36]. In a preliminary, but rigorous study, albeit one with acknowledged limitations, they presented an interesting idea. Students were given a term from which they were instructed to enter all the words that came to mind. This was repeated using a problem and a general topic area as different

types of prompts. Students were also asked to draw a concept map. The results indicated that the concept map usually contained more words, but most of the words picked up by the prompted term were included on the concept map. They conclude from this that the term prompted activity seemed to be able to probe the core conceptual knowledge, while the concept map collected a greater spread of structural information [36]. With only 16 participants they acknowledge that the ordering of the activities could be important, such that participants on the later test could be affected by previous tests. Nevertheless, Beatty and Gerace conclude that this research [36] is a step towards testing students' knowledge structures, with the acknowledgment that these are preliminary results.

Leonard, Dufresne and Gerace created a framework of expert and novice behaviour. The framework models the strength of students' cognitive links in relation to conceptual knowledge, procedural knowledge and problem state knowledge [37, 38]. Problem state knowledge relates to recognising the correct area of physics needed to solve the question. This is not always straightforward, as it can look like one area of physics is relevant when in fact it is not. This is demonstrated in the questions chosen in Chapters 3, 4 and 5. Conceptual knowledge refers to the physics content knowledge required to solve the problem, for example, an understanding of Newton's second law. Procedural knowledge is how to implement this concept in order to arrive at an answer, for example, working out a value for force by putting numbers into  $F = ma$ . This model [37] is based on results from problem solving studies and was devised to explain differences between expert and novice problem solving. Although they focus on the teaching and assessment of physics concepts, they also include operational and procedural knowledge in their domain knowledge model, which allows their model to be applied to a wider skill set. Students may need to be explicitly taught, and practice with, all the necessary skills (conceptual, procedural and problem state), as experts have many past examples upon which to practice in all of these areas. Further validation of this model may be needed, but it is useful for opening up discussion and so is referred to in sections of analysis in this thesis.

Redish [39] developed a theoretical framework which draws on areas such as cognitive science, neuroscience and education and applied them to teaching and research Physics Education contexts. Redish notes that students' internal resources can be explained by modular theory or model theory. Modular theory is

when internal resources are weakly associated and easily changed. Model theory [39] is where there are strong associations between resources which are resistant to change, but that do not necessarily agree with the theory the student is trying to learn. It seems the modular theory agrees most with the novice structures described by Gerace [37], where a novice has weaker links within and between different types of knowledge. Gerace's model does not account for model theory where students incorrectly have very linked knowledge. Redish speculates at the end of his paper that both model and modular theories of internal resource association are needed, as students need to be able to have stable knowledge in some situations, whilst still being flexible to respond to new situations [39] requiring a more modular approach.

A final area worth mentioning is phenomenological primitives or p-prims, another type of mental resource based on experiences in the world, which may affect the way a person approaches a problem [40]. Redish describes a simple definition of a p-prim as follows:

*..basic statements about the functioning of the physical world that a student considers obvious and irreducible. p.20 [39]*

DiSessa who invented this concept explains how to recognise p-prims in a study applying this theory to MIT students [40]. This includes principles for judging whether a p-prim exists, such as the principle of diversity where it is proposed that many p-prims may exist and no attempts should be made to unify them. An example of a p-prim from DiSessa's study is 'continuous push', which originates from experiences such as moving an object with continuous force, for example pushing a book along a table. This causes difficulties in analysing motion where there are no forces acting in the direction of motion, but the motion continues. For example when a ball is thrown in the air the only force acting when the ball is moving upwards is that of gravity. P-prims are referred to briefly in the analysis in Chapter 8.

To summarise, knowledge structure can be thought of in terms of access to mental resources and links between these elements of knowledge [39]. Expert-like behaviour exhibits strong links to accessible and relevant knowledge [16, 37, 38]. The methodology used in the research study can also influence the participants' ability to access mental resources [36].

## 1.6 Problem Solving Models

This section considers a variety of problem solving models, their strengths and weaknesses and a discussion of whether they were considered successful in improving problem solving, followed by implications for teaching. Jonassen explains that problem solving models are important as they provide strategies to help the solution process. These strategies are usually recommended by researchers and are mostly heuristic strategies [26]. A heuristic strategy provides a set of rules intended to increase the solvability of a problem. This section focuses on the Minnesota Model and Schoenfeld's model, as they are used in the analysis of data in this thesis. These models were selected as the Minnesota Model is used in instruction at the institution involved in this research and Schoenfeld's model, adapted by Kohl and Finkelstein [2] was considered a good general model used in the past to analyse student behaviour [2, 24].

### Schoenfeld's Model

Schoenfeld studied expert problem solving behaviour on previously unseen, difficult problems in mathematics. He was looking for emergent patterns from good problem solvers which could be turned into a prescriptive strategy [24]. Schoenfeld designed a problem solving model based on these expert approaches which included five behaviours:

- Analysis
- Design
- Exploration
- Implementation
- Verification

Details of each behaviour are given in Section 3.6.2 where this problem solving model was applied to the data collected. To test the effect of teaching the above strategies or heuristics, Schoenfeld split novice participants into control and experiment groups. The control group were given their problems in a mixed order, whereas the experimental group were given problems grouped according

to the same heuristic strategy that was useful to apply. The aim was to not provide any training for the control group and see if their past experience before the experiment helped them develop their heuristics. It showed that teaching heuristic skills allowed students to pick heuristics effectively. This study sets a good grounding for this type of work, but had very limited participants and heuristics. There were only 7 participants in this study which raises questions on generalisability. Students were trained in only 5 heuristics with 20 training questions (4 per heuristic). In real life, there are a lot more heuristics and they are un-contained. These studies were developed in and for mathematics, which requires different domain knowledge from physics. However his strategies are likely to be general enough to be applied in physics.

The strength of Schoenfeld's study regardless of its flaws is his recognition of the importance of giving unseen difficult problems to experts and novices to ensure the problems were 'real' (as described in Section 1.1 ) and to ensure useful comparisons. Furthermore, the problem solving model was not intended to be prescriptive; if students saw a different approach they were encouraged to use it, as he concedes that every problem and person is different. The emphasis of the problem solving model, and indeed the whole course through which his students were taught, was on overall control over the solution process. Even when expert solution processes were perhaps perceived to be unstructured, he said they showed constant control throughout. Schoenfeld [24] advocated having control on two levels: a strategic level and a tactical level. The strategic level is more global, such as deciding on future moves. The tactical level relates to decisions made at a more detailed level and includes immediate decisions, such as choosing the next step in re-arranging an equation. For example, determining what approach to use and where that might lead are decisions on a strategic level. Figuring out how to set up an equation of motion calculation requires decisions on a tactical level. This differentiation is an important one, as stepping back to look at the whole solution is a different skill from figuring out what can be done next. This differentiation is explored further in Chapter 4.

Simply applying Schoenfeld's model, without discussion of control, leaves out one of the main points that Schoenfeld was trying to make. He thought that control or executive behaviour was the reason studies before him which had tried to teach heuristic strategies had failed; that heuristic strategies were too complex.

As mentioned by Adams, students can demonstrate mastery of the individual problem solving strategies, but this does not clearly make them better problem solvers [5]. Perhaps the lack of emphasis on control and meta-cognition during the problem solving process is a contributing factor.

### **Kohl and Finkelstein's version**

Kohl and Finkelstein adapted Schoenfeld's model to include one extra code: 'translation' [2]. The added translation category codes for the stage after reading, where the student is translating the question onto paper. As they were studying representational use translation was a particularly relevant code to add. They used the problem solving model to determine students' behaviour during representation use, which is discussed in the previous Section 1.4.

### **The Minnesota Model**

The Competent Problem Solver method, developed by Heller et al. [3, 4] at Minnesota University consists of the following stages:

- Focus on the problem
- Describe the physics
- Plan your strategy
- Execute the strategy
- Evaluate the answer [3, 4]

Each point expands to a set of guidelines for a student to follow whilst solving a problem. Details of each behaviour are given in Section 3.6.1 where this problem solving model was also applied to the data collected. This model was influenced by Schoenfeld's model [24] and by work from Reif and Heller [35]. It applies to multi-step, contextualised problems. The Minnesota model is not so applicable for problem type at the very ends of the spectrum of well-structured to ill-structured problems (Section 1.2), such as exercises or dilemmas.

The problem solving strategy above was fully integrated into a reformed course. Reformed courses include active learning techniques, as explained in Section 1.4. Group work was used on contextualised problems and students



worked through the model together. Solutions were also based on this model. Although students solved problems in groups, their individual problem solving skills were also shown to have improved [3]. Whether or not students actually used the Minnesota model to help them when they were not required to use it was not determined in that paper and is something that will be explored in this thesis.

Heller et al. [3] used group work, problem solving strategies and contextualised, multi-step problems. However as they used them all together it is unclear whether the model was useful to students, or whether improved results were the result of other instructional changes, such as the use of group work. Perhaps it was the contribution of a variety of factors, as the model may have helped students be more aware of their approach [24] and when stuck it is likely that another group member could help [41].

### **Other Problem Solving Strategies**

Nokes et al. [16] found in their review that most studies deduced a problem solving theory which included some execution of the seven stages below (p.266):

1. Problem categorisation
2. Construction of a mental representation of the problem
3. Search of the appropriate problem-solving operators (e.g. strategies or procedures)
4. Retrieval and application of those operators to the problem
5. Evaluation of problem-solving progress and solution
6. Iterating stages (1-4) above if not satisfied with progress/solution, and finally
7. Storage of the solution

Nokes et al. [16] approached this topic slightly differently from the other studies mentioned above. They assumed the above are the general problem solving stages and that experts and novices differ within these stages. Nokes et al. discuss in detail how experts and novices differ at each of these stages.

Nokes et al. [16] associate prescribing these problem solving techniques with improved problem solving ability. However it is not proven to be so simple, as it is difficult to find a study which has shown this without areas of weakness, such as other factors being changed at the same time. Furthermore, some people may find it difficult to learn these skills and the model does not acknowledge feelings, beliefs and motivations, which are also considered general problem solving skills and play a role in the solution process [5].

### **Problem Solving Models: Further Analysis**

There are some common themes across all the problem solving models [3, 4, 42, 43]; notably the evaluation or checking stage is included in every one. All models have problem categorisation, identifying fundamental principles or more general analysis stages, which resonates with Chi et al.'s work [20] described previously. Teaching categorisation and problem solving strategies may provide a structure for students in difficulty, but Adams suggests that they are not the magic wand for improving problem solving ability [5]. These models are a simplification of what actually occurs and as mentioned above, the representation of these models should reflect the dynamic and flexible process that an expert with high control over the solution exerts. This simplification also misses other skills necessary in the solution process, such as those found by Adams [5].

Instead of applying problem solving strategies, Adams aimed to list many generic problem solving skills in order to evaluate solution processes. This could aid and target instruction. In her PhD Adams [5] described 44 problem solving skills, and adds that weaker students may not have some other essential skills to solve the problem. She found skills unrelated to content knowledge in areas such as beliefs, expectations and processes. Adams assumed that content knowledge would not affect the general problem solving skills she was looking for. She also argued that many studies focus on content knowledge, including students' conceptual understanding [44]. However, theories such as those from Redish [39] acknowledge the importance of content knowledge and that the context may activate different parts of the brain and therefore different chunks of knowledge. Although Redish theorises that context and knowledge can be kept separate, it is likely that they both play a large role in solving physics problems. It is hypothesised, for this thesis, that physics knowledge is interwoven into problem

solving. Although Adams experiments successfully evaluated a number of general problem solving skills, there needs to be further investigation extending into current areas of research, and consideration of the role of content knowledge if the result are to be applicable specifically to physics problems.

Applicability to physics aside, the studies examined in this section and in fact most of the literature on problem solving have been conducted in a different context to that studied in this thesis. Usually these studies are conducted in a North American University, which has different student abilities, backgrounds and instructional programmes compared to the UK. Unlike Adam's research, the aim of this thesis is not to uncover every possible skill students use. Instead its aim is to investigate the type of behaviours exhibited and their relation to success whilst solving physics problems in the context of students who attend a Scottish Russell Group University.

Finally, problem solving strategies provide a platform to aid the solution process and bring key skills such as evaluation into students' awareness, which could encourage general monitoring and control behaviours. This control over the problem solving process is one of the elements that distinguishes experts from novices [24]. Metacognition is a common theme across these problem solving models. Metacognition is defined as thinking about ones own thinking [45]. It often appears in problem solving models such as those discussed above as experts tend to demonstrate metacognitive skills [24].

Metacognition includes activities such as planning, evaluation and reflection. Promoting reflection can assist learning. This was demonstrated by May and Etkina who showed a link between reflection quality and performance on a physics course [46]. These metacognitive skills are encouraged and studied in order to improve problem solving [47, 48]. Therefore perhaps it is the improvements in metacognition that partly explain student improvements in problem solving, even when problem solving strategies are applied.

## **1.7 Working in Groups: Use of Language and Scientific Reasoning**

Cooperative group work has many reported benefits; students provide support to each other such as motivation and encouragement, and spread the cognitive load by sharing the management of explanations and argument construction. Group work also forces some students to explain their answers more fully, as they are asked questions or to elaborate their reasoning [49]. Vygotsky [41] argued that group problem solving with an adult or “more capable peers” allows students to work in a zone of mental development that is more advanced than if they were working individually. He called it the zone of proximal development.

Howe et al. [50] reviewed three key themes in relation to group work; structural features, such as number and composition of groups; role of the teacher, and forms of pupil interaction. These three areas are considered below. Although not the focus of Howe et al.’s work these areas are discussed with respect to University level research where possible.

Structural features can relate to factors such as gender, the number of people in the group or location of dominant members. The construction of groups in relation to gender has been [51], and continues to be, an area of investigation into group composition [52]. Harskamp et al. [52] found that females in mixed dyads did not perform as well and exhibited more questions and less solution seeking behaviour when compared to a male dyad. All male and all female dyads were more balanced in their behaviour and post-test performance. Even though these studies were in relation to school students with an average age of 15.6 years, it is possible that some gender bias would exist at University. Although not related to gender, Nielsen et al. studied University students in Norway discussing physics concepts in small groups in lectures [53]. He hypothesised that dominance in peer discussions (the students spending the most time arguing) may have been related to their seating position. Students with the most argumentation time were more likely to be found sitting in the middle of a group of three.

The role of the teacher is also important. Howe et al. suggested in their review of school studies that the role of the teacher is to guide and monitor, rather than control [50]. Similar considerations of the role of the instructor were made in a different but more relevant University physics context by Turpen and

Finkelstein [54, 55]. They considered how the class norms were associated with the instructors' approaches, and concluded that different class norms possibly encourage different skills, such as scientific reasoning and debate. This study used observation of lectures, where students discussed solutions to questions in small groups and with the whole class. Their studies are discussed further in Section 1.8.

Pupil interaction relates to the types of talk students engage in. According to Mercer [56] language is important in group work as it allows individuals to make better sense of the world together, although there are other forms of communication for students to discuss problems, such as body language and diagrams. Barnes [57] argued that the flexibility of talk allows great freedom in organising thoughts out loud and then adjusting those thoughts if necessary to improve understanding [57].

Mercer [56] defined three types of talk in group work: cumulative talk which builds understanding but lacks critical analysis; disputational talk where group members work in more of an individual competitive manner than cooperative; and exploratory talk, where students work together with appropriate reasoning and critique, by exchanging creative ideas to make sense of the problem. Exploratory talk is the goal in problem solving group work. Mercer admitted in his book that this is a simplistic model, which will need to be adapted as more research on group work is conducted [56], though Barnes made a similar distinction presenting exploratory talk and presentational talk [57]. He defined exploratory talk as students testing out ideas with each other; it is hesitant and incomplete as these ideas are shaped. Presentational talk is the final product, constructed after some thought and requiring only minor adjustments. These definitions show the prevalence, and importance of the role that different authors have given exploratory talk. Scientific reasoning at University, defined by Osborne [58], seems closely related to Mercer's definition of exploratory talk [56]. The essence of scientific reasoning is covered extensively in Osborne's review, which he related to argumentation, reasoning skills and critical thinking; the sort of research practices which students will experience after Undergraduate study. He concluded that reasoning skills and critical thinking should be taught explicitly and included in education in order to enhance learning.

Researchers have attempted to change instruction in order to improve

the quality of dialogue between students [49, 59]. For example, Christie et al. [59] used structured observation in Scottish primary school classrooms to look at classroom dialogue and the learning environment through an ongoing intervention. The intervention consisted of participating teachers who were encouraged to use group work in their lessons. The research team provided them with materials and suggested activities. Having examined student discourse during group work, they found an improvement in the quality of dialogue. This was evidenced in the form of more propositions, explanations and instructions. They relate these to Mercer's desired exploratory talk [56], as students use these forms of interaction to work together.

Although the observation methods used by Christie et al. [59] seem appropriate, even with quite precise observation windows of 16 seconds, the weakness in their results was a lack of argumentation observed. There were low amounts of questioning/ prompting and resolution/compromise, which they attributed to students being encouraged to share ideas without the need to resolve or compromise on any issues [59]. However, as mentioned above by Osborne [58], questioning and argumentation are considered important traits at university and are fundamental to academic study. This perhaps shows a weakness in their study, or the difference between primary and university education. Although there will be many parallels with group work studies focussing on school pupils, one has to be careful in generalising these results to university students and specifically physics students. The students are at different levels of maturity [60], are being taught in very different ways and university students have deeper knowledge of their subject.

Wellington and Osborne claim it is important to understand the language of the subject [61]; however understanding technical language is not simple. Farrell and Ventura [62] found that students in sixth form struggled to understand and use technical language. Itza-Ortiz et al. [63] showed that students confused 'everyday' meanings of words with their technical definition, such as 'force'. Itza-Ortiz et al. [63] got students to differentiate physics meaning from everyday meanings in certain sentences containing the words force, momentum and impulse. They found that most students (around 64% taking the word 'force' for example) were unable to discern everyday meanings from physics meanings. Although these studies do not analyse students usage of these words with peers,

they indicate that students can use the same words to mean different things and that students can struggle to grasp the meanings of technical vocabulary.

All the above studies have a common theme. Students engaging in active, exploratory talk have the opportunity to construct new knowledge by testing and refining their ideas with peers. Fewer studies focus on University students. Group work is increasingly becoming more integrated into a physics degree, including in lectures [3, 4, 64, 65]. Peer discussion is when students work in groups in a lecture to discuss conceptual problems. This is discussed in the next section.

## 1.8 Peer Instruction

Active learning techniques form a large part of PER and it has largely been accepted that they improve learning gains compared to traditional methods. Traditionally lectures have been modelled on a didactic style of teaching whereby the lecturer speaks and students listen. However this does not mean that students are understanding and analysing the information they are being taught. Hake [66] found that students across 62 introductory physics courses performed better when taught with interactive engagement methods, such as group work, rather than the traditional lecturing style. Although this result was not consistent across all courses Hake studied other authors, in other disciplines, have found similar trends [67, 68].

The active learning technique, Peer Instruction (PI), has been shown to produce high conceptual learning gains in introductory physics courses, especially when used with other active learning methods [69, 70]. Initially developed by Eric Mazur at Harvard University [64], the process of PI consists of: presenting the class with a question; a period of individual thinking followed by a vote on the answer; peer discussion amongst small groups of students; a re-vote in the light of the peer discussion, and then either further discussion from the class or (if necessary) explanation of the correct answer from the instructor. Voting is usually conducted with clicker technology providing the instructor with an instant way to assess the understanding of the class and adjust the lesson accordingly.

### 1.8.1 Theoretical foundations

Different models of peer discussions are emerging from discussions of PI. One such model is that of transmission, which according to Smith et al. [71] is given value in discussions on PI. It seems that many of the reported benefits of PI relate to the transmission model, where one student, knowing the correct answer, explains it to a weaker student. This includes the instructional expectation by Mazur that students convince others of the correct answer [64]. The explanation from the stronger student is likely to be in a language that their peers can understand [64]. The plausibility of students' justifications may assist in convincing weaker students, which may lead to basic constructivism where students question each other and discuss the reasoning to ideally create a stronger and more complete argument. Another theory is that of constructivism. Socio-Constructivism is a learning theory where peers construct knowledge together, integrating what they already know with new information [72]. There is currently little evidence to support which of these processes actually occur; transmission or the group constructing knowledge together. These theories are used in the discussion of the results from analysing PI in this thesis (Chapter 8).

### 1.8.2 Studies of Peer Instruction

There have been many studies investigating peer work, especially in primary and secondary education, such as those reviewed in section 1.7 on group work. However fewer studies examine Peer Instruction in the context of University physics lectures, particularly in the UK. The papers that do examine PI in Physics are the focus for the literature review below and current gaps in this literature, explained at the end of this section, justify the need for a study of this kind.

Studies have started to look at student discussions in more detail in order to further investigate the PI technique, and to inform practitioners of what students actually discuss during PI. The Mazur group have done an extensive amount of research in this area, providing a platform upon which to build on in instruction and research [64, 69, 70, 73].

Mazur [64] summarised the technique and the basis for using it. He discusses students' improved conceptual ability after using Peer Instruction [64, 73]. A subsequent report by Crouch and Mazur [69], taken from ten years of data,



confirms the success of PI more thoroughly. A more recent report by Crouch et al. [70] summarises the Peer Instruction technique, with detailed information for instructors on using Peer Instruction, to a review of student and staff satisfaction and improvements in student learning. In these studies it is generally assumed that if students answer the question correctly after discussing it with their peers, then PI has been successful. However a correct vote does not mean the student has completely understood the concept and that their reasoning is complete and correct. Nielsen and Stav found that some arguments in the group discussion were not based on the correct principles [74].

The instructor can begin to address the above problem in a small scale way in class. Aside from looking at the voting statistics, the instructor can ascertain the quality and content of discussions by walking around the lecture theatre and listening to what students are saying [70]. Crouch et al. [70] and Beatty et al. [75] emphasise instructor participation in discussions and present multiple benefits of this approach: the instructor can gain insight into student conversations and listen out for key characteristics to help assess the success of discussions. For example, the approximate level of understanding can be determined, instructors can listen for new perspectives, and it provides a means for the instructor to encourage and facilitate student discussions where necessary. The question then is what should they be listening for, what type of conversations are the students engaged in and what are the characteristics of a successful and unsuccessful discussion?

James and Willoughby [76] have begun to investigate student discussions in order to clarify unsuccessful conversation types, which are discussed below. They looked at conversations that deviated from an instructors' idealised standard [76]. They defined a standard conversation as one where the student talked about the distractors given in the multiple choice question, so the distractors presented could be used as a proxy for student ideas. 136 of 361 conversations were classed as a 'standard conversation'. The remaining 225 conversations were the focus of their analysis. They used the technique of constant comparison (also referred to as grounded theory) to find emerging conversation types from the non-standard data. Three broad categories emerged [76], for example, one category included conversations about a distractor that was not given as a multiple choice option. Though it can be argued that importance should be placed on whether the student

learnt the concept correctly, and not whether they discussed distractors that were presented in the question. Weaknesses of this study included that they did not look into the standard conversations in any detail. They considered mainly non-physics majors and suggested it should be repeated with physics majors. Also they did not do the silent thinking stage. They posed the multiple choice question and then went straight to group discussion before entering individual responses, so they could not compare students' answers to the clicker questions before and after the peer discussion phase. Both of the authors James and Willoughby [76] have done other research in peer instruction, focussing all their studies in astronomy.

James et al. [77] conducted a study of PI in introductory astronomy classes. They measured discourse bias by taking the difference in contributions from partners participating in PI. Discourse bias relates to the dominance of one student's contribution, compared to another. The contributions by each partner were measured by two methods: coding ideas, and counting the number of words spoken per person. James et al. [77] coded each idea (or phrase) into 10 categories adapted from Kaartinen and Kumpulainen's techniques for analysing discourse analysis [78]. These included categories, such as stating answer preference, providing justification for a way of thinking and stating agreement with a partner's idea. They reported this technique as being very time consuming and tried to find alternative ways of measuring discourse bias. By counting the number of words spoken per student, amount of time spent talking, number of turns and average number of words per turn, they could compare this to the results of discourse bias coded by idea. They found that both categorising student's ideas and using a word count per student found the same results in terms of discourse bias. They state that calculating the total number of words is probably the more reliable and replicable of the two methods. Their study opens up the potential of word counting as a viable method to study peer discussions.

James et al [77] used the method explained above to look at how grading incentives influenced peer discussions. They found that in high-stakes courses, that rewards high credit to correct answers and little credit for incorrect answers, that conversations are dominated by a single member. In low stakes grading they measured significantly less discourse bias and significantly more disagreement compared to high-stakes grading. In high-stakes grading it is suggested that conversations are dominated by a single group member in an attempt to get a

correct answer and therefore a high score [64]. The emphasis may be displaced to credit-scoring over learning in a high-stakes system.

Smith et al. [71] were concerned that some students may not have been fully engaged during peer discussions as they answered under the influence of fellow students, as opposed to fully understanding the concept themselves. To test this they presented students with two isomorphic problems. An isomorphic problem tests the same concept or principle, but is set in a different context. The first question was run with full PI (individual response and then post discussion response). Students were then presented with the isomorphic question which they had to answer individually. The study was repeated on 16 occasions throughout the semester. They found that the average percentage of correct answers for the second isomorphic question was higher than either pre or post discussion responses from the first question. Students who answered correctly were likely to stay correct on the second question. There also seemed to be a delayed effect, where students who were initially incorrect answered the second question correctly. Considering the distribution of students in the class statistically, Smith et al. [71] proposed that there was a probability that some groups were answering the second, isomorphic question correctly even after being in a group with no correct votes on the first isomorphic question. The conclusions of their study are that PI works to promote understanding and engagement, rather than through only peer influence. They related their results to the theoretical models mentioned above in section 1.8.1. They suggest that PI is not solely based on a transmission model of instruction, as a naive group, where none of the students knew the correct answer, could vote correctly on the second isomorphic problem. Although these findings seem to support a socio-constructivist theory of learning, whereby students construct knowledge together, one should keep in mind that other factors may have influenced the data, such as students interacting with groups other than their own. Although these findings may be generalisable in terms of continual learning after the standard PI process is finished, it is noted that Smith et al.'s [71] study was conducted in a genetics course for biologists, not physics, and in the US, so a different cohort in terms of subject and location compared to that examined in this thesis.

Implementation variations of PI have been recognised in studies by Turpen and Finkelstein [54]. They argued that PI conducted in different ways constructed

a range of classroom norms, and specifically encouraged different scientific skills and practices [54, 55]. Classroom norms are the cultural expectations in the lecture theatre. The authors said that facilitating discussion and engaging with students during the discussion stage would encourage interactions with physicists including practising the ability to formulate and ask questions. Depending on the encouragement of the lecturer to facilitate peer debate on the correct answer they said that students were encouraged to develop a broader or narrower range of desired physics skills, such as evaluating the correctness of a response and communicating publicly.

Crouch et al. [70] note the difficulty in constructing good questions for Peer Instruction. Beatty et al. [75] presented a paper with recommendations for constructing appropriate clicker questions. They present three instructional goals, including improving the particular physics knowledge, improving the students' skills or processes needed to get to an answer and awareness of the students' perception of physics as a subject. They suggest some tactics for achieving these aims, such as removing nonessential information or constructing a question which reveals a better way of solving it from one that the student may initially try. Beatty et al. [75] have considerable experience in constructing clicker questions and finish by emphasising the role of the instructor and formative assessment of teaching and question construction through listening to discussions and viewing student responses. Beatty et al.'s recommendations are considered in the case studies in Chapter 8.

The majority of the publications mentioned above have been in introductory astronomy, or introductory physics courses in the US [76, 77, 79, 80]. These can be taken as relevant to introductory physics courses, where PI was used at Edinburgh University. However many of them explicitly missed out the individual thinking phase and pre-discussion vote [54, 55, 76, 77, 79]. Without this they were unable to assess students' change in understanding according to their vote. Being aware of the voting statistics before discussion may also help the analysis of student conversations, as students may be more likely to operate in a transmission model when many of them have the correct answer before discussion. Matching student discussions with responses to the questions pre and post discussion appears to be an unexplored area in current literature and one that research in this thesis aims to address. Further research is also needed on the characteristics of success in PI

and the kind of reasoning students are engaged in. Ideally results from these will inform and improve instruction during PI.

## **1.9 Problem Solving in Summative Assessments: Open and Closed Book Exams**

Assessment is an important part of learning, as it places importance on areas for students to learn during the course. Students can be assessed in a variety of ways, such as homework questions during the semester and project reports. One of the main methods of assessment are final exams, which is the focus for this section, including support for open and closed book exams.

Williams and Wong emphasised that exams should test and therefore encourage the skills required of students after University and facilitate deep learning [81]. They advocate the use of open book exams to do this, though they acknowledge that there is usually some resistance to changing exams. Many people believe that closed book exams are the only way we can really assess if the student knows the material. Arguments supporting open-book exams are not a recent trend. Frank Bacon in 1969 [82] looked at open-book exams as a ‘forward-looking’ approach. Furthermore, a more recent guide written for the UK Higher Education Academy to improving assessment in the Physical Sciences [83] confirmed that if instructors wish to focus on how well students can use and apply information then open-book exams are preferable over closed-book.

The studies investigating open-book exams have been positive. Brightwell et al. found that open-book exams discriminated students’ abilities as well as closed book exams [84]. Williams and Wong [81] found open-book, open-web exams were received positively in terms of learning outcomes. To minimise the potential of cheating they used contextualised, ill-structured problems which meant that students could not simply find the answer, or model solution, on the Internet. Overall Gupta [85], recognised that there is a lack of research informing recommendations and that much of the advice is based on ‘experience and belief’. The main weakness in studies investigating open-book exams is the inability to look directly at students solving problems in an exam.

Finally, although open-book exams require higher cognitive abilities than closed-book exams, they are not suitable in all cases. For example, testing

extensive factual knowledge would be best tested with a closed-book exam. Race [83] makes the point throughout his guide that a range of assessment procedures are best to suit the learning outcomes of the particular course, to diversify assessment and give students a variety of opportunities to demonstrate their learning. As with closed-book exams there still needs to be the construction of reliable and valid questions for open-book exams [82, 83].

## 1.10 Summary

The section above reviews relevant literature on problem solving to provide a general overview and context for the areas studied in this thesis. This summary pulls together some of the areas of research to extend in these fields. Expert and novice behaviour has been well explored and a number of problem solving models suggested. It is unclear whether students use the recommended problem solving strategies when working on their own unprompted. Aspects of expert behaviour need further investigation, as it is still unclear how past experience on similar examples influences tasks, such as categorisation of physics problems. There is still progress to be made in relation to investigating expert behaviour on true ‘problems’.

Conceptual understanding and qualitative reasoning using appropriate concepts are important steps in the physics problem solving process [3, 21, 22, 23]. A largely unexplored area of research is a study of qualitative reasoning and its relationship with the chosen voting option following peer discussion in PI. Also research is presently constrained by the limited availability of tools to investigate large amounts of data in a realistic timescale so limiting the number of participants it is possible to analyse. Further investigation of a word counting approach [77] may provide such a technique.

Therefore the main research questions that arise from studying the literature on problem solving are:

- How do experts and novices use problem solving models when they are not explicitly asked to apply them?
- What problem solving strategies do experts use on difficult physics problems?

- What are the characteristics of successful and unsuccessful Peer Instruction conversations?
- Does using a word counting technique of some kind provide a quicker way of analysing peer discussions?

This thesis aims to further understand all the areas mentioned above with regards to successful and unsuccessful problem solving. The structure of how this will be achieved is set out in Section 2.3.

## Chapter 2

# Methodological Approaches

This chapter investigates methods to collect and analyse data on students' problem solving processes and to determine the kinds of data collection and analysis suitable for this thesis. Qualitative research is a key tool in understanding how students solve problems in physics. As stated above studies have discussed: how experts differ from novices in problems solving [20]; modelled the problem solving process [86]; and created problem solving criteria for assessment [87]. However, few of the studies discussed in the previous chapter give the methodological details of the collection and analysis of the qualitative data, such as the choice of data collection and analysis used and an assessment of the methodological strengths and weaknesses associated with this choice. Flick [88] supports qualitative research and says it gives practical results. Qualitative analysis is appropriate as human behaviour can be dependent on the situation and the person [88]. These variables cannot necessarily be controlled for and instead are best discussed qualitatively.

The data collection methods explored below include the think aloud technique, retrospection, structured questions, observations and using a smartpen. The data analysis methods reviewed include quantitative approaches, such as comparing problem solving behaviour to problem solving models and qualitative analysis, such as thematic analysis and grounded theory. The final section summarises and justifies the techniques used in this thesis.



## 2.1 Collecting Qualitative Data

### 2.1.1 The Think Aloud Technique

Think aloud is a technique well suited to examining cognitive processes and is now an established technique used in investigating problem solving strategies. Ericsson and Simon [6] proposed that cognitive processes could be investigated and not themselves affected by the think aloud process. Think aloud requires the student to talk through the problem out loud whilst solving it, providing information on the processes used by the student to get to a solution that a written answer may not.

Ericsson and Simon stated with confidence that cognitive processes were not affected by this thinking aloud [6]:

*...cognitive processes are not modified by these verbal reports, and that task-directed cognitive processes determine what information is heeded and verbalized. p.16 [6]*

They postulate that the subject would be too occupied in speaking and doing the problem to have any room in their working memory to adjust their cognitive processes. This itself is based on the information processing model. This model compartmentalises the brain into different areas, such as the working or short term memory and long term memory, each of which works in slightly different ways and can store information for varying amounts of time. The working memory holds a smaller amount of information for a short period of time, whereas the long-term memory retains large amounts of information for longer periods of time. This model is one of the least refuted of all the cognitive models, however it is likely to be a simplified version of what actually occurs. The information processing model does not, for example, consider other more complex cognitive theories where there are connections between parts of the brain in order to construct knowledge. Another drawback with this cognitive model is the assumption that everything a student knows passes through their short term memory and so they become conscious of it. They then think aloud these thoughts. It is unlikely people will be aware of, or say, everything they think and the cognitive load from speaking their thoughts aloud whilst also solving the problem will be increased. Some students have reported that their verbalisation could not keep up with their

cognitive process [89] and that their thought processes are more complicated than they could verbalise [90]. This leads to shortcomings in the data where cognitive processes have occurred without being verbalised.

A more recent review of think aloud agreed with Ericsson and Simon [89]. Someren et al. [89] found that emotional and motivational factors can influence task performance during think aloud, but stated that there is little evidence that think aloud influences performance other than that which occurs through creating a situation constructed to emulate reality, but which can never be the real thing. Someren et al. [89] recommend not using *tell me what you think* phrases, as opinions are not desired nor is an evaluation of their thoughts.

Instructions should not be too long. Laurillard [91] warns of contextualisation where the student is affected by the context of the study, such as what they think the researcher wants from them, without concentrating on the problem itself. As with any experiment, what naturally occurs is influenced by taking a measurement of it. Whether the subject solves the problem using think aloud as they would normally without altering their cognitive processes is uncertain. The students may provide a more structured account of their thoughts than they would normally do, just by vocalising them. This structuring could help them determine what to do next or could hinder this process. Schoenfeld [24] expressed similar concerns more generally and said awareness of influences was required in the analysis of data collected via think aloud.

To summarise think aloud seems to be a viable method of analysing problem solving processes, however there are possible weaknesses in the cognitive model it is based on. The process can only ever approximate reality as there is the possibility that thought processes may be changed by vocalising them.

### 2.1.2 Retrospection

Retrospection involves questioning the participants after the problem about the thought processes they used [89]. This means asking questions such as *Talk me through how you solved that problem* or *Tell me the solution to that problem*. Retrospection needs to be completed soon after the solution process to aid memory. Usually their solution is placed in front of them, or a video of their working is played back and students are asked what they were thinking at each point of the interviewer's choice. The basis for this is explained by Ericsson and

Simon[6]:

*A durable (if partial) memory trace is laid down of the information heeded successively while completing a task. Just after the task is finished, this trace can be accessed from STM, at least in part, or retrieved from LTM and verbalized.* p.16 [6].

(STM and LTM refer to short-term memory and long-term memory respectively.) Retrospection has a degree of interpretation and invalidity associated with memory errors, especially if reports are based on long term memories. Someren et al. [89] used the information processing model to explain why results from retrospection may not be valid. They state that not all information which students had in their working memory may be retrieved and it is possible that false information that was not there may be remembered as if it was in the working memory. The student may also construct a different picture of what they did, such as making their thoughts processes sound more coherent and structured than they actually were [89]. Furthermore, students can be easily influenced by expectations [89]. This relates to Laurillard's [91] discussion on contextualisation mentioned above. Students may be affected in their retrospection by trying to give the interviewer what they think they want to hear, instead of reflecting directly on their problem solving. Nevertheless, an advantage to retrospection is that the questions are asked after the cognitive process is complete and so there is no disturbance of the solution process [89]. It also reflects what students thought they did when solving problems.

To summarise, retrospection will have less disturbance of the thought processes during the solution process than think aloud, but errors are possible in the reflection of their approach, including the potential difference between what participants think they did and what actually happened.

### 2.1.3 Structured questions

Asking pre-defined questions is another technique. This would include asking questions such as '*How would you approach this problem?*' as the student is solving it. This approach is well-structured and due to the pre-defined questions and the direct requests for information, it will provide relatively complete and relevant data. For example, it could group responses by students' preferred

approach to a problem [89]. A disadvantage is that data beyond these questions is not necessarily collected. It also requires the researchers to set the questions, which in itself will influence the usual cognitive process and therefore the information obtained [89].

This technique affects the cognitive process, but collects data in specific areas of the chosen investigation.

#### **2.1.4 Observations and Field Notes**

Observations are a data collection method conducted in the participant's own setting [92]. The level of participation to obtain these observations depends on the observer's stance, for example, a passive participant observes without interrupting, whereas an active participant joins in the activities. Observations are important in collecting non-verbal, as well as verbal information. This is particularly useful in problem solving in groups, where interactions between members can be observed, including body language and behaviour during periods of silence. However, choosing what to observe may be difficult and gaining trust and entry into the groups is potentially crucial, especially when acting as an active participant. Furthermore, active participants will influence the thinking of the group. Technology, such as the smartpen (see below) brings a new dimension to observational studies and the ability to observe with some detachment, so potentially having less direct impact on the group as the observer.

#### **2.1.5 The Smartpen**

The smartpen is the main method of data collection used in this thesis. A Livescribe smartpen [93], referred to here as a smartpen works as a normal pen, with a video camera to record pen strokes and an audio recorder integrated into the design. Used on special Livescribe paper, the smartpen is able to capture audio and pen strokes in synchronisation. The audio can be played directly from the notes, by a control panel on the notepad itself, or files can be downloaded onto a computer.

The smartpen, referred to above in the observation section, is an un-obtrusive recording device and as it is designed as a pen, it is taken to be more easily 'normalised'. Any notes taken by the pen holder are recorded, which adds

to the richness of data collected. This means that students can be recorded authentically, to see how they communicate with each other and learn in a real academic environment.

Although a relatively new technology, there is evidence from Schaack [94] that smartpens may be useful in education, especially by aiding note taking, as notes and audio can be specifically connected to different parts of a lecture. Authors point to the lack of evidence showing the connection of smartpens with achievement, specifically in students with learning disabilities [95]. Use of smartpens to study problem solving is less reported.

## 2.2 Analysing Qualitative data

The data can be studied in more than one way to validate or construct theories and models, and considered using different paradigms in order to provide a variety of insights. Ideally these insights can inform pedagogy in order to improve students' problem solving processes. A number of qualitative analysis techniques are discussed below. There is no "correct" approach, but being aware of the variety of approaches to analysis will help determine the best approach in relation to the data collected and the research aims of the study.

### 2.2.1 Quantitative Methods

Data analysis methods in this section are referred to as quantitative methods [96], as the data that emerges can be counted quantitatively and percentage agreement between two coders counted numerically. Flick refers to applying set categories to data as qualitative content analysis [88], but in this thesis these techniques are referred to as quantitative methods to distinguish them from other qualitative approaches.

Schoenfeld suggested that data can be split into small units called episodes which can be categorised and coded according to an existing model. Flick [88] summarises content analysis in terms of categorisation:

*One of its essential features is the use of categories, which are often derived from theoretical models: categories are brought to the empirical material and not necessarily developed from it, although they are repeatedly assessed against it and modified if necessary. p.323 [88]*

To utilise previous research in the area of problem solving, as opposed to developing categories from the data itself, pre-defined coding schemes were used. The models used in this thesis macro-code the data, looking at behaviours more generally rather than analysing each statement or phrase verbalised in detail. Three behavioural models are used to investigate individual problem solving in chapters 3, 4 and 5. Schoenfeld's model [24], the Minnesota Model [3, 4] and representation use [2] are discussed in the previous chapter, see Sections 1.6 and 1.4. Within each model the criteria for identifying each behaviour is covered in the methodology Section 3.6.

Another method of quantitative analysis is counting the number of all or specific words in a conversation. Word counting was conducted by James et al. in Peer Instruction [77] and is used in Chapter 7.

### 2.2.2 Qualitative Methods

An alternative method of analysis is to develop theory from the data, as opposed to imposing pre-made models [88]. One popular method of analysing transcripts to develop new theories is by grounded theory. First developed by Glaser and Strauss [97] the emphasis of this is to generate a coherent theory from the data, without any preconceived ideas about what it shows. This approach is by no means straight forward: even after publication of their widely cited book [97] the authors Glaser and Strauss disagreed with each other about the details of this approach [98]. After reading of the transcript, the data is first coded on a detailed level encapsulating all ideas that occurred to the researcher. The open coding procedure is followed by a construction of concepts from these codes [98]. The aim is to keep a close relationship between the data and the concepts derived from the data and to build a coherent, overarching theory which encompasses systematic links between the concepts developed from the coding. Grounded theory is not without its limitations. Researchers have questioned whether it is possible for the analyst to prevent their awareness of current concepts and theories in literature from influencing their coding [98].

Thematic analysis developed by Flick (p.318)[88], involves first sampling data to develop major themes which start to address the research aims. These themes are then studied in more detail and comparisons made across cases. It is similar to grounded theory in that the theory emerges from studying the data, but unlike

grounded theory it does not require researchers to develop themes purely from the data without referring to past experience or other literature.

Although not used in this thesis, discourse analysis can also be used in analysing data. Discourse analysis is interested in the subtle differences in language, and the creation and construction of social interaction [99]. It is concerned with analysing the use of language. Discourse analysis may be used to reveal information on the student's attitudes and beliefs about problem solving, which have been proven to affect how well the student completes the problem [100]. Pajares and Miller linked students' judgments about their capability to solve mathematics problems to their problem solving performance [100]. Hogen et al. [101] and other authors used several kinds of discourse analysis in their papers studying scientific reasoning, conversation types and discourse in peer and teacher-guided discussions. These authors used discourse maps, to track the conceptual journey of the students through the process. However discourse analysis is time consuming, difficult and requires training, as the intention of the student is not always clear.

The above perspectives do not provide a complete account of all the paradigms or models, but are considered to be the most relevant to this thesis. Although it is acknowledged that selecting one approach over another may influence what emerges from the data.

### **2.2.3 Validating and Consolidating Coding**

Reliability in quantitative or content analysis determines whether the same categories would be assigned to the same sections of data if coded again. Validity measures whether the section of data is being described by the correct category. In coding with respect to problem solving models, a way to test validity and reliability of the categories applied is to have a second, independent coder. The percentage agreement between coders can show whether they are both interpreting the categories in the same way and applying them at the same time.

However, having two coders is not always feasible and so a different stance is taken if there is just one coder interpreting the data. Constructivist grounded theory recognises that the researcher interprets the data within their own views and experiences [102]. There is no single correct interpretation, but the interpretation has to be based on the data and transparent.

Using more than one type of analysis on the same data set is a way to validate

or enhance the results, with any number of coders [89, 90]. Triangulation is possible using multiple techniques with the aim of strengthening the validity and reliability of the data by using the results of one technique to verify the results of another method. Data can also be compared between techniques, for example, collecting or analysing data in two different ways to ascertain if the results agree.

## **2.3 Structure of the Thesis**

The research of this thesis is focussed on three main sections:

- Individual Problem Solving (3 chapters)
- Peer Instruction (3 chapters)
- Experts (1 chapter)

The three main sections of the thesis align with the topics reviewed in Chapter 1, including individual problem solving, expert versus novice behaviour and working with peers. This work extends the literature in these areas. The overall aim is to develop a greater understanding of student problem solving and investigate characteristics of problem solving success in different situations. Each section of research is briefly introduced below and the methodology chosen for the particular research focus summarised. More in-depth introductions with detailed aims and research questions are presented at the beginning of each set of chapters (Chapters 3, 6 and 9). The final conclusions summarise work from each chapter and present an overall picture.

### **2.3.1 Individual Problem Solving Chapters**

The aim of this section is to provide an insight into students' problem solving strategies with and without the availability of resources, such as a textbook. The first of the three chapters investigating individual problem solving (Chapter 3) describes how the data was collected, the implementation of the study and the methods of analysis. The second chapter (Chapter 4) analyses student solution processes with respect to problem solving models, such as those discussed in Section 1.6. The results from looking at the data this way are presented. The final section in this set (Chapter 5) looks at the data qualitatively and in more



depth, developing arguments from the previous chapter 4 and bringing the results together for discussion. Students solving problems individually, with and without resources, can provide a means to further understand students' problem solving behaviour and examine how to inculcate expert-like approaches. As mentioned above, there is a lack of research into use of resources, such as using the Internet and a textbook during individual problem solving. Implications for the skills needed for open and closed book exams are also discussed. This is original work and extends the research of this kind in the UK.

### **2.3.2 Peer Instruction Chapters**

Three chapters are dedicated to understanding more about Peer Instruction. The first of these, chapter 6, explains how the study was implemented. The second of the Peer Instruction chapters (Chapter 7) gives the method of analysis, results and discussion about the technique of counting the number of technical words used in peer conversations. The final chapter in this set (Chapter 8) describes the method of qualitative analysis used to further examine Peer Instruction conversations and the results and discussion from doing so. There is very little research which has investigated Peer Instruction in Physics using the data collection methods in this thesis or counted technical language use in these conversations.

### **2.3.3 Expert Group Work Chapter**

Comparing to expert behaviour is an important part of understanding how to enhance novice performance. Little research has been done studying experts working in groups and this chapter begins to approach this area. Initial results are presented from expert problem solving behaviour in groups. Results from expert behaviour in this final chapter are also compared to the novice behaviour exhibited in other chapters of this thesis.

### **2.3.4 Chosen Methods**

There are many possible ways of collecting and coding qualitative data and the ideal approach depends on the study's research aims. The aim of the individual problem solving chapters was to further understand students' individual problem solving processes. A think aloud interview with a smartpen is likely to be best

suited to doing this, as it records what they write and say simultaneously as they progress through the problem [6]. It is arguably a more realistic account of the solution process when compared other methods, such as retrospection, though it is not without its limitation, as discussed above.

When students are talking through a problem together, observational studies are appropriate, although the assumption is that they are vocalising the majority of their thoughts about the problem. Again a smartpen is useful to capture written as well as spoken data in synchronisation. This methodology is suited for analysing Peer Instruction and the approach chosen for this thesis. It was conducted in-the-field, to investigate what happens during problem solving in the environment where it actually occurs.

Multiple methods of analysis were used to investigate the data in a variety of ways. The chapters investigating individual problem solving and Peer Instruction have the same overall structure in terms of analysis. Each study begins with quantitative analysis, either by modelling problem solving behaviour or counting the number of technical words used. As this only provides limited information on student behaviour, further qualitative analysis is then conducted to develop the understanding of student behaviour. Thematic analysis was chosen as the qualitative analysis method (Chapters 5 and 8) to pick up major themes emerging from the data relating to the research questions. Discourse analysis was not chosen as a method of analysis due to the difficulty in interpreting the way students were speaking to each other and the time-consuming nature of this analysis. Full details, including the observed strengths and weaknesses of using each of these approaches are given in the relevant chapters.

## Chapter 3

# Individual Problem Solving: Aims and Methodology

Students are required to solve problems individually for the majority of their Undergraduate degrees, including in the final exams. This demands a large amount of the students' time and it seems appropriate that an investigation into student problem solving in physics starts with that of individuals. Furthermore, establishing what students do whilst solving problems could provide further insights into successful and unsuccessful behaviours of students. This in turn, could inform instruction of problem solving.

This chapter is the first of three chapters investigating student problem solving with and without access to resources. Students have access to multiple resources, including textbooks and information available on the Internet, in many problem situations, for example when solving problems at home, or after finishing University. Closed-book exams are perhaps the exception to this, as students have to sit these exams without help from external resources. A study to investigate authentic problem solving will include analysing behaviour with access to resources which will then be compared to behaviours without access to these sorts of materials.

Findings from pilot studies supported an investigation into use of resources. Pilot studies using the think aloud technique were conducted with two Physics Masters students solving problems on introductory physics topics. The key finding was that lack of domain knowledge was a limiting factor, as the participants were given problems from introductory physics to which they could

no longer remember the equations. This raised the question, should students be given the necessary equation to see if they can understand it and apply it appropriately to the question, or should equation retrieval be expected as a skill which they need to be able to demonstrate?

Two studies were conducted with first and second year Undergraduate Physics respectively to collect data at different levels of expertise. Both groups were given the same two physics problems, except one group only had a calculator, pen and paper, while the other group also had access to a textbook and the Internet. The two groups are called 1B and 2B named after the courses the students were participating in. Physics 1B was a first year, second semester introductory course and Physics 2B was a second year, second semester physics course. These courses are summarised in Table 3.1. This chapter details the aims and implementation of these studies.

Physics 1B	A first year, second semester introductory undergraduate physics course. It was mandatory for students studying physics. It covered a range of topics, such as introductory quantum mechanics, particle physics and nuclear physics and involves lectures, tutorials and lab work.
Physics 2B*	The second year and more advanced equivalent of Physics 1B. It covered subjects, such as thermodynamics, electricity and magnetism, and quantum mechanics. It also involves lectures, tutorials and lab work.

Table 3.1: Summary of first and second year courses in Physics at Edinburgh University. \*Scottish Universities have degrees lasting four years, with an extra year at the beginning of the degree compared to English Universities. Very well qualified students can ‘fast-track’ by entering University straight into second year, missing the introductory first year.

### 3.1 Aims and Hypothesis

There were two aims for these studies:

1. To start to determine the behaviours of successful students compared to unsuccessful students. Success is defined as students’ quantified mark according to a standardised marking scheme created for the question.

2. Investigate any fundamental differences in approach or skill utilisation when students are given access to resources and assess the implication of the findings for open and closed book exams.

From the first aim above there were four research questions:

- First year students are given the Minnesota problem solving model, but do they go through each stage unprompted? Will high scoring students move through the problem solving models in a linear or iterative fashion?
- Whilst solving problems, what transitions do students make between problem solving behaviours? (This will give an indication of their executive decision making at both tactical and strategic levels [24])
- Are there any skills that students lack according to these problem solving models?
- Will higher scoring students exhibit more standard expert-like behaviour as defined in the literature, even though they are an advanced-novice?

With regards to investigating differences in approach when given access to resources (aim two) there were five research questions:

- What specifically are students saying or doing when consulting a resource?
- When are students choosing to access a resources; at what stage in their problem solving process?
- How does their use of resources influence their progression in the problem?
- Is the core, conceptual physics needed to solve the problem critical? Are students more likely to access this core physics if they have resources available?
- What areas gave students difficulties in progressing with the problem? Do these areas change when access to resources is given?

## 3.2 Open and Closed-Book Exams

As reviewed in Section 1.9, the main weakness in studies investigating open-book exams is the inability to look directly at students solving problems in an exam. Nevertheless individual problem solving investigations, such as those in this thesis can start to examine what students would do with resources (open-book exams) and without resources (closed-book exams).

## 3.3 Data Collection

Each cohort, for both the 1B and 2B studies, was divided into quartiles according to their score on the previous semester's Physics test. From within each quartile; two students were randomly selected and sent an invitation to attend the problem solving interview. This was to ensure a spread of abilities attending the sessions. The study was blind in that the interviewer who collected and analysed the data was not told which quartile the students were in. If there was no response from the selected students another two people were emailed from each quartile. They were given two main incentives including feedback / feedforward on their work and a £10 Amazon voucher for their time. Feedback focussed on what they did or did not do in their solution process, whereas feedforward recommended what they could do differently for future problems.

In all research conducted, each participant was informed that no data, in any form of write up, would be personally identified to them. Reassurances were given around protection of data, anonymity and confidential handling of data. Participants gave their verbal or written consent after a clear explanation of the study (Appendix A.5). Short instructions for the think aloud approach were given in order to keep the process clear and simple. Students were asked to vocalise everything they thought or did whilst solving the problems. Students were recorded using a Livescribe smartpen digital voice recorder. As discussed in Section 2.1.5, this recorded what the students wrote and said in synchronisation. Capturing their writing was particularly appropriate for this study, because it captured their solution process in real time and occasionally students worked through the problem on paper without vocalising it. It was also important in capturing representational use.

The first recommendation by Someren et al. is that the setting should assist

in making the subject feel at ease, especially as think aloud is potentially a long process, with the potential for high cognitive load [89]. Quiet rooms within the Physics building were selected for both studies.

Study	Number of participants	Number of Males	Number of Females	Number of Majors	Number of Non-Majors
1B Study	6	4	2	3	3
2B Study	7	1	6	7	0

Table 3.2: Demographics of students participating in the 1B and 2B studies.

The number and main demographics of students who participated in each study are shown in Table 3.2. In the 1B study there were two students from the top quartile, three from the second quartile and one from the third quartile. No students responded from the lowest quartile. In the 2B study there were three students from the top quartile, two students from second quartile and two students from the third quartile. The 2B study was conducted a year later; with the same cohort of students, but selection ensured different participants from the 1B study. As with the 1B study students in the lowest quartile did not reply. Random selection resulted in more females being asked in the 2B study. This study was not concerned with gender differences, but with student strategies using resources whilst solving physics problems. Finally, there were multiple differences between the 1B and 2B data, such as a change in interview protocol, year group and availability of resources, which are mentioned further in the following sections. Therefore it was not intended that these two cohorts would be directly compared, but similar patterns were observed which are discussed together.

### 3.3.1 1B Study Design

In terms of interview protocol, the first part of the interview was to explain the purpose of the research in general terms and the plan for the session. A detailed protocol was taken into the interviews by the researcher. This is given in Appendix A.4. A warm up question was presented first, followed by two physics problems. Section 3.4 gives the reasons for choosing the specific questions given.

### 3.3.2 2B Study Design

A similar overall structure to the 1B study was used, except the 2B study was the second in a two part study. Students started on ill-structured problems for the first part of the study (not relevant to this research). It was assumed that due to the very different nature of the task before the study reported here that students would be largely unaffected by this prior activity. The same protocol and main questions (Appendix A.4) were used with the exception that, as this study was the second part of a two part study, students were not given a warm up question. It was presumed that they were already accustomed to thinking aloud and the researcher did not wish to make the interview too long in order to minimise the possibility of fatigue.

Students in both the 1B and 2B studies attempted the drawbridge and army cadet questions which are the areas for analysis in Chapters 4 and 5. The 2B study differed in design in that students were given access to resources whilst solving these physics problems. The resources allowed in the 2B study included a Halliday, Resnick and Walker textbook [103], the specified textbook for introductory physics courses at the institution, and the Internet. Students were told they could use these resources to help them, but it was not a requirement, and indeed some students did not use them all at. An interviewer had to remain in the room to collect information on the students' use of the textbook. The interviewer noted textbook use manually, noting the page number and behaviour along with other relevant field notes with a smartpen, so the notes could be matched up to the audio. There was no recording device that captured students' use of the textbook, as having a video camera was considered too distracting for the students. Internet use was recorded using Camtasia, the screen capture software which also records audio.

### 3.3.3 The Role of the Researcher

The role of the researcher is considered important in the think aloud process [89]. The interviewer is advised to keep disturbance of the think aloud process to a minimum to avoid influencing students' thought processes [89]. From the pilot think aloud study and from sitting in on the warm up questions it was clear that the interviewer was having an affect on students, and this was different for



different students. This included students trying to engage with the researcher by asking a question or ignoring the researcher completely. This was taken to be an indication that the researcher would have an influence on the students when in the room and thus should not be present. Therefore the interviewer was not in the room for the 1B study. The role of the researcher was evaluated at the end of this study, lessons were learned from this approach and the 2B study adjusted. Details of these implications of this are discussed in Section 3.6.7.

The main problem was long periods of silence and no writing. In light of a review of this, the interviewer remained in the room in the 2B study. This was to try and minimise long periods of silence like those that occurred in the 1B study. The interviewer used prompts to encourage the student to vocalise their thoughts, such as “What are you thinking?”. The current method of prompting students was considered more successful in gaining an insight into student problem solving, as there were fewer occasions of silence when analysing the protocols. However the students’ natural solution process may have been altered by this. This is discussed further in Section 3.6.7.

## **3.4 The Questions**

The questions presented to students are explained below. These physics problems and their answers are also given in Appendices A.1 and A.2. Students were given 15 minutes on each question. Unless the student wanted to continue, the researcher moved them onto the next question. This was because most students should have made substantial progress by then and will have displayed key problem solving traits by this point that can be analysed. It was hoped that limiting time on each question would also prevent fatigue.

### **3.4.1 The Warm Up Problem**

A warm up question at the beginning of the experiment was used in order for students to begin thinking aloud automatically and without difficulty. Van Someren proposed that practice is essential, not just to train the student in the think aloud protocol, but for the investigator to correct the subject if they begin to interpret their thoughts as opposed to verbalising them [89]. Though this did not seem to be a problem, some students had to be encouraged to speak more

during the warm up question. The warm up question was chosen from Halliday, Resnick and Walker [103] the students' course textbook. A simple mechanics question was chosen to settle the student and help them gain confidence in the procedure and environment. The question was:

*When the legal speed limit for the New York City Thruway was increased from 55km/h to 65km/h, how much time was saved by a motorist who drove the 700km between the Buffalo entrance and the New York City exit at the legal speed limit?*

### 3.4.2 Main Question One: The Drawbridge Question

*A man is trying to push a large stone block across a uniform drawbridge which has a mass of  $m = 200\text{kg}$ . The mass of the large stone block and the man combined is  $M = 300\text{kg}$ . The drawbridge is of length  $L = 5.00\text{m}$  and is held up by a wire. The wire can withstand a maximum tension of  $5000\text{N}$ . The vertical distance between the drawbridge and the point where the wire is attached to the wall is  $h = 4.00\text{m}$ . How far along from the wall can the man push the stone block before the bridge collapses? Assume  $g = 10\text{ ms}^{-2}$ . From Halliday, Resnick and Walker p.323, question number 27 [103]. The diagram in Figure 3.1 was provided with the problem statement.*

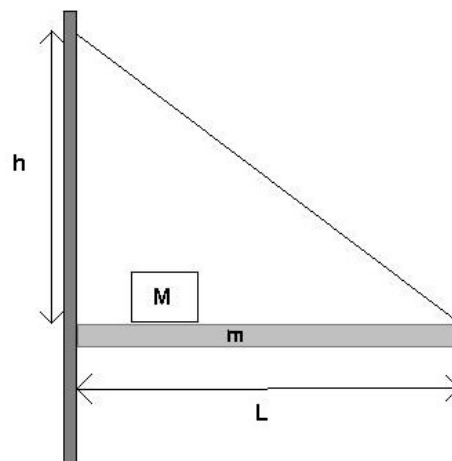


Figure 3.1: The diagram provided to the students in the drawbridge question

This question was adapted slightly with minor adjustments made to the

diagram. The diagram was given to students deliberately to assess whether a representation assisted their solution process, as it would be expected to help the students describe the situation [104].

This question was chosen, as it was a multi-step, reasonably well-structured problem, with a final correct answer, much like the type of question students would be expected to do as part of their weekly assignments. It was also not necessarily clear from the question which physics concepts would be most useful. The best approach to this question is to balance torques, with contributions from the drawbridge and man and block downwards and the vertical component of the tension upwards.

### 3.4.3 Main Question Two: Army Cadet Question

*An army cadet uses a rope swing to cross a river. He starts on a high platform with his centre of mass 7.3 metres from the ground and grabs onto the rope which is 6m long, to swing. He lets go when the rope is vertical and his centre of mass is 1m off the ground. He needs to travel 5m horizontally to clear the river after letting go of the rope. If he successfully clears the river what is the tension in the rope just before he releases? The cadet weighs 80kg. (Inspired by Van Heuvelen's paper [22])*

A contextualised, multi-step question with a final correct answer was chosen as the second main question. This question was adapted from an article about a course using case studies by Van Heuvelen [22]. One approach to solve this question is to calculate the velocity when the cadet leaves the rope by using his projectile motion towards the ground. This value for velocity is required when calculating the tension. The centripetal force is the sum of the tension in the rope and the cadet's weight (both of which act in opposite directions). This equation can be rearranged to calculate the tension. A full solution is shown in Appendix A.2.

Both of the main questions were selected so as to sit between well-structured and ill-structured problems [26] (Section 1.2).

## 3.5 Analysis

In order to investigate problem solving approaches and success there had to be a quantitative measure of student success on the two questions set. As the aim of the study was related to informing research on open and closed-book exams, the student's working was marked in an exam style. Each question was marked out of 10 marks. The breakdown of how the marks are allocated is given in Appendix A.3.

The methods employed for analysing the data collected from the think aloud interviews are discussed below. An overview of the structure for this research is shown in the Table 3.3.

Data Collection (Chapter 3)	Think Aloud with Smartpen
Data Analysis 1 (Section 3.6)	Behaviours coded according to problem solving models. Transitions made between behaviours counted.
Data Analysis 2 (Section 3.8)	Thematic analysis
Results 1 (Chapter 4)	Results from applying behavioural coding.
Results 2 (Chapter 5)	Results from thematic analysis.
Discussion (Chapter 5)	Results from both sets of analysis discussed with respect to knowledge structure theory (Section 1.5).

Table 3.3: Summary of the structure of the individual problem solving research, including data collection and analysis techniques employed for both 1B and 2B students.

## 3.6 Quantitative Coding of the 1B and 2B data

As reviewed in Section 1.6 the models selected for use in this study were:

1. The Minnesota Problem Solving Model
2. Problem solving behaviours from Schoenfeld's [24] and Kohl and Finkelstein's [2] research.
3. Representations

### 3.6.1 Model 1: The Minnesota Model

The Minnesota model is a general problem solving model and was introduced to students in their first year and is intermittently reinforced in solutions to problems. The stages are Focus on the problem, Describe the physics, Plan your strategy, Execute the strategy and Evaluate the answer (given the acronym FDPEE). Applying this model to the analysis can investigate whether students use the strategies recommended through instruction and gain an insight into their own problem solving strategies. Each part of the transcript can be broken down into the points at which the students exhibit each FDPEE behaviour and the order they used these behaviours throughout their solution. The coding for the Minnesota Model was based on the titles and sub-titles in the Physics 1A: Foundations 2012/13, Student Course Handbook [105]. The sub-titles from the course handbook are shown in italics to differentiate them from additional comments added with respect to how the codes were used in the study.

#### **Focus on the problem (understand the problem)**

*Draw a sketch* - Any evidence of a diagram being constructed (free body diagrams were coded as ‘describing the physics’)

*Choose a sensible notation* - Choosing symbols to represent physical quantities and indicating their definition.

#### **Describing the physics (analyse the problem)**

*Identify the principles* - Identifying which physics principles to use in the question. Any use of a free body diagram or other tools to solve the problem are included here.

*Formulate the equations* - Translating the problem into a mathematical form using symbols (when numbers are used to calculate a value, this is usually execution).

#### **Plan your strategy**

*Devise a strategy* - Evidence of planning including how principles, equations, diagrams etc would be manipulated. Planning more than one step in advance.

*Do not substitute numbers until you must* - This is a guideline in the book and does not refer to the planning stage in this context.

#### **Execute the strategy**

Execution is any rote or mechanical working which proceeds rapidly without much conceptual or planning thought, for example, solving a quadratic equation by standard techniques. Includes basic calculations including calculating the

weight.

*Use quantities in consistent units* - this is the guideline given in the course handbook, but explicitly checking the consistency of units would come under evaluation.

#### **Evaluate the answer or result**

*Asking if the answer or sub-answers make sense* and are physically reasonable. Thinking critically about the stages or final stage of the problem. Other recommendations in the course hand book include *checking the units, appealing to special cases and be prepared to look from a different angle*. Evaluation includes using another approach to solve the problem to see if the results are the same.

### **3.6.2 Model 2: Problem Solving Behaviours**

The seven codes of this model are Reading, Translation, Analysis, Exploration, Planning, Implementation and Verification. Kohl and Finkelstein's codes [2] are heavily quoted in order to apply the same coding scheme to this study. Quotations are indicated by italics. Since their interviews were videotaped some of their categories were set by physical gestures or looks. Videoing students was not done in this study, so some of the descriptions of each category have been adapted as appropriate.

#### **Reading**

Reading or re-reading the problem statement aloud in some way at any point throughout the solution. It was coded as silence if students appeared to be reading, but were not vocalising anything, as there was no way to know what they were thinking at this time. A second coder who was not part of the interviewing process would be unable to discern when silence was reading, so silence was not coded as reading at any point.

#### **Translation and Analysis**

The direct quote from Kohl and Finkelstein was used consistently as the criteria for the translation and analysis behaviour codes.

***Translation:** Taking information directly from the problem statement and representing it. This includes writing numerical data or the quick construction of a diagram on which to place data from the problem statement. This does not include substantial work independent of the problem statement. p.1557 [2]*

**Analysis:** Represents a directed, systematic attempt to understand or solve the problem. It can include constructing supplementary representations such free-body diagrams or pictures once the reading and translation phases are over. It can include talking out loud about their understanding of the problem or calculating intermediate quantities that need to be obtained before a final answer can be reached. Reasoning does not have to be correct to be coded as analysis. p.1557 [2]

### **Exploration**

Exploration was a difficult behaviour to code. It can be difficult to differentiate between exploration and analysis. Kohl and Finkelstein defined it as:

**Exploration:** A less-structured version of analysis. The student was searching for options or trying things out with little direction. Examples include a student searching through equations in the book, listing remembered equations without making use of or reference to them or cycling through their previous work out of apparent inability to proceed further. p.1557 [2]

Exploration is an undirected, less structured version of analysis. From the definition of a true problem, as defined in the literature review Section 1.1, a degree of exploration could be expected if students do not immediately see how to solve the problem. Exploration was defined as when the student had no idea of how to proceed and was just trying or saying anything even remotely associated to the problem.

### **Planning**

Planning was defined as the student explicitly planning more than one move in advance. For example, saying 'I am going to draw a diagram' is not planning, as this just describes their immediate intentions. Saying 'I am going to draw a diagram and then balance the forces acting on either side of the pivot' was coded as planning. Students implementing their plan was coded as implementation.

### **Implementation**

Implementation includes activities such as entering numbers into a symbolic equation to solve it and typing on a calculator. It is usually after the student has constructed the symbolic equation (analysis) and are in the stage of putting numbers into the equation to find an answer.

### Verification

The description for verification was unchanged from Kohl and Finkelstein's study.

***Verification:** Follows the discovery of an answer (right or wrong) and involves taking some kind of step to check the correctness of the answer (a final or intermediate answer). If a student decides outright that an answer is wrong and begins work again, this is analysis, exploration, and/or implementation rather than verification. This includes reflection on whether the answer obtained is reasonable or expected.* p.1557 [2]

### Additional Code

An additional code of 'silence and no writing' was created, as there was no indication as to what the student was thinking in these times. Note that none of the codes require students to be correct.

### 3.6.3 Model 3: Representations

As discussed in the literature review Section 1.4, representation use is an established area of research [2, 30, 31, 32]. The use of the smartpen enabled the real-time capture of representation use. The initial coding used the same codes as those in the literature [2], that is for students' use of Equations, Numbers, Diagrams and Words. However, from initial analysis there was insufficient clarity between Numbers and Equations, so these were further expanded. Therefore the final codes were: Equation symbolic, Equation numeric, Free numbers, Diagrams and Words. Equation symbolic was an equation written out symbolically. A numeric equation was an equation where the students had put numbers into the equation, as opposed to symbols. Finally, free numbers were values such as "d=5cm" which were not written in equations, but numbers which perhaps came from the problem statement and written to stand alone on the page.

### 3.6.4 Applying the Models

Kohl and Finkelstein [2] split the data into 10 second blocks and coded the dominant activity or representation used in each block. Using this method, students were classed with the dominant behaviour in each 10 seconds. There



was rarely a problem of two representations or behaviours being used in the same 10 seconds. The Livescribe episodes were coded straight from the recordings.

### 3.6.5 Inter-rater Reliability: Validation of the Coding Scheme

Coding was completed independently by two coders. One researcher coded all the interviews (two questions for each participant). A second researcher<sup>1</sup> coded a random selection of interviews in order to check inter-rater reliability. Neither coder knew the student's quartile, and therefore the expected student ability whilst coding. Coding was compared between two coders in order to validate the coding scheme. For each 10 second slot it was noted whether the coders were in agreement as to the behaviour at the time or not. In times of disagreement the default was to coder one, as this coder had analysed all episodes and so results were consistent with the other episodes that were not coded by two coders. There had been no discussion of the coding system between coders before the first round of coding. The codes were refined after the initial coding and all the data re-coded in light of this. The specific discrepancies in each 1B and 2B study are discussed below and how the coding was then refined.

#### 1B Study

Coder two coded one question from every student using each of the three coding schemes mentioned above. The initial inter-rater agreement results are shown in Table 3.4. Some iterations were conducted during this process in light of the inter-rater scores. The main adjustment to the coding scheme was the introduction of sections of silence and no writing, where student behaviour was indecipherable. With regards planning there was confusion between when students were planning and when they were just saying what they were going to do next. Thus the criteria for planning was updated to: *Planning - When students show their intentions for more than the immediate next phase of the operation i.e. what they are planning to do a few steps ahead.*

Student 2 in Table 3.4 had low inter-rater agreement due to multiple and complex behaviours exhibited. Student 6 had multiple disagreements on

---

<sup>1</sup>Many thanks to Ross Galloway for assisting in the coding of this data.

representation use due to insufficient clarity with the representation coding, as discussed in Section 3.6.3. Another iteration was therefore conducted to modify the representation use in light of this initial coding to separate equation numeric, equation symbolic and free numbers. Students were re-coded by Coder one in line with the new coding scheme. Overall these were initial coder agreements and after further refinement this process improved, as demonstrated in Table 3.6. The average agreements for the 1B data are shown in Table 3.5. The final average results in Table 3.5 showed on average a reasonable agreement between coders.

Student	1 (Army)	2 (DB)	3 (DB)	4 (DB)	5 (Army)	6 (Army)
Minnesota % agreement	90	56	88	82	82	91
Schoenfeld % agreement	80	55	74	74	78	90
Representations % agreement	82	82	80	67	87	56

Table 3.4: The percentage agreement between two coders analysing the same question for each student. The question that was double coded for each student was randomly selected from the two questions the students were given. The question used to determine inter-rater reliability is shown beside the student number with ‘Army’ to represent the army cadet question, or ‘DB’ for the drawbridge question.

Model	Percentage agreement
Minnesota	82%
Schoenfeld	75 %
Representations	76 %

Table 3.5: Average percentage agreement on 1B episodes double coded

## 2B study

Five students on a randomly selected question each were coded by coder two. After the first two students were analysed by both coders, the Minnesota and Schoenfeld codes were discussed in order to establish the agreed coding for these students and iron out any misunderstandings of the model. There was initially a low agreement between the two coders for the first two students with respect to the Minnesota model, with 40% agreement for Student 1 and 63% agreement for Student 7. These episodes were discussed and coding agreed by both coders. One major discrepancy was whether certain sections were classified as silence or another behaviour. Unlike the 1B study, the interviewer was one of the coders and included direct observation of the students' behaviour, so marked on short periods of silence less regularly than Coder 2. For example, the interviewer could see that the student was silently reading the question at the beginning so coded it as such. Using this method of quantitative coding, information is lost in this way, because it has to be classed as 'silence', as the second coder brings no knowledge of the interview to the coding. The other main disagreement was whether checking the Internet was 'describing the physics' or 'execution'. Internet searches were classified as 'describing the physics' where students used it to find information rather than working out quantities.

After discussion the two coders reached complete agreement for these two students. All the remaining episodes were re-coded by coder 1 in light of the clarifications. The clarifications and adjustments to the coding scheme are included in the previous Section 3.6 which describes the models. Further double coding was completed to check coding reliability. Table 3.6 shows the inter-rater agreements for each student with respect to each coding model. The subsequent inter-rater agreement percentages were in an acceptable range, with average agreements shown in Table 3.7. Episodes where a joint conclusion was reached were excluded.

Student	1 (Army)	2	3 (Army)	4 (DB)	5	6 (DB)	7 (DB)
Minnesota % agreement	100*	/	82	79	/	83	100*
Schoenfeld % agreement	100*	/	78	86	/	90	100*
Representations % agreement	88	/	82	76	/	75	82

Table 3.6: Coding record of the 2B recordings. The percentages show the inter-rater agreement on the particular student, question and model. An asterisk at those at 100% agreement indicates when coding was agreed by both coders and a joint conclusion was made. The question used to determine inter-rater reliability is shown beside the student number with ‘Army’ to represent the army cadet question, or ‘DB’ for the drawbridge question.

Model	Percentage agreement
For three students	
Minnesota	81%
Schoenfeld	85 %
For five students	
Representations	81%

Table 3.7: Average percentage agreement on 2B episodes double coded

### 3.6.6 Sources of Error

The smallest time slot for each behaviour was 10 seconds which meant there was a degree of error in recording the length of time on each behaviour. For this reason, exactly how long students were exhibiting each behaviour was not the focus of analysis in this study, but rather whether the behaviour existed and in which order the student transitioned between behaviours.

If students were speaking or writing they were coded as one of the behaviours, even if they were not specifically doing any of the activities in the coding scheme, for example by saying ‘hmm I don’t know’ was coded as analysis, as they were engaged and thinking about the problem. This happened infrequently, but meant that information was lost in that no new categories could be created to encompass different behaviours.

It was occasionally quite difficult to determine whether the student’s behaviour was analysis or exploration in the behavioural model used by Kohl and Finkelstein. For example, a student that was unsure and speaking slowly or in a

stilted fashion could still be doing physics in a directed way and have behaviour categorised as Analysis rather than Exploration. Students were likely to lack certainty in their answers, as they did not know in advance how to solve the problem. This shows the exercise was a true problem for them.

The coding did not indicate why students chose a particular approach. Coding with pre-set codes missed a large part of students' problem solving behaviour. Although there are some insights into student behaviour from the coding, Flick [88] emphasises that quantitative research is restricted in the interpretation.

Finally, the models are quite simplistic and do not seem to adequately describe student behaviour, as they miss perhaps essential information. A qualitative approach, discussed in the next section 3.8 was used to triangulate findings from the quantitative methods and to illuminate students' problem solving processes that were not captured with the quantitative approach.

### 3.6.7 Methodological Assessment

Time spent silent was studied in order to evaluate the think aloud method. This assessment was based on the reasoning that silent students are likely to have been thinking about something, but not vocalising it. The ideal think aloud process would capture everything that students are thinking without cognitive overload. Long periods of silence could indicate cognitive overload or an inability to vocalise what they were thinking about. A concern was that students when left alone would not speak as much as they would if someone else was in the room, so would spend a longer time silent on average compared to prompted students. This was not numerically justified. A Mann-Whitney test was used, due to the non-parametric nature of the data [106], to compare the percentage of time that students in years 1 and 2 were silent during their solution attempt. No significant differences between years 1 and 2 on the percentage of time spent silent were found, with a U value of 62.5 and  $p=0.269$ . This means that the un-prompted group did not spend on average a longer time silent and not writing than the prompted group. This however does not account for the length of time of each silence, for example perhaps 2B students had less periods of extended silence and more shorter chunks. Nor does it account for other factors, such as the possibility of the 2nd year students' ability to vocalise their solution being different, by being a year further on in the University system.

From the interviewer's impressions, prompting seemed to initiate students to vocalise what they were thinking when they may not have been inclined to do so without a prompt, however this did not significantly decrease on average the percentage of time they spent silent when compared to the 1B study. When prompted out of silence this may have changed what the student was actually going to do. In externalising their thinking, students may form more of a structured plan than they would have originally. Prompting may have encouraged more strategic thinking, as students when asked 'what are you thinking now' had to then ask themselves this question and think about what they were thinking. Perhaps this encouraged more meta-cognitive thinking, but there is no method in this study to prove this speculation. The 2B students seemed to exhibit more metacognitive behaviours than the 1B students. Nonetheless, learning what their thoughts were after prompting in the problem solution was considered more important than the interruption.

A mix of the two methods was considered the optimal procedure for future work. Although there was no statistical benefits from prompting students in terms of the amount of time they were silent, there were observed additional benefits to the researcher being present in the room. It meant that the researcher could take field notes on the interview and immediately begin to understand students' problem solving procedures which could be used for targeted questions later in the interview. Perhaps having the researcher in the room and only prompting after periods of extended silence would be the best way to have the benefits of both methods.

### **3.7 Analysis After Coding**

From the coding of student episodes, temporal maps were created, which indicated the students' behaviour in each 10 seconds of their interview. Overall structural patterns were sought that could indicate problem solving archetypes.

In order to try and capture students' strategic control [24], transition maps for each students were created for each model. Behaviour over the course of the solution was tracked by numbering the transitions. Vue concepts mapping technology from Tuft University [107] was used to map transitions between behaviours. This concept mapping software allowed descriptive nodes to be

created and linked to other nodes, whilst also allowing links to be labelled. Transitions between activities in the Schoenfeld model were presented and used to study the data. Studying transitions reduced the effect of errors due to the 10 second timing slots, as only whether the behaviour occurred and in which order was needed to create the transition map. The temporal coding maps could be used to show if any activities were maintained for a particularly long or short time.

The total number of transitions each student made were counted. A template was made in Excel to code automatically for transitions. Some transitions were also counted manually to check that the coding process was going correctly. The number of transitions made were calculated as a percentage of the total possible transitions in order to normalise for time spent on the problem. The total number of possible transitions was  $N-1$  where  $N$  was the number of 10 second slots. For example, if the student had 10 transitions out of 20 possible transition points ( $N-1$ ) then their percentage number of transitions they made was 50%. This was called the normalised number of transitions. All the transition data presented in the results and discussion is without silences, so the final analysis did not include transitions made in and out of the code of “silence and no writing”. This assumed that students will have been in, or moved to, the code that was noted before or after the period of silence. A focus on externalised information was taken, as there are a number of possibilities of student behaviour doing in periods of silence, including not thinking about the problem at all. Counting transitions was not conducted for representation use, as usually students had long gaps between representation use and this did not provide any further insights into the data.

The temporal and transition maps were grouped per question and organised according to the students’ performance on the question. After looking at all the transition and temporal maps, for the 1B and 2B students, there seemed to be differences in complexity. The high scoring students appeared to have simpler transition maps than the lower scoring students (under 50% correct). Quantitative scores were devised to measure for the complexity. These are called complexity scores 1 and 2.

Complexity score 1 counted the number of different types of transitions made. A transition type was defined as a transition between two behavioural codes, for example, reading to translation. This transition was only counted once,

even if the student made this transition many times. This was to capture and more quantitatively measure the differences observed between the successful and unsuccessful students. Transition types were considered in order to quantify the types of transitions students were making, rather than their frequency. Thus an episode with transitions to and from a variety of different behaviours was considered more complex with respect to this measure.

Complexity score 2 looked at changes to new behaviours. The count is increased by one if a student moves to a new behaviour, which was different from the previous behaviour. For example, starting in behaviour A and changing to behaviour B, would increase the count by 1, but from B back to A would not increase the count (as A was the previous activity). This is another measure of change in behaviour as it counts for changes in behaviour without over-counting when many transitions were made back and forth between two codes. It is another measure of the complexity of transitions. For statistical analysis it was normalised by the number of possible transitions to account for length of time spent on the question, as the longer the student spent on the problem the more possible transitions they could make according to this measure.

Once complexity scores were calculated statistical analysis was conducted in IBM SPSS statistics version 19 to see if there were any significant correlations between complexity and score on the question. Two-tailed partial Spearman's correlation tests, controlling for question and year, was conducted to test statistical differences between complexity scores and mark, as well as the normalised number of transitions made and mark. The data was interval, as there were equal increments between complexity scores and number of transitions [106]. Although significance results were the same across two tests, Pearson's and Spearman's, the results of the Spearman's test are presented, as it does not assume normality. This is due to the relatively low data numbers and non-normality of all the data when plotted on a histogram.

### **3.8 Qualitative Analysis of the 1B and 2B data**

Conducting qualitative research on the same data was to triangulate results, look for other successful and unsuccessful traits that were not perhaps picked up in the quantitative coding, and give further insights into the solution procedure.



All 12 questions (6 students, 2 questions each) from the Physics 1B investigation and 14 questions (7 students, 2 questions each) from the Physics 2B investigation were transcribed for analysis. The transcription protocol was to use minimal punctuation. A short gap was denoted by .. or .... and new sentence was a new line. A long gap was noted as such in brackets. As the data was transcribed, the large amount of other data collected was also noted in the appropriate section of the transcripts. In both the 1B and 2B studies this included what students wrote or drew whilst solving the problem. In the 2B study, what students looked up either in a textbook or online, and any field notes written about them at the time or shortly after the interview were also recorded. After an initial look at the data, by full transcription and annotated reading, themes emerged relevant to the research aims of the study. The aims of this section of research, shown in Section 3.1, include: investigating the characteristics of the successful and unsuccessful students; studying their physics approach and exploring what happened in both correct and incorrect approaches. The emergent themes [88] related to students' approach to the problem and the skills they appeared to need in order to solve the question successfully. These themes were set so as to structure the qualitative analysis to gain insight into these particular areas of student problem solving, but were not intended to be conclusive. The themes were:

1. Ability to recognise an appropriate underlying physics approach (see Appendix A.2 for appropriate solutions to the problems set),
2. Ability to look up or remember information,
3. Ability to apply this information correctly.

All relevant codes were used as characteristics of high and low scorers to study the students' interactions with the correct approach when they were successful or unsuccessful in solving the problem. An assumption could be made that if students did not use the correct approach then they would not get the question right (and vice versa), but it was the process with respect to their approach, that was of experimental interest in this study. For example, did the students with the right approach know it instantly, or did they establish the approach through the course of the problem? Results are presented with supporting quotes as evidence.

One researcher coded the data, which is valid, as discussed in Section 2.2.3, but should be taken as that researcher's interpretation of the data. Quantitative

coding was also conducted - as discussed above - so results can show agreement or not between the quantitative and qualitative coding methods. Thus the study has multiple methods of analysis to support its claims.

### **3.8.1 Data Analysis 2B resources specifically**

In addition to the qualitative coding described above, the data analysis of the 2B data had an extra stage of analysis where specific points on how students used resources were noted. Resource use from the points mentioned below was then compared to students' success when solving the problems. The points of interest were:

- How many separate times the students accessed resources.
- Which resource was accessed i.e. was it the Internet or the textbook.
- The type of consultation the resource was used for e.g. checking an equation.
- Why they consulted the resource e.g. trying to find a similar problem (usually indicated from what they said before they looked up the resource).
- Whether the information looked up was useful and helped them progress in the problem.
- Whether they were able to use the information they found.

### **3.8.2 4th and 5th year students on the Physics Skills course**

Students on a 4th and 5th year general physics problem solving course, called Physics Skills were given the opportunity to take a smartpen home to use for physics. The Physics Skills exam tests knowledge from the first three years of students' undergraduate degree, with short written exam questions on a variety of topics. Its focus is on problem solving and coverage of previously learnt knowledge. This is a required course for Physics students and they were given two workshops in order to work through past papers and ask questions.

While they had the smartpens at home they were asked to record some think aloud problems on a physics skills past paper question that they had not seen

or attempted before. Two students responded to this request and each student provided solutions (or attempts at solutions) to two past paper questions. As they were doing these problems in a location of their choice, both students had access to the Internet whilst solving them, and made use of it. It is acknowledged that there are only two students in the sample, but discussion of these is added to the qualitative arguments already established from the 1B and 2B analysis. This also provided a means to understand what students in the older year groups were doing and how they compared to the 1B and 2B students.

# Chapter 4

## Individual Problem Solving: Modelling Problem Solving Behaviour

The previous chapter explained the implementation of this study with first and second year undergraduates. The first analysis of this data involved coding students' behaviours with respect to existing problem solving models. This chapter details the results from applying these models.

### 4.1 1B Quantitative Study Results

Table 4.1 shows the 1B students' mark on each question and their quartile placement (quartile placement being dependent on their first semester course results). No students from the bottom quartile attended either the 1B or 2B interviews. Only two students (Students 5 and 6) achieved marks over 50% on either question, with the remaining students scoring less than half the available marks.

Four of the temporal maps for the three coding schemes (Minnesota, Schoenfeld and representations) are shown below for each question for high and low scoring students. These demonstrate typical features found. The remaining maps can be found in the Appendix A.6. When studying these temporal maps the students' success should be kept in mind (by considering their mark on the question), as well as the order in which they progressed through the behaviours

Quartile	Major			Non-Major		
	Student	Drawbridge	Army	Student	Drawbridge	Army
Q1	5(M)	9	10	3(M)	3	2
Q2	6(M)	10	8	1(F) 4(F)	2 1	1 5
Q3	2(M)	2	5			
Q4						

Table 4.1: Student marks out of a possible 10 marks on each of the problems according to quartile placing and whether they were majoring in physics or not (non-major). The number in the student column identifies each student. These student numbers are used in discussion. M or F represents whether these students were male M or female F.

on both the Minnesota and Schoenfeld models. The temporal maps provide an indication of the processes that the student went through whilst solving the problem and show that students rarely followed the behaviours in order. There was lack of implementation near the end of the solution process in 5 out of the total 8 unsuccessful solutions on both questions (see all maps in Appendix A.6). Although, having implementation at the end did not always indicate success as the remaining unsuccessful students implemented to get an incorrect numerical answer.



### 4.1.2 Temporal maps from the army cadet question

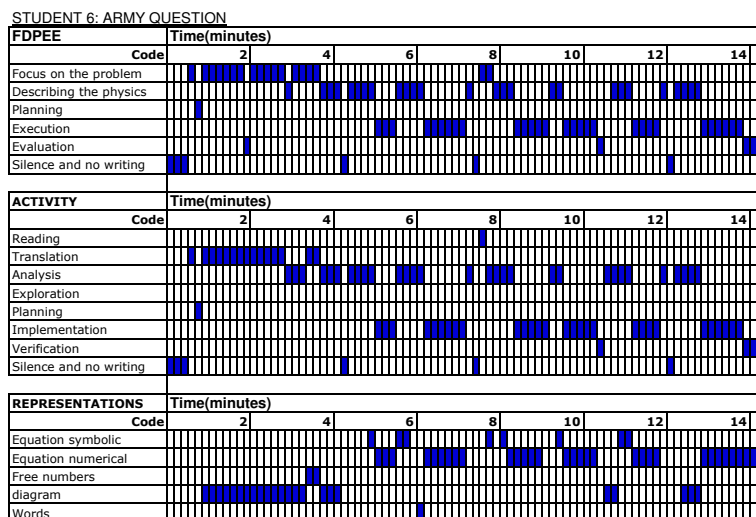


Figure 4.3: 1B Student 6 Army Cadet question - all models. Scored 8/10

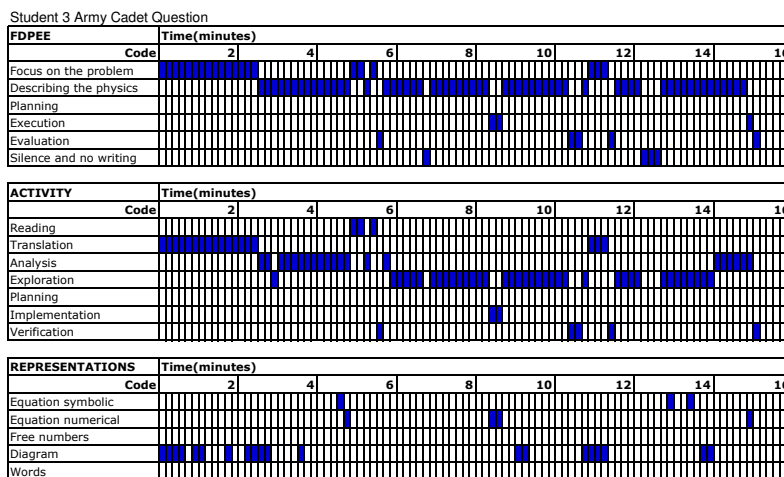


Figure 4.4: 1B Student 3 Army Cadet question - all models. Scored 2/10

## Representation Use

As mentioned in Section 1.4, the use of multiple representations is encouraged [22, 25]. Students in this study used multiple representations, as shown in the temporal maps here and in the 2B study. There were no obvious differences between successful and unsuccessful students, so this area of research was not pursued further.

### 4.1.3 Transition Patterns

Using temporal maps it was not possible to differentiate easily between successful and unsuccessful students, so the data was configured using transition maps. Examples of successful students' transition maps are shown in Figures 4.5 and 4.6. These are on the whole simpler maps than some of the maps for unsuccessful students, as shown in Figures 4.7 and 4.8. The full set of transition maps is shown in Appendix A.7. Using Schoenfeld's model, three types of transition maps emerged:

1. Successful students - Simple picture - as shown in Figure 4.5 and 4.6
2. Unsuccessful students - Complex picture - as shown in Figure 4.7 and 4.8
3. Unsuccessful students - Simple picture - Student 2 on army cadet question, as shown in Appendix A.7.



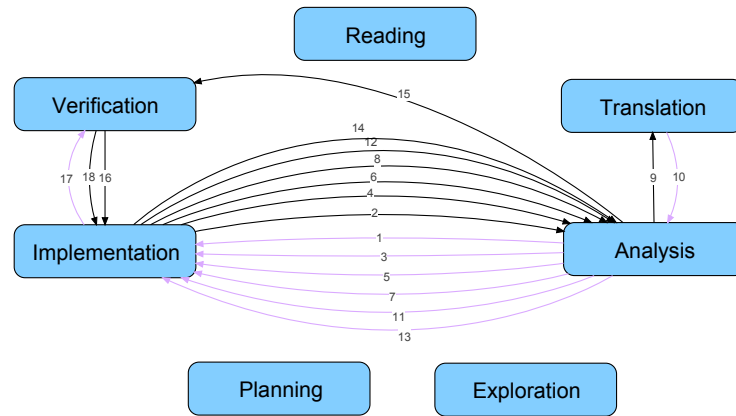


Figure 4.5: 1B Transitions Student 5 Drawbridge question - Schoenfeld Model. Scored 10/10. The numbers show the order in which the transitions were made over the course of the solution. The colour represents the direction of the transition.

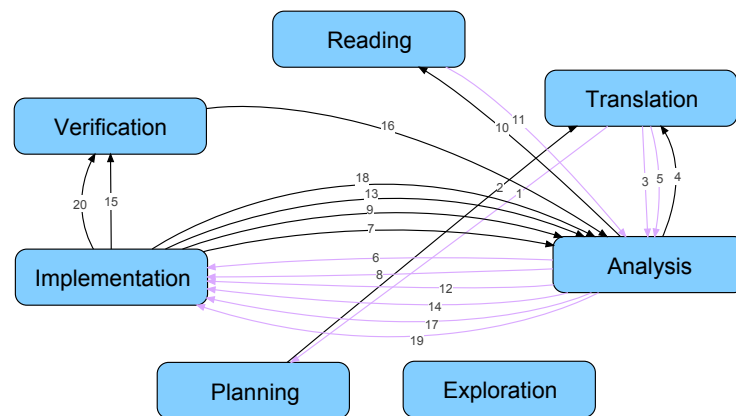


Figure 4.6: 1B Transitions Student 6 Army Cadet question - Schoenfeld Model. Scored 10/10

Unsuccessful students - Transition Maps

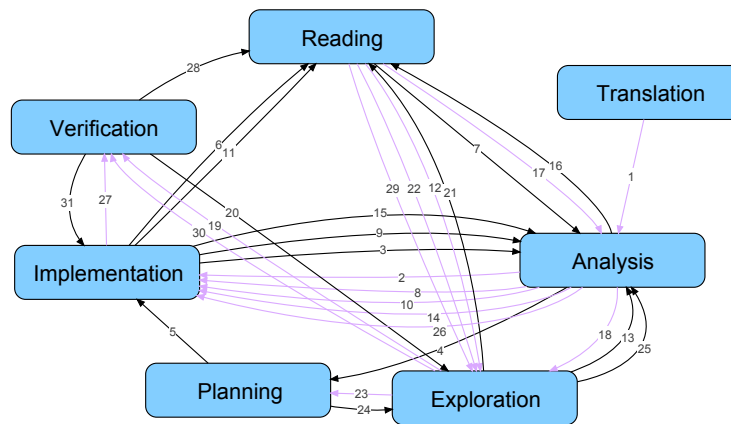


Figure 4.7: 1B Transitions Student 4 Drawbridge question - Schoenfeld Model. Scored 1/10

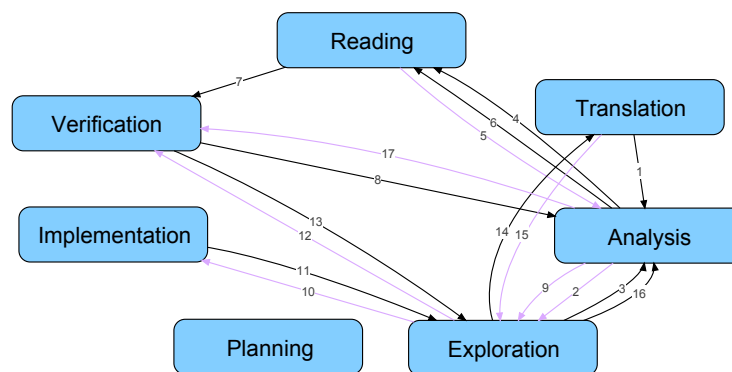


Figure 4.8: 1B Transitions Student 3 Army Cadet question - Schoenfeld Model. Scored 2/10

Complexity ratings were calculated on the number of different types of transitions made. Examples of complexity scores include the student in Figure 4.5 who had a complexity score one of 7, including translation to analysis, analysis to translation, implementation to verification, verification to implementation, analysis to implementation, implementation to analysis and analysis to verification, as the different transition types. The student in Figure 4.5 had a complexity score two of 5. Table 4.2 shows that students with, on average, a lower number of transitions types tended to do better on the problem. There are exceptions including Student 1 and Student 2 on the army cadet question which are examples of unsuccessful students with simple transition maps. Section 4.3 shows a statistical analysis of these results, which are discussed in Chapter 5.

Student number and Question	Mark	Complexity score 1	Complexity score 2
6 DB	10	6	4
5 Army	10	7	10
5 DB	9	7	5
6 Army	8	10	9
4 Army	5	14	13
2 Army	5	4	3
3 DB	3	10	12
1 DB	2	11	12
2 DB	2	12	17
3 Army	2	14	11
1 Army	1	9	6
4 DB	1	19	23

Table 4.2: Mark on the question and number of transition types, using the Schoenfeld model, for each 1B student. As discussed in Section 3.7, complexity score 1 counted the number of different types of transitions made. Complexity score 2 looked at changes to new behaviours.

## 4.2 2B Quantitative Study Results

Overall, as shown in Table 4.3, 3 students achieved a high score (over 50%) on both questions, with the remaining 4 students scoring 50% or below on both questions. Two of the students in the high scoring group were in the top quartile, with the third student in the second quartile.

Student	Nresource		Mark	
	Drawbridge Q	Army Q	Drawbridge Q	Army Q
Student 6, Q2	None	1I	10	10
Student 4, Q1	1B	None	10	6
Student 1, Q1	1I	1I	6	7
Student 7, Q3	2B	1I*	3	5
Student 2, Q2	4B	2B+2I	2	4
Student 3 Q1	None	1B	0	1
Student 5, Q3	2B	1B+2I	0	2

Table 4.3: Each student, their quartile (Q), the number of separate times they looked up a resource (Nresource) and their mark on the question. Q1 is the highest quartile and Q4 the lowest. In the Nresource ‘B’ means the students consulted the textbook and ‘I’ refers to the Internet, with the number preceding these letters the number of separate times they used that resource. \*Student 7 on the army question tried the textbook before consulting the Internet, but this was not counted, as she could not find what she was looking for and did not actually stop at a page in the textbook, just flicked through.

A few exemplar temporal maps are given below, with the full set in Appendix A.6. The blue band on the temporal maps show the approximate time that students accessed the internet as a resource, with the yellow band indicating when students accessed a textbook. 2B students seemed to have some verification and planning behaviour, but this was not consistent across all students. Again it is interesting to see what behaviour the students end on. Unsuccessful students on the drawbridge question tended to implement and then return to analysis at the end of the solution process. There was a lack of implementation or verification at the end of the solutions of 6 out of the total 8 unsuccessful solutions on both questions (see all maps in Appendix A.6). However, similar to the 1B students, implementation at the end did not guarantee success as some students progressed incorrectly in order to get a numerical answer.

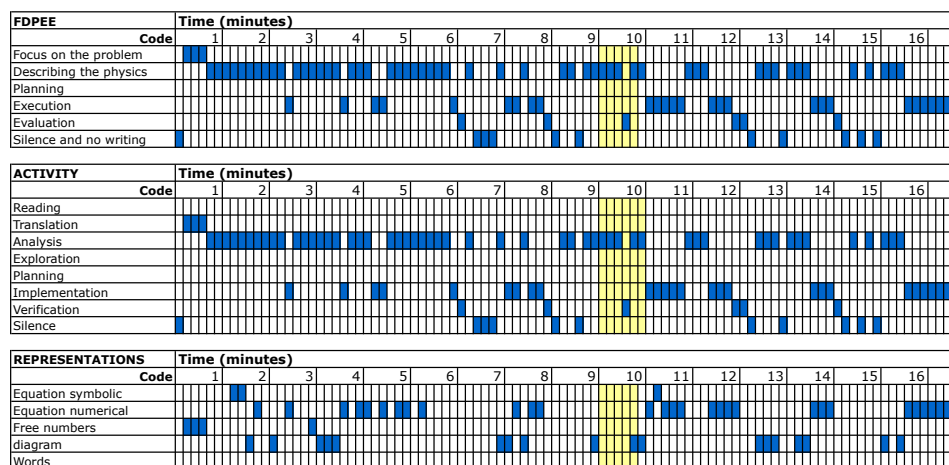


Figure 4.9: 2B Student 4 Drawbridge All Models. Student scored 10/10. Vertical yellow bars indicate use of a textbook and vertical blue bars the use of the Internet.

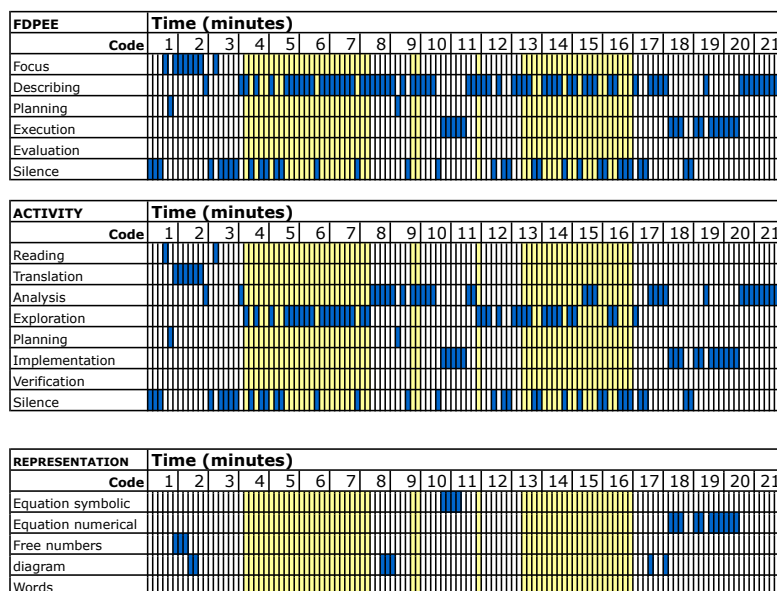


Figure 4.10: 2B Student 2 Drawbridge All Models. Student scored 2/10. Vertical yellow bars indicate use of a textbook and vertical blue bars the use of the Internet.

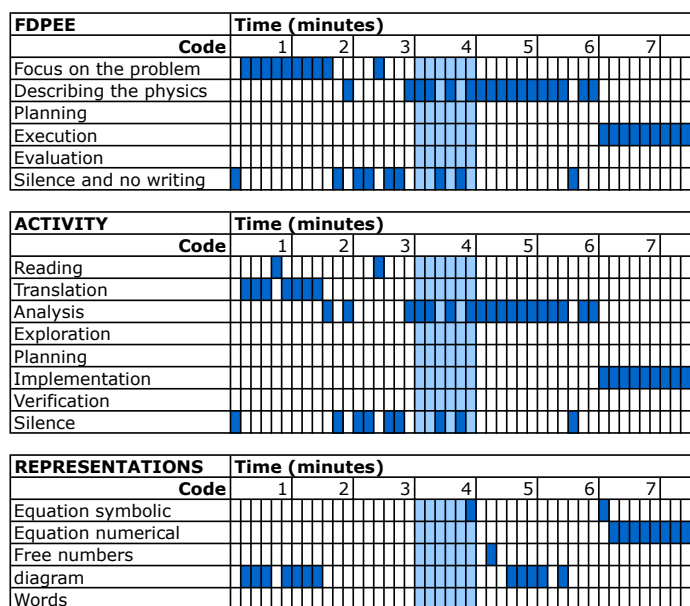


Figure 4.11: 2B Student 6 Army All Models. Student scored 10/10. Vertical yellow bars indicate use of a textbook and vertical blue bars the use of the Internet.

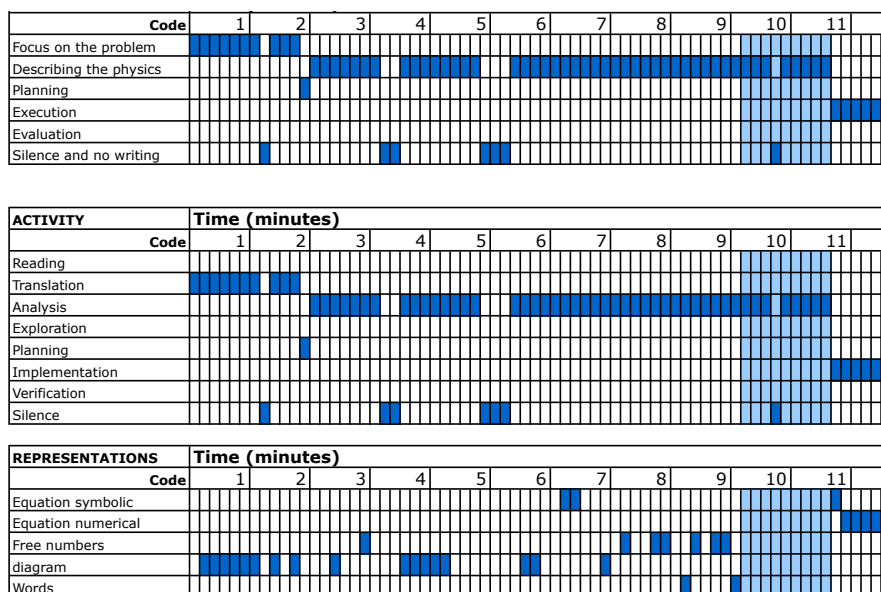


Figure 4.12: 2B Student 7 Army All Models. Student scored 5/10. Vertical yellow bars indicate use of a textbook and vertical blue bars the use of the Internet.

### 4.2.1 Transition Patterns

A similar set of transition maps were found in the 2B study, as occurred in the 1B study, with some differences. Again three types of transition maps were found. Relatively simple maps from the successful students and mostly more complicated maps from the unsuccessful students, see Figures 4.13 to 4.16. The full set of transition maps is shown in Appendix A.7.

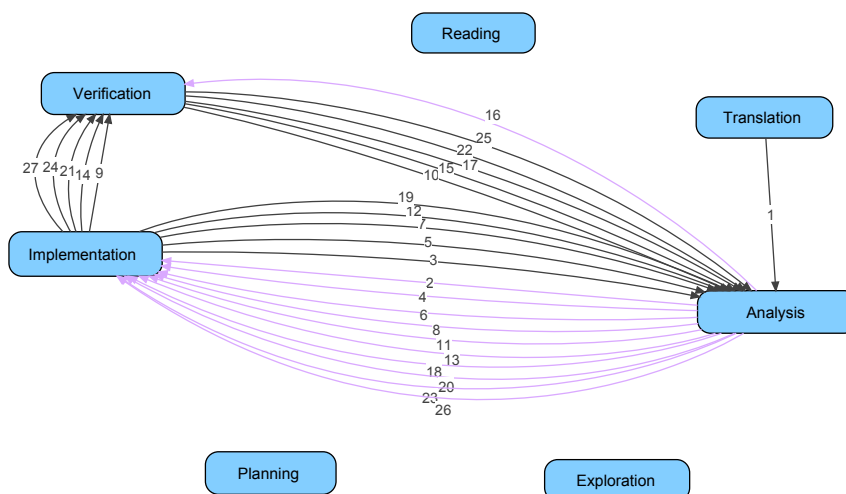


Figure 4.13: 2B Student 4 Drawbridge question - Transitions on the Schoenfeld Model. Student scored 10/10.

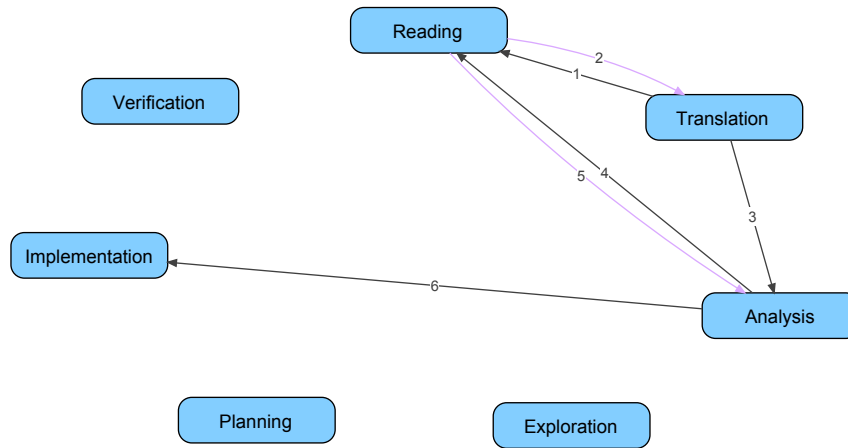


Figure 4.14: 2B Student 6 Army cadet question - Transitions on the Schoenfeld Model. Student scored 10/10.

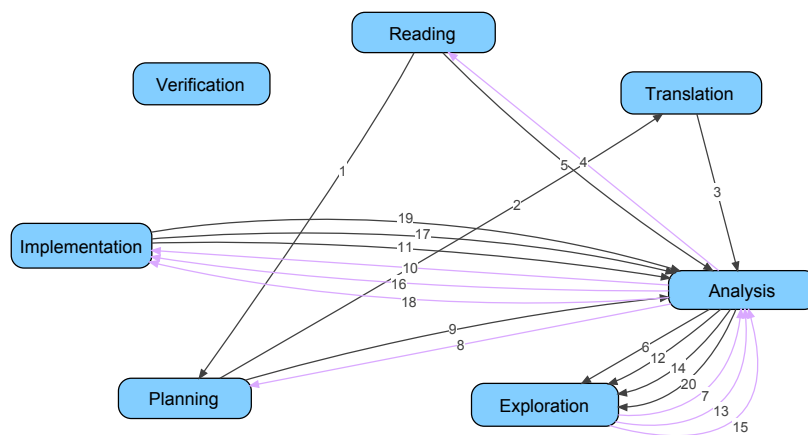


Figure 4.15: 2B Student 2 Drawbridge question - Transitions on the Schoenfeld Model. Student scored 2/10.



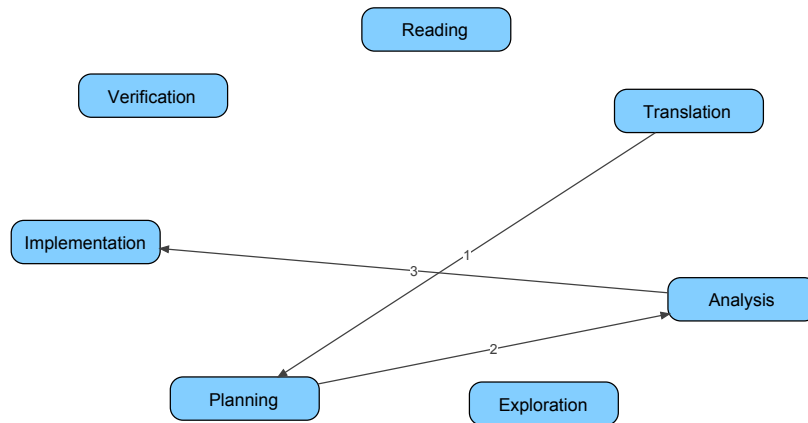


Figure 4.16: 2B Student 7 Army cadet question - Transitions on the Schoenfeld Model. Student scored 5/10.

As shown in Figure 4.16 there was one unsuccessful student with a very simple map. As discussed later Student 7 mentioned correct areas of physics on the army question, but did not apply them correctly. Perhaps lack of verification meant that she did not self-correct. The simplicity of this map is unusual, even for a successful map there seem to be too few transitions. This suggests that although successful students had simpler transition maps, they should not be too simple and some minimum threshold may apply.

Using the complexity scoring it can be seen that the number of different connections between behaviours to be on the whole higher for unsuccessful students, as shown in Table 4.4.

Student number and Question	Mark	Complexity score 1	Complexity score 2
4 DB	10	6	15
6 DB	10	6	5
6 Army	10	6	4
1 Army	7	6	3
1 DB	6	6	4
4 Army	6	6	5
7 Army	5	3	3
2 Army	4	12	9
7 DB	3	7	13
2 DB	2	11	10
5 Army	2	7	5
3 Army	1	12	9
3 DB	0	15	14
5 DB	0	6	6

Table 4.4: Mark on the question and number of transition types, using the Schoenfeld model, for each 2B student. As discussed in Section 3.7, complexity score 1 counted the number of different types of transitions made. Complexity score 2 looked at changes to new behaviours.

### 4.2.2 Resources Use

The approximate duration of resource access was included on the 2B temporal maps, to indicate resources use in the solution process in relation to their problem solving behaviours. The approximate length of time the resource was used was noted when the focus was on the resource and not on working. There was some error in the timing of resources use, because it was difficult to note exactly when they were used. However, one can hear the pages turning or Camtasia picked up use of the computer, so along with things the interviewer or the student said and the field notes written at the time, there was usually a reasonable indication of when participants started and stopped looking up a resource. The approximate percentage time that students spent using resources as a fraction of the total time on the problem (per 10 seconds) is given in Figures 4.17 and 4.18 for the drawbridge and army cadet questions respectively. They show that the more successful students (nearer the top of the figures) used resources, in general, less and for less time than unsuccessful students. Students that used resources more scored lower than those who did not use resources or used resources less. This

confirms the hypothesis that students cannot just use a resource to be successful. The implications of this for open-book exams are discussed later.

Another finding that is clear from studying Figure 4.18 is that none of the students looked up a resource in the first 40% of their time spent on the army cadet problem. This may be explained by a longer period of time needed to set up the problem when compared to the drawbridge question (Figure 4.17), as there was no diagram provided from the army cadet question and it had multiple stages. The temporal maps also verify this theory (Appendix A.6). These show that students seemed to spend longer ‘focusing on the problem’ than in the drawbridge question.

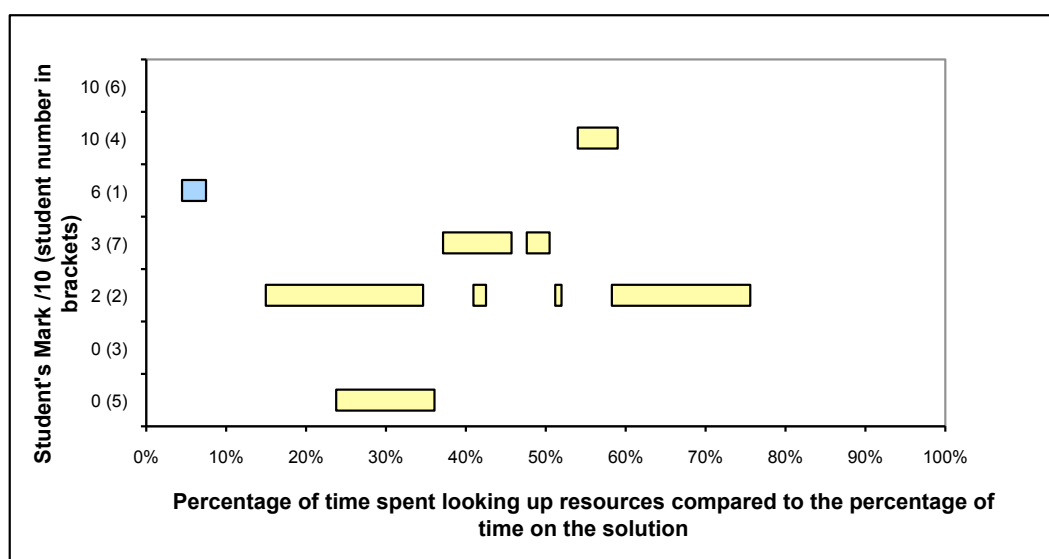


Figure 4.17: Percentage of time that students spent looking up resources (coloured bars), in terms on the percentage of time they spent solving the drawbridge problem. A yellow bar indicates when a textbook was consulted and a blue bar the Internet.

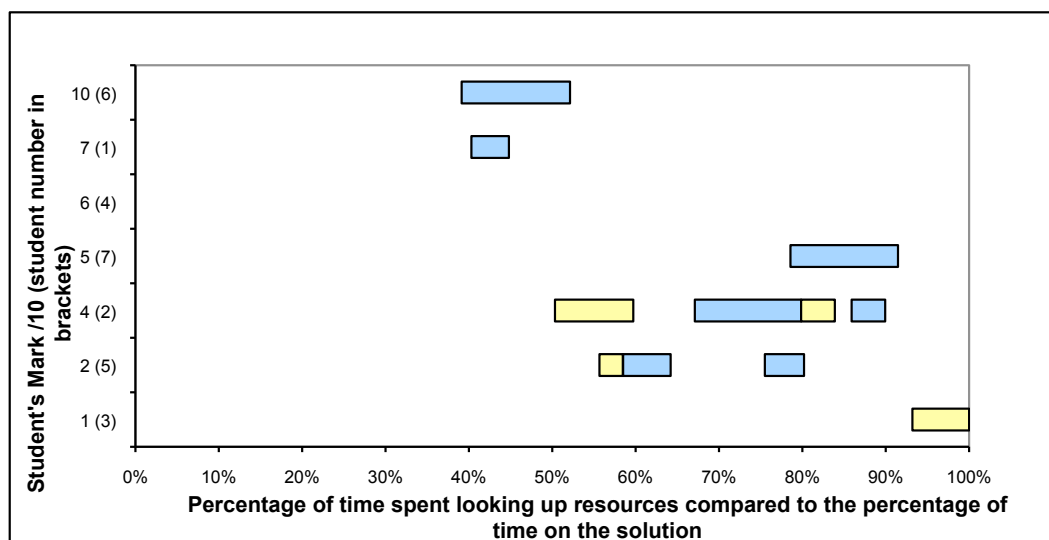


Figure 4.18: Percentage of time that students spent looking up resources (coloured bars), in terms on the percentage of time they spent solving the army cadet problem. A yellow bar indicates when a textbook was consulted and a blue bar the Internet.

### 4.3 Results From Both 1B and 2B Studies

The results from both the 1B and 2B studies were aggregated. Figure 4.19 shows the complexity score 1 against mark and Figure 4.20 gives complexity score 2 normalised by the total number of possible transitions against mark. Figure 4.21 shows the number of transitions students made in total, also normalised by the total number of possible transitions (equivalent to time spent on the problem).

Although one should be careful in comparing two data sets from different year groups, comparisons show general impressions. Furthermore, the results cannot be generalisable to all students due to the nature of the study, such as the small sample number. Instead, they are useful for providing further understanding on these problem solving models.

	Complexity Score 1	Complexity Score 2
Score on question	$r=-0.609$ , $p(\text{two-tailed})=0.002$	$r=-0.082$ , $p(\text{two-tailed})=0.703$

Table 4.5: Partial Spearman correlation coefficient values and p values for complexity versus score controlling for question type and year group.

The normalised number of transitions was multiplied by 100 to express it as

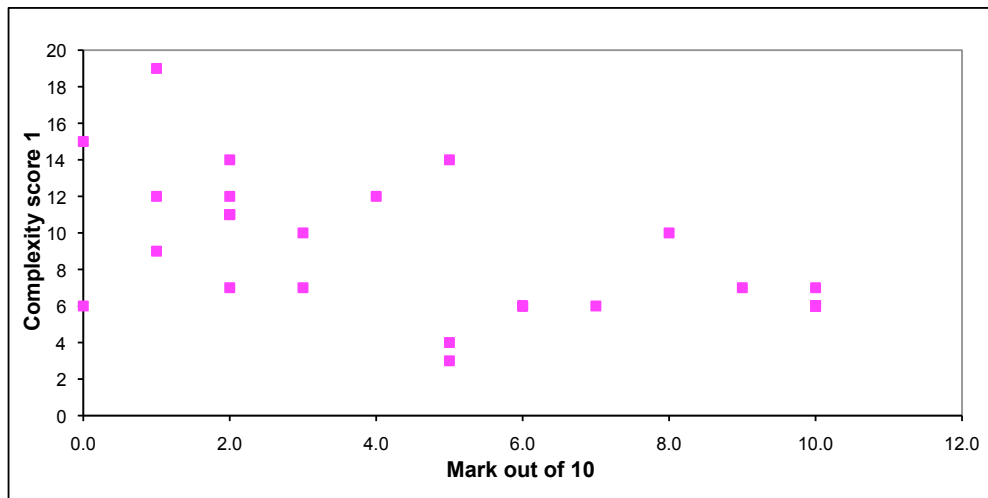


Figure 4.19: Complexity Score 1 and exam score /10.

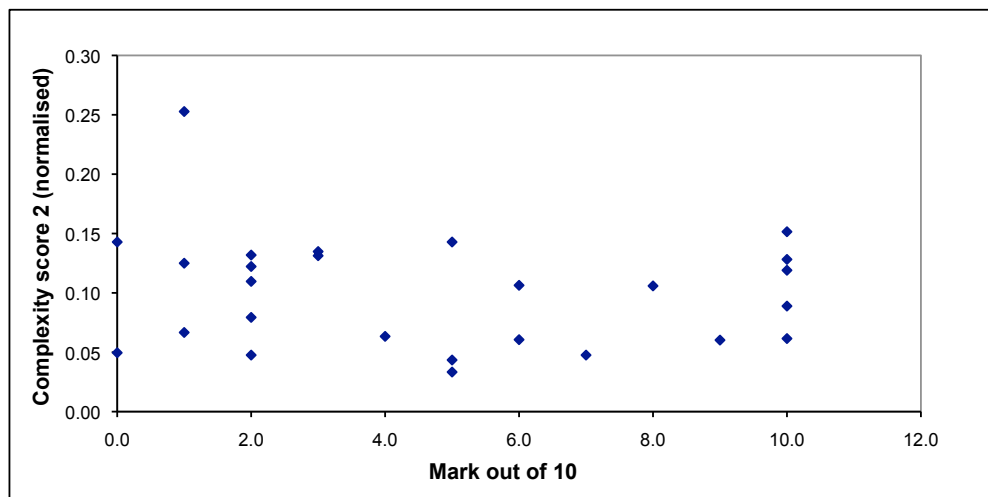


Figure 4.20: Complexity Score 2, normalised by number of possible transitions and exam score /10.

a percentage. The correlation coefficient values and their p values are shown in Table 4.5. Partial Spearman correlations were calculated in SPSS controlling for question and year group. These controls did not change the significance of the correlations. As shown in Table 4.5, there is a significant correlation between complexity score 1 and exam mark, but no significant correlation with complexity score 2 and exam mark. The partial correlation of the normalised number of transitions made, with score, controlling for question type and year group was  $r=0.077$ ,  $p(\text{two-tailed})=0.721$ , showing no significant correlation between these

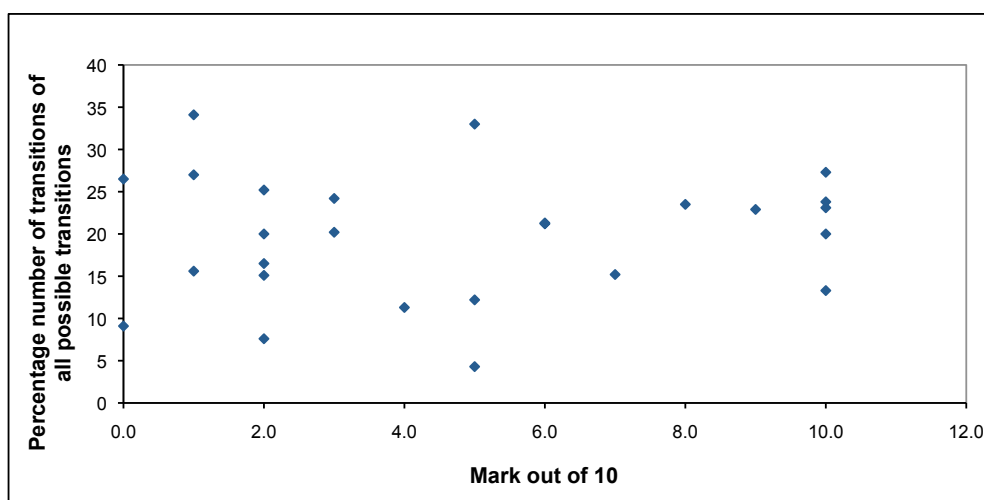


Figure 4.21: Number of transitions made (not including silences) against exam score, normalised by the number of possible transitions.

factors. Therefore there was no significant correlation between the normalised number of transitions made and exam mark.

These results indicate that successful students made fewer different types of transitions than unsuccessful students (complexity score 1), but did not make significantly fewer transitions in total (number of transitions made). There were, what seems to be an optimum number of transitions in and out of different behaviours (neither too many or too few). With about 6 or 7 types of transition successful students tended to progress steadily to a correct final answer. These results are compared to the qualitative results before being discussed further in the next Chapter (Section 5.5).

## 4.4 Summary

This chapter shows results which give an insight into problem solving behaviour according to the problem solving models and use of resources. The main findings are that successful students tend to have fewer different types of transitions between problem solving behaviours, though not necessarily fewer transitions in total. Successful students also appear to use resources less and for shorter periods of time. These findings are compared to the qualitative results in the following Chapter 5. See Section 5.5 in Chapter 5 for a full discussion of these results.

# Chapter 5

## Individual Problem Solving: Qualitative Analysis

The previous chapter studied students' solution processes using behavioural models. This chapter details the results and discussion from analysing the data qualitatively. This chapter also brings together the results from this and the previous chapter to discuss the whole picture of problem solving behaviours of students in this study.

### 5.1 1B Qualitative Study Results

The results from analysing the 1B data are presented below split by question and high and low scorers. A student was a high scorer if they achieved over 50% of the total mark on the question.

#### 5.1.1 The Drawbridge Question

##### The Low Scoring Students

Four out of the six students had low scores with 30% or below on the drawbridge question. Three of these students mentioned the word 'pivot' at some point in the solution, which is close to the correct approach of balancing torques. Their approaches are considered below.

##### Student 1

Student 1 mentioned pivots in the following context:

*the drawbridge is on a pivot against the wall....So the rope isn't supporting..the whole 200 grams, the 200 kilograms of the drawbridge.*

*There is an equation you could use. I can't remember it.*

She tried to remember an equation, but it is unclear which equation she was trying to recall. Providing this student with a textbook may have clarified this. Although she mentioned distances, weights and a pivot, she did not engage too deeply with the physics concepts and these variables were not connected with torque. There was a recognition that the distance from the end of the drawbridge was important, but the final answer was calculated using proportions. The final answer was incorrect and Student 1 blamed her memory of equations for her lack of progress. When asked how she did:

*I might of managed it, but I can't remember all the equations I needed*

### **Student 3**

Student 3 also mentioned a pivot point:

*so the pivot point is when  $L$  equals zero, so as  $L$  increases these will increase*

This student also mentioned distance from the wall, weight and force, but connected them all to the centre of mass (CoM) without considering torque. When thinking about the CoM he said twice that he could not remember the equation:

*the force, the block. I just can't remember this eh, the centre of mass equation*

Providing this student with a textbook may have shown him that CoM was not relevant here and a new approach was needed, or allowed him to pursue the CoM equation to apply it incorrectly to get an incorrect numerical answer. These are speculations, but show the potential use for a textbook in this experimental setting.

### **Student 4**

Student 4 also realised that distance was important, mentioning pivots after thinking of the moment of inertia. Although she did not know how to relate distance to the tension or weights. Later in her working she said:



*let's look at the calculation and see if I can factor in anywhere..eh..distance*

She then associated moment of inertia with pivot points and said:

*you know it was like the problems where you had to, it was like pivoted,  
there was a pivot and yeah*

*I didn't understand those questions*

She immediately rejected the idea of a pivot due to her lack of understanding of those type of questions. Perhaps she was inhibited by this belief. Beliefs, expectations and motivations are important skills in problem solving and considered part of the set of skills that students require to solve problems successfully [5]. In the end she applied a similar process to Student 1, using proportions, but realised that this approach may be flawed. She reduced the vertical component of tension (which was calculated incorrectly) to half, if applied half way along the drawbridge. She worked out the distance it would be for 5000 Newtons and got a final answer of 3.13 meters. She acknowledged potential flaws in this argument:

*but I don't know if I can do that, because tension may not be  
proportional to the distance kinda thing*

Associating torque with quantities mentioned by Students 1, 3 and 4, would have helped them solve the problem, but as discussed later, this information did not seem to be mentally cued even though they mentioned all the components which make up torque [27].

### **Student 2**

The final low scorer was Student 2. This student did not mention pivots or torque and initially considered the wrong part of the problem. Placing importance on the wrong part of the problem is something he also did in the army cadet question. He started by applying Newton's 2nd law and incorrectly assumed the block was moving and so there was acceleration in the x direction. He then realised that he needed to consider the problem at one point, when the block is stationary, so he correctly calculated the vertical component of the tension, but around 20 minutes he said:

*yes for this part I guess I'm complete lost*

At the end he wondered if he could use linear momentum, but was interrupted when the interviewer re-entered the room. At this point the student still seemed quite far from applying torque.

### **The High Scoring Students**

#### **Student 5**

Student 5 achieved a mark of 9/10, but given an extra minute or two would have solved the problem correctly, as he proved by quickly working through the rest of the solution at the end of the interview. Just before 4 minutes he mentioned torque in the middle of a sentence, as something he seemed to have previously decided to do without saying so explicitly. You can see this in his first mention of torque in the following context:

*so as we are considering the component of the force acting perpendicular to the bridge the torque due to this force due to the tension is simply this component of the tension multiplied by the length of the bridge...*

Therefore his planning seemed to be executed in periods of silence. As with other high achieving students, he did not question his approach (i.e. using torque), but did question his implementation of it. He found an answer and realised correctly that it was likely to be wrong, so repeated his working symbolically instead of numerically and spotted his mistake. Aside from this, he appeared very confident and did not express any major uncertainties over his working.

#### **Student 6**

Student 6 was awarded full marks on this question. He started with a free body diagram, perhaps because use of diagrams was encouraged in the Physics 1B course, but after a period of no writing or speaking he changed tack. As with Student 5, he decided on his approach in a period of silent thought, without planning it explicitly. Around 4 minutes, after being silent, he said:

*oh right okay take the moments around the origin*

Although he did not identify the approach of using moments straight away, he did not question this approach once he had taken it. When implementing this approach he considered the concepts he used and did not just plug numbers into

a formula. For example, for each contribution to the torque he drew a circular arrow in the direction that it was acting (clock-wise or anti clock-wise). He also made small checks throughout his solution which allowed him to spot a small mistake in calculating the angle.

### 5.1.2 The Army Cadet Question

As in the drawbridge question the same four students (Students 1-4) were in the low scoring category (with a score of 50% or below) and Students 5 and 6 were high scorers (above 50%).

#### The Low Scoring Students

##### Student 1

Student 1 recognised the key physics in the two important parts of the question (projectile motion and circular motion), but this did not help her in the solution. She was thinking of the problem in an incorrect way, as she thought that an acceleration acted in the horizontal direction when the cadet was at the bottom of the rope, so tried to apply  $F=ma$ . Around 3 minutes 30 seconds she said:

*you could work out the acceleration he needs to have at the bottom using  $F$  equals  $ma$*

This incorrect way of thinking continued at around 9 minutes to the cadet leaving the rope:

*the cadet's acceleration will be perpen. No will be parallel to the ground when he leaves the rope...because... the acceleration's tangential to the... rope*

She realised at around 10 minutes that the cadet was performing circular motion, but then considered an inappropriate area of physics to apply to the problem when she said:

*so you could use...angular momentum or something like that*

At the end of the problem she confirmed that the solution should be related to angular momentum due to the circular motion of the rope. Although both projectile and circular motion were mentioned, given more time, or resources, it

is unlikely that she would have been able to apply projectile motion correctly, as she thought there was a horizontal acceleration. She would not have been able to apply circular motion, because she thought it was related to angular momentum.

### **Student 2**

What is interesting about Student 2 is that he correctly recognised the pendulum motion of the cadet very early on, but did not connect this to the correct physics concept. At the beginning of his problem attempt he said:

*so I'm going to make a diagram that looks like a pendulum for simplicity and also make a free body diagram of all the forces acting on the body*

Projectile motion was then mentioned and he correctly stated that there was no acceleration in the x direction at the moment the cadet released the rope (unlike Student 1). However, he was unable to use projectile motion in order to find the cadet's velocity. After he was unable to do this there was a gap in writing and speaking and he said:

*so at this point I'm not exactly... I'm.. I don't exactly know what to do, so I'm just going to.. write down the Newton's second law and express tension and see what I know and what I don't know.. and see if it helps me*

As with the previous problem, Newton's 2nd law seems to be the thing he uses when he does not know what to use. He applied Newton's 2nd law to the wrong part of the problem, by taking the vertical and horizontal acceleration of the cadet from when he catches the rope to when he releases it. Although he correctly calculated the cadet's velocity from energy conservation, he applied Newton's 2nd law to find an acceleration. Student 2 recognised the correct pendulum motion, but used an approach which was not suitable for the situation. Perhaps the wrong area of knowledge was activated for this problem.

### **Student 3**

Student 3 was convinced that the final tension would only be due to the weight of the cadet. He even mentioned the correct area of physics, but did not pursue it:

*don't know which bit it's into. The one where it says centripetal and stuff, but it wouldn't be that..would it?*

*it wouldn't matter, because when he's, when he's vertical the only force acting on the, the rope...would be the weight of the man*

In the end he settled on  $T=mg$ , missing the centripetal component of motion. He said:

*I don't think there is going to be anything else*

It may have been difficult for this student to pursue centripetal motion when it seemed so unlikely to him.

#### **Student 4**

Student 4 expressed uncertainty right at the beginning of the problem saying “*oh no*” when first reading the problem statement, followed shortly by “*completely forgotten about this, okay*”. She wondered if projectile motion was a good approach around 2 minutes in:

*I don't know should I approach this like a projectile question*

She did not pursue this until later. At around 5 minutes she said:

*if it would be maybe... central force*

But she did not do anything with this information at this point. After this, she thinks of equations of motion, but could not figure out which equation to use. At about 12 minutes 30 seconds she said:

*it's got forces and stuff I'm think central force, but it's not going round in a circle so that can't be right*

Then at 12 minutes 50 seconds:

*Unless it is a circular question because your kinda going like that*

Having reached the correct motion she said that  $F = \frac{mv^2}{r}$  and tension  $T = \frac{mv^2}{r}$ . She was then still unable to find the release velocity, because she could not work through the equations of motion. As she said later when the researcher was talking her through the solution, she forgot about splitting the motion up into x and y components. She finished at 15 minutes by saying:

*I think though if you could get v squared here then you could find the tension here*

She seemed quite sure of the circular motion by the end of the problem even though she could not find the velocity. However, she missed out  $mg$  in her equation for tension.

## The High Scoring Students

### Student 5

Student 5 used a number of approaches in this question (including using energy conservation, centripetal force, weight, and projectile motion) and did so without making any mistakes. Perhaps this shows that this was more an exercise than a problem for this student. As with the drawbridge question this student appeared to think strategically in his silences. After setting up the problem he was quiet and then said:

*so we need to find the tension in the rope which is centripetal force since he is going to be moving in a circular path*

He checked the velocity calculated from using energy conservation was the same as that using projectile motion. He also evaluated his answer at the end. In both questions it seemed he paused the recording himself and re-started it later, which would explain why his recording did not end at 15 minutes, even when the interviewer timed 15 minutes separately. In the drawbridge question he appeared to pause the recording after realising he had made a mistake in his working. He switched the pen back on knowing what approach he would use to try to rectify the situation by saying:

*I'm going to repeat the calculations doing everything symbolically...*

In the army cadet question it appeared that he paused the smartpen after finishing the problem. He then turned the pen back on to follow up with an evaluation of his answer. Although there did not seem to be a great amount of information lost, it showed that this student was unable to, or choosing not to, vocalise some of his thoughts.

### Student 6

Student 6 spent around 4 minutes constructing the problem space. This question seemed to demand a large set-up time by most students compared to the drawbridge question. This maybe because it was a more complicated problem, but also there was no diagram provided. This student started with a free-body diagram, a common theme for these first year students. He also worked out the velocity using both energy conservation and projectile motion and found the velocity to be the same in both cases. He recognised that it was circular motion at around 10 minutes 40 seconds when he said:

*it's going round in a circle motion so force inwards is going to be given by.. the mass and acceleration with the eh.. centripetal force, eh centripetal acceleration being  $v$  squared over  $r$*

However he set up the equation in the wrong order  $T = mg - \frac{mv^2}{r}$ . This could have been a mistake, or an underlying misconception about how to apply centripetal force to problems i.e. that centripetal force:  $F_c = \text{net forces}$ .

## 5.2 1B Discussion

The higher scorers chose the correct physics concept usually quite early on in the problem. In the army question, both the high scoring students mentioned the correct motion followed immediately by the correct physics. There appeared to be strong associations between the problem type and the applicable physics for these students [39]. Both Students 5 and 6 made a mistake, but both students also corrected them. Although many students had a link between Analysis and Implementation these connections were particularly clear in the successful students, again perhaps reflecting that they knew which approach to apply and then worked to implement it. Unsuccessful students, on the other hand, could not decide which approach to use, or applied an incorrect approach. They did not seem confident about which general approach to take or were over-confident about an incorrect approach.

Lack of planning and verification was observed in many of the students, which are both classified by Meijer et al. [108] as metacognitive activities and by Adams as beliefs [5]. These are arguably control type activities. Perhaps they do not occur explicitly or planning may occur in a fleeting moment which was hard to vocalise. These are activities frequently recommended to improve problem solving [24, 109], so it is interesting to speculate why students in this study do not exhibit these behaviours. There are maybe two main reasons. Students were novices and so have not developed metacognitive thinking in their solutions, or these behaviours were hard to observe using the think aloud protocol. Both these factors perhaps contributed to the lack of metacognitive behaviours observed. Successful students in this study seemed to be planning in periods of silence. Perhaps it was difficult for them to vocalise this thinking, or they were not doing it consciously.

A common question arose; what would these students have done if they had access to resources, such as a textbook or the Internet? Would they have been able to solve the problems? On the drawbridge question Students 1 and 3 blamed their memory for their lack of progress and it would have been interesting to see what the 'lost' Student 2 would have done with resources. Student 4 on the army question may or may not have benefited from the availability of resources. Further study with resources would be required to investigate this further.

## 5.3 2B Qualitative Study Results

The results from analysing the 2B data are presented below, split by question and high and low scorers (above and below 50% score on the question respectively).

### 5.3.1 The Drawbridge Question

#### The Low Scoring Students

All except one student mentioned the correct approach in the drawbridge question (three out of the four low scorers and all three high scorers mentioned or actually used torque). Each of the three low scoring students who mentioned torque or moments also had the correct reasoning for why this would be the right approach. However, the low scorers did not pursue this correct approach. Their quotes are shown below for when they first mentioned the correct approach of torque. 'E' represents the voice of the interviewer.

Student 3 around 14 minutes:

*Do I want to be doing something to do with moments rather than what I'm doing?*

*E: What makes you say that?*

*because obviously the further out the bigger the moment around here is.*

Student 5 around 4 minutes:

*I'm looking up, I think it's called like moment or s..oh no, oh yeah I'm going to try work out that*



*E: Why do you choose that?*

*just because I recall that like it's to do with the mass being in a different position..*

Student 7 around 6 minutes:

*I'm thinking I should maybe be using um torque instead*

*E: What makes you say that?*

*because there's like..this is almost like the radius, the distance that he has got it pushed out at*

### **Student 3**

The next key question was why students did not pursue this approach. The quotes below support a dynamic interaction between choosing an approach and being able to see how to apply it. Student 3 was not confident that torque was the correct approach and saw no easy way to implement it. Immediately after saying why moments would be a good approach this student said:

*I don't really know how to do that*

After this she said:

*Okay I've been having..I definitely did want to resolve tension, cause that's the only way*

And in doing so reverts back to looking at components and working out what she can without a general approach. Student 3, finally used an invalid approach where she tried to correspond the forces on each side of the triangle to the length of that side in meters.

*so if 3000 newtons corresponds to 4 meters. Then 4000 newtons corresponds to*

She stopped soon afterwards saying she was unable to finish the problem in the last few minutes. She did not use any resources for the drawbridge problem, even though she mentioned the appropriate approach of using moments, it did not occur to her to look it up.

### **Student 5**

Student 5 thought of using moments, but was not confident that this was the right

approach, and evaluated too early in the procedure. After mentioning moments she said:

*Is the only thing I can remember from like physics off the top of my head that looks similar to this.. problem*

*But I don't know whether I'm thinking in the right way*

*em I can't remember what it's called..*

*okay I'm just going to look at forces*

She started fairly confidently referring to a similar past problem, but subsequently questioned her thinking and changed approach to use proportions, such as setting components of force to zero at a certain distance. Perhaps because forces were a more familiar area of physics from previous study.

She looked up the textbook twice in the drawbridge question. The first time to look up moments, which she did not find and the second time immediately after looking up moments, she searches for forces. She was:

*just looking up anything that looked like it would help me, to do with balancing forces*

This indirected search was not successful. Being unable to look up the required information, such as using the correct search term in the textbook index, is likely to have contributed to Student 5's failure on this question.

### **Student 7**

The third low scoring student who mentioned torque was slightly different, she pursued it further than the other low scorers mentioned above, but with an incorrect implementation of the approach she did not get to the correct final answer. She said:

*I want to be balance...the torque and the tension*

However to obtain a final answer, she incorrectly balanced forces. She used the textbook once to look up the definition of torque, but did not apply the equation correctly.

### **Student 2**

Student 2 did not mention torque at any point, trying instead to use centre of mass and forces. Student 2, from quartile two, was marked with 2/10 on the

drawbridge question. She had a low score, because she used an invalid approach, could not explain her reasoning and was stopped at around 21 minutes without a final answer. This student tried unsuccessfully to use centre of mass arguments for part of the question.

She used the textbook and consulted it on four separate occasions. She tried to find a similar problem, evidenced by her saying:

*I'll see if I can find a similar problem*

However, the way she tried to find a similar problem was by matching surface features, a typical novice approach [20]. For example she said:

*I'm reading a problem with a mass on a slope with a wire*

which she said might be similar. This is problem 5-7 from page 103 of HRW [103] and is not the same kind of problem in terms of the underlying physics concepts, as it is about tension in a wire, attached to a block, on an inclined plane, and is best solved using free-body diagrams and Newton's second law.

The relationship between selecting the correct approach and implementation was demonstrated in reverse. In other words a few students implemented what they could first, in order to then see if that helped them progress in the problem. Their lack of overall approach, or control [24], is apparent in what they say. For example, Student 2 read about a similar problem and said there were going to be vertical and horizontal components of the tension, but soon after said:

*I don't really know what I'm doing*

Furthermore, this student worked out a value and then wondered what she did to get it:

*I'm trying to work out what I've done*

In another example, Student 5 implemented a process before deciding on the next step. She worked out the component of tension in the vertical direction and expressed uncertainty as to whether that had helped her solution. Working from implementation is a valid problem solving approach [16], as students tried to close the gap between their current state and the goal state. However in these examples, this style of approach did not help, as they did not have a clear idea of why they did it, or what would follow.

### The High Scoring Students

All three high scoring students recognised the correct approach very early on and did not question its validity.

#### Student 1

Student 1, a quartile one student, performed similarly on both questions with 6/10 on the drawbridge question and 7/10 on the army question. This was a reasonable performance compared to the other students. Marks were deducted as he missed out key physics in each question. Student 1 at around 16 seconds said:

*So this question it would just be torque*

#### Student 4

Student 4, also a quartile one student, performed very well on the drawbridge question with a mark of 10/10. This student identified the correct approach of torque after about a minute into the problem. At this point she said:

*It's going to be balancing..torques*

After implementing this approach she evaluated her answer and realised it was not correct. Her initial thought was she had used the wrong angle between the radius and the force. She checked this equation and where the angle should be taken from in the textbook. In fact she had missed out the mass of the drawbridge, but recognised the error and corrected for it. The contrast is observed with Student 1, who did not account for the weight of the drawbridge.

#### Student 6

Student 6, a quartile 2 student, performed very well, scoring 10/10 on both questions. She got both questions correct and explicitly evaluated her answer on the drawbridge question. Student 6 also around 1 minute said:

*So the idea is about moments and stuff*

She did not use any resources on the drawbridge question. She found an incorrect answer initially, using the right approach (torque). She got to the correct answer after evaluating her working. Her implementation of this approach was fine or had minor errors that were corrected along the way.

This shows that these successful students had access to the appropriate knowledge for this problem and that, there is a question from the perspective of the physics approach, as to whether this was a true problem for these students.

### 5.3.2 The Army Cadet Question

The army cadet question revealed slightly different results. This question was not so insightful into the relation to the three themes that emerged from the data, as discussed in Section 3.8. Only one student actually used the correct approach, with none of the other students mentioning it at all. Student 2 nearly got to the right approach using resources as support, but took a long time over it. Three students used  $T = mg$  missing  $\frac{mv^2}{r}$ , two students set  $T = \frac{mv^2}{r}$  without accounting for weight and one student used an equation she found on the SQA website  $Ft = mv - mu$ . Finally, Student 6 correctly used  $T = mg + \frac{mv^2}{r}$ . Each of these approaches is discussed in more detail below.

#### The Low Scoring Students

##### Student 2

Initially Student 2 said that the tension was equal to the cadet's weight and expressed her uncertainty about why it would be different from this. At around 6 minutes she said:

*I thinking about em.. he's going to be traveling in like an arc, so it might have something to do with centripetal acceleration or something, not acceleration, velocity. Cause he is going round a bit em*

Student 2 then exhibited similar traits in this question to the drawbridge question. She calculated what she could, not seeming to know what else to do or an overall approach to use. The student noticed that projectile motion applied to part of the problem and tried to look up equations of motion in the textbook. For example she said:

*Although I'm not sure if this is the right way to go about it, because I am trying to work out the tension. But...I'm not sure how to do anything else.*

Around 17 minutes, Student 2 searches the Internet for “tension in a rope swing” after noting that she is trying to work out what difference the tension makes to how far the cadet travels. This did not relate to the shape of the cadet's motion, but she found a similar question online and noted that the cadet will experience centripetal motion at the lowest point. Near the end of the problem she

connected centripetal acceleration to the velocity. She realised that she needed to find the release velocity with equations of motion, but did not manage to calculate this. Then she said:

*the acceleration inwards is... I was going to find out what, how that relates to the tension, but I am sure that it does...em.. Yeah*

The information from the Internet had shown her that centripetal force was important, but not how to connect it to the tension in the rope. By not knowing how tension and centripetal force were connected she demonstrated weaker knowledge; these are connected by only one concept, as the definition of centripetal force is the sum of all forces acting on the cadet.

She consulted the textbook twice and the Internet twice, looking at resources on four separate occasions. She was not sure what to do, but found a similar problem online to point her in the direction of centripetal motion, which she then looked up in the textbook. As discussed above she was unable to apply or implement the equations of motion from either the textbook or Internet to the problem given. This meant she could not find the release velocity to calculate the centripetal acceleration and therefore could not find the tension in the rope.

### **Student 7**

Student 7 mentioned the correct physics concepts including weight, tension and centripetal force, but put them together incorrectly. In the army cadet question she connected tension to the centripetal force, but did not account for the cadet's weight, similar to the mistake made by Student 1. However from what she says, as quoted below, there seems to be some deeper misunderstanding.

*..like I mean he'll have a force in this direction, but like tension will be going up so it seems like the only force that needs to be balance it is the weight*

*um..I suppose that will be given by the centri...centripetal force, like the em... that's what's holding him in his orbit almost*

*so I'd probably want  $um..m v \text{ squared over } r$  to be equal to the tension*

Saying centripetal force is the thing 'holding' him in his orbit shows a misunderstanding of this concept. She also incorrectly calculated the cadet's velocity when letting go of the rope.

Student 7 used the Internet once after opening the textbook and did not find the information on the equations of motion that she was looking for. This was because she had forgotten what the equations of motion were called, so instead she typed in ‘displacement speed acceleration formula’ into Google which gave her what she wanted. Again she applied this information incorrectly, because she had already set the situation up incorrectly before finding the equations of motion online. For example, she set the final vertical velocity to be zero.

#### **Student 5**

In the army cadet problem Student 5 inappropriately applied incorrect equations without considering the physics involved. She tried to implement equations without thinking about the physics concepts they are based on. She did so by applying an incorrect equation, because the variables matched those in the problem:

*I’m just looking at that formula there (on SQA website) the Ft equals mv minus mu and thinking I may be able to do something with that, but I’m not too sure*

She found ‘t’ by using the horizontal distance the cadet had to travel after letting go of the rope, divided by the release velocity she had previously calculated. The release velocity was calculated using equations of motion to find the velocity of the cadet from starting to swing to letting go. All known values were put into  $Ft = mv - mu$  to find F as her final answer.

This was a form of plug and chug, she scrolled through the list of equations until there was one which looked like it had the variables from the problem and then she plugged in the numbers. She expressed uncertainty over this method, but pursued it none the less and did not comment on how it could have been improved. It was this same uncertainty that had steered her away from the correct approach in the drawbridge question (see above).

#### **Student 3**

Student 3 used  $T = mg$ . Student 3 did not know what to search for and said:

*he is letting go here and wanting to reach...here. Em. I’m trying to work out how any of this is relevant to what the tension in the rope is, which again I am sure it probably is, but I don’t know how*

$T=mg$  was the only thing she could come up with.

She could not find another value except that of the army cadet's weight to contribute to the tension. She justified her incorrect approach several times knowing perhaps that the projectile part was probably relevant, but she could not connect that to her approach.

*em, yeah as far as I can tell, the only thing I can come up with is that is it going to be 800 newtons, just because it has got to balance his weight and I don't think it could be...*

*...because if it was, the rope would physically be pulling him upwards*

When reminded that she could use the textbook or internet she said:

*well if I knew what to look for I would do, but I don't really know what is the relevant, I don't really know how to connect...the two*

On the army cadet question again she used no resources until reminded of them by the interviewer when she said:

*I don't really know what I'd look for to be honest*

She was unable to look up the relevant physics and there was very little working on the page besides multiple diagrams. In both questions, use of resources did not seem to occur to her as being helpful, as she was unable to vocalise the underlying physics which would have directed her to the correct approach.

## **The High Scoring Students**

### **Student 6**

Student 6 scored full marks on this question and did so by using the motion of the rope swing in the Internet search, by searching for "tension in a pendulum". The key motion was recognised i.e. that the cadet is swinging in a circle, arc or pendulum motion. On this particular question this seemed to be a necessary stage before recognising the physics approach. Recognising the motion of the rope could potentially activate the correct memory retrieval either from the student or from an external resource e.g. online. In these particular aspects a linear progression was necessary, as the student needed to first identify the appropriate motion in order to apply circular motion principles.



Student 6 looked up the Internet once in the army cadet problem. By recognising the correct motion of the rope swing, she googled ‘Tension in a pendulum’ which gave her a worked example which used  $T - mg = \frac{mv^2}{r}$ . She realised from this equation that she had to find the velocity to get the tension, which she did and it resulted in her solving the problem correctly. Student 6 therefore used the information found online as a problem solving tool. This student correctly used a worked example by finding it via the correct conceptual physics and not the surface features of the problem [20]. She was also able to use the information she had found by applying it appropriately to the problem given. There was a dynamic interaction between approach and looking up the resource. She did not check her final answer or sub-answers as she was happy with them, but seemed aware of when an answer was completely unrealistic.

#### **Student 1**

Student 1 scored 7/10. He missed out the weight of the cadet in the final calculation for tension. He vocalised in under 2 minutes the link between centripetal acceleration and tension, but part of the solution was missing.

For both questions, Student 1 accessed resources in a similar way. He looked up the Internet once on each question for reference to equations only. He then used this information in order to help him solve the problem. It was clear that he was only looking up equations to aid his memory. In both questions, he found a page on the Internet with the relevant equation on it, read it and then clicked back to the Google homepage. Without copying these equations down while they were on the screen he proceeded in solving the problem.

Student 1 had a good performance where resources provided aided his memory, but parts of the solution were missing, with similar types of mistakes on both questions. This perhaps occurred due to his lack of evaluation of either question.

#### **Student 4**

Student 4 said  $T = mg$  and wondered if it was a “trick question” perhaps because much of the information from the problem statement had not been used. Even with this feeling of uncertainty Student 4 maintained that  $T = mg$ . She did not perform as highly on the army cadet question as the drawbridge question, with a mark of 6/10. Student 4 used energy conservation to find the release velocity. She did not look up any resources to help with her solution and she could not see any forces other than the weight to contribute to tension, even though she

worked out the release velocity (which is part of the way there). She realised that there was something more to the question, but stuck with her answer that it was just the cadet's weight that contributed to the tension. As with the drawbridge question, there were some evaluation skills demonstrated, as shown in the quote below, but the solution was not corrected.

*I think that if someone is hanging from a vertical rope the only force is going to be their weight. But.. it doesn't make sense.*

## 5.4 2B Discussion

The main points from the 2B qualitative results are highlighted here. High scoring students on the drawbridge question used only one approach. They saw the correct approach of applying torque almost immediately and did not deviate from it. They were also able to implement this approach to reach a final answer. In the army cadet question these insights were not so striking. Although it was clear that students progressed when they recognised the motion of the army cadet (e.g. circular, pendulum or arc) many struggled connecting the two parts of the question or left out terms when finding the tension.

Although finding the necessary physics principles is a key factor, the process of selecting which approach to pursue appears from this study to be more interactive and dynamic than one might expect. This includes interactions as to the approach at a tactical and strategic level. For example, even if the student mentions the correct approach (as happened for 3 of the 4 incorrect students), if they could not find it in the textbook (Student 5, drawbridge) or could not see how to implement it (Student 3 and 7, drawbridge) then they would move to another approach. Students do not know which is the best approach to use and perhaps have not yet developed a 'feel' for what the right approach is, as experts have developed from doing many previous examples [29]. The low scoring students did not recognise the correct approach when they came across it and difficulties with the implementation of this approach often steered them away from it. Students also tried to implement what they could to progress the solution without seeming to have a clear idea of where it was heading (Student 5, drawbridge). In this way, analysis and implementation skills seem inextricably linked. Alternatively, this suggests that control on a tactical level interacts with that at a strategic level. If

students could not see how to make the next step (tactical), the overall approach was discarded and another one applied (strategic).

### 5.4.1 Resource Use

In terms of how high and low scoring students accessed resources, two of the highest scorers, Student 1 (on both questions) and Student 4 (drawbridge question), used resources for reference to check an equation. Resources assisted by providing knowledge and these students had no need to re-understand these equations, they simply needed reminding of them. The other high scorer, Student 6, did not use any resources on the drawbridge question. On the army cadet question, she figured out the right motion (pendulum) to look up the correct area of physics (an example problem), which she could then appropriately apply to the problem.

Lower scoring students had certain characteristics. In terms of accessing resources they either did not look up a resource to help (Student 4, army; Student 3, drawbridge), or did not know what to look up (Student 3, army). Difficulties with what to look up included unsuccessfully finding similar problems by matching the surface features of the problem set to those in the textbook [20] (Student 2, drawbridge), or just looking up anything they thought might help (Student 5, drawbridge). Student 5 on the army question tried to inappropriately match unknowns to a list of equations. Even if they did find some information online or in the textbook, unsuccessful students could not apply it to the problem given (Student 2, army), or could not apply the equation/concept they found correctly (Student 7, both). It appears that for these students, providing access to resources was not enough [?]; they needed to understand the underlying physics in order to search for the correct concept.

It can be postulated that giving students access to the Internet would mean they would not only be able to look up facts and equations, but also how to implement these equations. However, there was no evidence of students, who did not know the area of physics, managing to solve the problems correctly. Even when the correct deeper structure of the problem was mentioned, (such as torques by Student 7 and moments by Student 5 on the drawbridge question) students did not use this approach or did not use this approach correctly in order to solve the problem.

This study had potential for students to use resources in a true problem solving fashion to achieve the correct approach, as they may not know initially how to solve the problem. For example, worked examples could have been looked up. This is potentially a very useful technique and one that contributes to the fear that open-book exams will not test students' knowledge. Student 6 successfully found a worked example in the army cadet question. She used the underlying motion, as opposed to surface features to work out what to look up and the resource online showed her partly how to solve it. Student 2 was unsuccessful because she compared worked examples by their surface features and not the central physics involved (in the drawbridge question). This is in agreement with research from Chi et al.'s study [20], whereby experts categorise problems by their principles and novices categorise by surface features. In this study there seem to be expert-like students, as student 6, for example, recognised the underlying physics concepts. This perhaps is expected after almost two years of University tuition, however this study shows weaknesses in classifying whole cohorts as 'novice' when there is a spread of abilities. This also shows the strengths of a small scale study, such as this one, where students are considered individually.

Time was a critical factor in the army cadet question. For open book exams this is worth noting. The results from this study indicate that for students who know the material well, a time limit will not be an issue. For students who have strong problem solving skills, but cannot remember a particular area of physics, it will take them longer to find an answer, if at all. Finally, students who are not well prepared or have forgotten the material will not know what to look up or how to implement it, so are likely to be unsuccessful. It is unclear whether exploring with resources will ever lead the student to the correct approach. The student would have to recognise the correct approach, even if they chanced upon it and then also be able to apply it, something that students in this study struggled with (for example, Student 2, army).

The textbook is more restrictive than the Internet and harder to access if the terminology cannot be remembered. This happened to Student 7 who in the army question had forgotten the name of the constant acceleration equations, so could not find them in the textbook. Instead, to obtain the necessary information, she looked up 'displacement speed acceleration formula' online. The textbook is more reliable in some ways, although may be less up-to-date. The quality of

information consulted online is not guaranteed and evaluation of the quality of online material was not observed here, but would be a necessary instructional point if students were using the Internet regularly.

Students in the generation under consideration can be referred to as digital natives, as they grew up with technology and are familiar with computers in everyday life including the use of search engines [110, 111, 112]. However, this does not mean that they know how to use the technology appropriately to solve physics problems. This has interesting implications for the use of open-book exams. With the appropriate questions [85], some students were still unable to solve problems, even with the potential to ‘look up the answers’. In fact in general more time on resources in a less directed fashion meant less success. The study findings support the use of open-book exams as a valid way of testing students’ physics knowledge and skills.

However, this study may not be relevant in the future, when it may be easier to find appropriate worked examples just by describing the physical problem by its surface features as the data base of physics questions grows and technology develops. This would suggest the need to stick to paper resources, such as textbooks or the students’ notes for open-book exams, so as to minimise the likelihood of students cheating by looking up an entire worked example online.

## 5.5 Discussion

This section compares and discusses the qualitative and quantitative analysis of the 1B and 2B data, covered in this chapter and chapter 4. The quantitative and qualitative results tell the same story in different ways. Successful students made fewer different types of transitions than unsuccessful students, but there was no difference in the number of transitions, or the number of transitions to different behaviours (complexity score 2). In other words successful students with complex transition maps were not observed. This indicates that a directed approach correlated with success, while task switching does not. There is an optimum level of complexity of behaviours (around 6 different transition types), and successful students do not seem to have to make further major strategic control decisions. From the qualitative analysis, it appears that successful students chose the correct approach first time, and then only had problems at a tactical level. Perhaps a

more complex problem would be required to test strategic decision making for these successful students. For example, Student 5 showed the best strategic control. He verified all his answers and even repeated a section of the problem without putting in numbers to successfully isolate where he went wrong. Complex transition maps are perhaps explained by the lack of strategic control. It is not guaranteed that a student will find the correct approach to the problem first time. However, students with more complex maps did not recognise or choose the correct approach, for some students this was even after mentioning it and they often turned to an incorrect approach. There may have been other factors affecting their progress, such as p-prims as discussed below.

It is not possible to determine from the quantitative data the causality of the significant relationship between complexity score 1 and mark on the question (Section 4.3), but it seems clear from the qualitative data that being a successful student is associated with a simpler transition pattern. This can be seen by the students spotting the correct approach to use very early on and having sufficient physics knowledge to have confidence in their approach.

Considering the unsuccessful students, lack of an overall approach and lack of confidence are potential explanations for their higher number of different types of transitions when compared to successful students. Their lack of confidence in choosing the correct approach was evident and it contributed to students lack of strategic control, as discussed in the qualitative analysis sections. For example, many unsuccessful students mentioned the correct approach and did not pursue it, especially in the drawbridge question. This raises questions related to student success. Why do the successful students choose the correct approach? Why did successful students have both strategic and tactical control that seemed to be missing for the unsuccessful students?

Redish [39] states that it is important to relate results to theory as this can develop hypotheses that otherwise may have been missed. He was referring to his own theoretical framework in physics education, but this point is applied here and different (but not necessarily conflicting) theories of knowledge structure are discussed in relation to this data. Firstly, some of the results are interpreted in terms of Redish's theory [39]. Redish [39] chose p-prims, discussed in the literature review Section 1.5, as an example of internal resources at the students' disposal. Many students mentioned that distance was important in

the drawbridge question and activated the p-prim that *more is more*, in that moving the block a greater distance from the supporting wall means a greater pull on the supporting wire. However, these students were unable to connect this information to a useful physical model to then pursue i.e. the principle of torque. What is of more interest is why students did not pursue the correct physics model even when the words torque or moment were mentioned (some students pursued incorrect models based on proportions instead).

Redish discusses strength of connections between internal resources, which is in agreement with literature from neuroscience where if connections are reiterated they are stronger. Once one mental resource is activated, students need to be able to activate other relevant mental resources [27, 39] without having to go through many behaviours in different orders (complexity score 1). For example, if they mention distance and force around a pivot this should activate torque. Similarly if circular motion is described this should activate centripetal force equals the sum of the forces on the body. These are the activations needed for these particular problems and different activations are needed for the variety of problems in the same general areas of physics.

We are reminded by Redish and Sabella that knowledge depends on context:

*..if a student does not use a particular fact or process in a given situation, that does not mean the student doesn't possess that knowledge. It may mean that the student has not correctly associated the knowledge with the conditions and circumstances relevant to its use. All knowledge is context dependent; the critical factor is whether knowledge is activated in appropriate contexts. p.1027-1028 [27]*

Redish and Sabella suggest links need to be built to cues, so students can learn when to activate the appropriate knowledge. It may be in the army cadet question that students were not cued to activate the appropriate knowledge, especially if they did not see that the motion of the army cadet was circular before the rope was released. However in the drawbridge question, it could be that students had weak connections between the approach, torque, and the procedural techniques for using torque (or perhaps they could not remember the procedure itself). If students do not have the connections between the bits of information they cannot access the necessary knowledge when required. For example, if they cannot see how to implement an approach (even if it is the correct one) then they will not

use it, the connections to the appropriate knowledge are needed.

Although Redish discusses connections between internal resources, the model by Gerace [37] is also considered here, because of its clear diagrammatic representation (shown in Figure 5.1) and distinction between problem state, conceptual and procedural knowledge. Although p-prims [39, 113] assist in the understanding of how the student could be thinking about the problem, it was the links between different aspects of knowledge that students in this study appear to be missing.

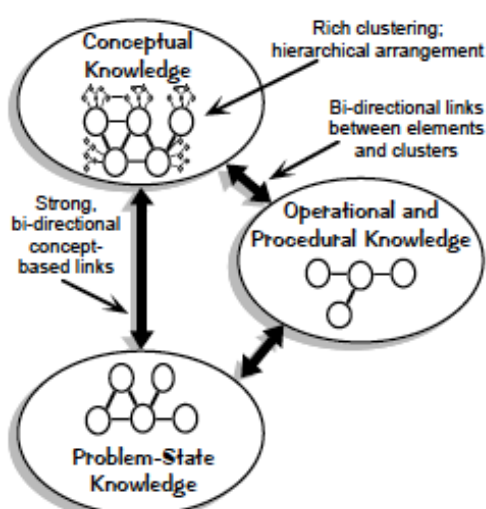


Fig. 1: Expert's knowledge store.

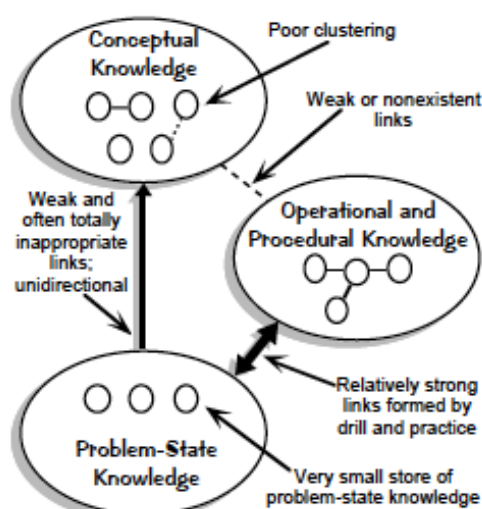


Fig. 2: Novice's knowledge store.

Figure 5.1: Gerace's knowledge structure model [37]. Reproduced with permission.

Gerace's model is one of knowledge structures [37]. Experts have conceptual knowledge in an hierarchical arrangement, and strong operational, procedural and problem state knowledge. All three are strongly connected both bi-directionally. Novices are weaker in terms of conceptual, procedural and problem state knowledge, but also have weaker links between them. Problem state knowledge includes recognising what the problem is looking for, beyond the context, and understanding which conceptual knowledge can be applied to the problem. Conceptual knowledge covers the relevant concepts and principles used to solve the problem, for example conservation of angular momentum. The authors describe a hierarchical arrangement of conceptual knowledge, meaning that the person has a well organised and coherent conceptual knowledge structure.



This is explicit in tools such as concepts maps where students have to relate topics they have learnt to current material to build a picture of how the topic fits together. For example, on the topic of kinematics, force and mechanics conservation laws can be related to other areas of mechanics including torque and Newton's laws. Finally, procedural knowledge is the ability to work through the equations and apply mathematical relations in order to get to a final answer. For example, if the concept was conservation of momentum, procedural knowledge would be used to arrange and rearrange symbolic relations in order to find the appropriate value.

One can speculate as to the condition of the knowledge structures of participants in this study. It seems Gerace's model is relevant as unsuccessful students seem to exhibit partial or weak knowledge structures, mentioning an area of physics and then struggling to connect anything else to it. For example, mentioning the correct approach in the drawbridge question, but not pursuing it, or not being able to see what approach applied to an object moving with a circular motion (1B study).

Not cueing appropriate knowledge [27, 37] (army cadet question) and weak links between conceptual and procedural knowledge [37] (drawbridge question) have emerged as being important factors in unsuccessful students solutions. Cueing of knowledge is covered in Gerace's 'problem state knowledge' to 'conceptual knowledge' link and was the focus of Redish and Sabella's study [27]. This may explain why students were unable to progress in a simple fashion through the solution and jumped to and from different types of problem solving behaviour.

The reverse is also consistent with the knowledge structures model. Successful students were easily cued by the problem as to what approach to use and had strong links between problem state, conceptual and procedural knowledge in order to carry out the solution successfully. Student 6 in the 1B study was not cued as quickly to recognise the forces on the cadet, but managed it after drawing a free body diagram. Students could use resources, such as the Internet, to cue knowledge that they had not previously considered. For example, Student 2 on the army cadet question, found that centripetal force was relevant from searching the Internet, however she did not have strong links of centripetal force to the tension, so was unable to progress further.

Thus both cueing appropriate mental resources and having strong links between these resources are important in solving these problems successfully. External resources, such as the internet have the potential to cue knowledge that might not otherwise have been used, but students need to be able to find this knowledge and then know how to apply it, all within the time allowed.

## 5.6 Further Investigation: 4th and 5th Year Students

In order to further consider the arguments of knowledge structure and their applicability to the data, they were applied to data from senior students solving more difficult problems and comparisons made to the 1B, 2B data. Two students tried two problems each. Student A did two questions A6 and A7 from the 2010 past paper and Student B completed two questions B4 and B5 from the Astrophysics section of the 2008 paper.

### 5.6.1 Student A

6. The interaction energy of two Kr atoms has the form

$$V = 4V_0 \left[ \left( \frac{\ell}{r} \right)^{12} - \left( \frac{\ell}{r} \right)^6 \right]$$

where  $r$  is their separation and  $V_0$  and  $\ell$  are constants. Sketch the variation of interatomic *force* with  $r$  and comment briefly on the physical significance of the main features of your sketch. [5]

7. To summon my dog, I blow a whistle that has a frequency of 500 Hz. My dog runs towards me with a speed of  $10 \text{ m s}^{-1}$ . Unfortunately he sees another dog in the distance and runs past me. What is the change in frequency that the dog perceives upon passing me?

[The speed of sound in air is  $330 \text{ m s}^{-1}$ .] [5]

Figure 5.2: Questions A6 and A7, 2010 paper

The two problems Student A submits are shown in Figure 5.2. She spotted

the correct physics approach for both questions straight away. In question 6, after reading the question, at 36 seconds she immediately said:

*that looks like Lennard-Jones potential*

She then drew this correctly. Although she recognised that the given equation was the Lennard-Jones potential, she missed the minus sign in the equation for force ( $F = -\frac{dU}{dr}$ ), so although she reasoned the gradient from using her equation very well, the graph was incorrectly flipped.

In the Doppler question (Question 7, 2010) after reading the equation she said:

*right so this is going to be looking at the Doppler effect*

She knew she needed to use the Doppler effect equation, but could not remember it. Once she looked it up online, she had no problems working through the implementation of this equation. She explicitly said that she had to look it up and acknowledged that she would not be able to do this in the exam.

In both questions, there were strong links between the problem statement and the relevant physics for this student. In question 7, expertise was demonstrated by having links to the relevant equation, which was looked up and applied correctly, as she knew exactly what to search for and how to apply it. In question 6, strong links were made to the correct diagram for the Lennard-Jones potential. In both questions, without accessing external resources, she missed an equation, or the correct version of an equation. This could be classed as procedural knowledge, as the correct concepts were applied to both questions. Thus this student had strong links between problem state knowledge and conceptual knowledge, but weaker procedural knowledge, which was supplemented by the use of the Internet on one question.

### 5.6.2 Student B

The first question Student B attempted is shown in Figure 5.3. For this student there was evidence of monitoring throughout the solution process, including planning and evaluation, so it was an expert-like approach in this regard. The student was also pattern matching to the required equation.

He mentioned three main conceptual points upon which to base the solution near the beginning of his solution:

- B.4.** Show that the Lagrange point between the Earth and Sun, where there is zero radial acceleration, has an orbital radius  $r$  which satisfies

$$\frac{1}{r^2} = \frac{m}{(r_E - r)^2} + \frac{r}{r_E^3}$$

where  $r_E$  is the orbital radius of the Earth, and  $m = M_{\text{Earth}}/M_{\odot}$ .

[5]

Figure 5.3: B4 question, 2008 paper

*okay so it's the Lagrange point so it's going to be orbiting at the same speed\* as Earth (1). Hmm. so there's going to be centripetal acceleration term (2) as well as..the forces coming from the sun and the earth (3). So..okay I'll write the forces down.*

\*It is likely that he meant angular speed at this point, as he said it correctly later in the solution process and worked it out correctly.

It was a “show that” question which seemed to aid his monitoring throughout the problem. Near the beginning he said:

*..but I can see already is looking vaguely correct*

then later

*oh I'm not sure..well okay I'll assume that's right*

He expressed his uncertainty, but unlike the 1st and 2nd year students he continued anyway. This ‘show that’ question demanded different skills from other types of problems. Student B knew he needed to find ‘m’ which is mass of the earth over mass of the sun, because it was in the final equation.

He was fluent in maths and seems to effortlessly think through equations. With regards to Gerace’s model [37] he had very strong links between procedural knowledge and problem-state knowledge. In other words, he compared the procedure he was using to the final answer in the problem statement to assess whether he was on the correct track. In this question the problem-state knowledge contains extra explicit information with regards to the final answer. The strong link may be due to this additional problem state information supplied in the “show that” format of the question. This information was solid and reliable, so he could be confident it was correct. The 1st and 2nd year students did not have

a “show that” question and some were perhaps not as confident in pursuing the correct approach.

Future work could investigate “show that” problems. Students perhaps need different skills in these style of questions and have the opportunity to practice them. These questions allow students to constantly pattern match and evaluate their answer with respect to the given answer. Confidence in approach is therefore perhaps less of an issue. This is a skill that stands them in good stead for becoming expert like, as they will be getting knowledge through doing lots of worked examples and matching these to future problems [29].

**B.5.** A  $1M_{\odot}$  Helium White Dwarf has a radius of 5000 km. Estimate the momentum of its most energetic electrons. [5]

Figure 5.4: B5 question, 2008 paper

On the B5 question, shown in Figure 5.4, this student was unsuccessful. The equation needed was for the Neutron star degeneracy pressure (non-relativistic). The crux of the problem was recognising that it was related to degeneracy. Unfortunately he decided not to use degeneracy at the beginning of the problem. Near the start of his solution he said:

*yeah it's not necessarily degenerate, because it could be very old and really cold and it wouldn't be degenerate*

This is, in fact, the opposite of what is true, these conditions make it very degenerate. What he does after this is the derivation for a star (the sun, not what is wanted in the question). He did this reasonably well, but it was not relevant to this problem. He did not manage to solve this problem, but made a good attempt.

According the Gerace's model he has adequate problem state knowledge, as he recognised that degeneracy might be relevant, but there are some ‘malfunctions’ with the conceptual knowledge. His procedural knowledge was all related to the central pressure problem so activated the right area of physics, but it was applied to the wrong type of problem. He also questioned whether the velocity was relativistic, which was relevant as well. His knowledge store was reasonable, according to Gerace's model, but missed some essential conceptual knowledge

parts. This is still progress from the type of knowledge and connections between the approach and implementation that the unsuccessful 1st and 2nd year students seemed to exhibit.

### 5.6.3 Summary

Both students showed more expert-like approaches than the first and second year students. They evaluated their answers throughout and obviously had experience in knowing how to do certain procedures. Both students had more sophisticated and connected knowledge structures, although they were by no means perfect. Both students had full access to resources, and used them. Although they could not necessarily solve all the questions correctly there were strong and correct links between the approach they decided to use and their implementation. Activation of knowledge generally helped rather than hindered their progress [27], and connections were made to the correct physics concepts.

The obvious improvement between year groups demonstrates the validity of the knowledge structure model as expertise and knowledge organisation improves in later years.

## 5.7 Open and Closed-Book Exams

Although an investigation of this kind looks at the skills required in closed and open book exams, there were some limitations to doing this. Providing students with resources to analyse their problem solving behaviour in this study is different from an open-book exam; the pressures on the student are greatly reduced, they volunteered to attend and there was no credit obtained for completing the questions. Although these interviews were not the same as exams, the questions set were of the same style and difficulty as exam questions and the students had reasonably strict time-pressures (about 15 minutes) in which to do them. Furthermore, it is not intended that rigorous comparisons can be made between the two studies with and without access to resources. Even though they were solving the same questions, both the year group and the availability of resources changed. These studies were intended to stand alone, however some similar patterns were observed in both studies. Some comparisons are made above, but one should keep in mind the change of multiple variables.

There are implications from the 2B study and more senior students, with access to resources, for the skills students need in open book exams. It is possible that even with a book of notes, students will not be able to answer an appropriate level of exam question. In fact, students who spend longer looking up resources may in fact do worse on the problem, as indicated in the resources in Chapter 4. Applying the knowledge structured model [37], perhaps students who do not have strong links between the required conceptual knowledge and problem state or procedural knowledge will be unable to know what to look up or how to apply it. It is possible that students could find a similar worked example to ascertain the main conceptual knowledge to apply, but this only happened with one student in this study and they were unable to easily implement this process, also running out of time and motivation. Procedural knowledge was also necessary even with resources provided. Therefore, conceptual and procedural knowledge are required, as well as links between them. Without these skills it is unlikely students will reach the correct answer even if given access to resources during the exam.

These results can be tentatively related back to the transitions maps produced in Chapter 4. Perhaps students with more complicated transition maps do not have the required knowledge and will be unable to use resources to assist them. Students with simpler transition maps may indicate stronger knowledge structures, as discussed above and may mean that resources can be used by these student productively, if they are needed at all.

Finally, the type and level of problems set is important in open book exams. Exercises are likely to be relatively straight forward with a textbook available, as the equation for a certain quantity can be directly looked up. However, it seems that harder, multi-step problems, where the physics concepts required to solve it are not immediately obvious, similar to those used in this study, are not necessarily made any easier with access to resources.

## 5.8 The Minnesota Model

The Minnesota problem solving model's steps are intended to be taught in their linear order to encourage qualitative analysis of the main principles or concepts at the beginning [3, 4], followed by a plan of how this approach will be implemented. Describing the physics is an important first step, [3, 21, 22, 23] however it does

not seem to be the natural behaviour of 1st and 2nd year students to work linearly through the model. Most started by describing the physics, but then even students who successfully solved the problem tended to work through the stages with strong links between Describing the physics and Execution. Perhaps this was because students worked out what they could and then thought about the next part of the problem. This approach may have helped cue information that they had not thought of at the beginning of the problem. Successful students also evaluated their solution at points throughout the process, not just at the end. Other research also supports continual evaluation, as experts spend more time evaluating their working throughout the problem, for example even evaluating the initial framing of the problem [16].

A re-ordered representation is proposed to accommodate non-linear problem solving behaviours and the iterative nature in which evaluation is carried out. The stages are presented in a circle and with evaluation in the background to show the non-sequential and iterative nature of the model, as shown in Figure 5.5.

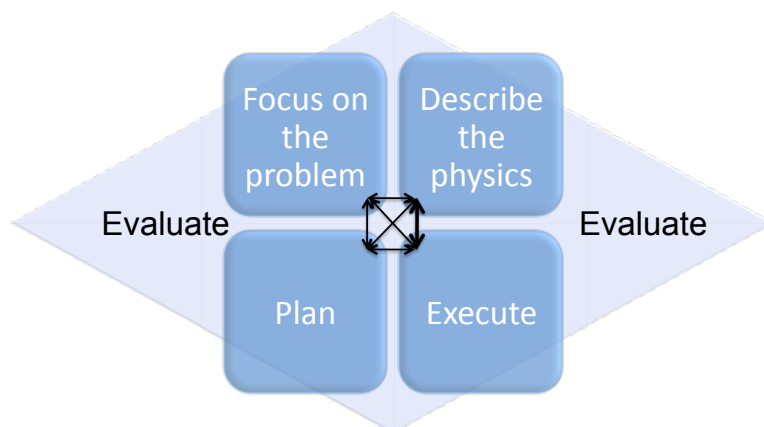


Figure 5.5: Re-ordered Minnesota Model [3] problem solving strategies

This discussion does not invalidate the Minnesota model as a learning tool. Firstly, it was not examined from this perspective and secondly, the model was not fully integrated into the Edinburgh course, as recommended by Heller et al. [3]. Problem solving models provide novice students with a starting point that they may not have had otherwise and demonstrate expert problem solving behaviours to students. Problem solving models, specifically the Minnesota model and how it compares to observed expert and novice behaviour in this study, is discussed



further in Chapter 10.

## 5.9 Conclusions and Implications for Teaching

There were key differences between successful and unsuccessful students. Unsuccessful students had more complicated transitions to and from different types of problem solving behaviour. Successful students had relatively simple transition maps, as they seemed to know how to approach the problem on a strategic level. In addition, successful students appeared to have stronger links between conceptual and procedural knowledge and were able to cue appropriate mental resources. Redish and Sabella [27] say knowledge structures need to be organised so that students can recognise the appropriate cues. The question arises of how this can be taught. Students need practice with both strategic and tactical decisions. Pointing out problem cues to certain approaches and practising procedural techniques could help. This could assist students to reflect on where they are weaker, whether it is knowing which approach to apply or knowing how to implement it. Other metacognitive skills, such as planning and evaluation are also development areas for novice students. Giving marks for evaluation skills, such as applying limiting cases may help develop this skill and re-enforce its importance. Encouraging students to reflect on their problem solving process may be another way to help develop these problem solving skills [26]. For example, encouraging them to reflect on how well they solved the problem, list assumptions they made and things they would do better or differently next time.

These results were considered with respect to open-book exams. Understanding key physics concepts and organised and rich conceptual knowledge seem to be necessary even when resources, such as the Internet, are available to the student. Therefore appropriate open-book exams could bring the exam closer to assessing application and understanding of physics, as well as skills which students will require after university, such as applying information to new contexts. Strong knowledge structures are needed to anticipate, and be flexible to, new problem contexts.

The methods of this study were authentic, as it is likely that students will use a range of resources when solving problems at university and beyond. Resources could include colleagues, textbooks and technologies, such as the Internet.

Although familiar with technology through everyday experience [110, 111, 112], students may not have the skills to search for and apply information from the Internet in order to solve physics problems. Perhaps students need to be taught how to use resources, such as the Internet, to optimise their use, especially as this is a skill which will be required after University. Assessment of the quality of the information found is also an important skill and students should not take worked solutions found online at face-value.

Peers can also be a resource, used in a more complicated way, not just to solve problems, but also to transmit or construct knowledge. These aspects are incorporated into Peer Instruction [64], discussed in the next section. The aim of the following chapters is to investigate characteristics of a successful peer discussion and explore the depth and type of reasoning students exhibit whilst learning about key physics concepts.

## 5.10 Future research

It is clear that further research into resource use is needed, both for investigatory purposes and to consider implications for open-book exams and more generally for learning and teaching. This would involve similar studies to the ones described above, but in a more explicitly time-pressed environment and on the same cohort. This would confirm the results from this study and base them in a more exam-like environment. Full investigations into open-book exams, and possible other examination techniques, would further expand an area that has been behind developments in teaching [81]. The potential outcomes of this work could be to evaluate the use of open book exams, and to confirm which skills these exams are actually testing, so instructors can choose the appropriate test style for the skills they wish to examine. The wider question of whether open-book exams develop true physics competency could also be examined. Defining student competencies would be the first task and one which relates to graduate competencies, for example those defined by the University or bodies such as the IOP [1], and local competencies, such as the learning outcomes on the specific course that open-book exams would be used. Finally, increasing the sample size would make the results more generalisable.

## Chapter 6

# Peer Instruction: Aims and Methodology

Although exams are usually completed by students individually, learning at University can be conducted in a variety of ways, using different techniques. Active learning techniques have been shown to result in higher learning gains than traditional methods [66]. Lectures in particular have traditionally been conducted in a didactic manner and active learning techniques are encouraging a more interactive lecture structure. For example, feedback on student understanding in lectures can be collected by instructors. The student is likely to be engaged in order to answer the question and the instructor can assess whether students have understood the concept covered and adjust the lesson accordingly. Learning with peers also has many benefits and can potentially push students further than they could go individually, as they work in the zone of proximal development [41], as defined in Section 1.7. Peer Instruction (PI) combines both peer work and student feedback usually in a lecture setting. The stages of Peer Instruction as recommended by Mazur are given below, however Turpen and Finkelstein [54] suggest that this model of PI is not used ubiquitously, as some instructors deviated from the recommended stages below.

1. The class is presented with a conceptual question and asked to think individually on how they would answer.
2. Pre-vote: the class votes individually. In the case of this study, the class voted by using electronic clickers [114].

3. Results from the pre-vote are available to the instructor, but not to the class, as this could affect their subsequent vote [115]. The instructor decides on the next step; class discussion is used if around 30-70% of the class are correct [64].
4. Having chosen to initiate the class discussion stage, the lecturer instructs students to turn to a neighbour and explain their reasoning.
5. Post-vote: the class re-votes on the same question in light of their discussions.
6. The lecturer explains the solution, either using whole class discussion or by didactic methods [54].

As well as using feedback from students via their response to the question, the lecturer is advised to wander around the lecture theatre listening to conversations and engage with students [54, 64, 70]. This is in order to assess aspects of their discussion, such as their methods of explanation and the level of student understanding [70]. In this way, a fuller picture of student understanding can be developed.

This is the first chapter in the set of three investigating Peer Instruction (PI) to understand more about the PI process. This research could be used to inform instructors in terms of what to listen for and to illuminate the types of conversations occurring in their class. The investigation presented in these chapters on PI is based on the implementation of PI at the University of Edinburgh, but its findings are likely to be of wider relevance within the field of active learning and to other institutions. This chapter presents the aims of this research, data collection and an analysis of the methodology.

## 6.1 Aims

The aim of this research was to investigate characteristics of success in PI conversations. Success was classified in the first study (Chapter 7) as a correct vote by the student with the smartpen (penholder) after the group discussion. It is recognised that this may not be the case (i.e. a correct vote does not mean the students understand the concept). Conversations are considered in more detail by qualitative analysis in Chapter 8 to examine this assumption.

The research questions were constructed in the order shown below, as results from one stage of research informed the next. This is discussed further in the subsequent chapters and several case studies are used in Chapter 8 to illustrate the emergent themes and bring the results from both Chapters 7 and 8 together. These aims were studied through three stages of research, as described in Table 6.3.

1. Does use of technical language by students during PI correlate with the pen holder's success on the clicker question? This research question emerged from a pilot study which is discussed further in Chapter 7.
2. What features emerge from the data to characterise a successful or unsuccessful conversation and provide further insights into students' discussions?
3. On a smaller case study basis, did students have complete and correct reasoning? Correctness of reasoning is often missed out in problem solving models [2, 3, 24].

## 6.2 Cohort Background

Any study of student learning depends on the context within which it is based [3]. Therefore the cohort and course is explained here. The cohort selected for the study was the first year undergraduate core physics course students, enrolled on a course called Physics 1A in Semester 1 of the academic year 2011/12. Physics 1A is mandatory for any physics majors student, but is also taken as an optional course by students from other majors. Just over 200 students were registered on the course with appropriately 50% of them majoring in a physics subject. Physics 1A mainly covered linear and rotational motion, energy, work and simple harmonic motion. The course had a history of instructional improvements and used 3 hour, weekly, group tutorials, solving a range of problem types to supplement lectures. The study was conducted the first year that lecturers on the Physics 1A course had used PI extensively in lectures along with demonstrations and whole class discussion. In each lecture, between 1 and 3 full PI cycles were completed. In occasional lectures there were no peer discussions when a large portion of students answered correctly after individual thought and so the instructor did not spend time on peer discussions. Students are assumed

to understand the material if they vote correctly, so after explaining the answer the lecturer moved on, as recommended by Mazur [64].

## 6.3 Data Collection Methodology

PI is a relatively new area of research. As reviewed in Section 1.8, studies have recorded students using observational measures [54, 55] or audio recorder [76]. So far very little research has been able to connect students' notes with audio in an unobtrusive manner, to minimise an alteration in their behaviour. Although PI discussions have been connected to whole class results on the clicker question [74], there have been no studies, to the author's knowledge, that match conversations to clicker votes for a specific student in the group. This study addressed this gap by connecting the pen holders' clicker responses to their conversation. It was also an authentic study in that smartpens were used in an ongoing ethnographic style study conducted in the field.

In order to investigate peer instruction used in the first year undergraduate physics course at the University of Edinburgh, students were given the option to use a Livescribe smartpen to record lectures. The smartpen was used as the data collection method and is discussed in Section 2.1.5. 50 smartpens were bought, including paper, spare ink and headphones and were available for 1st year students to use voluntarily. A smartpen provides a useful notation device for the student volunteers, as they can listen to the lecture again when studying at home. For example, when a student wrote a specific word they can listen back to what was being said about it at the time. Students were given minimal instruction about what to record and were under no obligation to record specific items.

Ethical approval was obtained from the institution. For the PI study, a code of conduct and consent form (Appendix B), which set out the responsibilities of the researchers in terms of anonymity of data and the responsibilities of the students, was signed by volunteers. The students were free to discontinue from the study at any time. The whole class was informed of the study and what a smartpen was on multiple occasions and students were under no obligation to keep recording during a discussion if they or others felt uncomfortable. Students also had time between recording and uploading the pen to the University laptop

to delete any audio they did not want heard. They were encouraged to delete any non-physics subject recordings from the smartpen before uploading. No attempt was made to identify non smartpen holders and details of smartpen users were kept anonymous in any form of report.

### 6.3.1 Constructing the Study

When constructing the logistical plan for the study, a risk assessment was first constructed which helped in the development of other necessary documents. The main risks and how they were mitigated against are discussed below. A range of documents was developed to give to the students, tutors and the research group. These included documents available to students:

- Code of conduct and consent form - Emailed to students who said they wanted to use a smartpen and printed out so there was a paper copy to sign when they picked up the pens. Shown in Appendix B.
- Information on the study.
- Short Instructions - A printout was given to students with their pen.
- Instructions and FAQs.

For the research team additional documents were available:

- Risk analysis and what to do to reduce or prevent these risks.
- Logistics documents - what needs to be done and when.
- Advert - what the students are shown and told in their second lecture to advertise the pens.

For the tutors there was a document created explaining what they should expect and this was emailed to them at the beginning of the semester by the head of the 1A course. A code of conduct for coders was created, which was read by anyone coding the data. The content of this was designed specifically to show the coders what the students were told about the study and the confidentiality agreement, but also general points that needed to be adhered to whilst in contact with the data.

### 6.3.2 Deployment of the study and Timeline

In the second lecture of the first week of lectures, the smartpens were advertised to students. The lecturer also reminded students of the study and availability of the smartpens in subsequent lectures. Students could then email the research group asking to participate. After a student emailed they were told a time and place to pick up their smartpen in week 2. Students kept the smartpen for the whole semester, but could return it at any time, ending their participating in the study. This meant that the smartpen could be used regularly by participating students as a learning tool, rather than distributing and collecting smartpens for each lecture.

Instructions were kept deliberately vague, students were asked to record one physics ‘thing’ a week, for example, a lecture, workshop, or individual problem solving thought aloud.

### 6.3.3 Running the Study and Collecting Data

Uploading the data from the pens was done in physics workshops. First year students attended one of four three hour workshops a week. From week 3 the University laptop was taken into workshops, so students could upload the data from their pens. Students were encouraged to upload every 2 or 3 weeks, but highest upload numbers were when the researcher was present to encourage students to do so. Analysis was conducted as the data came in. The quality and type of recordings was checked, and a list of categories of students’ smartpen use developed and updated over the course of the Semester. This included what students were recording and whether it was group work, individual work, or lectures.

## 6.4 Risks

Risks such as loss of data were minimised by students regularly uploading to the University laptop and this data was backed up onto a separate USB memory stick weekly. In order to keep track of the smartpens and avoid students incorrectly labelling them, each smartpen was labelled physically and electronically before deployment to the students. Students were instructed not



to change the smartpen's name, so the pens could be connected to the student using them, essential for data management and analysis purposes.

A risk that occurred that was not anticipated was that the writing of some students was not detected clearly by the smartpen. A written page when viewed on the computer appeared as scribbles, because the smartpen had not detected when the student had raised their pen to begin a new letter. Students with this problem were encouraged to swap their smartpen for another one. Students not uploading during the semester was another concern, as research data may not be captured. However, the smartpens of the students who did not upload their data was examined on return of the pen and the smartpen memory was not full, so no data was lost. The smartpens could hold a lot of information before the memory became full.

## **6.5 Limitations of this Method**

As students were not filmed, a certain amount of data was lost with regards to who was speaking, in what order each student contributed an answer and where they were sitting in relation to each other was not collected. Two main approaches could have been used to look at this and that is using a video camera and taking field notes. Neither approach was deemed suitable for this study, as the point of this approach was to research authentic PI discussions. The smartpen was available when students chose to use it and was encouraged as a learning tool. A researcher's presence in a lecture and especially a camera would have undermined the aims of this particular study. However these are areas that would be very interesting to look into in the future. Finally, as this was the first time that full PI had been used, the instructors may have conducted this pedagogical technique differently from future years, as the approach becomes more normalised for them.

## **6.6 Statistics of Smartpen Use**

This section provides an overview of smartpen use in the study in order to investigate PI in lectures. 27 students volunteered to use a smartpen. Two of these students were discounted from any analysis, as they did not keep the pen for longer than a week. Therefore 25 students were in possession of a smartpen

throughout the semester of 2011/12. In total 19 students used their smartpen to record a physics activity, which included at least one lecture for each of these students and a PI episode. A total of 162 PI episodes were captured throughout the semester. Taking a PI question as a session where a clicker question was presented and followed by full PI (i.e. discussion and then re-vote), 29 out of a possible 31 different PI questions were recorded by at least one student. Two PI questions from the first week were missed, as students only received their smartpens in week 2 of the semester. This meant 12 out the 14 questions were captured from the first part of the course of mechanics and circular motion and 17 out of 17 questions were captured from the second part of the course on work and energy. The minimum number of students recording a PI question was two and maximum ten students. Though as discussed later, the quality and audibility of recordings was not consistent across all episodes.

## 6.7 Investigating the Data Collection Method

Students' statistics of those volunteering to use a smartpen were compared to the rest of the cohort, using scores on a conceptual mechanics test and the course exam. This was to determine whether there was a reasonable distribution of smartpen users with regards to student ability and test for equivalent learning gains after a semester of using a smartpen.

The Force Concept Inventory (FCI) [44] is used to measure conceptual understanding in mechanics and was administered at the beginning and end of the semester. Matched data was used to calculate learning gains on the FCI, taking data from students who completed both the pre and post tests. The normalised gain was calculated as:

$$\text{Normalised gain} = \frac{\text{Post}\% - \text{Pre}\%}{100 - \text{Pre}\%}$$

Pre-test is the results of the FCI conducted before instruction on that topic, post-test represents test scores after instruction. Post % and Pre % in the equation above represents the students percentage mark on the post and pre tests respectively. There was FCI matched data for 18 of the 19 students who used the smartpen. All 19 students were awarded a mark for the course, as they completed the final exam.

On Figure 6.1 the vertical lines represent the pre-test FCI scores of smartpen

users. These were compared to the whole class distribution of marks also shown on Figure 6.1 to demonstrate the range of pre-test smartpen scores. This indicates that volunteers for the study were at different levels of conceptual mechanics understanding before instruction. It also disproves selection effect theories where an already successful student may volunteer for a learning tool.

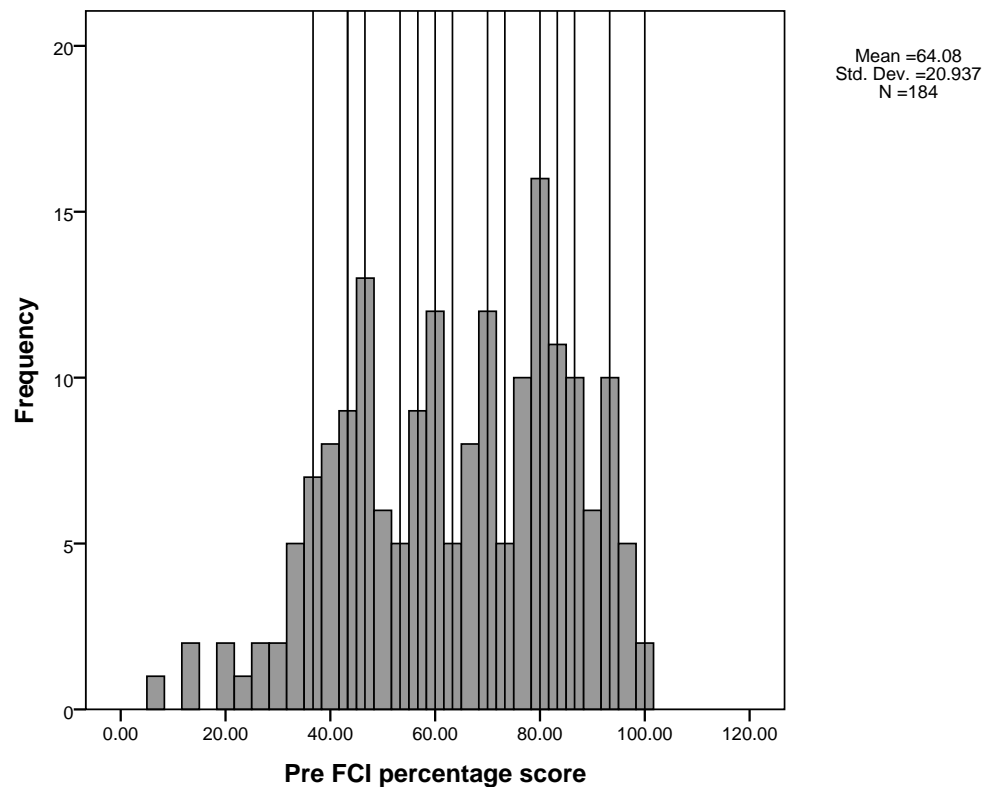


Figure 6.1: Shows FCI pre-test histograms for all the students who completed this test at the beginning of Physics 1A. The scores of students who volunteered to use a smartpen and recorded peer discussions on it, are shown by the vertical lines

Students' gain on the FCI and final course marks were also compared to smartpen use in order to compare learning gains of students who used a smartpen to those that did not. In order to determine whether comparisons could be made between smartpen and non-smartpen users, an independent samples t-test and Pearson chi-squared test were conducted to test for differences between smartpen and non-smartpen users on both the FCI and the final exam. The results are

shown in Tables 6.1 and 6.2. A null-hypothesis would mean that there were no differences between smartpen and non-smartpen users on tests measuring understanding. Statistical significance was set at the  $p < 0.05$  level, where any comparisons between groups with a  $p$  value greater than 0.05 meant a null-hypothesis. SPSS statistical software was used to analyse the data and it also conducted a Levenes test for equality of variances to determine whether equal variances should be assumed or not for the t-test.

As shown in Table 6.1 the FCI scores of smartpen versus non-smartpen users were not significantly different. Furthermore, no significant differences were found comparing course marks of smartpen and non-smartpen users, as shown in Table 6.2. Although there are suggestions in the literature that the smartpen aids studying [94], there is no evidence to suggest, from this data, that the smartpen users had higher learning gains than non-users. The use of the smartpen as a learning tool is discounted, but means that the peer discussions recorded during lectures provided a potentially representative sample of the whole year group. However, one should be careful to generalise too much, as the numbers in one group are quite low. Nevertheless, these results enable the study to proceed without major concerns over student selection effects or bias due to the use of a learning tool.

Students	Number in each group	Mean FCI Normalised gain (Standard error of the mean)	Sig (2 tailed) from independent samples t-test	Pearson Chi-squared p value
Smartpen users	17	0.73(5)	0.07	0.14
Non-smartpen users	142	0.54(3)	equal variances assumed	

Table 6.1: FCI gain for the whole class depending on whether the students used a smartpen or not. Although the number of smartpen users was 19, one student who used a smartpen did not have matched FCI data (they only did the post-test) and another student achieved 100% pre and post, and so did not have a normalised gain score.

Students	Number in each group	Mean course % mark	Sig (2 tailed) from independent samples t-test	Pearson Chi-squared p value
Smartpen users	19	66(2)	0.20	0.24
Non-smartpen users	180	62(10)	equal variances assumed	

Table 6.2: Mean course percentage mark of smartpen users compared to non users.

## 6.8 Peer Instruction in Context

Research was based on the assumption PI works as a learning process, both generally and for the cohort of students studied. In order to test this assumption a number of measures were used. Firstly, students' normalised gain on the FCI were compared to previous years and to other studies. Secondly, the percentage of correct votes for each student on the PI questions, after the discussion stage, was compared to their post-test FCI score, as shown in Figure 6.2.

The normalised gain of 0.56 on the FCI was the highest mechanics diagnostic score measured for this course at the University of Edinburgh, which had been integrating other active learning techniques for a number of years, but was the first year of full PI. Prior to this a mechanics diagnostic similar to the FCI had a typical average gain of around 0.4. Crouch and Mazur's [69] FCI normalised gains at Harvard University, started at 0.49 when introducing PI into their calculus courses in 1991 and improved to 0.74 in 1997. Comparing to other Universities [116], normalised gain in this study for a class of majors and non-majors is reasonably high. As shown by Hake [66], average normalised gain for courses using interactive-engagement methods was  $0.48 \pm 0.14$  (std dev) compared to traditionally taught courses when the average normalised gain was  $0.23 \pm 0.04$ .

The end of course exam was open-book which tested understanding and higher-order thinking, over memorisation of facts, as there were no questions asking for factual recall. The cohort was successful in that pass rates and average course marks were in an acceptable range according to College guidelines for well-functioning courses (since using a flipped class, mean exam marks have been in the range of 55-65% and course pass rates above 85%). This is comparable to previous years, which were also within college guidelines, though the open-book

exams now tested further higher-order skills. This suggests that students had learnt how to apply their knowledge rather than memorising it.

Figure 6.2 shows the percentage of PI questions students answered correctly compared to their post FCI percentage score. Matched PI questions were selected, when a student submitted both a pre and post-discussion vote and therefore fully participated in the PI process. Also only matched data from the FCI was used in Figure 6.2, where the student had answered both the pre and post FCI tests. Only the PI questions relating to force and motion were selected, as those were the concepts tested in the FCI.

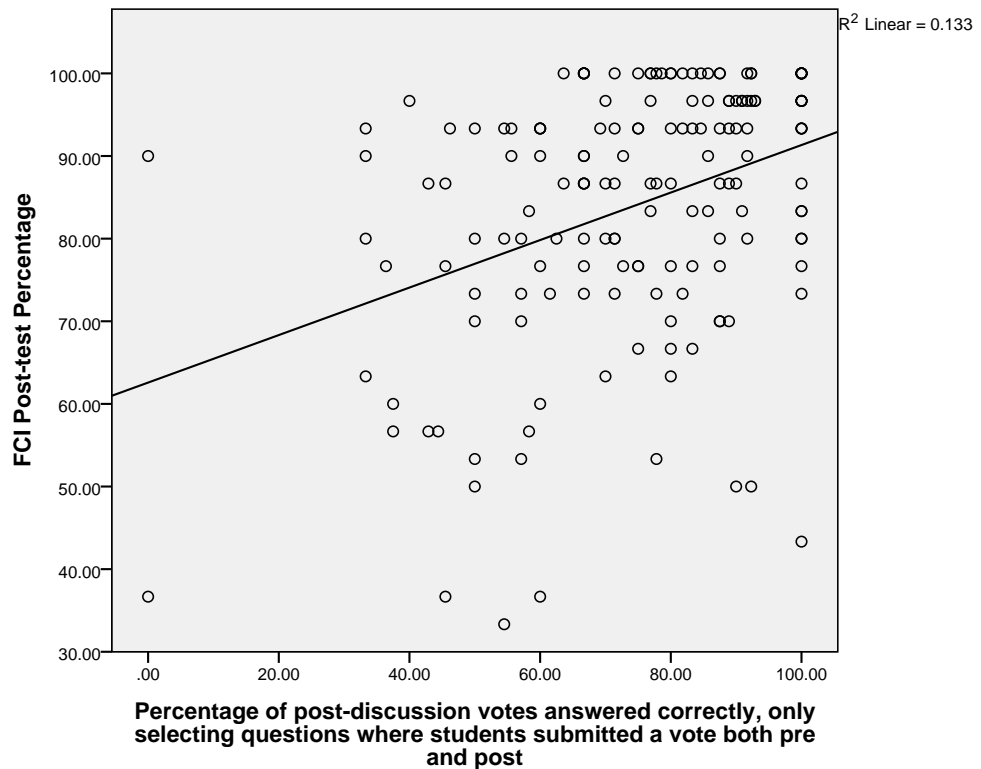


Figure 6.2: Percentage of PI questions students answered correctly out of all matched questions they attempted, against Post FCI percentage score. This data was taken only in the mechanics section of the course to relate to the FCI. A matched question was when the student submitted both a pre and post-discussion vote for a clicker question.

The Pearson's correlation coefficient for the relation shown in Figure 6.2 is

0.364 and has a highly significant value with  $p < 0.0001$ . The  $R^2$  value on Figure 6.2 is 0.133, meaning 13% of the variation in FCI post percentage mark was explained by answering PI questions correctly after discussion. This means that out of the students who participated fully in Peer Instruction by submitting both pre and post discussion votes, those who voted correctly in the post-discussion vote were more likely to get a higher mark on the FCI post test. Although causality of this relationship is unknown it shows the strong link between success after PI discussion and responses to the FCI at the end of the course.

## 6.9 Summary

This section summarises the smartpen study in general and sets the context, in terms of the course, cohort and use of smartpens. The aim was to authentically study student discussions whilst they participated in PI in lectures. The smartpen volunteers did not appear to be significantly different from the rest of the cohort in terms of FCI score or course marks. PI was considered successful in terms of improving understanding after examining students' FCI marks and percentage of PI questions answered correctly.

Three methods of analysis were used to examine this data. Table 6.3 shows an overview of this research including the methods employed for analysis to investigate PI. The methods for analysis are discussed in the relevant chapters.

Data Collection (Chapter 6)	In-the-field observations with a Smartpen
Data Analysis and Results 1 (Chapter 7)	Counting the number of technical words used and matching to pen holders' votes on the PI question.
Data Analysis and Results 2 (Chapter 8)	Thematic analysis and case studies. Includes use of technical words for further consideration with the qualitative analysis.
Discussion (Chapters 7 and 8)	Discussion from both methods of analysis are contained in the relevant chapters.

Table 6.3: Summary of the Peer Instruction research including data collection and analysis techniques employed for first year students.

## Chapter 7

# Peer Instruction: Use of Technical Language

The data analysis method stemmed from a pilot study, described below, where students with qualitatively less sophisticated conversations and on the question with a lower whole class gain and persistently low percentage of students voting correctly, appeared to use less technical language than students on a high gain question. This type of methodology (counting words spoken) has been used in prior Peer Instruction (PI) research [77]. In this research technical language was targeted specifically and matched with the students' responses to clicker questions. This research was also used to determine whether the methodology of word counting straight from the recording was viable as a faster method of analysis. The method of analysing the data, creating the list of technical words and the results and discussion are covered below.

Firstly, technical language was defined as subject-specific words which have a specific meaning when used in a physics context. They are used to represent or describe physics models, situations and movement. This included words such as 'force', 'Newton's laws' and 'momentum'. Scientific words were also classed as technical in this study. This included words, such as 'maximum' and 'minimum', which have a precise meaning in science, but may not be unique to physics.



## 7.1 The Pilot Study

Included in this section is a condensed and edited version of a pilot study which produced a proceedings paper for the 2012 HEA STEM Conference written by M. Wallace and R. Galloway. This shows the rationale behind choosing technical words as an area for investigation.

There were 162 PI episodes in total. An analysis of a small sub-set of these conversations was used in an initial study to help shape the direction of analysis of this data. From the data set collected, as described in the chapter 6, eight discussions on two different PI questions were transcribed for the pilot study. The two questions selected had very different whole class gains from before and after the peer discussion stage.

### 7.1.1 Pilot Study Method

Unsuccessful and successful PI questions were identified using the whole class gain from pre to post discussion (i.e. what proportion of the class went from initially holding the wrong answer to having the right answer after discussion) and the percentage of students voting correctly. Individual success was measured by the students' post discussion response and the correctness of their argument. Notes were made on the transcripts using a grounded theory approach [117], where concepts are extracted directly from the data. General characteristics of the conversations were noted, as well as their sophistication with regards to whether students correctly reasoned with a suitable physics justification. Success was then related to technical word use during the discussion stage.

Two questions were chosen for analysis, one classed as a successful PI question and the other unsuccessful. One question had a whole class gain of 0.58 with 14.2% of students voting correctly before discussion, improving to 64.3% after discussion and was considered as an example of a successful PI question. The other question had a whole class gain of 0.09 with 34.8% of students voting correctly before discussion and 40.9% voting correctly after discussion and was therefore considered an unsuccessful question. Four student conversations recorded by the smartpens and reflecting the range of successful and unsuccessful peer instruction interactions were analysed in detail for the two questions to examine why students were successful or not on these questions.

A list of appropriate technical words was developed from frequently-appearing words in the course handbook. The course handbook was chosen as the source for appropriate technical language as every student was given a hard copy of the handbook and it set the expectations of the language to use during scientific reasoning on the course; this is linked to the enculturation process discussed by Schoenfeld [118], as students begin to learn what skills the community of physics values.

The total number of technical words used in each conversation,  $N_{tech}$ , the number of distinct technical words used, the resulting  $h_{index}$ , and whether the question was answered rightly or wrongly pre and post discussion was determined for each student recording. The  $h_{index}$  is usually used to quantify research output and citation rates [119]. In this study the  $h_{index}$  equals  $n$  if  $n$  different technical words are used  $n$  times and the remainder of the words have been used no more than  $n$  times. Simply counting the total number of technical words used would give a high weighting to conversations where the students used only a small number of distinct words many times over: the advantage of the  $h_{index}$  is that it could identify conversations where students made extensive use of a variety of ideas and concepts during their discussions, i.e. it more heavily weights rich discussions rather than simply lengthy ones.

## 7.1.2 Pilot Study Results

### Case Study 1: A Successful Peer Instruction Episode

The clicker question: Suppose a ping-pong ball and a bowling ball are sliding towards you along a frictionless surface. Both have the same initial linear momentum, and you exert the same force to stop each. How do the distances needed to stop them compare?

1. It takes a shorter distance to stop the ping-pong ball
2. Both take the same distance
3. It takes a longer distance to stop the ping-pong ball

The correct answer is option 3 and the whole class gain on this question was 0.58. The conversations of the four students analysed are summarised in Figure 7.1.

Student	$N_{technical}$	h-index	Number of different technical words	Right (R) or wrong (W)	Conversation
A1	31	4	10	WR	Both right and wrong arguments. Used KE and forces.
A2	44	4	16	WR	Correct persuader. Used work energy theorem.
A3	14	2	5	WR	Persuaded correctly (but not totally convinced). Used forces.
A4	11	2	6	WW	Persuaded incorrectly. Used forces.

Table 7.1: Discussion parameters for the whole conversations during the successful PI episode;  $N_{technical}$ , h-index and number of different technical words used, each penholder's votes (right or wrong pre and post discussion) and the general type of conversation.

Although all discussions were on-topic, student A2 was part of the only group to use the work-energy theorem to solve this question, which was the most efficient approach. Other groups A3 and A4 discussed kinematics, but not energy. A1 discussed kinetic energy briefly, without pursuing this further.

A2 had a higher  $N_{tech}$  value when compared to the other groups and also high  $h_{index}$  and  $N_{diff}$  values. Their reasoning was complete and correct which perhaps indicated a relationship between sophistication of argument and frequency of technical word use. This was supported by the other groups. A4 had comparatively low technical word use and this student answered incorrectly after discussion (WW). Qualitatively student A4 did not engage with the discussion, as they said they had done this question before. Furthermore, Student A3 voted incorrectly and then correctly after discussion (WR), but as they did not summarise their understanding at any point in the conversation, it was unclear whether they fully understood the question. Finally students talking with Student A1 did not fully explain their reasoning either, but started with a discussion of kinetic energy making their discussion more on track in the beginning than A3

and A4. The only student to answer incorrectly after discussion had the lowest use of technical language. Success and sophistication of argument could relate to frequency of technical word use, according to this case study.

### Case Study 2: An Unsuccessful Peer Instruction Episode

The clicker question: Which one of the following is NOT a true statement about the frictional force acting on a block on a rough surface?

1. The frictional force is given by  $\mu_k F_N$  if the block is accelerating
2. The frictional force is given by  $\mu_s F_N$  if the block is stationary
3. The frictional force can be less than either  $\mu_s F_N$  or  $\mu_k F_N$

The correct answer is option 2 and the whole class gain on this question was 0.09. The four student conversations analysed in detail are summarised in Figure 7.2.

Student	$N_{technical}$	h-index	Number of different technical words	Right (R) or wrong (W)	Conversation
B1	7	1	6	W-	Partly incorrectly persuaded.
B2	5	1	5	WR	Persuaded correctly at end.
B3	20	3	6	WW	Neutral.
B4	3	1	3	WW	Incorrect persuader.

Table 7.2: Discussion parameters for the whole conversations during the unsuccessful PI episode;  $N_{technical}$ , h-index and number of different technical words used, each penholder's votes (right or wrong pre and post discussion) and the general type of conversation.

B2 was the only student to successfully answer this question. Unfortunately some of this conversation was inaudible, so much of the reasoning was not heard, which may have affected the number of technical words counted. Students in B1 and B4's groups did not consider the correct answer at any point in their conversations, as they considered option 2 as being a true statement. B3's

discussion included a conceptual discussion, but as they attempted to consult a textbook, they ran out of time.

As shown in Figure 7.2, two students who either did not answer (B1) or answered incorrectly (B4) did not make progress on this question, mostly focussing on irrelevant details which were tangential to the main point of the question. They also did not use many technical words and had an  $h_{index}$  of 1. B3's group frequently used many technical words, however they did not make tangible progress.

Overall, on this unsuccessful question (Figure 7.2), there was a lower use of technical language, when compared to the previous successful case study (Figure 7.1). In the first case study, students were on topic and mostly answered the question correctly after discussion. In this case study, students were mostly incorrect and focussed on irrelevant details.

### 7.1.3 Conclusions of the Pilot Study

The case studies presented here suggested that pursuing technical word use had potential in defining characteristics of successful and unsuccessful conversations. The number of distinct technical words used seemed to be higher when overall class gains were higher. Higher quality conversation also seemed to have a greater use of technical language. Mazur argues that students can convince others in their own language [64], without the need for technical jargon. However Wellington and Osborne [61] suggest that students will understand a subject fully if they understand the language of that subject. From the results of the pilot study, the initial hypothesis was that students who used more technical language would have a greater mastery of the physics and be more likely to answer the question presented correctly.

This hypothesis also has theoretical foundations with regards to the constructivist theory of enculturation [118]. Students are expected to develop and use technical language in order to construct clear and cohesive arguments. By becoming part of the physics community of practice in their course and institution, it was hypothesised that students would develop and practice their use of technical language, as is used in that community. As Schoenfeld wrote:

*learning is culturally shaped and defined: people develop their understandings of any enterprise from their participation in the 'community*

*of practice' within which that enterprise is practiced* p.22 [118]

Lastly, the methodology used worked well in the pilot study. Parameters were accessed relatively easily from recordings of student conversations, without requiring full transcription. The approach was thought to offer an efficient and effective way to analyse large data sets of discussion recordings. This was also intended to be something an instructor could listen for when walking past PI discussions, however it does not assess whether technical language was being used correctly, this is discussed in more detail in subsequent sections.

## 7.2 Methodology

Having conducted the pilot study, the subsequent aim was to count technical word use in all 162 PI episodes and relate it to the success of the conversation as determined by the penholder's post-discussion vote. This methodology is described below.

### 7.2.1 Developing the List of Technical Words

#### The Initial List of Technical Words

Unlike the pilot study, an initial list of technical words was developed by looking at the physics problem statements presented to students to initiate the PI conversation. The problem statements were chosen instead of the pilot study's method of using frequently cited words in the course handbook, as it created a more refined list in the first instance, which was specific to the questions presented to students in the lecture. The problem statements were also one aspect of the course that set the expectations for the use of technical language.

To validate words that the coders classed as technical, two coders independently coded the PI problem statements to develop an initial list of technical words, before coding the PI episodes. The technical words used in the problem statement, but not their frequency, were noted. An average 84% agreement (there was 83% agreement of Coder 1 with Coder 2 and 85% agreement of Coder 2 with Coder 1) was found. Coder 1 had 13 words that Coder 2 did not, and Coder 2 had 14 words that Coder 1 did not. All 27 additional words were added to create the first draft list of technical words.

A second draft list was then constructed after discussion. To create this version of the list three of the words were removed completely: spring (from Coder 1); and ahead and behind (from Coder 2), as they were not deemed to be ‘technical’. Undecided words were then analysed within the context they were used and assessed whether they were being used technically, for example ‘upwards force’. Coders were unsure about words such as ‘upwards’, ‘downwards’, ‘increase’, ‘decrease’, ‘less than’, ‘greater’ and ‘static’, but these were all kept in. ‘Ahead’ and ‘behind’ were deemed too elementary to be classed as technical words for university students, while a word such as ‘increase’ carries more information than ahead. It became apparent that different lists of technical words would need to be developed for different set of questions and different levels of students.

### **Finalising the List of Technical Words**

After the initial list of technical words was devised from the problem statements, ten PI episodes across the first two sections of the course were chosen to test reliability and further develop the list. As additional technical words would be spoken by students that were not in the problem statements, both coders coded all data looking for all technical words they could find whether on the list or not. A word was added to the list of technical words if both coders picked it up and it was confirmed as technical through discussions with the research group. Words where coders were unsure as to whether to classify them as technical were kept in this initial stage of coding (to create a richer data set) and could be discounted later in the process. Therefore, only once all 162 sessions had been coded by both coders was the final list of technical words completed. The finalised technical word list was essential in establishing the final count of technical words heard and was therefore necessary for the final inter-rater reliability checks. The list of technical words is given in Appendix C.

### **7.2.2 Criteria for Coding**

Criteria were established for coding. Firstly, the technical words had to be used in a conversation that was on-topic and relating to the clicker question presented in class. Conversations about previous questions were not included. The conversation also had to be one in which the penholder was considered to be an active or passive participant. This was obvious in most discussions due to the

clarity of the recording near the smartpen. The main conversation, which was the closest to the microphone, was coded. Coders listened to all the PI episodes for each individual pen holder separately, so that a familiarity was developed with who was usually talking and whether it was likely to be the right conversation to listen to.

For all PI episodes coded, in order to take down as much information as possible, and to try to later validate the method of coding for technical words, qualitative information was also noted. This included whether the coder thought the group was audible, whether it was easy to hear what was being said and any other information that emerged from the data which the coder thought was relevant in providing an understanding of that conversation and its clarity. For example, the coder may have noted that they were unable to distinguish which student was speaking. Trying to distinguish which individual student said which words was discarded for this reason.

### 7.2.3 Determining percentage agreement between coders

There were two reasons to compare analyses between the coders <sup>1</sup>:

1. To check the validity of the list of technical words. Were both coders in agreement with what was deemed to be a technical word? This refined the list of technical words, as explained above.
2. To check the reliability of coding process. The percentage agreement was determined between coders according to the finalised list of technical words.

This section considers how the inter-rater agreement was determined. A list of technical words was created by each coder, in approximately the order they were spoken, for example:

Coder 1: velocity, speed, distance, time

Coder 2: velocity, time, distance

As students often spoke over each other, some words were picked up at different times by different coders. Therefore, the words were ordered alphabetically. Before this, the order of the words coded were checked manually to reduce the error of a technical word spoken at the beginning of the conversation may

---

<sup>1</sup>Many thanks to Alison Kay for assisting in the coding of this data.



have been incorrectly matched with the same word picked up at the end of the conversation by the other coder. This created a second list in alphabetical order, continuing the example from above the two lists would now be:

Coder 1: distance, speed, time, velocity

Coder 2: distance, time, velocity

The words were matched up as demonstrated in Table 7.3.

Words from coder 1	Words from coder 2	Coder 1	Coder 2	Agreement
distance	distance	1	1	1
speed		1	0	0
time	time	1	1	1
velocity	velocity	1	1	1

Table 7.3: Inter-rater agreement method

The percentage agreement was then calculated by taking the sum of the final column in Table 7.3 divided by the number of rows, and multiplying by 100. In this example, the percentage agreement was 75 %. If either coder did not pick up a word it was given zero agreement, but used as a row to calculate the final percentage agreement.

Before multiplying by 100, the agreement calculated above was the same as Jaccard's coefficient [120]. Jaccard's coefficient is a measure of similarity and can be represented as  $\frac{p}{p+q+r}$ .

p is the number of words found by both coders,

q is the number of words found by Coder 1, but not Coder 2 and

r is the number of words found by Coder 2, but not Coder 1.

It ignores words that are absent for both coders, which in this case would be the other words on the final, full list of technical words.

The inter-rater agreement level was set at 70% and conversations below this level of agreement were discarded. 70% was chosen as an appropriate level of agreement to take into account the low quality of some recordings due to the background noise created when the whole lecture theatre was engaged in discussions.

### 7.2.4 Counting Technical Words and other variables

All words picked up by both coders were noted and two methods of counting the technical words coded were considered. The first method was to use overlapping words, which counted words which only both coders picked up. This eliminated any false positives whereby one coder picked up a word incorrectly that the other did not. The second method considered was to use all words coded by both coders in order to eliminate possible false negatives. A false negative is potentially a word that one coder found correctly and the other coder missed. In the example in Table 7.4, the overlapped list is distance, time and velocity. A longer list in this example, which includes all words found by both coders, is distance, speed, time and velocity. Initial analysis was conducted on the data in order to determine which of these measures to use, as discussed in Section 7.2.6.

Words from coder 1	Words from coder 2
distance	distance
speed	
time	time
velocity	velocity

Table 7.4: Overlap and All words example

### 7.2.5 Matching to Responses for Each Question

Once both coders had coded all 162 episodes each penholder's pre and post discussion vote was compared to the number of technical words used in each conversation. Each conversation was split up into the various pre and post discussion response types, including wrong to right (WR), right to right (RR) etc. Conversations where students did not vote on one or both of the opportunities to vote were separately labelled as 'none'. This was because students who did not vote may have been thinking correctly, or incorrectly, so could not be categorised as right or wrong. For example, no vote pre or post, but a recorded discussion on the smartpen may have meant that the penholder had forgotten their electronic clicker that day.

### 7.2.6 Data analysis

Three different measures of technical words were calculated. Students' total number of technical words used ( $N_{tech}$ ), h index ( $h_{index}$ ) and the total number of different technical words used ( $N_{diff}$ ). Values were calculated for both the overlapping words found by coders and all the words found by coders. As discussed in Section 7.1, the  $h_{index}$  [119], was used to identify conversations where students used a variety of ideas and concepts and hence technical language during their discussions. It meant that a conversation where students used one word frequently was weighted less than a conversation using a range of technical words. The  $h_{index}$  is defined here as n, when 'n' number of words are used 'n' times.

It was necessary to determine whether 'overlapping' or 'all' words picked up by coders would be used in the analysis.  $N_{tech\ overlap}$  and  $N_{tech\ all}$  words were compared using two methods; firstly, the comparative shape across individual clicker responses WR, RR, RW, WW for  $N_{tech\ overlap}$  and  $N_{tech\ all}$  were examined and secondly,  $N_{tech\ overlap}$  and  $N_{tech\ all}$  distributions (irrespective of voting response) were compared.

The shape of the  $N_{tech}$ ,  $h_{index}$  and  $N_{diff}$  values for 'overlapping' words, and 'all' words, were similar for clicker response types, but the distributions were significantly different, as many more words were picked up when counting 'all' words found by both coders. In this chapter numerical  $N_{tech}$  values were not considered, but rather the relative frequencies across clicker responses pre to post discussion. It was found from this perspective that the results were the same whether overlapping or all words were coded. As there was a higher degree of certainty that the overlapping words existed, having been heard independently by both coders, this method was used in the results and discussion.

The next stage of analysis was to determine the normality of the data. The overlapping data of  $N_{tech}$ ,  $h_{index}$  and  $N_{diff}$  was broken down into response type WW, RW, WR and RR. Taking each response type (RR, WR etc) of  $N_{tech}$ ,  $h_{index}$  and  $N_{diff}$ , a Kolmogorov-Smirnov test was conducted in order to compare the deviation of the data from a normal distribution [106]. Normality of each distribution was tested in order to determine whether to use statistical tests which assumed normality. It was found that none of the distributions of response type in  $N_{diff}$  deviated significantly from normality using this model. For  $N_{tech}$  only RR (correct pre and correct post) deviated significantly. All distributions for

the different response types measured with the  $h_{index}$  were significantly different from the norm. This may have been related to the fact that there were very few possible values of the  $h_{index}$  obtained by students.

The final stage of analysis was to test for statistical differences of technical word use between response types. There was some confidence in the normality of the distributions for response types in  $N_{diff}$ , so a one-way Independent ANOVA was conducted in SPSS to test for differences in technical words use in each response type. Due to the non-normal nature of the  $N_{tech}$  distribution for the RR response and the lack of normality for the  $h_{index}$  results, a Kruskal-Wallis test was conducted to see if there were any differences between response groups for  $N_{tech}$  and  $h_{index}$ . This measure does not require the data to have met the assumptions for parametric tests, as it ranks the data in order to determine differences between groups [106].

### 7.3 Strengths and Limitations of the Study

This section discusses some of the strengths and weaknesses of methodology employed. Two coders were used to code all 162 discussions, so there was no need to establish a high inter-rater agreement to determine confidence in a single coder. This was part of the process for validating the list of technical words. In order to add new technical words to the list, both coders had to have noted the particular word in the same conversation. The coders came from different backgrounds; one had a background in physics and one in primary school education. This potentially made the research more rigorous, as both coders had to agree on what was classed as a technical word. Having two differing perspectives to look at the data could provide a less biased approach and more potential words were presented as technical.

An advantage to the coding technique was the time it took to do. Listening for technical words straight from the audio was approximately three times faster than transcribing the recordings. Technical word counting as a means of classifying the data was therefore a relatively fast (compared to transcribing) method for this type of research. It could be improved by faster analysis techniques, such as voice recognition and word counting by a computer.

Ideally, all actual technical words spoken would be picked up by both coders

and there would be high inter-rater agreement across all discussion, however due to the authentic nature of the recordings, situated in a noisy lecture theatre, some conversations were inaudible and difficult to hear and therefore not coded. The qualitative comments on the coding pro-forma and discussions between coders determined whether episodes were discarded due to the lack of clarity of the recording. For example, many of the episodes with low agreement had comments relating to the problems in distinguishing who was speaking or that it was difficult to hear, so were discarded from the final data set.

With regards to the study being conducted over the course of a semester, although students may have been becoming familiar with the PI process, the material presented was new every week, so it was likely to be the same type of challenge each time.

PI episodes could have been normalised for time allocated to different questions. Each individual question had the same amount of time allocated for all students, but this time varied for different questions. This was not calculated, because there was only a slight variation between question length, so they were all considered as equal PI episodes. The data presented here was not normalised by the total number of words heard in the recording to account for the length of the conversation. Ideally all words used in each PI episode would be used to normalise the data. This was beyond the scope of the study and inconsistent with one of the aims of the study which was to find a faster way of analysing a large amount of data.

## 7.4 Results

### 7.4.1 Inter-rater Agreement

The inter-rater agreement on all 162 episodes is shown in Table 7.5. There were 63 episodes discarded from the data set. Analysis was conducted on the remaining 99 episodes which had a Jaccard's coefficient of over 0.7. This agreement level of 0.7 could also be described as a percentage agreement of 70%.

Data	Number	Jaccard's coefficient
Average agreement of all data	162	0.71(2)
Average agreement of episodes with over 0.7 agreement	99	0.86 (1)

Table 7.5: Inter-rater agreement and data discarded. Standard errors are shown in brackets.

### 7.4.2 Technical Word Use

The mean number of technical words ( $N_{tech}$ ) used for each response type (i.e. WR, RR etc.) were plotted and compared statistically. This is shown in Figure 7.1. There was no significant differences between any of the groups with a significance value of 0.607 using an independent samples Kruskal-Wallis test. Of particular interest were the four main response types RR, WR, WW and RW. There were no significant differences between these four groups, which meant that students who answered correctly after discussion were not on average in a group that used a significantly high number of technical words. This disproves the study's hypothesis.

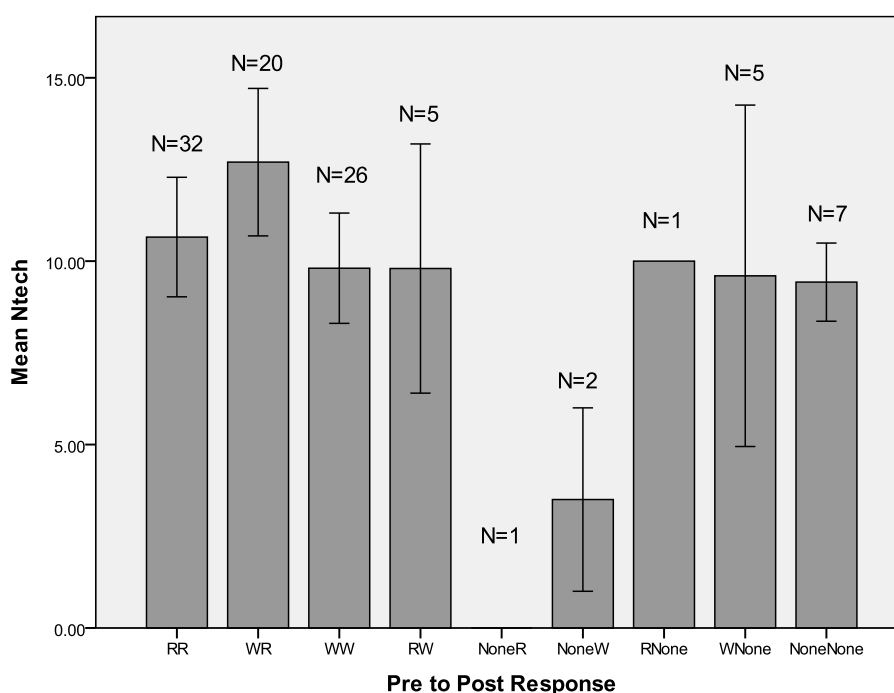


Figure 7.1: Mean Technical word use for each clicker response type. The number of episodes for each response is shown above the bars. The Number of episodes in each response is the same for Figures 7.1 to 7.3. Error bars are  $\pm$  one standard error of the mean. W means the vote was incorrect, R means correct and None was when a student did not submit a vote. The clicker responses read left to right, so WR means a student voted incorrectly in the pre-vote and then correctly in the post vote.

Figure 7.2 shows there was no significant differences between response type and mean  $h_{index}$  score with a significance value of 0.445 using a independent samples Kruskal-Wallis test. This meant that students in successful conversations (RR or WR) did not use a greater repertoire of technical language, more frequently. Again this disproves the study's hypothesis.

Finally, the mean number of different types of words used ( $N_{diff}$ ) was compared to response type, as shown in Figure 7.3. Although a slightly higher numerical value for the WR group, there was no significant difference between groups using this measure. The significance value using an ANOVA was 0.537. As mentioned in the methodology, an ANOVA was used in this  $N_{diff}$  analysis and not a Kruskal-Wallis, as it did not violate the parametric assumptions according

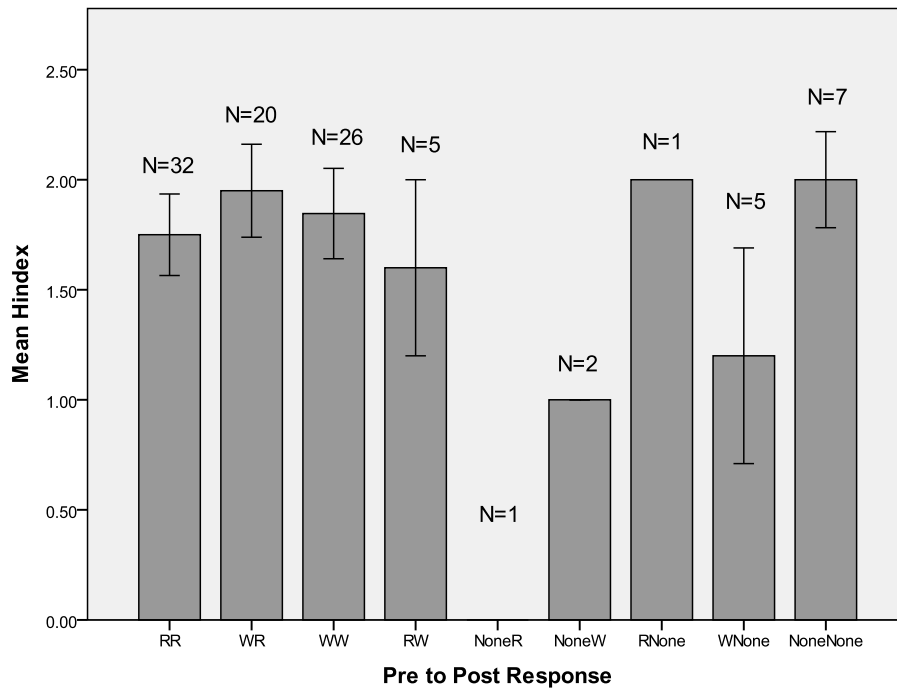


Figure 7.2: Mean h index value for each clicker response type. Error bars are +/- one standard error of the mean. W means the vote was incorrect, R means correct and None was when a student did not submit a vote.

to a Kolmogorov-Smirnov test. Once again the hypothesis was disproved in that students who answered correctly did not use a significantly higher number of different types of technical words.



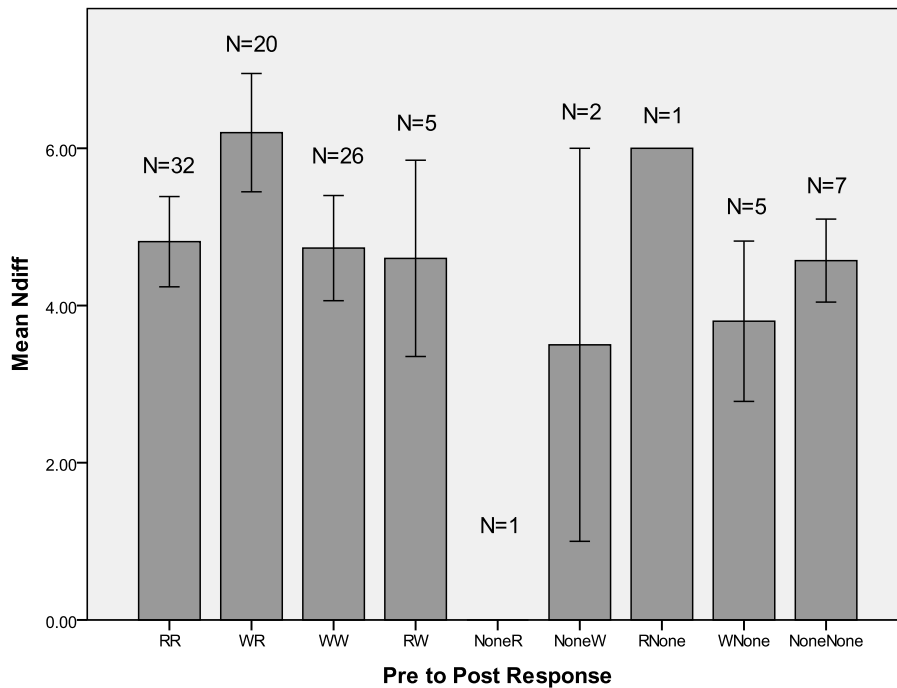


Figure 7.3: Mean number of different technical words used for each clicker response type. Error bars are  $\pm$  one standard error of the mean. W means the vote was incorrect, R means correct and None was when a student did not submit a vote.

There was a large spread in students' responses and their use of technical language. The large variation of technical word use for each clicker response is indicated by high standard deviation values, which ranged from 2.8 to 10.4 words for  $N_{tech}$ . The standard errors, which divides the standard deviation by the square root of the number in the sample, are shown in Figures 7.1, 7.2 and 7.3.

Figure 7.4 shows the mean number of technical words used for each clicker question. As a guide the line shows the un-matched class gain on each question. This shows the disparity between mean technical word use and whole class gain on each question. In other words there was no corresponding pattern between whole class gain on the question and use of technical language. There were only a sample of students contributing to the mean technical word scores, but this adds to the evidence that success on the question does not appear to be related to technical word use.

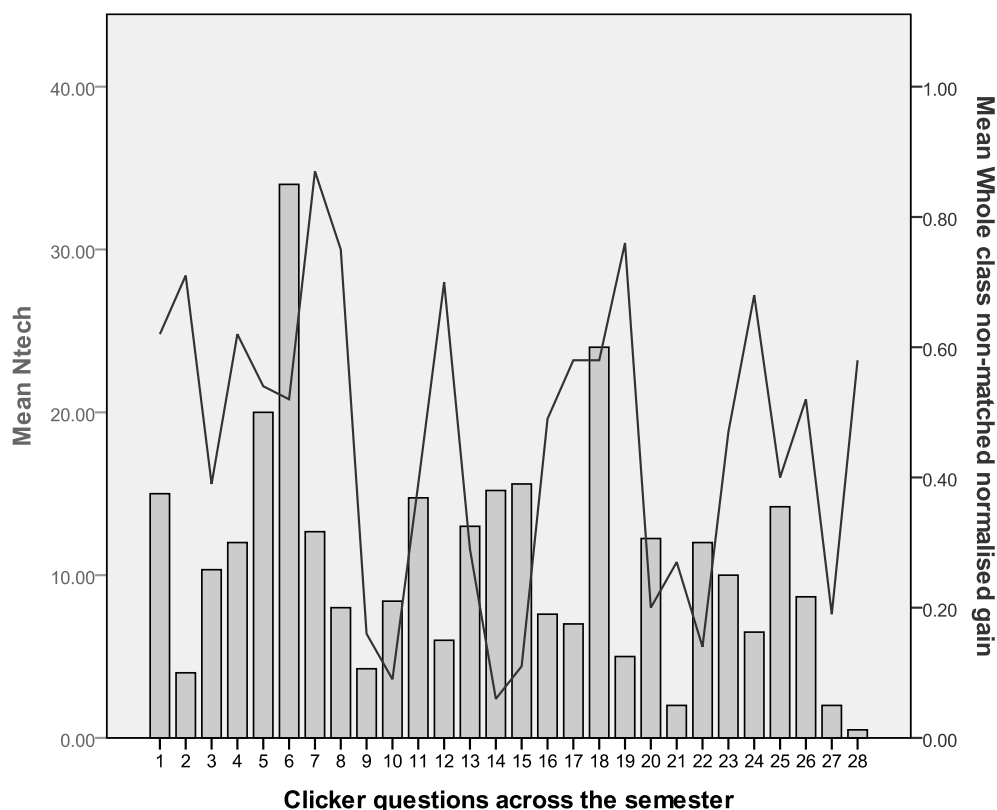


Figure 7.4: The bars show the mean number of technical words used in each clicker question recorded and the line shows the whole class normalised gain on each question. The line is a guide and the graphs are superimposed only for comparison of the peaks and troughs.

Figures 7.1 to 7.4 consider technical word use during the semester on the relevant clicker question. Figure 7.5 was plotted to study the relation of the use of technical words to final mark on the course. This was to investigate whether students needed time for learning between responding to clicker questions and using those concepts to answer questions in the exam.

A 2-tailed Pearsons correlation gives a Pearsons correlation coefficient of 0.14 and  $p=0.16$ . Although not significant, this shows a trend in the expected direction, that is that the more technical words used in PI discussions, may tend to be connected to better performance. The lack of significance again suggests that technical word use does not correlate with success.

To summarise, it seems technical language use is independent of the success for the question, as both WR and RR results for  $N_{tech}$ ,  $h_{index}$  and  $N_{diff}$  are not

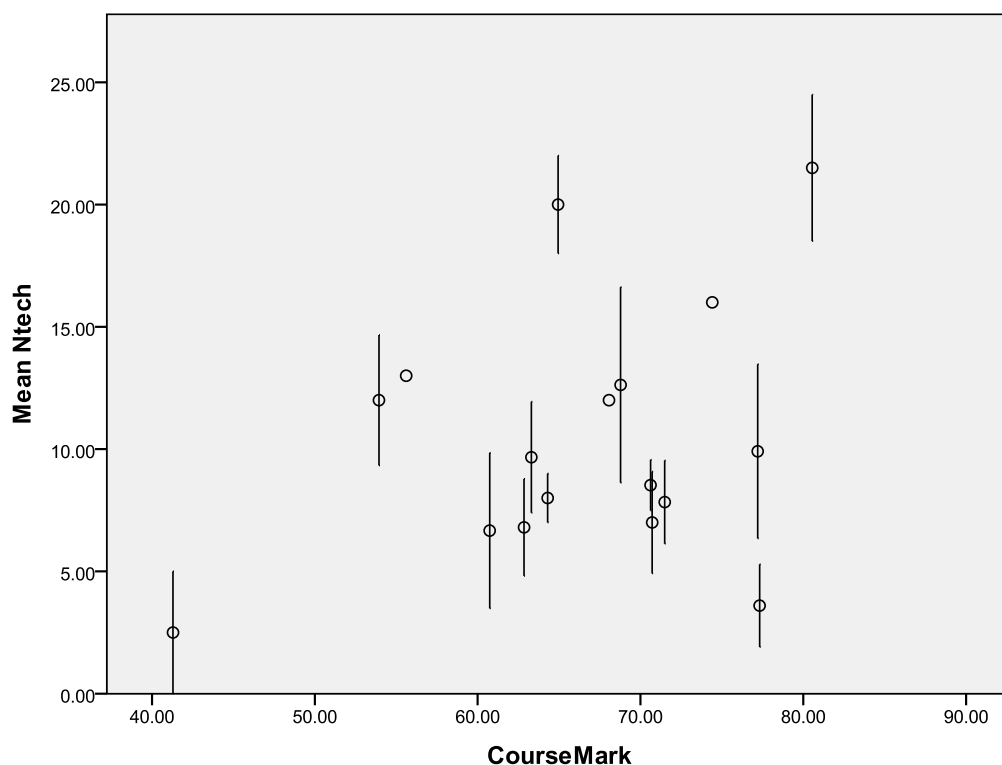


Figure 7.5: A plot of the average number of technical words used for each final mark in the physics course.

significantly different from WW or RW responses. Looking at the discussions in more detail verifies this result where there are many examples of RR and WW with high and low technical word use. Some students were correct, but did not use a large  $N_{tech}$ , and others were incorrect and did use a lot of technical language, meaning that technical language was neither demonstrating a correct response (as an effect), nor causing a correct response to occur. In the next chapter a selection of conversations are considered in more detail to see if there were other characteristics related to success.

## 7.5 Discussion

In agreement with Mazur [64], this translation of a concept into their own language to explain it, may be key to successful discussions in PI. Under the transmission model, which has been encouraged in PI [64, 71], students were

able to convince others of the correct answer, or be convinced, without the need for consistently high technical language use. Interpreting the results using the constructivist model, students discussed ideas together in their own language, without the use of more technical language in successful conversations. Perhaps they were shaping their learning together in a way that they could understand it. In some cases this might be through a greater use of technical language and in others less. Neither learning model demands the high use of technical language.

The results indicated that frequency of technical word use does not provide a means to characterise success of conceptual discussions. Discussions to understand and answer conceptual questions are thought to improve students conceptual knowledge [69]. Therefore technical word use is not related here to success in building conceptual knowledge structures [37, 39]. It was expected that good conceptual knowledge structures would help students in the open book exam, and use of technical language does not appear to be a way to assess this. That is consistent with the results from Figure 7.5 where there was no significant correlation between success on the course exam and use of technical language. However, the exam questions were different to the conceptual questions posed in the clicker questions and so students may have required different skills and use of language to answer them.

The use of technical language indicates that students were discussing the concepts in their own language (whether that was higher or lower use of technical language) regardless of their response to the question. This is related to the community of practice [118] in which the students operate, and shows the importance of this developing community, which is not yet at the level of experts. This community of practice is further exemplified by the differing frequency of technical words used, the lowest being 0 technical words in a conversation and the highest 41 words. The range of conversations indicated the real and authentic nature of the conversations collected and also shows the range of conversations used to discuss physics concepts in this cohort.

The use of technical words does not seem to be related to success, but students should still be encouraged to use and be comfortable with technical language. In fact some believe that a thorough understanding of a topic only comes from a mastery of the words used to convey it [61, 63, 121]. Wellington and Osborne said that:

*learning to use the language of science is fundamental to learning science* p.6 [61]

The use of technical language and scientific reasoning are important graduate skills [1, 54]. These authors proposed that low use of technical language suggests incomplete understanding, regardless of the post discussion vote. Students should be encouraged to use technical language, whilst ensuring they understand the meaning of the words and the content of the discussion. Work in student use of language in Physics at University has described [121] and shown [63] student difficulties in distinguishing everyday meanings of words, such as ‘force’ and ‘momentum’, from their specific meanings in physics. This may be what is occurring with some students in this study, however as their understandings of the words were not probed further it was hard to tell whether they were using the words in an ‘everyday’ or ‘physics’ context. In this case student understanding of the technical vocabulary used was unclear, and specific definitions that students had on certain words were not investigated. It is possible that the students in this study had, or were at least, developing a reasonable grasp of the language used, especially in mechanics, as unlike the students in the American studies [63, 121], these students had studied physics at school to a reasonably high level to achieve admission to the course. The students in the class in this study performed well on the FCI post test suggesting they understood and could apply concepts around the technical words used in mechanics.

As the discussions were complex and varied, none of the hypotheses for explaining the students’ use of technical language are ruled out. The next chapter aims to explore this more fully and look at the possible reasons behind correctness and whether it is related to reasoning or other factors.

## 7.6 Conclusions

Technical word use and scientific reasoning are valuable graduate skills [54], but it does not mean that students who use more technical language answer the question correctly. There was no correlation between technical language used and success on the questions after discussion. It seems that successful students use their own language, which can include high or low use of technical language, to discuss conceptual physics questions in PI. Assuming that students who answer correctly

are developing correct knowledge structures, technical language cannot be used as a proxy to identify students who are successfully developing these conceptual knowledge structures. The next chapter explores other possible characteristics of successful and unsuccessful students and discusses the PI episodes in more detail. Smartpens were used successfully in this study and provided a large amount of information on student behaviour during PI in an authentic manner.

## 7.7 Future Research

Future research could investigate whether students were using technical words correctly, as in this study it was unclear whether students understood fully the words they were using. Wellington and others have presented strategies for researching the understanding of language in science [61, 122]. This includes a taxonomy of scientific words, so that instructors are aware of the language that students have to learn. Their taxonomy has four levels: level 1 - naming words; level 2 - process words; level 3 - concept words; level 4 - mathematical words and symbols. As you go up the levels there is a greater level of abstraction. For example, words in level 3 include ‘work’ and ‘energy’, where words no longer refer to physical objects. Looking at the list of technical words from this research, there are words from all levels of their taxonomy. A large number of words are concepts, such as ‘momentum’, ‘moment of inertia’ and ‘energy’. There are also words which Wellington and Osborne classed as ‘non-technical, but widely used in science [122]. These include words, such as ‘maximum/minimum’, ‘initially’ and ‘increase’. Asking students about particular word meanings, similar to part of Farrell and Ventura’s methodology [62], where they asked students the meaning of a word in context, would provide a means to assess their understanding of these words. Therefore, student understanding of their varied technical word use at university could be examined and broken down into a taxonomy of terms would help separate semantics of conceptual terms, such as ‘work’, from general science terms, such as ‘maximum’, to assess where students have the greatest difficulty. Furthermore, listening to students’ explanations of particular words may also indicate the strength of their conceptual understanding and therefore provide an insight into student knowledge conceptual structures, as discussed in Chapter 5. Consistent gaps or difficulties could then be addressed in instruction.

## Chapter 8

# Peer Instruction: Further Analysis

As discussed in the previous chapter, there seemed to be no association between technical word use and the penholders' success in answering clicker questions during Peer Instruction (PI). Exploratory research was therefore required to further investigate PI conversations and study characteristics of success. The aim of this study was to examine and explain the different characteristics of successful and unsuccessful student conversations during PI. This research was intended to be hypothesis-generating, as opposed to hypothesis-testing, to generate insights from the data. If appropriate, these results can be used to create models of student PI discussions, that can be tested with multiple coders. Qualitative research was chosen for this study, as it is one method to generate theory from the data [88, 97, 102]. It is more suitable than quantitative methods for analysing contextualised cases which do not fall into a standard model.

Success in the previous chapter was defined as the penholder voting correctly post peer discussion. However, previous research suggests that some group members do not necessarily have correct reasoning even if they voted correctly [74]. Furthermore, the instructor will perceive that the student understands the concept by their correct vote. It is therefore instructive to determine what happened in the conversations where the penholder voted correctly, compared to those where they voted incorrectly. In this study 'success' could be examined more closely and whether students who were voting correctly were also reasoning correctly.

As discussed in Section 1.7, there have been many investigations into group discussions: classifying them into different types of talk [56]; forms of pupil interaction [50] and strategies to scaffold group work [49]. However very few studies [123] have studied characteristics of success, as according to clicker vote, choosing to look at deviations from a ‘standard’ discussion, based on misconceptions expected by the lecturer [76]. As reviewed in Chapter 6, no studies have been conducted in Physics PI at a UK University.

The literature from Chapter 1 [74, 75] informed the analysis of the data and is incorporated into the discussion of this study to examine student understanding and success in PI, in particular the models of transmission and constructivism, as discussed in Section 1.8.1. The use of technical language, as coded in Chapter 7, is also included in order to add to the discussion and further validate the findings from the previous chapter. The results from the analysis of this chapter provided potential instructional improvements for PI.

## 8.1 Method

### 8.1.1 Qualitative Analysis

The process of analysis is summarised in Figure 8.1. Qualitative analysis of the data followed a primarily thematic analysis approach, as described by Flick [88] and discussed briefly in Section 2.2.2. Firstly, for familiarisation, all the data (162 episodes) were listened to and 120 of these episodes were transcribed. Episodes were randomly selected after dividing them into piles separated by clicker post-discussion vote response, as the aim of the study was to look at conversations with right and wrong post-discussion votes. Transcription was important as it meant the conversations could be easily accessed and analysed. It was also less likely that different analysis of the data would occur each time the researcher read the transcript, rather than listening to the audio. Themes were developed through an iterative process of analysing around 50 of the selected transcripts. The main ‘areas’ within each theme emerging from the data were then studied in detail. The themes are listed below with the reasons for choosing these particular ones.



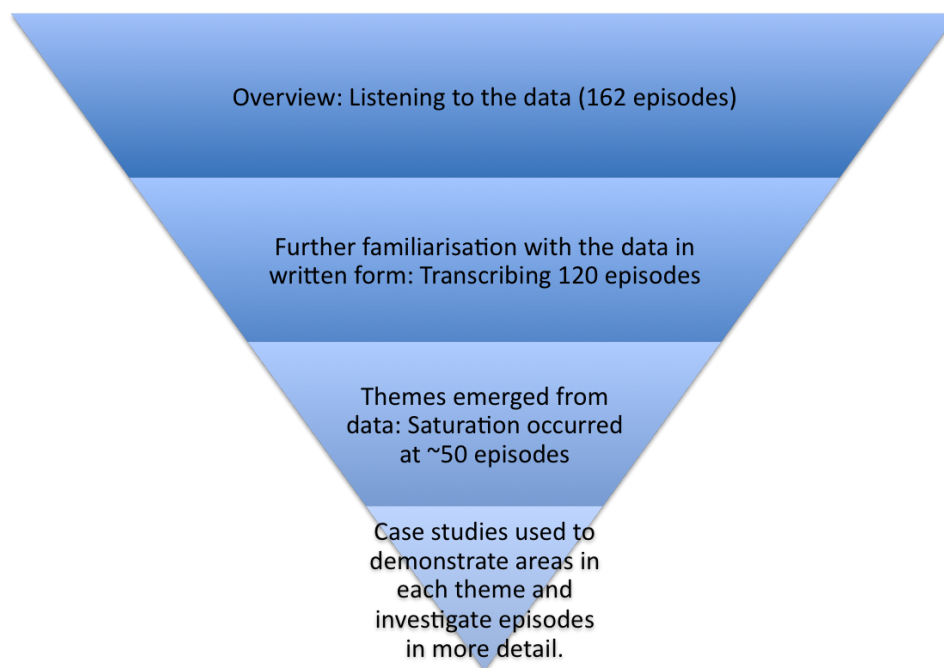


Figure 8.1: Stages of Qualitative Analysis

- **Completeness and correctness of reasoning** - As discussed in Chapter 1, qualitative reasoning is one of the main characteristics of problem solving. The relationship between reasoning and voting option could also be examined.
- **Whether the group was in agreement** - As only the vote of one person was recorded it was necessary, where possible, to keep a track of whether the group was in agreement and therefore whether it was possible that they might have chosen the same voting option at the penholder. Mainly, it could indicate whether the group was in agreement over their reasoning.
- **Transmissionist and constructivist conversations** - Whether conversations were perceived to be exhibiting a transmission or constructivist model of interaction to try to provide further insights into the theory discussed in Section 1.8.1. Under the transmissionist model one student explains the correct reasoning to their peer. The constructivist model allows for students to construct knowledge together.
- **Other prominent emergent features e.g. difficulties in making assumptions** - These characteristics could possibly explain lack of reasoning

and / or an incorrect vote.

- **Other reasons** - for example students not engaging in the process, so not reasoning at all.

Four case studies were selected to demonstrate the main areas that emerged within each theme. Each theme had different emergent areas. Each case study was based on one clicker question and included student conversations recorded during this time (each conversation referred to here as an episode). There were three additional areas which emerged from analysing the data with respect to the above themes, which were not covered in the case studies (for example, lack of confidence), so these are discussed after the case studies (Section 8.4.1). The case studies gave an opportunity to look at a smaller section of the data in depth, to discuss the main areas which emerged from applying the themes above and to expand on students' use of technical language as counted in the previous chapter. The context of each specific clicker question is considered in the analysis as well as possible conceptual difficulties. Results are separated per question, as conceptual difficulties in one question may be different from another.

This exploratory work was not constrained by pre-established codes, such as the problem solving models used in Chapter 4. Although many models of student discussions exist [51, 56, 57, 74, 76] there were several reasons for not using pre-established codes in this study. Firstly, as mentioned above, this is a new area of investigation and it was considered important to investigate potentially new aspects of student conversations with respect to the research aim. Secondly, coding with respect to pre-defined codes loses information on particular aspects of conversations not in the coding model. Finally, many pre-established codes focus on group dynamics or interactions (see Section 1.7), which are not relevant to the aims of this particular study and the thesis overall, which focusses on problem solving and themes relating to qualitative reasoning.

### 8.1.2 Inclusion of Technical Language

In order to tie in work on analysing technical word use, students use of technical words was also included. Students total number of technical words used ( $N_{tech}$ ), the h-index ( $h_{ind}$ ) and the total number of different technical words used ( $N_{diff}$ ) values, as defined in Section 7.2.6, are included in the results. This could

validate results from the previous chapter, where use of technical words was not correlated with success. With this aim in mind the technical word count for each conversation determined in the previous chapter was used, as opposed to being recounted from the transcripts. Only episodes which had high coder agreement of over 70% on the frequency of technical words used were included in the case studies in order to make comparisons to the previous chapter. Although there may be some discrepancies between the two-coder count of technical words and the number of technical words on the transcript, only macro-codes of the episodes were studied here and there were no comparison of the details, such as the total number of words used. A new consideration in this chapter was words students used that were not in the problem statement. Qualitative remarks were made, supported by quotes, to compare students' technical word use to the literature and what emerged from analysing their discussions more qualitatively.

Tables 8.1 to 8.4 present the qualitative reasoning and technical word use per group and the correctness of the pre to post votes per individual penholder for each of the questions shown in Figures 8.2 to 8.5. Penholders are represented by a letter which corresponds to the question and a number to identify different students. For example, A1 represents Student 1 on Question A. Two penholders could have been in the same group, for example in Table 8.1, A1 and A3 were in the same group, but had different voting profiles. Any discrepancies in technical word use between two smartpen holders in the same group, may have either been due to coder error from the method in the previous chapter, or the different position of one student with a smartpen in the group in relation to the other, so that more or less of the conversation was captured by one of the smartpens. Finally, within each group, students quotes are used as examples, these are stated as M1 and F1 to represent male student 1, female student 1 and so on. These are not to be confused with the group names and are at an individual student level. The penholder is not labelled explicitly as A1, A2 etc. in the quotations, but as M or F and where possible it is indicated qualitatively which student was the penholder (though this was not clear in all cases).

## 8.2 Strengths and Limitations

One of the main issues in qualitative results is ensuring validity, for example ensuring that the researcher is coding the right thing in the right way. When areas emerge from the data, the main question on validity is whether these areas are based on the data. This can be demonstrated by its visibility of the area to the reader [88]. With this in mind, areas discussed here have supporting quotes to show the link made by the researcher between the data and the theory. As with the experts study (Chapter 9), one researcher conducted this qualitative, exploratory study. The main conclusions are therefore one researcher's presentation of reality.

With regards to validity of methods, the PI discussions were not directed by the researcher, therefore themes emerging are likely to be student, rather than researcher, generated.

A limitation of this approach is that although group reasoning was related to the penholder's vote through what they vocalised, it was unclear what the penholder was actually thinking. For example, perhaps they were voting correctly, in a group with incomplete reasoning, but actually knew the correct approach without vocalising it. Ideally, and in order for PI to work, students should be vocalising all their reasoning, so that they can learn from each other. From examining the case studies below, in difficult problems all group members seemed to struggle with the problem explicitly, so appeared to vocalise what they knew. However, there was a possibility that some thoughts, essential to the problem, were not vocalised.

Finally, there was no video analysis of students during PI, so data on body language and facial expressions was missing. This did not affect the data discussed in the this chapter, but meant that it was not possible to make comparisons of content discussed to where students were sitting in relation to each other, as Nielsen did in his studies [53].

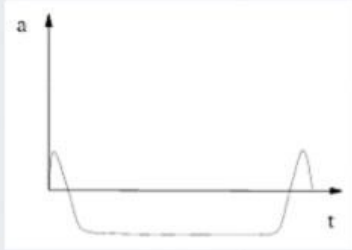
## 8.3 Results

### 8.3.1 Case Study 1: Acceleration of a ball thrown vertically

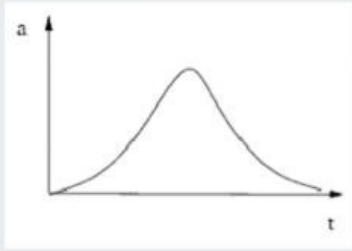
A ball initially at rest in the hand, is thrown upwards, comes back down and is caught.

Which of the following represents a plausible graph of vertical **acceleration** against time?

**1**



**2**



**3**

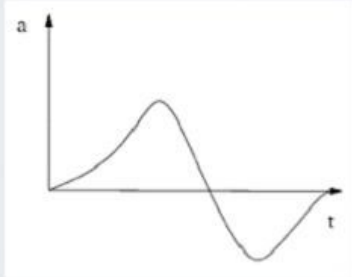


Figure 8.2: Clicker question set for students, examined in Case Study 1. The correct answer is option 1

The question shown in Figure 8.2 had a high whole class normalised gain of 0.62. Table 8.1 shows the technical word use for each group and qualitative comments, based on the themes, which emerged particularly prominently from each episode. The qualitative results are expanded below. Penholders labelled in groups A1 and A3 were working together (i.e. they were in the same group), so

Group	$N_{tech}$	$h_{ind}$	$N_{diff}$	Response PrePost	Conversation
Group A2	26	3	9	RR	Complete and correct reasoning and in agreement, but one student was unsure about peaks at the beginning and end even after discussion.
Group A3	8	1	5	WR	Correctly persuaded. Complete reasoning and general agreement.
Group A1				RR	Correct persuader.
Group A4	18	2	8	RR	Correct persuader with correct argument presented after one student was incorrect.

Table 8.1:  $N_{tech}$ ,  $H_{ind}$  and  $N_{diff}$  for each student on the Acceleration of a ball thrown vertically question. The PrePost response shows how the pen holder voted before and after discussion respectively and the conversation column provides some qualitative notes on these conversations.

are treated together in the analysis.

The correct reasoning was presented in all conversations, with incorrect students changing their minds to agree with this. It appears that enough students knew the answer to persuade the others. Students with incorrect reasoning were explicitly corrected in conversations A1, A3 and A4.

A1 and A3's group used the lowest number of technical words compared to the other groups in the question. One student questioned the peaks at the end of the correct graph which was explained by others. One question was:

*M1: But why would it accelerate right at the end?*

To which the other student replies:

*M2: Because he's caught it, so it's stopping, so the acceleration is up.*

This is followed by further persuasion from M2, as he seemed to clarify the reasoning as to why there were peaks at the end of the graph.

A4's group had a very similar conversation, but incorrect reasoning was presented by a student before being corrected by others. The incorrect reasoning being stated by student M1 was:

*It's just the acceleration is highest at the beginning, and then it's the lowest at the top and then it becomes higher again as it goes down.*

This student therefore suggested option 3, to which student M2 who was the penholder said:

*The acceleration should be constant at minus 9.81 shouldn't it.*

And student M1 replied:

*The acceleration stays the same.*

So the correct reasoning was stated, but it is unclear whether M1 completely understood the concept, as they seemed to eventually agree on the correct reasoning relating to the acceleration, but did not explicitly correct themselves. This A4 group also did not explicitly state why it had constant acceleration (i.e. constant force due to gravity was the only force acting on the object).

In penholder A2's group there was no need to persuade another student in the group of the correct answer, as both students had the right answer. This group also had the highest use of technical language from the three groups recorded. Beyond those presented in the problem statement, there were 7 different types of technical words used, including 'positive', 'downwards', 'gravity' and 'force'. Although they had agreed on the correct final voting option finishing with

*I'm still not quite sure*

in reference to the peaks at the ends of the correct graph. It may be surprising that although A2 had no one to persuade (there was group agreement on voting from the start), they used the highest number of technical words. This is because one student explained it a couple of times and once especially for the smartpen recording. This perhaps shows the students' need for a full explanation of the problem, even after agreeing on the correct answer.

## **Discussion**

In this particular case study these were all standard conversations according to James and Willoughby [76], because students discussed the intended features of the problem and their discussion was reflected in their vote. Some students persuaded others unsure to the correct answer.

As shown in the quotes above, this problem was successful in eliciting explicit student misconceptions in A4's group, in that the graph in option 3 plotted a line

that passed through zero at the top of the motion. Student M1 demonstrated this misconception when arguing that the acceleration changed throughout the motion in the air. Good distractors is one of Beatty et al.'s recommendations for promoting articulate discussion [75], in that it provokes argument and makes students re-think their understanding, as their particular answer turns out to be incorrect. The reasonable number of technical words used during the conversation ( $N_{tech}=18$ ) supports the conclusion that the questions elicited student discussion and misconceptions for the A4 group. However, it is unclear at the end of the discussion whether M1 really understood the correct reasoning or not.

A1, A3's group used less technical language compared with the other two groups. Only five different technical words were used, two of which were in the problem statement ('velocity' and 'accelerate'). The lack of technical language in penholders A1, A3's group could be indicative of at least two things. Firstly, that students were easily persuaded and therefore technical language was not needed. Alternatively, that the lack of technical language could mean that students persuaded their peers more easily when using simpler language. The former is likely to be the case in this question. There was a lot of agreement throughout the conversation, with the one query answered, so the students asking the question seemed easily persuaded. Nevertheless a discussion, in their own words, regardless of how easily persuaded their peers are, is likely to be important in understanding and making sense of particular concepts. It may be however that a more in-depth discussion can occur when there is less agreement. A2's group had two explanations which included technical language, one of which was explicitly for the smartpen. Although the technical language used in the repeated explanation could be discounted, it goes against the aim of the study, which was trying to find a simple counting algorithm that did not require qualitative analysis, therefore it is acknowledged here as an unexpected source of error.

Given the ease of persuasion and the confidence of the correct students in convincing their peers in groups A1, A3 and A4, it appears that the transmission model of learning is evident here. This is when one or two students correct their peers without the need to construct knowledge together. The transmission model was working successfully in this example, as students were correctly convinced (as opposed to incorrectly convinced). This may have been assisted by the higher number of students who seemed to already understand the concept presented,



which was indicated in groups with a smartpen. However there was still a large gain, so there must have been students changing their vote to the correct one, but perhaps there were students with correct reasoning in each of these groups.

That all groups had complete and correct reasoning, with apparent agreement on the right option by the end of the conversation, demonstrates two things. Firstly, in general, conceptual knowledge was clear and complete for most students on this problem, with the main area of difficulty focussed around the peaks at the beginning and end of the graph. It appears that students were not building new conceptual knowledge structures, but refining or reinforcing existing ones. Secondly, the level of the question was straightforward for these students. Many of these students will have covered the topic in school and/or the FCI pre-course test, which supports the conclusion relating to their well constructed knowledge structures. It is worth noting that conceptual change may not occur over the course of one conversation. Student M1 in group A4 is an example of where a student had not demonstrated complete and correct reasoning on the difference between acceleration and velocity by the end of the conversation.

The varied use of technical language for all correct post-discussion voters already validates the finding in the previous chapter, in that students' technical language use is not associated with success on answering the question. Students can vote correctly, but use a variety of technical language depending on the group they are in and the level of explanation they or others need to satisfy their understanding. Unlike this question, many subsequent conversations involved naive groups, where no one in the group knew the answer. The case studies below can provide an insight into what happens in the conversations where no one in the group knew the correct answer.

### 8.3.2 Case Study 2: Rocks Item A

Another question which had a very high whole class gain of 0.75 is shown in Figure 8.3. A full summary of all the groups studied in this question, including the frequency of technical language used is shown in Table 8.2.

Four of the five groups studied here, R2, R3, R4 R5, had incomplete or partially correct reasoning. Although in most conversations there was some discussion of forces and net forces, this was not extensive and in general students did not know which forces should be included on the final diagram. These groups

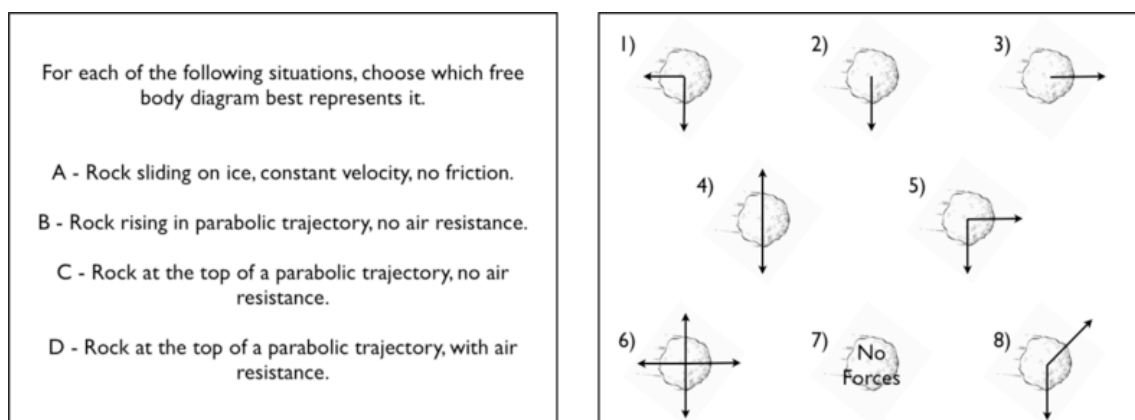


Figure 8.3: The clicker questions set to students, examined in case study 2. Students were asked to answer part A. The correct answer is option 4.

Group	$N_{tech}$	$h_{ind}$	$N_{diff}$	Response PrePost	Conversation
Group R1	16	3	8	WR	Reasoning complete and correct after contributions and discussion from 3+ group members. Group was in agreement.
Group R2	4	1	4	WNone	Difficulties in making assumptions with vertical forces and whether normal force and gravity should be included.
Group R3	3	1	3	WR	Difficulties in making assumptions with horizontal forces and whether there was friction.
Group R4	13	3	4	NoneNone	No agreement on one option, the reasoning presented was correct.
Group R5	4	1	4	WW	Difficulties in making assumptions with vertical forces and whether normal force and gravity should be included.

Table 8.2:  $N_{tech}$ ,  $H_{ind}$  and  $N_{diff}$  for each student on the Rocks Item A question. The PrePost response shows how the pen holder voted before and after discussion respectively and the conversation column provides some qualitative notes on these conversations.

also did not agree on a final voting option. Difficulties in making assumptions about either the vertical or horizontal forces acting on the object seemed to be the reason for their partial reasoning and lack of group consensus. In group R2,

one student said:

*..are we taking into it the normal contact or gravity, or.. hmm it's hard to know really.*

In R2's group there were two potential answers presented at the end of these conversations (options 4 and 6), one of which was the right answer (option 4). These options both have the normal contact force and gravity, which the student quoted above from R2's group was unsure about. This student remained unsure even at the end of this problem.

In R5 and R3's groups it was unclear what the final consensus was for the main group members, but students seem to be divided. In R5's group, when talking about whether there is a contact force or not, one student said:

*It depends on what's being included*

Group R3 discussed the presence of friction.

*yeah say there was friction and there was something like (?). Like say, I don't know, a car going along a road, that would be right.*

One student asked what could be propelling it forward, if friction existed and to keep it at a constant velocity, to which a student responded 'wind'. As one student in the group answered this question correctly it is possible that this was just speculation on what would happen had there been friction included in the question.

In R4's group, students did not come to a consensus, but rather a range of voting options. They mostly discussed whether friction was present. At the end, they did not seem to be in agreement, with one student giving options 1 or 6 "because there are no forces acting apart from friction" and another student opting for voting options 4 or 7.

In contrast to the other groups, R1's group involved the majority of students contributing to the argument and helping each other to the correct answer. They also used a lot of technical language. The entire conversation, with the contributions of each member is shown below:

M1: *Well I said no forces, but it is going to be number 4 isn't it.*

M2: *I said number 7, because it is constant velocity meaning no acceleration meaning no forces.*

F1: *I don't know I thought there could be forces, as long as they were equal...*

M1: *Excuse me (didn't hear)*

F1: *I thought there could be forces, as long as they were equal. Like equal and opposite.*

M1: *yeah..that's a good point too*

F1: *I put 7 as well, but like I thought there could be..*

M1: *but there are no forces acting on it at that point from what he told us*

M2: *gravity acting on it. And there would be a normal force acting on it.*

M1: *that's right*

M3: *did you go for 4*

M1: *yeah 4*

Although all students in the above conversation may be capable of reasoning to the correct answer, not all students necessarily participated in this exchange. For example, student M3 above.

Two groups R1 and R4, displayed greater use of technical language than the others. Table 8.2 shows the use of technical language by these groups. R1's group used 6 different technical words that were not included in the question, such as 'acceleration', 'acting', 'equal' and 'normal force'. Group R4 just used 'acting', 'force' and 'weight' that were not in the problem statement, and used the words 'friction', 'force' and 'weight' more extensively.

## Discussion

Although three out of the five penholders did not vote, or voted incorrectly in this case study, there was a high whole class gain, indicating that many students were answering correctly after PI discussion. While physics reasoning relating to unbalanced forces and acceleration was not discussed extensively, the greater difficulty in this problem seemed to be around making assumptions, such as which forces to include and whether friction existed. Group R3 was the only

conversation out of the four groups that highlighted students' difficulties in making assumptions where the penholder voted correctly. Students' difficulties in making assumptions confirms the need for an explanation of the correct answer by the lecturer after discussion, not just for clarifying understanding, but also for setting the frame of reference of the problem and the standard assumptions made in the course.

Perhaps the lack of technical language could be related to the inability to apply the correct physics model and less complete reasoning. For example, many students presented answers without reasoning (see group R5). For a deeper discussion, perhaps students should be encouraged to use the terms 'normal force', 'gravity', 'friction', 'Newton's second law', 'acceleration', 'forces'. This may also help them use the correct model, i.e. that normal force and gravity apply in this question.

No one in R1's group seemed to initially vote correctly (before discussion) and they spent time adding to each others statements and constructing meaning together. This supports socio-constructivism as argued Smith et al. [71], as this naive group was able to construct the correct reasoning together. This constructivist approach demonstrated in group R1 is in contrast to groups R2 and R5. It was unclear whether R2 and R5's groups fully understood the arguments, as at no point did they talk through their reasoning explicitly and there seemed to be a few conversations going on at the same time. Students also used less technical language.

This question had successful distractors which students used to reason about the question [76]. It provided a way for some students to practice eliminating potential incorrect answers to narrow down their choices. This question also involved the use of interpreting representations and promoting qualitative reasoning, all recommended by Beatty et al. [75].

### 8.3.3 Case Study 3: Friction

This question on friction had a low whole class gain of 0.09 and is shown in Figure 8.4. Only one student answered this correctly after discussion, as shown in Table 8.3. Each group is discussed below.

N2's group lacked reasoning on the intended concepts in the problem and demonstrated a lack of knowledge on the fundamental concepts in the question.

### About the force of friction

Think about the question below. Only one of the answers listed is correct.

We'll do some (but probably not all) of these questions in the lectures when you will all be asked to vote on which answer you think is correct.

Which one of the following is NOT a true statement about the frictional force acting on a block on a rough surface?

- 1** The frictional force is given by  $\mu_k F_N$  if the block is accelerating
- 2** The frictional force is given by  $\mu_s F_N$  if the block is stationary
- 3** The frictional force can be less than either  $\mu_s F_N$  or  $\mu_k F_N$

Figure 8.4: The clicker questions set to students examined in Case Study 3. The correct answer was option 2.

Group	$N_{tech}$	$h_{ind}$	$N_{diff}$	Response PrePost	Conversation
Group N1	0	0	0	NoneR	No discussion
Group N2	5	1	2	NoneNone	Little reasoning relating to the question. Two students lacked foundational knowledge, as to the definition of $\mu$ .
Group N3	12	3	4	WW	Some partially correct reasoning, but spent a lot of time working out the definition of $\mu_k$ and $\mu_s$ .
Group N4	13	3	5	WW	
Group N5	12	2	4	WW	

Table 8.3:  $N_{tech}$ ,  $H_{ind}$  and  $N_{diff}$  for each student on the Friction True or False question. The PrePost response shows how the pen holder voted before and after discussion respectively and the conversation column provides some qualitative notes on these conversations.

The penholder did not commit to a vote on either option or perhaps had forgotten their clicker that day. Compared to others on this question, group N2 used a relatively few technical words, though they did pause the recording just before the end of the discussion time. This is an extract:

M1 *What is the coefficient of friction?*

M3 *..like the number 0.6*

M1 *Ah you see I didn't understand what that meant.*

F1 *me neither, I just don't get it*

and later F1 asked:

*where does that number come from?*

There were three penholders in one group, N3, N4 and N5. Unlike students in N2's group, they knew that  $\mu_k$  and  $\mu_s$  were coefficients, but they were unsure of their separate definitions. They pursued an irrelevant path where they considered if the block was moving or accelerating. An extract to show this follows:

F2: *are the top two like the same coefficient or are they different?*

F4: *No we have different coefficients depending on whether it's stationary or it's accelerating.*

F1: *It is accelerating or is it moving? I just thought, I thought it was the first one, because that was just moving not accelerating.*

F4: *The first time it is accelerating and eh and I don't know if it's stationary.*

F1: *Let's look at a book.*

All three members voted incorrectly both pre and post discussion. The group N3, N4 and N5 used a variety of technical words a moderate number of times, indicating engagement with the question. Yet they did not solve it correctly. They tried to look up a book, probably in the hope that understanding the definition of terms would help in the solution, but they ran out of time.

Finally, N1's group had no audible discussion, but voted correctly after the discussion stage.

## Discussion

All audible conversations in N2-N5 groups align with one of James and Willoughby's non-standard conversation types [76], which may help explain the low whole class gain on the question and lack of progress in reasoning on the question. The relevant code from James and Willoughby [76] is called 'unanticipated ideas regarding fundamental science content knowledge'. It was not anticipated by the instructor that students would not know what the coefficient of friction was, as this was covered, at least superficially, in their pre-reading before the lecture.

N1's group also aligns with one of James and Willoughby's [76] other codes, where students were unable or perhaps unwilling to start a discussion, even though the penholder voted correctly. Possible reasons are that the penholder already knew the correct answer and did not need to discuss it; the conversation was not recorded properly; or they were not near other students to engage in a conversation. Possibly the student had an idea as to the correct answer in the pre-vote, but after further individual thought actually voted in the post discussion vote.

There was confusion over the meaning of the symbols and students focussed on factors, such as whether the block was stationary or moving which were not relevant to solving the question. Moreover a lack of foundational knowledge impeded students' progress on the problem. Perhaps these factors made students revert to a transmission style approach, where students asked their peers directly for information, including what the coefficient of friction was. A textbook was also a source of information, directly consulted where students wished to look up what the symbols in the problem statement meant. Students were unable to correctly construct knowledge together in these cases. Case Study 2 and Smith et al. [71] reason that naive groups can still work their way to the correct answer. This implies that the construction of the correct answer does not depend on someone knowing the correct reasoning (transmission). The fact that these naive groups did not get to the correct answer suggests there could be factors impeding students' progress, such as lack of the required knowledge or the construction of the problem statement, as discussed below. Furthermore, of the conversations where technical language was heard (N2-N4), the only words heard, beyond those given in the problem statement were 'coefficient' and 'kinetic'. This showed a minimal use of technical language by students beyond that given to them in the question. This supports the argument that students were unable to engage in the question and access appropriate knowledge about this topic.

As penholders N3, N4 and N5 were in the same group, the number of technical words counted in the three instances could be compared to check coder consistency on this. Only one additional technical word (the  $\mu$  symbol) was required in conversation N5 to bring the  $h_{index}$  up to three and to make it the same as N3 and N4. 12-13 technical words were counted consistently for all three recordings. The slight variation in results may be explained by the different

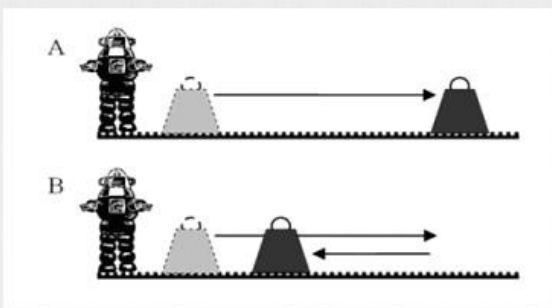


locations of each student with a smartpen. There were occasionally times in the PI recordings where a student was too far away for the smartpen to pick up the conversation clearly. Errors in coding are another possible reason for a lack of agreement in counting the number of technical words used. However, these results were reasonably consistent considering the nature of the data.

Lastly, this question failed to stimulate the appropriate conceptual conversations from students. The question was intended to test students' understanding of friction before the point of slip. One weakness is that it was a negatively stated question, so the correct answer was actually not true. This may have been confusing and created a high load on the working memory, which can only hold smaller amounts of information at a time. It is also possible that the symbols in the questions activated a formula-based approach, as opposed to conceptual reasoning. The question is also flawed in the sense that there is a point where frictional force does equal  $\mu_s F_N$  (on the point of slip), which would eliminate option 2, however students did not appear to pick up on this or use it to inform their voting.

### 8.3.4 Case Study 4: Work done by a robot

A robot can move a heavy weight by applying a constant force to push it along a rough surface. The figure shows two paths along which the robot can push the weight. On which path is the work done on the weight by the robot the greatest?



1 Path A

2 Path B

3 Both are the same

Figure 8.5: The clicker questions set to students, examined in Case Study 4. The correct answer was option 2.

The question in Figure 8.5 had a very low whole class gain of 0.06. As with

Group	$N_{tech}$	$h_{ind}$	$N_{diff}$	Response PrePost	Conversation
Group W1	21	2	6	RW	Partial reasoning. Discussed as displacement not distance, but do not discuss force as a vector.
Group W2	13	2	6	WW	Similar to above, incorrectly persuaded.
Group W3	23	3	8	WW	One student presents correct reasoning, but their peer did not agree.
Group W4	3	1	2	RR	Not much discussion heard
Group W5	16	3	7	RR	Engagement by one student, but not the other.

Table 8.4:  $N_{tech}$ ,  $H_{ind}$  and  $N_{diff}$  for each student on the Work done by a robot question. The PrePost response shows how the pen holder voted before and after discussion respectively and the conversation column provides some qualitative notes on these conversations.

the previous case studies, Table 8.4 shows  $N_{tech}$ ,  $h_{ind}$  and  $N_{diff}$  used by each group and qualitative comments on their discussion. The penholder's pre and post discussion responses are also given.

Both W1 and W2's groups had partially correct reasoning. Their conversations were spent deciding on whether the symbol 'd' in  $W_d = \vec{F} \cdot \vec{d}$  was distance or displacement. Students decided that path B was correct if 'd' represented distance, whereas defining 'd' as displacement was linked to a correct choice of path A. There was no discussion of force as a vector, or that work done is a scalar quantity. There was also no discussion of the dot product and the resultant scalar quantity because of this. In W1's group it was decided that:

*d is displacement not distance.*

which meant that they voted for the incorrect path A. Force was mentioned, but not discussed further:

M1: *is it the same force?*

M2: *I don't know looks like it.*

In W2's group one student tried to teach another. The transmission model was demonstrated, but the persuading student argued for the incorrect voting option, as shown in the quote below.

F1: *is energy a scalar?*

M1: *yes*

F1: *so is it B?*

M1: *No*

After this part of the conversation student F1 clarifies her previous thinking. M1 asks her the definition of work done. Applying light touch discourse analysis here, it is apparent that this student is trying to teach F1 and not ask a question for an unknown answer. Student F1 defines work done as force times displacement, and then agrees on path A. This is contrary to M1's definition of work being a scalar, as quoted above, as considering work done as a scalar would immediately lead to path B. In this naive group they all incorrectly agreed on path A. One student insightfully said:

*That happened to me once, everyone around me had the same answer and it was wrong.*

A participant in W3's group presented the correct reasoning (M1), but the other student (F1) did not seem to be convinced:

M1: *you can't get negative energy*

F1: *no, but you can get negative force*

M1: *yeah but then it would be a negative force and a negative displacement..be happening*

F1: *I don't know, I just don't know*

M2: *what would happen when you bring it back, it will be a negative force and negative displacement*

M1: *I'm sticking with B.*

F1: *Aw I still think it might be A*

Perhaps M1's lack of depth and sophistication of reasoning, even though correct, was why Student F1 was not convinced. They later talk about the results in terms of the 'clicker game', which is discussed further in Section 8.4.1. W3's group used a large amount of technical language, with 7 different types of technical words

that were not included in the problem statement. These words included ‘positive’, ‘negative’, ‘displacement’ and ‘distance’.

In W4’s group the correct voting response is presented by one student (Path B), with agreement from another. However, after this they discussed the symbol ‘d’ in the same manner as groups W1 and W2.

M1: *I’m not sure if it is displacement or distance..is important.*

Unfortunately their subsequent discussion is inaudible. Left at this point it was unclear whether they had full reasoning even though they answered correctly pre and post discussion.

W5’s group demonstrated a new area, which was a one-sided conversation, where one student was unwilling to engage in a conceptual discussion. Although technical words were spoken by the student trying to initiate the discussion, there was at no point conceptual discourse or equal reasoning [76]. This extract shows the deliberate lack of engagement from M2:

M1: *work done is force over the distance.. you’ve done displacement, that has a further oh yeah that has a bigger displacement part A then part B but..that’s gone a further distance.*

M2: *Yeah I know your argument it’s just I don’t think that it’s distance or displacement we’ll see. We will just have to wait and see.*

M1: *Or we could just ask someone*

M2: *No just wait and see, because they could be just as wrong as you, they will be just as wrong as you...or just as right as me, or just as wrong as me, it is the very truth.*

The lack of engagement of M2 is analysed in the discussion below.

Use of technical language can indicate what students were or were not discussing during these conversations. Groups of W1, W2 and W3 used a moderate to high number of technical words compared to the other groups, but all penholders voted incorrectly. None of these groups used the words ‘vector’ or ‘dot product’. Both groups, W1 and W2, were heard saying the word ‘scalar’ once each. W1’s group said the words ‘displacement’ 11 times, ‘distance’ twice and ‘force’ four times. W2’s group said ‘displacement’ five times, ‘distance’ twice and ‘force’ only once. Supported by their lack of use of technical language and the

qualitative analysis, in groups W1 and W2, the argument was incomplete opposed to incorrect. Furthermore, in groups where the penholder voted correctly both pre and post discussion, there was still no mention of the ‘dot product’. W5’s group used the word ‘vector’ once, though one student was not engaged. All five conversations used the words ‘displacement’ and ‘distance’ a number of times even though they were not in the problem statement.

## Discussion

W1’s group was a standard conversation according to James and Willoughby [76], because they discussed a distractor from the voting options in their conversation and voting reflected their discussion. W2’s group seemed confused, one student defined work done as a scalar, but as they also defined ‘d’ in the equation for work done as displacement. They finally decided that path A was the correct answer. Perhaps these students have not yet consolidated their learning on this topic, but they were trying to think about it more deeply than more distance meant more work.

The transmission model did not work in W2’s group, as student F1 was persuaded incorrectly. The transmission model, where one student teaches another, has its weaknesses. The student in W2 accepted the incorrect reasoning too easily, whereas F1 in W3’s group did not believe the correct reasoning of her peer. Some skills in discussion may have been useful for student F1 in W2’s group, so she could have argued further with the incorrect persuader, for example, to point out the incorrect logic of having work as a scalar, but the answer as if it was a vector. A certain degree of conceptual knowledge is required to argue in this way. Nevertheless, there is a degree of foundational understanding that first year undergraduate students would be expected have from previous study, such as the definitions of a vector and scalar. Thinking critically may have helped them resolve their reasoning. In fact Osborne [58] suggests teaching arguing and critiquing skills to students.

Other groups attempted to construct knowledge together as shown by discussions on whether ‘d’ was distance or displacement (W1’s group). These were differentiated from the transmission style, as there was not one student explaining an answer confidently to others. As many of these students were participating in naive groups, where none of the students initially knew the correct answer, it

showed that they did not always construct knowledge together correctly.

Students W1 to W3 had incorrect responses, but their reasoning was categorised as incomplete rather than incorrect. Firstly, these groups were engaged and on-topic. Groups of W1 to W3 used a reasonable amount of technical language from within and out-with the problem statement. Technical language use can perhaps indicate the specific physics students are engaging in to answer the question. All three of these conversations were fully engaged in reasoning, but incomplete, because discussion of force as a vector was missing. The infrequent use of the term ‘dot-product’ and the word ‘force’ compared to the word ‘displacement’ in most conversations supports the incompleteness of their arguments regardless of response type. This may mean that students were trying hard to apply what they had learnt to this question, but it is important that the instructor completes this reasoning by explaining the concept after the post-discussion vote. Furthermore, asking students to use certain technical words, such as ‘dot product’ and ‘vector’, that were missing in these conversations, would be an interesting test to see if this encourages a complete conversation.

Comparing incorrect to correct post-discussion responses was difficult, due to the nature of conversations by groups W4 and W5. In W4’s group the inaudible parts of the conversation make it difficult to know whether the group had only partial reasoning, or completed the remaining reasoning in the inaudible section. In W5 there were no conceptual contributions from the other group member. In W5 the disengagement of one student from the PI process did not make it a successful interaction. It is not clear why the student held the distrust in himself and his peer. Perhaps he thought it not worth doing, or that the correct answer could only reside with the lecturer and not with a peer. This could mean he was in Perry’s dualism stage of intellectual development where answers are right or wrong and students receive the correct answers from authority [60]. It is clear from both groups W4 and W5 that a correct response does not necessarily mean complete and correct reasoning.

Finally, there was a reasonable amount of technical language used in all incorrectly answered and partially reasoned conversations, which once again validates the results from the previous chapter that technical language does not correlate with success. However, there was clearly success in initiating engagement and discussion on the relevant ideas. As discussed above, use of

technical language may indicate engagement on the question and point to specific areas of weakness, such as the lack of a discussion on force as a vector. This would not be suitable for a time-limited analysis technique, but could be something that instructors could listen for when walking around the lecture theatre.

## 8.4 Summary

A summary of the case study results is shown in Table 8.5.

Case Study 1	Most students knew the correct approach and easily convinced others
Case Study 2	Students had difficulties in framing the problem successfully by deciding on the important factors in the question and what could be neglected. They also had difficulties in making assumptions, for example, whether to include the normal force or not.
Case Study 3	Students lacked foundational content knowledge to start the problem
Case Study 4	Partially developed conceptual understanding. No complete and correct reasoning heard.

Table 8.5: Summary of main findings from the four Case Studies

### 8.4.1 Other Categories

As mentioned above, there were three main areas that emerged from applying the themes for analysis that are not explored in the case studies above. These categories were: lack of reasoning, the clicker game and lack of confidence. These are discussed below, as they were general enough to discuss without specific examples relating to a certain questions or correctness of response and occurred in a number of questions and conversations.

#### Lack of Reasoning

Lack of reasoning was one of the categories that emerged where students did not discuss the concepts in the question. Lack of reasoning is closely related to lack of engagement in that in some cases there was lack of reasoning, because students were not trying. Other potential reasons for lack of reasoning could be a lack

of understanding, lack of group communication, off-topic for the whole time or inaudible. An extract showing lack of engagement in a PI conversation is shown below.

M1: *So we got.*  
OFF-TOPIC by others in group  
OFF-TOPIC by M1 as well  
M2: *what did you vote for?*  
M1: *eh 5, its the right answer believe me*  
M3: *we are meant to be talking about the question*  
M2: (inaudible)  
F1: *what did everyone put?*  
F1: *3?*  
M2: *yeah 3*  
F1: *3*  
OTHER group members talking in background inaudible  
F1: *we all have to vote again*  
OTHER group members talking in background inaudible  
M1: *oh come on people*  
OTHER group members talking in background inaudible  
END

Student M3 knew what they should have been doing, as he says “*we are meant to be talking about the question*”. Although votes were presented, there was no discussion about why students thought it was that option, hence the lack of engagement and reasoning. Lack of reasoning and other categories presented in this section are discussed further in Section 8.5.

### **The Clicker Game**

Although infrequent, there were other reasons that students used to select their votes or discuss voting options. One of these reasons is referred to here as the clicker game, which occurred in both correctly and incorrectly voting groups. The clicker game usually included students choosing a vote for strategic reasons, either to beat their peers or improve their overall score on all the clicker questions. The clicker game could occur for the majority of the conversation or just at the end



after some conceptual discussion. For example, after some discussion on topic, this group said at the end:

- M2: *I saw ..(student's name) vote 2 so I'm not going to vote the same*  
M?: *be quicker (?) teacher*  
M?: *oh no I went for 3 last time*  
M2: *he is spreading his votes, that's how he is planning to win.*  
*(name) is planning to win by spreading his votes.*  
M3: *he does that all the time, he is tactical (?) It's not what a true*  
*champion would do*  
M2: *not a true champion, stick to your guns*  
M4: *he is strategic, he is strategic*

A label of 'M?' means it was unclear which student was speaking at that time. Another slightly different example, is demonstrated below.

- M1: *I hope you're wrong*  
F1: *why because then you're right*  
M1: *yes*  
M2: *well one of us..*  
M1: *I would even mind if it's 3 as long as you're wrong*

These students had already decided on their vote through their conversation (including reasoning), but made it a competition of who would win. It seems that one student did not mind being wrong, as long as the other student was also wrong. Perhaps this is so they would finish with the same number of correct votes by the end of the process. Some students also discussed their overall clicker scores, which were stored online, which was also considered part of the clicker game. The clicker game has parallels with game theory, where the choices of each decision are weighed up to think strategically about winning or gaining the advance over your opponent. Game theory is a whole topic of research in itself [124], and beyond the score of this work. However, this shows the competition that students have with each other and themselves (to create a good clicker scoring profile) whether consciously or not. The clicker game was obviously important for some students and occasionally influenced how they voted. Unfortunately in some cases their votes do not reflect their reasoning and if correct, may have

mislead the instructor and possibly the students themselves. This may have been an unexpected negative effect of the clicker leader board in which students could anonymously see how many questions they had answered correctly compared to others.

### **Lack of Confidence**

Lack of confidence was found in the episodes with an incorrect post-discussion vote, though it may emerge from other voting outcomes. Attitudes and beliefs in problem solving is considered important in theory [12] and instruction [22]. In this example, one student was put off by a previous question that she got wrong. At the end of the discussion when the participants seemed quite unsure, she says:

*F2: I don't know though, I did this something similar, somewhere else  
and I got it wrong, but I don't know what I did*

Students were not usually put off by having done a similar question, in fact they used them as a means to convince others, or be more certain of their answers. This is an example where this has not worked, and the student is confused from doing this question in the past.

## **8.5 Discussion**

One of the main aims of PI is to improve students' conceptual understanding and it is assumed that reasoning relates to understanding of the intended concept. These case studies provided an opportunity to look in detail at PI conversations from this perspective. Ideally students would understand the concept in more detail by the end of the PI process, but should this understanding and therefore reasoning be complete? Many conversations did not demonstrate complete reasoning even though the participants voted correctly. The disparity between voting choice and understanding was also found by Neilson and Stav [74]. A problem arises if some participants with partial reasoning, and therefore potentially partial understanding, vote correctly and think that they understand the concept from seeing that they voted correctly. This reinforces the need for an instructor explanation of the answers after discussion.

This research showed there were factors which were related to partial reasoning; students got stuck making the assumptions for the question (rocks Case Study 2); lacked the expected knowledge for that question (friction Case Study 3); could not access the required concepts to complete their reasoning (work done Case Study 4); or did not have the confidence to proceed. The findings demonstrate James and Willoughby's types of non-standard conversations [76], for example 'unanticipated student ideas regarding fundamental science content knowledge' p.126 [76] is similar to the discussions in the Friction question. This shows that the reasons for partial reasoning are more universal than this study.

The above factors involved in partial reasoning fall into two of the knowledge structure categories set out by Gerace [37] and shown in Figure 5.1. Difficulties in making assumptions can be related to problem state knowledge, as the students did not know how to apply their physics understanding to that problem. Conceptual difficulties, such as lack of foundational knowledge and the inability to access the required concepts (case study 4) can relate to the conceptual node. This means that PI is therefore involving more than just students' conceptual reasoning, as it also covers problem state knowledge on some questions. It is complicated, as lack of conceptual knowledge limits discussion, but is also what PI is trying to improve. Osborne concluded that:

*essentially that the limits on student's capability is attributable to their lack of knowledge rather than their reasoning capability* p.465 [58].

Areas of weakness in students' conceptual knowledge structures are different from factors such as lack of engagement and reasoning. With lack of reasoning it was hard to tell if there were any conceptual or problem state difficulties. Additionally, Nielsen et al. [125] said that students did not feel like they needed to 'win' the argument. The majority of conversations were in agreement with this research, with the exception of a few, where the clicker game was apparent.

Taking knowledge and reasoning as connected there is a question as to whether reasoning 'should' be complete and correct as students try to understand potentially unfamiliar concepts for the first time. It seems from these case studies that there is room for students to construct and discuss knowledge together, but it did not mean that all students presented full and correct reasoning. This may be the ideal from a constructivist perspective when students co-construct knowledge together, but this is perhaps not so applicable from the transmission

model perspective when one student explains the complete and correct reasoning to another. It is not surprising that students have partial reasoning in that they are grappling with potentially new concepts for the first or second time.

There were many types of naive groups in this study, for example students constructing knowledge correctly together (see Case Study 2, group R1), students that were partly correct (see Case Study 4, groups W1 to W3) or students that discussed the wrong part of the problem (see Case Study 3, group N3). Perhaps given a different question on the same concept students would have subsequently learnt from their discussion. Smith et al [71] found that incorrect groups after discussion were often correct on a subsequent, individually attempted, isomorphic question. This is only indicative, as they established that naive groups existed statistically rather than knowing the responses and reasoning of each group, but suggests that learning can continue after the PI process is completed.

It seems that developments in conceptual understanding can be made by beginning to tackle these concepts through short physics problems, such as those presented in PI, as student often changed their vote to the correct option and discussed topics that they may not have already thought about. This is supported by the case studies above and whole class gains. This is where PI has its strengths. Students start reasoning with concepts with their peers [49] as early as possible in the course (the actual lecture). Though, as students may not fully understand the concept even if they voted correctly, the lecturer needs to provide and explain the right answer afterwards to reinforce the correct conceptual model. Ultimately the aim was to encourage complete and correct reasoning, but perhaps not expect it, and minimise factors that might lead to partial or incorrect reasoning. Potential solutions for these factors involved in partial reasoning are discussed in the following instructional strategies section.

### **8.5.1 Intellectual position of the student**

Students have to accept that they or their peers are able to work towards the correct answer during PI discussions, in order to engage properly in the process (unlike W5's group in Case Study 4,) [60]. The authority for solution does not lie with the lecturer or a textbook during this discussion phase, which some students may find difficult to accept. Therefore in considering learning theories like transmission and constructivism it is also worth considering the intellectual

position of the student [60]. In both the transmission and socio-constructivist models it is assumed that the student is willing to believe that his or her peers can reason correctly, or that the group is able to construct knowledge together correctly. This means accepting that the final authority on whether the answer is correct no longer rests with the lecturer. Students who hold authority figures as keepers of the correct knowledge, operate in a position of duality [60]. To these students it is unlikely that they will believe that their peers will lead them to the correct answer.

The intellectual position of the student can be considered with regards to the implementation of PI. Applying Perry's theory to PI indicates ways that PI should be implemented from this perspective, as it is possible that students in the introductory years will be operating in the dualist mode. This would mean that even if they engage with the process they do not fully believe what the correct solution is until verified by the instructor. This reiterates the necessity of the instructor confirming the correct solution at the end of the process, even if most of the class have got it correct. This confirmation will allow students who do not trust the reasoning of their peers to consolidate and confirm the correct answer and reasoning by hearing it from an authority figure.

### **8.5.2 Technical word use**

These findings are consistent with the results from the previous chapter. Technical language use is not correlated with success in answering the question, nor does technical language seem to be related to the correctness of discussion. There were conversations with a high use of technical words, but an incorrect vote (for example, group W1). Students may have used a large amount of technical language to only partially reason. Conversely some successful voters were in groups with low technical word use (for example group A3). However, the type of words used may provide information on which physics concepts groups are or are not discussing. The case studies examined here show that students can have a variety of conversations. These results confirm that students reason in their own language to create understanding in introductory physics. Nevertheless, it is still important for students to develop an appropriate use of technical language, so that they can reason in a scientific manner as this is expected of them later on in academic and non-academic careers. The findings from the case studies indicate

that perhaps PI is not the appropriate place for explicit instruction on technical vocabulary, as students are learning through discussing new physics concepts in their own language.

### 8.5.3 Instructional Strategies

Potential instructional strategies became clear after factors that influenced the conversational reasoning were determined. From considering the results and discussion above, some speculative instructional strategies are discussed here to work on resolving issues related to partial or incorrect reasoning.

To assess in the lecture whether students are having difficulties interpreting the question or making assumptions the lecturer could listen to student discussions during PI. The results from above indicate that the instructor could be listening for multiple factors, such as: whether students are discussing the intended concept, misconceptions, use of appropriate technical language, ability of students to reason and critique others, the key assumptions made in answering the question and their understanding of the question. This could also provide the instructor with an insight into the quality of reasoning or level of agreement. Crouch et al. [70], Beatty et al. [75] and Turpen and Finkelstein [54] all emphasise listening in to students' PI discussions. Each technique to engage with understanding student discussions has its weaknesses. If the lecturer listens or joins in, this can only cover a couple of groups in each lecture and could disturb their conversations. If a team of teaching assistants (TAs) circulated the lecture theatre there would be better coverage in terms of help given and less disturbance of conversations (as perhaps TAs are less intimidating than the lecturer), but they are unlikely to cover all groups and this may not be financially viable.

One reason for incomplete reasoning was students' inability to engage with material which they had already learnt. Technical language indicated what material students were not engaging with to answer the question. For example, the dot product was not mentioned in Case Study 4. A potential method to encourage full and complete scientific reasoning, could involve presenting students with a list of technical words with all the terms they should use in that problem. This would possibly encourage discussion on the 'dot product', 'force' and 'vectors'. A disadvantage of this technique is that it may provide too much help for students, so this scaffolding should be taken away eventually.

The prevalence of non-standard conversations, in Case Study 3 on friction may be improved by two possible interventions. Removing nonessentials, as recommended by Beatty et al. [75], by re-wording the problem statement, would have assisted in students conversing on the intended concept. Secondly, a warm up question on the basic knowledge regarding static and kinetic coefficients would check understanding before applying them to a question.

These results emphasize the need for a feedback loop in PI, where unsuccessful questions, with very low class gain, are revised. The factors mentioned in this work and others [75] may help this process.

Finally, as discussed above, clarification of the correct solution appears to be an important stage in the PI process, so should not be bypassed, even if the majority of the class have voted correctly.

## 8.6 Conclusions

The development of conceptual knowledge through reasoning with peers is one of the main intentions of PI. Students attempted to understand a concept explicitly with their peers, which both provided them with an opportunity to reason conceptually, but also placed importance on conceptual reasoning as a problem solving stage.

PI engaged most students in reasoning, whether this reasoning was correct and complete or not. Reasons for partial explanations included difficulties in interpreting the question, lack of required knowledge and difficulties in making assumptions. There were other factors associated with voting, which seemed to influence students' responses, including the clicker game where students were competing against each other to answer correctly rather than basing their votes on a conceptual argument. There was evidence of both transmission and constructivist models in PI conversations, but further research is required to determine if one is preferable over the other.

PI was not necessarily sufficient for students to reason correctly about a concept, even if they answer the question correctly. A natural conclusion from this is that clarifying the solution post PI is an important stage. There were also some factors which affected the way students reasoned, which may have been mitigated by changing the instructional approach, such as changing the phrasing

of the problem statement. This demonstrates the importance of a PI feedback loop, as emphasised by Beatty et al. [75] and listening to PI conversations to check students are on track.

### 8.6.1 Future Research

Although a difficult and time consuming area to study, future research could establish and refine a set of characteristics of successful and unsuccessful discussions, which would then be tested with multiple coders. Instructors listening to conversations could then be better able to differentiate successful and unsuccessful traits. However, it is likely that data will be more complicated than a clear, mutually exclusive taxonomy where the characteristics of successful and unsuccessful discussions are separated. It will also depend on the definition of success and whether that is a correct voting response post-discussion or complete and correct reasoning. The next stage would then be to implement and test changes in instruction that try to mitigate for factors in unsuccessful conversations. This would help determine whether minimising factors associated with unsuccessful conversations could help students reason correctly.

A separate area for future research in PI would be the investigation of the additional skills that students are developing, beyond physics conceptual understanding, for example group work, communication and critiquing skills. This could show how students develop these skills and perhaps how to further encourage their use. In order to do this a similar study to the one carried out in this chapter would need to be conducted. The data analysis would be based on demonstrable skills that students are developing in group work and communication. Showing the additional skills that students are developing could show that PI does not just improve conceptual understanding, but other skills that students will need after University, such as communication and group work.



## Chapter 9

# Expert Group Problem Solving

The use of groups in physics instruction is one of the main developments resulting from PER research and there have been many studies and observations that develop teaching tools for students solving problems in these environments [3, 4, 64, 65]. Students construct knowledge together in their zone of proximal development, at a higher level of operation than if they were solving the problem individually [41]. As discussed in the literature review, there have been interventions to improve group work [59], analysis of the types of talk students have [56, 57], and investigations on the construction of groups with respect to gender [52]. However, there is a lack of literature on experts working in groups to solve physics problems. Individual expert behaviour has been extensively researched for a long time, especially in making comparisons to novice behaviour on similar activities [16, 20, 126, 127]. See Section 1.3 of the literature review for some examples and discussion of these studies. One major constraint exists, which is that what would be classified as a ‘problem’ for a novice, may just be an exercise for an expert. This may be especially pertinent in groups of experts where they can work on harder problems together [41]. This was taken into consideration in the methodology of this study.

There are two ways that expert behaviour can be applied. In the review by Nokes, Schunn and Chi [16], they state that researchers use either absolute or relative methods to study expertise. Absolute methods focus on an in-depth analysis of a high-level expert, whereas relative methods involve the comparison of an expert and novice when completing the same task. In this chapter an absolute measure of expertise is assumed. The expert group were high-level experts (all

lecturers or professors) which justifies this approach. In this way their behaviour is compared to results from other chapters in the final discussion.

The aim of the study is to examine the macro-behaviour of experts and postgraduates working in groups. These behaviours can then be analysed for prominent characteristics, and compared. Comparisons can also be made to the undergraduate behaviour examined in previous chapters and speculations can be made on how to decrease the gap between expert and novice behaviour. There were a number of research questions in this area:

- What do experts do when put in groups and how does this compare to what PhD students and Undergraduate students do?
- Do experts follow the ‘expert’ stages presented in the problem solving models, such as the Minnesota model [3]?
- How do individuals within the group operate?
- Does the group split into sub-groups?
- How do experts use resources, such as the Internet or textbooks?
- Finally, what suggestions can be made from these observations, with respect to developing the teaching of students in group environments?

The abundance of research questions reflects the lack of research in this area of physics education research. This chapter aims to start answering these questions and develop an insight into expert behaviour in this environment.

## 9.1 Method

Lecturers and PhD students were invited to attend a problem solving workshop-style study. Workshops involve students working on physics problems in groups of around 3-5 students. Purpose built workshop rooms have tables designed for group work, with one computer and a white board per table. The computer can be controlled locally or by the instructor at the front of the room. These types of workshops are used with students in all years of their undergraduate degree at the University of Edinburgh, but especially at the Introductory level. An incentive for participants was to gain a perspective into student experiences

in workshop classes by participating themselves. The ‘expert’ group refers to a group comprised entirely of academics, that is lecturers, readers or professors. The PhD group consisted of three PhD students.

The methodology of the study progressed as follows. At the beginning of the session they were told not to ‘role play’, but behave normally. This was to try and observe expert behaviour in a student setting. The experiment was conducted in the same specially designed room as a normal student workshop. Two teaching assistants (TAs) were present to answer questions from either group when necessary. The groups were informed of the details of the study and assured of anonymity in any form of report. Each of the eight participants used a Livescribe smartpen [93] and notebook. Individuals within each group were asked to start recording at approximately the same time, which made coding their behaviour easier. The expert group was also recorded by a video camera to show body language and group dynamics. Finally, a couple of days afterwards all attendees were asked to fill in a feedback form about what they had learnt and any reflections on the session. This feedback is not the focus of this report.

## 9.2 The Questions

Participants were given two questions similar in terms of physics topics to those given to introductory physics students. The questions were made more challenging than introductory level to ensure that these would not be merely exercises for the experts, thus satisfying the definition of ‘problem’ as discussed in Section 1.1 of the literature review. Unfamiliar problems were sought, so as to minimise the likelihood that participants had seen the problem before and the possibility of participants spotting the solution straight away. This was because past experience may be an important factor in expert behaviour [29]. Novices on the other hand, in the same setting, with their comparatively limited practice in physics, are unlikely to be able to appeal to past experience.

The questions were presented on the students’ virtual learning environment WebCT, which is how students normally accessed and viewed questions in workshops. Each question came with a hint which was given if clicked on. After the designated tutorial time was up, the answers were revealed with full solutions. The questions are given below, with full solutions in Appendix D.1.

Questions one and two are shown below. The first question below was taken from a standard introductory course textbook by Halliday, Resnick and Walker p.302 [103]. There was a diagram in the book, but the groups were deliberately not given this image to see how they constructed the problem situation and diagram and to make the problem more challenging.

1. A small 50g block slides down a frictionless surface through height  $h = 10\text{cm}$  and then sticks to a uniform rod of mass 100 g and length 40 cm. The rod pivots about point O through angle  $\theta$  before momentarily stopping. Find  $\theta$ .  
HINT: You can assume that there is no energy lost in the collision of the block with the rod.

2. GRS 1915+105 (also known as V148 Aquilae) is a double-star system in the Milky Way Galaxy. One of the stars is a regular star but the other is a black hole surrounded by an accretion disk. The black hole is a microquasar, emitting intense X-rays and opposing jets of sub-atomic particles along the axis of rotation of the accretion disk. The speed of the ejected particles with respect to the black hole is 92.0% of the speed of light,  $c$ .  
Determine the speed of one jet relative to the other.  
HINTS: The speeds are a large fraction of the speed of light in a vacuum,  $c$ , so this is a relativistic problem.  
The speed of light in a vacuum has the same value in all directions and in all inertial reference frames.

## 9.3 Method - Question 1

### 9.3.1 Development of the coding scheme

Experts' behaviours from solving question 1 were coded with a coding scheme which is discussed here. Categories to code the data were initially created based on the Minnesota Model [3] (see Section 1.6) and those that emerged naturally from the data. Although an external model was used as a basis for analysis this was not used exclusively with additional codes included as they emerged from the data. This was important for a number of reasons. A rigid coding scheme restricts coding of the data to pre-determined areas, requires a second coder for verification and is insufficiently flexible to accommodate group behaviour as

distinct from individual behaviour. The data was considered at a macro-level, so details of specific behaviours were not included.

Each participant was coded for activity separately and the behaviour marked on a sheet for the times it was observed. For example,

0-1.40 - Off task,

1.40- 3.46 - Making sense of the question

and so on. These were then rounded to the nearest 5 seconds to put in a table of activities, such as the one shown in Table 9.1. Each participant was coded separately, but analysed together. Colour coding indicated the individuals working together. For example, if two group members were coded as procedural group work and the rest were not, then the different colour of these two pens during the time would immediately show this. The video of the expert group assisted this analysis. Each participant had a smartpen and so when they were not talking it was clear whether they were working on paper, for example on procedural activities, such as working through equations.

The first set of categories were created from listening to the expert group and were marginally added to when coding the PhD group. Two new codes emerged from studying the PhD group complete the first question, which were:

- Getting set up - Looking for the question on the Internet and discussion around getting set up and logged in, and
- Reflection- Reflecting on anything to do with the question task or as a result of doing the task.

At a later point the coding scheme was adjusted to include ‘Evaluation of final answer’. Evaluation was used as a code only when participants explicitly evaluated whether their final answer, or sub-answer, was sensible. It was not included otherwise, as it was often difficult to distinguish group work and comparing answers to evaluation. In order to maintain the macro-level of the coding some sub-categories of different types of evaluation, such as “is that what you got” were left out of the coding scheme. The codes were constructed to maintain the distinction between group work which may have had evaluative aspects and evaluating whether the final answer was reasonable.

Intra-rater coding was conducted on one participant, a year later, to validate the results. After re-coding one participant (labelled as Pen A) there was 91%

agreement with the year before. The discrepancies were mainly in transitions and the original codes were used as the default as they were coded with the video camera as well. The final coding system is shown in Table 9.1.

Category	Criteria
1. Off Task	Anything unrelated to solving the problem. This could be off topic conversations, or discussions on where the question is located.
2a. Getting set up	Looking for the question on the Internet and discussion around getting set up and logged in.
2. Worksheet focus	Reading the question and the instructions on the worksheet
3. TA focus	Any interactions with the Teaching Assistant
4. Making sense of the question	Understanding what the question asks and what it means, includes translation, focusing on the problem and defining notation.
5. Discussion of physics concepts	Working out the main principles and concepts that can be applied to the question. Exploring possible strategies.
6. Procedural work - Group	Applying the concepts to the problem to work through a solution, by working through tactical and strategic logistics in a group. Obvious group activity between more than one participant. Acting on the strategies whilst consulting each other.
7. Procedural work - Solo	Applying the concepts to the problem to work through a solution, but working through it independently, with no consultation with others. Acting on the strategies individually.
8. Giving or receiving help	Asking for, giving or receiving explicit help
9. Evaluation of answer	Explicit evaluation of the final answer or sub-answers and whether they are sensible. Does not include comparing answers with others or questioning strategies, as these were marked as procedural group work.
R. Reflection	Reflecting on anything to do with the question, task or as a result of doing the task.
NS. Not Sure	Unable to establish this person's behaviour.

Table 9.1: Categories for coding expert and post-graduate groups

### 9.3.2 Error in coding

There were two possible sources in error in coding. One error emerged from the smartpen of each member of the group being approximately 1 second out of sync with other group members. This was minimised by starting the pens at the same time and re-adjusting those that were slightly off, by finding a syncing point. Another source of error was when the time recorded was rounded up or down to the nearest 5 seconds in order to present the coding on a table. As macro behaviours, which usually last a couple of minutes, were the focus of the study these time errors were not considered major.

## 9.4 Method - Question 2

Question 2 was not coded for problem solving behaviours for a number of reasons. For two of the participants in the expert group question two was not a real 'problem', as they knew immediately which approach to solve it. This meant that it would not give an indication of problem solving behaviours on a true problem. Question 1 successfully provided an insight into expert problem solving behaviours, but as question 2 was not a true problem, more research was needed to confirm these findings. Therefore, different analysis was conducted on the second question for the expert group only. The expert group used resources to solve question 2, this meant their behaviour could be compared qualitatively to the introductory level students studied in Chapter 5. Question 2 was analysed qualitatively, with the same research aims as in Chapter 5 with regards to participants' use of resources such as a textbook or the Internet. These were:

1. Ability to recognise an appropriate underlying physics approach,
2. Ability to look up information,
3. Ability to apply this information correctly.

And specifically with regards to resources use:

- How many separate times the students accessed resources.
- Which resources were accessed e.g. was it the Internet or the textbook.

- The type of consultation the resources was used for e.g. checking an equation.
- Why they consulted the resource e.g. trying to find a similar problem
- Whether the information looked up was useful and helped them progress in the problem.
- Whether they were able to use the information they found.

This question was transcribed, as the group mostly worked together and mainly only one person spoke at a time. Although the PhD group completed question 2 and used resources, they were not used in the analysis, as there was no record of how and when they accessed the resources. For example, it was clear from the recording that they looked at a textbook and the Internet, but as there was no video of this group, it was unclear who was looking it up. When they searched the Internet there was no record of what they typed in, compared to what they were talking about typing in. Although speculations could still be made about this data, it was considered beyond the scope of the study.

## 9.5 Strengths and Limitations

A strength of the study was that subjectively question one was sufficiently difficult for all the experts. None of the participants seemed to ‘see’ the solution straight away and it took both groups over 30 minutes to complete. This satisfies the aim of setting a problem, rather than an exercise.

However one problem when comparing expert to novice behaviour in this thesis (for example to Chapter 5), was that experts were working in a group, whereas novices were working individually. Perhaps the expert group performed better, or at least differently, than they would do working individually [41]. This has to be kept in mind when making any comparisons to the novice students.

Due to the lack of another available coder, this study was exploratory and taken from the stance that one coder could still gain valuable insight into the behaviours of experts working in a group. The results are therefore the interpretation of the researcher and have to be taken as such. This method was similar to other studies in this discipline [123]. With respect to the final coding



scheme, separating individual and group coding did not give details of who was working with whom. This meant the group could split into two sub-groups and according to the coding still be working as one. It also does not include the extent to which participants were working together due to the subjectivity in coding this. For example, an in-depth discussion had the same code as asking a question and getting an answer, but not being actively engaged discussing the task. Finally, these codes only apply to the data they have been derived from. New codes would need to be developed for different groups participating in the same type of activity (for example with a novice group). The codes here however will form a good basis for initial analysis and reflect what occurred in the participating groups.

## 9.6 Results

### 9.6.1 Macrobehaviours - Experts, Question 1

The graph of the expert group's macrobehaviours is shown in Figure 9.1 and qualitative comments to support these behaviours provided below. The expert group tended to go through the Minnesota model linearly, waiting until each stage was completed satisfactorily before moving on and finally answered the question correctly. A large portion of the time was spent confirming the concepts to be used.

Starting with the first main code in Figure 9.1 of making sense of the question, the group made absolutely sure that they had interpreted the question correctly. Towards the end of this section of interpreting the question Pen D checks this with everyone:

*so has everybody got what they think is.. a realistic diagram?*

They agree with this and then work out an agreed notation. After being sure that they all had a working diagram, they began to think about which concepts would be applicable to the problem. This section starts with:

*Pen D: right so let's identify the principles, as we've been told to do*

*Pen E: so it's frictionless, so we can use conservation of energy when it's sliding down the slope*

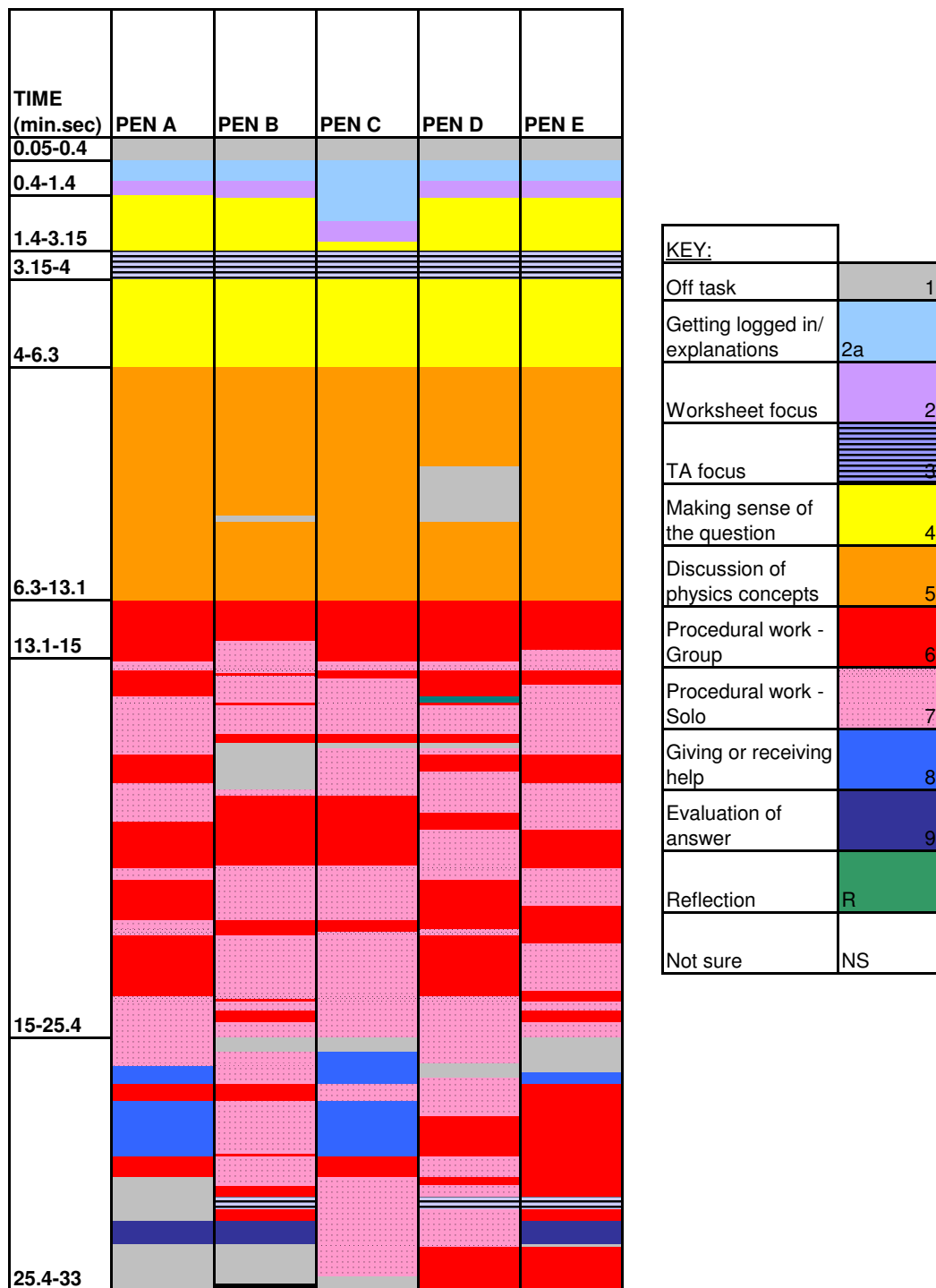


Figure 9.1: Macrobehaviours of the Experts on Question 1

The expert participants questioned different approaches, the assumptions required and even the hint given in the question. An example of this sort of critical thinking is the following comment from one participant below. This was not their final approach, but part of the discussion on the required concepts.

*I would argue that you conserve momentum not energy during the collision*

To which another expert replies:

*but that's not going to help we need to use conservation of angular momentum during the collision*

The participant making the initial comment agrees with this.

After a critical discussion of how to approach the question, the expert group agreed upon the correct approach. The solution to this question is shown in Appendix D.1. Once the expert group agreed on the approach they started to split up into individuals or sub-groups for the procedural stages. In other words, once they had decided to use conservation of energy, they worked through the equations individually. Some sub-groups were formed when one participant turned to the other to check their working. Participants checking working or using comparisons of equations occurred frequently, whether with the whole group or in sub-groups, and a lot of the red bars in Figure 9.1 show this. For example Pen D asks the group:

*so has anyone got an expression for the gravitational potential energy*

To which one participant replied and then wrote it on the white board. The equation on the white board was discussed and corrected by the group.

The blue bars at the end indicate participants helping or explaining a part of the solution to each other. After waiting for all group members to complete the question, they analysed whether their final answer was reasonable before moving on. The grey sections indicate when participants were off-topic and mostly likely these participants were finished or nearly finished.

Finally, planning is a prominent code in the Minnesota Model, but did not arise explicitly from group behaviour in these studies. A lot of planning-type behaviour was incorporated into the substantial time spent deciding on the appropriate concepts to apply. Planning more than one step ahead did not seem

to occur, but planning next steps, such as “so let’s identify the principles..” did occur. As in Chapter 4 stating immediate next steps were not marked as planning, as these were unlikely to reflect strategic decisions regarding the overall solution process.

### 9.6.2 Macrobehaviours - Post Graduates, Question 1

The graph of the post graduate group’s macrobehaviours is shown in Figure 9.2. The post graduate group also arrived at a correct approach, although their problem solving stages were less clear than the expert group. This group spent longer getting set up, and less time making sense of the question than the expert group. They mis-interpreted where the pivot point was placed, which was corrected by a TA later in the problem, during the discussion of physics concepts stage. The reoccurrence of the yellow, making sense of the problem after TA help reflects this moment. Similar to the expert group they spend a long time discussing the physics concepts before doing any implementation. Before writing down equations they confirmed the general approach.

Once an approach was agreed the participants worked through the procedural work individually at different speeds. The PhD group also checked working with each other, for example:

*Pen G: so have you got your speed after the collision as that?*

*Pen H: Yes*

Similar to the expert’s macro-behaviours, one participant helped another towards the end. One participant finished sooner than the others and was off topic for the rest of the problem. This group did not evaluate their final answer, but did check sub-stages with each other, as discussed above.

There were many similarities with the expert group, but there were more interruptions and more help from the TA for the PhD group. They also seemed to be more aware of being recorded.

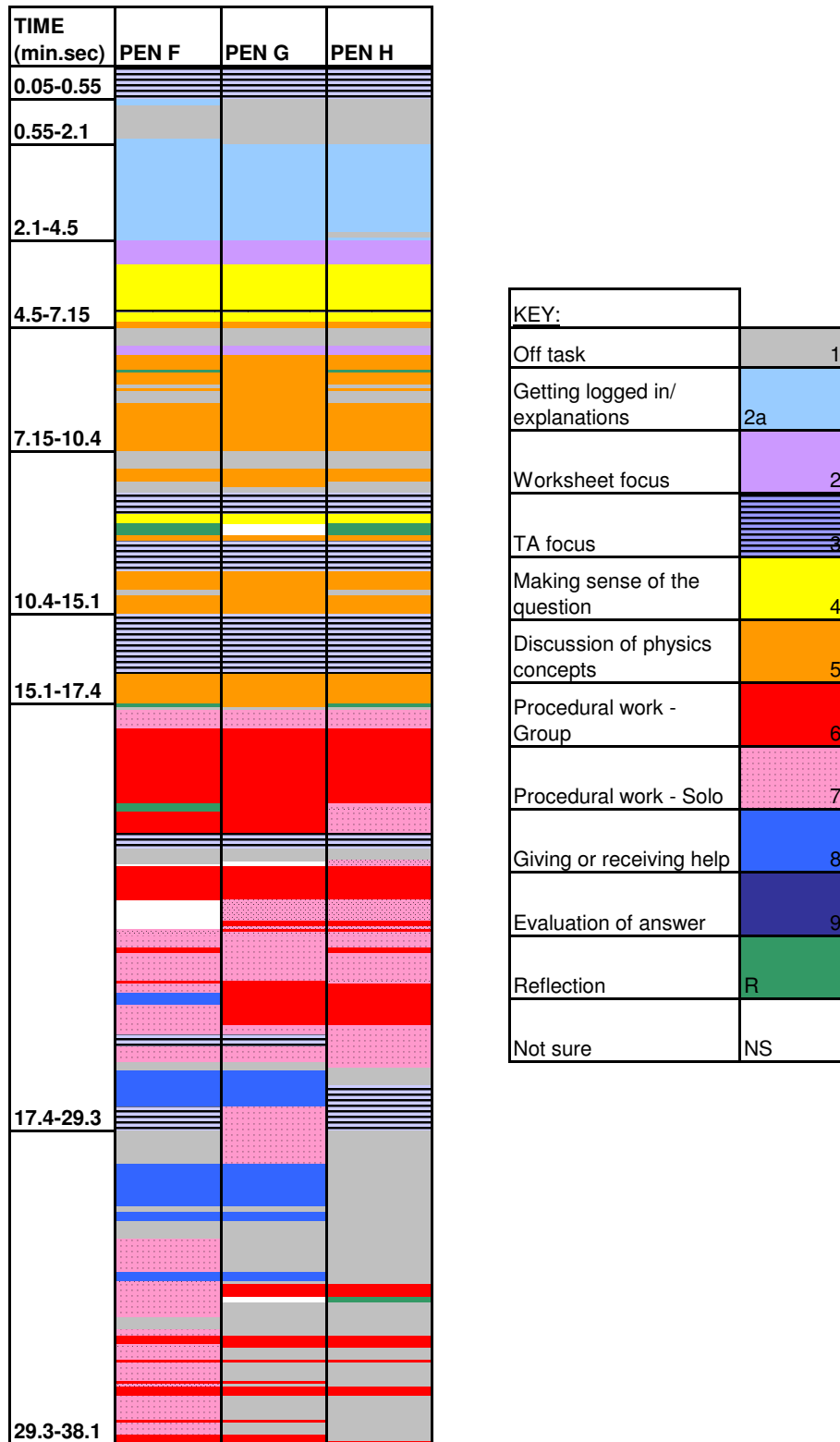


Figure 9.2: Macrobehaviours of the Post Graduates on Question 1

### 9.6.3 Resources Use - Experts, Question 2

The expert group solved this question correctly having looked up and applied the required information successfully. Participants in the expert group knew what information to select from the question and which concepts applied. No one could remember the equation, but they knew exactly what to look up.

*Person A: all you're doing is a transformation right*

*Person E: it's just velocity addition problem*

*Person A: yeah, right exactly, yeah and I just can't remember the equation*

They decided that a textbook would give this information. They could have searched online, as there was a computer at their table with Internet access, but the Internet was not mentioned at any point.

*Person A: so shall we go get a book, 'cause that will give us the equation*

*Person D: yeah, go get a book. [Person E: yeah I'll go Halliday] Lorentz transformations. Velocity additions*

They then spent some time figuring out what each symbol represented. After this part of discussion they were in consensus as to the notation in the equation.

*Person E:  $u$ , is the velocity of the particle as measured in frame  $S$ ,  $u$  dashed is the velocity as measured in  $S$  dashed, the moving frame*

*Person A: so are we solving for  $u$  dashed then*

*Pen D: yeah we do, yeah we do. We've got, we've got  $u$  and we've got  $v$*

*Pen B:  $v$*

*Pen D: because these are both in the inertial frame*

*Pen E: yes. It says determine the speed of one jet relative to the other, so we need to get in the inertial frame, so I think we want  $u$  dashed [Pen D: yes].*

*Pen A: yeah*

*Pen D: yes*

*Pen B: both jets have the same velocity yes*

There were a few problems in applying the Lorentz equation from the textbook. When two group members got zero as a final answer they realised they must have been doing something wrong. This showed evaluation and Person A re-read the textbook and said:

*Person A: we've got to be careful of the signs haven't we. In other words if we transform, if we are transforming in a positive direction we've got to make.*

*Person E: u and v have opposite signs*

*Person A: they do I think that's right*

*Person E: yes*

This was the moment they realised why their method had not been working previously. Therefore evaluation was followed up with an adjustment to the approach. The correction to the application of the equation led to the correct answer without any further major problems.

## 9.7 Discussion

Characteristics which emerged from the expert group are explored below. The problem solving behaviours are in agreement with problem solving models [16], even though there were a few new codes created that were different from the pre-established Minnesota model. Both expert and PhD groups were attempting problems that they did not initially see how to solve, so it is interesting that, with the exception of planning, both groups worked more linearly through the Minnesota model than anticipated (see Section 5.8). The expert group followed the more linear process in terms of recommended problem solving stages on question one. This started with the construction of the problem space and therefore visualisation of the problem. Discussion of the physics followed making sense of the question. The expert group spent a long time on qualitative reasoning, as can be seen by the orange bar in Figure 9.1. This has been considered as an individual expert trait [21, 22, 23] and seems here

to apply to group work too. The experts argued and reasoned their opinions very thoroughly and upon choosing one, made sure they completely understood why it was applicable before using it. The chosen concept was implemented individually or in smaller fractions. Sub-groups can be viewed from the coding (Figure 9.1). The individual procedural stages occurred at different times, but all expert participants worked individually at some point. In fact it was found in the first question that both groups did a substantial amount of work together, but then they split up especially when at the procedural stages. This indicates that group work is not always about working together. Individual stages were necessary, because equations had to be worked through in order to get a numerical answer. In this sense group work is preparation for individual problem solving. Development of procedural skills is also important, especially in an environment where group members can compare sub-stages with each other to check correctness and save time. The problem solution finished with verification and helping fellow group members.

This linear approach in terms of the problem solving model may have been due to the presence of Physics Education Researchers in the group who either knew the recommended approach and what they should be doing, or had perhaps internalised these approaches and did them naturally, though the model was not provided to them at any point. However, the non-linear nature of the Minnesota Model is supported by experts' use of evaluation. Although the experts were only coded with evaluation at the end, they were in a process of constant critique and questioning throughout the solution process. This means that although the Minnesota Model was followed linearly, there was a high-order of control throughout the solution process. Participants would not move on until they had explicitly mastered each stage to their satisfaction, such as confirming the appropriate principles were applicable, before thinking about how to apply them. It is possible this type of evaluation with peers (coded as group work) may be due to the solution being solved in a group, where they were able to constantly compare and question other group members.

The PhD group had some similarities to the expert group. They initially worked together and then individually in the procedural stages. The problem solving stages were less linear, for example they went back to 'making sense of the question' after discussing the physics concepts. They had more 'TA focus'



than the expert group. The PhD group did not evaluate their final answer, but did the same process of checking sub-answers with other group members. This type of checking may again be a product of being in a group. Their solution process could be viewed as becoming expert-like, but with more diversions and TA help.

In question two, a couple of group members recognised the correct underlying physics approach at the beginning of the solution and although they were unable to remember the relevant equation they knew exactly what to look up and found it easily. Procedural skills were also important. Their first application of the Lorentz equation was not correct, but after evaluating their final answer, the textbook was consulted. Its application was adjusted to take into account the different signs in the equation and they were able to reapply the information to reach a correct final answer.

### 9.7.1 Coding and the Minnesota Model

The Minnesota Model was designed for context-rich problems where there may be too much or too little information to solve the problem or a problem statement which does not always state the unknown variable [3]. The problems in this study were slightly different in that they were less based on real-life and the unknown variable was clearly stated. However, they were not simple exercises and the problem solving stages were useful. The method of creating a temporal coding map, mostly based on pre-established codes, worked well for question one. Coding could be done for each person, straight from the recording. The flexibility of being able to add new codes meant that behaviours which did not fit into the Minnesota Model were accounted for. This included splitting codes into individual and group work.

Another additional code was ‘Not sure’, where it was unclear what participants were doing. This seemed to occur more with the PhD group, perhaps because the PhD group were not also recorded by video. It was clear from the video of the expert group that some time was spent looking at other people’s work or listening in, even if not actively participating in the conversation. This will not have been picked up in the PhD group.

Planning was the only code of the Minnesota Model not to be explicitly found in either group. This does not mean that experts do not plan. There are a

couple of reasons why planning was not found. Planning was perhaps internal and second nature to the experts participating so that no one vocalised it. For example, they may have an intuitive sense on how to proceed on problems by applying a conceptual approach and working through the equations to solve the problem. It could also have been due to the nature of the questions. The questions were not ill-structured and the strategic decisions were relatively simple once the appropriate concepts were applied. Tactical decisions had to be made, as to the implementation of the approach, but no planning was needed in determining what to do overall. For example, after finding the appropriate physics concepts it was clear that they needed to implement this approach.

Finally, tactical decisions were needed in procedural stages to keep the solution on track. For example, on question 2 the expert group realised that they needed to evaluate their procedural work when they initially got zero as their final answer. On question 1, one member of the group wrote an equation on the white board which was then corrected by the group. Small adjustments to procedural work were also made as members of the group compared answers to each other, showing tactical control.

### **9.7.2 Recording Technologies in Group Work**

A short aside can be made regarding the recording technology. Providing each individual with a smartpen in a quiet room and video recording the group was considered a successful way of collecting data, as it made it easy to analyse what each group member was doing. There were no other groups near by, which made it a clear and audible way of recording, especially when compared to a lecture theatre, but it was less authentic, as it did not reflect a real life learning situation.

## **9.8 Conclusions**

Although this was a small scale study it provided many insights into expert behaviour whilst solving real problems. The behaviour exhibited by the group of experts was a linear progression through the problem solving model. They started with fully understanding the question, followed by a thorough discussion of the applicable physics concepts and the assumptions underlying them and finished with implementation and evaluation. It showed the coverage of all the steps

to a high level of completion. The concepts were fully understood before being applied and the procedure was checked and completed with evaluation of the final answer. However, the expert group did not establish the correct approach straight away and work towards the answer seamlessly, but instead went to and fro on the appropriate concepts and down a few dead ends in the procedural work. This, in a sense, is true problem solving, and expertise was demonstrated through their approach to the final and correct answer via a rigorous process and a level of strategic and tactical control [24]. It seemed that the PhD group were becoming expert-like, as similar patterns can be viewed in their coding and could be viewed as a point on a spectrum of developing expert behaviours. Solo work seems to be a necessary part of group work, as each member of the both groups carried out the procedural work individually. All group members arrived at an answer, having been supported in the solution process by the group, including discussions on appropriate physics concepts and the application of these concepts (e.g. checking sub-answers during the individual implementation stage). Conclusions of comparisons to novice behaviours are summarised in the final conclusions in Chapter 10.

## 9.9 Future Research

Further research into expert behaviour when solving physics problems in groups or individually would be informative. In order to compare to students at all levels in the University the same question could be used to show any similarities or differences in solution processes for different levels of education. The most difficult part of this process would be finding an appropriate question that would be a problem for all participants. Perhaps a non-intuitive question, such as the one chosen by Singh [29] and discussed in Section 1.3 would be appropriate here. Furthermore, a future study would ideally look for participants with no experience of Physics Education Research, to see whether they would follow the problem solving stages so closely. Similar data collection to this study would give clear and easy to analyse data. Where possible groups should also be video recorded, as more information is collected regarding their solution process and group work.

# Chapter 10

## Conclusions

### 10.1 Summary

This thesis has provided insights into student problem solving in various contexts. Student expertise whilst problem solving individually was considered in the individual problem solving chapters 3, 4 and 5, where students ‘thought aloud’ whilst solving multi-step, contextualised physics problems. Successful students had simpler transition maps between problem solving behaviours than unsuccessful students. They did not use a fewer number of transitions overall, but used fewer different *types* of transitions, so their approach was considered more directed. The qualitative results supported the quantitative findings; successful students knew the correct concept to use early on in the problem and applied it confidently, whereas unsuccessful students, with or without resources, struggled to identify the correct approach. The results suggested that a reasonable level of conceptual knowledge and procedural skills are required before successful problem solving is possible, even if students are given access to resources, such as the Internet. Furthermore, a link was shown to knowledge structures as a potential model to explain students’ problem solving behaviour. If students had good knowledge structures then information from resources could be easily accessed and applied. However, poor knowledge structures meant students struggled to know what the question was asking or apply the correct information, as they had not yet formed sufficient links between the problem statement, the problem type and the procedural skills necessary to solve it. This does not necessarily mean that students had not been taught or did not have these skills, such as particular

concepts, but perhaps that they were unable to cue this knowledge. Finally, there are implications for open and closed book exams. Even when resources are available, students need strong knowledge structures to solve physics problems, so it is suggested that appropriately constructed open-book exams would be able to test students' understanding and application of knowledge. Given questions that are not exercises, students are unlikely to be able to just 'look the answer up', as they need sufficient problem state, conceptual and procedural knowledge.

Peer Instruction (PI) is a method of setting students conceptual problems in order to improve their understanding. The resulting conversations were studied in Chapters 6, 7 and 8. Student conversations during PI were recorded in-the-field using smartpens to investigate characteristics of success when solving conceptual questions together. Despite a promising pilot study, there was no link between technical word use and success with the question. Technical language use is a valuable skill, but it cannot be used as a proxy to identify successful discussions, as students discuss and understand physics concepts in their own language. From the qualitative analysis, a strength of PI is that it encourages engagement with the material and most students showed some form of reasoning, whether complete or not. Students therefore have the opportunity to practice qualitative reasoning and conceptual understanding in PI, in their own community of practice, using their own language. When further examining PI discussions students were found to have a variety of reasoning depth and sophistication whichever way they voted, which may reflect that students are still learning these concepts. Instructors have already been encouraged to listen to conversations [54, 70, 75] and this work supports that. This is likely to elicit information on whether students are discussing the intended concepts, understanding what the question is asking for and skills such as critical reasoning. Careful construction of the problem statement is important, but an iterative edit of questions is also necessary, as students can tackle them in unintended ways. A final point is exemplified in the final case study (work done by a robot) where even though there was a very low overall class gain students had partial reasoning. Furthermore, students can answer correctly, but not explicitly have full reasoning. Therefore the clarification stage, where the lecturer goes over the reasoning for the correct answer, is a very important part of PI, especially if students are in a dualist Perry stage [60] and so need to hear the solution from an authority figure in order to believe it.

A group of experts was studied to establish and confirm expert characteristics during problem solving. It was found that experts understand the problem qualitatively before working through mathematical procedures (Chapter 9) and tended to work linearly through the Minnesota model, with the exception of planning, which did not occur explicitly and was perhaps conducted internally. Procedural work was also important. Even if a clear procedural strategy was presented, there were small details to clarify in actually working through the equations. Instructors should be aware that solo work in groups appeared to be an important part of the process, so each member could achieve the answer individually, whilst being supported by the group.

## 10.2 Comparing Expert and Novice Behaviour

Overall there appears to a spectrum in use of frameworks from an iterative style by novices - also reflected in complex transition maps - to a more linear style by experts with some evidence of progression along the spectrum by the PhD group. This is demonstrated graphically in Figure 10.1. Successful students identified the correct concept early on in their approach and were confident to progress using it. It indicated that they had more sophisticated knowledge structures than the unsuccessful novice students in order to be able to recognise what concepts were applicable to the question and apply them correctly. Identification of the correct concept and using it confidently without questioning it, is similar, but different to the expert group who spent a long time establishing and being confident about the correct concept before progressing with the problem. This agrees with other studies where qualitative reasoning is considered an expert-like trait [21, 22, 23]. Both successful students and the expert group used evaluation repeatedly throughout the problem solving process. Unsuccessful students either could not identify the correct concept or were not confident even if they did identify it to progress the problem solving. PI encourages the development of conceptual knowledge, but could also embed the importance of the qualitative reasoning stage, which is necessary for more complex problems. PI should be useful for students to identify, and be confident in, using the correct concept when solving a problem, but its success in terms of whole class improvements may be dependent on a number of factors, such as the way a question is

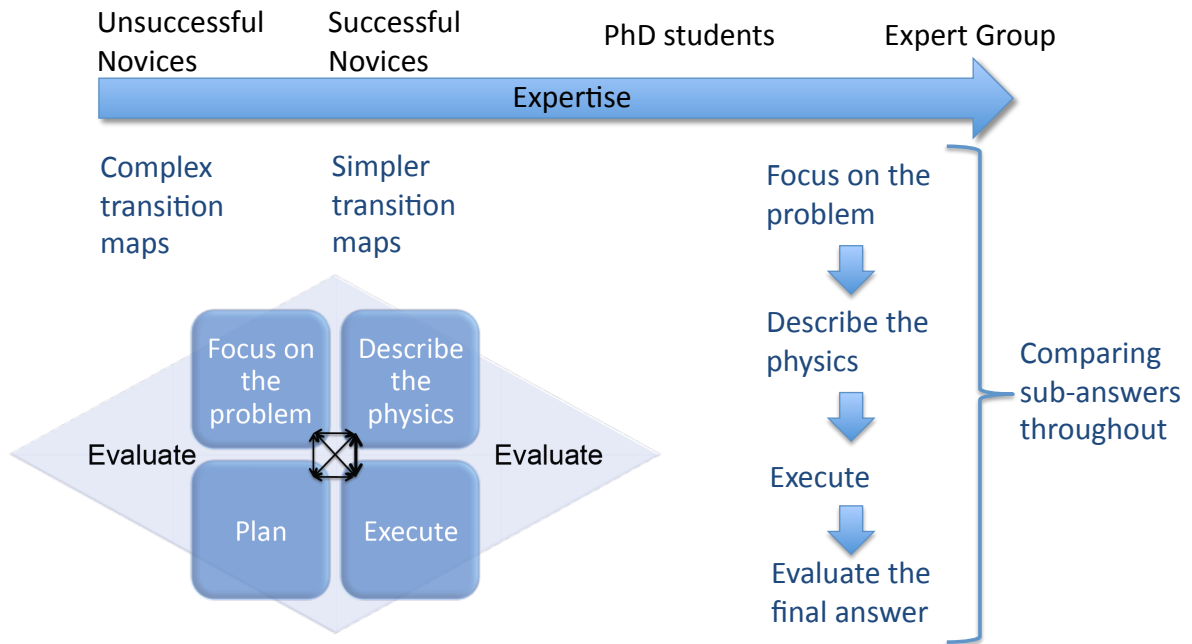


Figure 10.1: Summary of Expert and Novice Behaviour

phrased, distractors, and possibly other things not as yet known for example, the enthusiasm of lecturer and the ability of students to discuss problems.

The final recommendations as to the use of problem solving models, such as the Minnesota model are interesting. Without being explicitly asked to work through certain problem solving stages, even successful novices worked in a non-linear fashion through the problem solving stages. Experts, on the other hand, followed a linear approach. However, the expert participants were working in a group and constantly checking their reasoning and sub-answers with each other. Their behaviour may have been different if they got the answer wrong and had to go back to re-think about the physics. This, with the exception of the metacognitive behaviours of constant evaluation and lack of planning, would seem to validate the model to be used in novice group work [3, 4], as this is indeed what experts appear to do.

This research raises the question of should instructional recommendations be based on what experts do or on what successful novices do? Which is going to have to greatest learning benefit? There is no data to answer this here, but a new model is presented in the section below which incorporates the data from

this thesis. With regards to existing problem solving models, an iterative model was developed to allow for the flexibility in the use of the problem solving stages as shown in Figure 5.5. Perhaps novices could be recommended to work linearly through the stages when in groups, discussing and evaluating their solutions with each other during each stage. Used individually the behavioural stages seem to be important, as students use them, but perhaps the emphasis could be on the stages and not the order of them. Furthermore, enforcing linearity for novices, working individually, may not be helpful and may restrict cueing of knowledge that occurs from skipping to procedural work.

Access to resources by successful novices in previous chapters tended to be shorter and near the beginning of the solution. This was similar to the expert's use of resources. Perhaps successful novices are already becoming more expert-like in their use of resources. The results from studying expert behaviour agree with the resources study with Undergraduates, in that both conceptual and procedural knowledge are necessary for a correct answer. Experts exhibited these skills when using resources or not. Unsuccessful novices tended to have less direction in finding the appropriate information or could not apply the information they found. It also aligns with knowledge structure models [37], as access to problem state, conceptual and procedural knowledge and connections between them are likely to be needed for a successful solution.

### 10.3 A New Model

Bringing together the results from this thesis a new model of problem solving with resources can be created, this is shown in Figure 10.2. This model is based on the various studies presented in this thesis and is set in the context of the type of physics problems used in this research. This study of expert and novice behaviour with different types of external resources available to support conceptual and procedural knowledge form the foundations of the model.

The model in Figure 10.2 shows the necessity of *threshold knowledge*. This is the minimum understanding required to solve the problem regardless of the availability of resources. The requirement of this threshold knowledge is supported by the results in Chapters 4 and 5. Without threshold knowledge students are unlikely to be able to use resources to solve the problem, especially



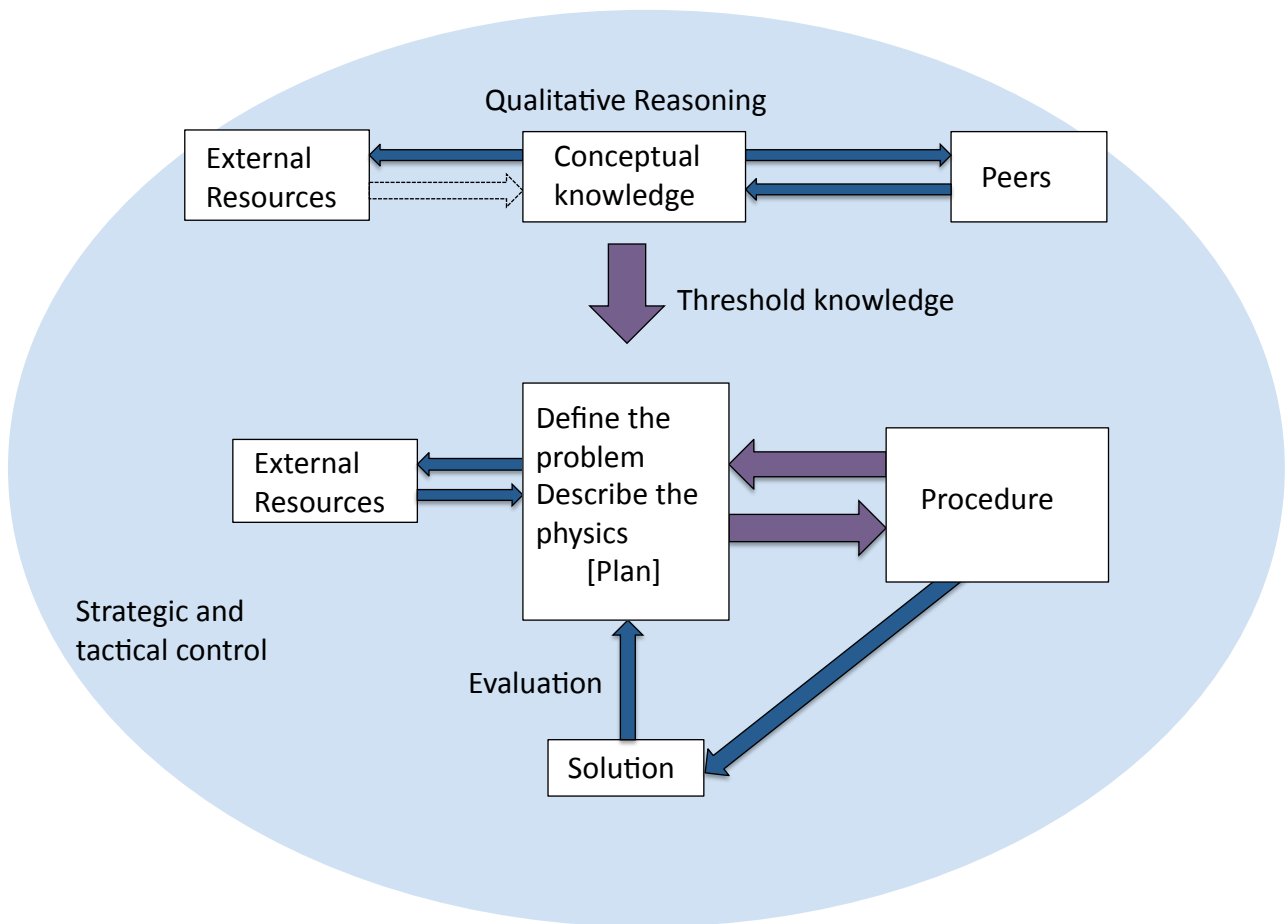


Figure 10.2: A New Model of Problem Solving Behaviour

within a time limit, as represented by the dotted arrow between the top ‘external resources’ and ‘conceptual knowledge’. Past the threshold knowledge these resources seem to become useful, for example for checking an equation is correct. Peers can also be used as a resource, as discussed in Chapter 8 and Peer Instruction as a technique can contribute to students’ conceptual knowledge to potentially help get them past this threshold knowledge for future problems.

The largest box in the model includes defining the problem and describing the physics, which represents the qualitative, more conceptual reasoning needed to solve the physics problem. The importance of qualitative reasoning or describing the physics is well known in literature [3, 21, 22, 23] and re-enforced from results in this thesis. Planning is inherent in ‘define the problem’ and not as a separate stage, as planning was not something that experts do explicitly (see Chapter 9). Furthermore, the purple larger arrows represent the importance of certain links; the strong link between conceptual and procedural work emphasises the importance of seeing how to apply a procedure in order to use a particular concept, as found in Chapter 5. Therefore, although not included explicitly in the top

line of Figure 10.2 threshold knowledge may also contain procedural work, as if students do not know how to progress in the solution then they may not apply the correct concept. In the background, as indicated by the blue circle, there should be constant strategic and tactical control, as demonstrated by the expert group in Chapter 9. In the model this involves tactical control within each area, for example checking the information from the Internet is correct and making sure procedural decisions are appropriate. Strategic control is necessary to monitor whether a transition to a different behaviour is necessary. This is separate from the evaluation stage where the final answer is checked as to whether it makes sense and the limiting cases work etc.

This model is based solely on the data from this thesis and does not account for a focus on the problem for example, in terms of agreeing what a successful outcome would be for ill-structured problems. Furthermore, many other resources can help students get past the threshold knowledge barrier, such as worked examples and intuition, which were not studied in this thesis.

## 10.4 Authentic Methods

The methods used varied from authentic, in the field data collection, to more audible, but less authentic lab-based investigations. Both have their strengths and weaknesses. Data from in the field was inaudible in parts and there is evidence of students being aware of the smartpen by turning it off or presenting their reasoning for the benefit of the smartpen, but it is perhaps more reflective of ‘real’ student conversations and experiences. Investigations in a quiet room have better quality data, but may not completely reflect true problem solving skills, as participants may change their behaviour when under more obvious observation and in a different location from their usual learning environments. Use of resources was authentic, as students will use a textbook or the Internet to solve homework problems and then after University. Following on from this, the observations for open and closed book exams are relevant for understanding more about the examination techniques employed at University.

## 10.5 Instructional Recommendations

This section summarises the main instructional recommendations that emerged from the thesis. Some of these are repeated from previous sections of the work, but are distilled here for use by instructors.

Individual problem solving and problem solving models:

- Successful problem solvers and experts have strong knowledge structures and, in general, selected the correct physics principles before doing the procedural stages. Emphasis on qualitative reasoning may help with the development of knowledge structures, so the use of Peer Instruction in lectures would emphasise its importance.
- Cueing of relevant knowledge is also important. Potential recommendations include: being aware of the wording of problem statements as it may guide students to the appropriate knowledge to solve the problem; and pointing out cues to certain concepts and procedures in worked examples to potentially make it easier to utilise this learning in future problems. Perhaps explicitly explaining the difference between surface features and conceptual knowledge as this stage may help students learn to cue the appropriate knowledge.
- Awareness of approach and general strategic and tactical control over the solution are important factors [24]. Evaluation was demonstrated throughout the expert group work and this process should be encouraged rather than evaluation solely as a final step.
- Students are unlikely to work through recommended problem solving models linearly, while a small scale study suggested that experts work through the stages linearly when working in a group.

Assessment:

- Appropriately constructed open-book exams are likely to test understanding of the relevant knowledge, as students are unlikely to be able to successfully look up the required solution without having a good understanding of the topic.

- Although there is limited evidence from this work, there is an indication that students use different skills in ‘show that’ problems compared to questions where they do not know what the final answer will be. Instructors should be aware of this potential difference when constructing summative assessments.

Peer Instruction in lectures:

- The phrasing of the problem statement is important, for example, unnecessary distractors can lead students to discuss unintended concepts. In particular questions with a low whole class gain should be considered and revised. Recommendations from Beatty et al.[75] would be useful here.
- Instructors should listen to student discussions in lectures. Listen for: misconceptions, use of appropriate assumptions, ability of students to reason and critique others, their understanding of the question and whether they are discussing the intended concept.
- Technical word use was not found to be correlated with success on PI, but a list of technical and non-technical language to use in tackling particular conceptual questions may help students to access and use the appropriate conceptual knowledge and encourage a fuller, more complete argument.
- The explanation of the solution after student discussions seems particularly important to clarify reasoning and reinforce the correct conceptual model.

## 10.6 Future Research

Future research in each area investigated is described in detail in the relevant chapters. More research is needed with experts individually and in groups on difficult problems to compare expert and novice behaviour. For PI, recommendations included research into students’ understanding of technical language and establishing and refining characteristics relating to successful and unsuccessful conversations. Testing instructional techniques which mitigate for unsuccessful characteristics may help refine and develop the PI process. Further investigation into the additional skills students need in PI, such as critical reasoning and argumentation skills would show if other skills are being developed.

One particular topic of future research on which to focus is further testing of the skills required for open and closed book exams. This is a current topic which could be extended in terms of evidenced-based analysis. It relates to current teaching practices that are being used in instructional settings. Further investigation could allow instructors to make informed decisions when setting exams as to what skills they will be testing and how relevant these skills are to the expertise needed after University in this increasingly digital world.

To conclude, this thesis develops a better understanding of student and expert problem solving behaviour in individual and group environments, including characteristics of successful and unsuccessful solution processes. From further understanding problem solving processes instructional recommendations have been presented, as well as avenues for future research, to continually improve and understand student problem solving in physics at university.

# Appendices

# Appendix A

## Individual Problem Solving and Resources Use

### A.1 Physics Questions

These questions were given one at a time to the student.

**The warm up question was:**

When the legal speed limit for the New York City Thruway was increased from 55km/h to 65km/h, how much time was saved by a motorist who drove the 700km between the Buffalo entrance and the New York City exit at the legal speed limit?

**The first main diagram question was:**

A man is trying to push a large stone block across a uniform drawbridge which has a mass of  $m = 200\text{kg}$ . The mass of the large stone block and the man combined is  $M = 300\text{kg}$ . The drawbridge is of length  $L = 5.00\text{m}$  and is held up by a wire. The wire can withstand a maximum tension of 5000N. The vertical distance between the drawbridge and the point where the wire is attached to the wall is  $h = 4.00\text{m}$ . How far along from the wall can the man push the stone block before the bridge collapses? Assume  $g = 10 \text{ ms}^{-2}$ .

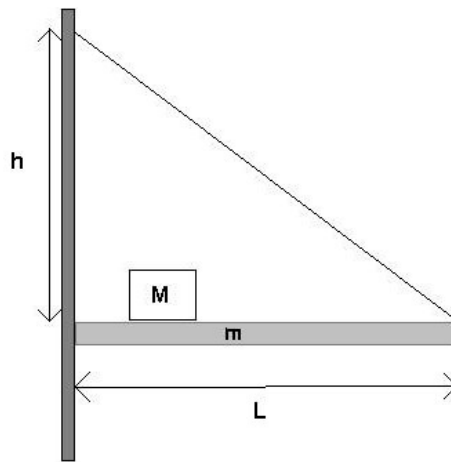


Figure A.1: Drawbridge question

**The second main question was:**

An army cadet uses a rope swing to cross a river. He starts on a high platform with his centre of mass 7.16 metres from the ground and grabs onto the rope which is 6m long, to swing. He lets go when the rope is vertical and his centre of mass is 1m off the ground. He needs to travel 5m horizontally to clear the river after letting go of the rope. If he successfully clears the river what is the tension in the rope just before he releases? The cadet weighs 80kg.



## A.2 Answers to Physics Questions

The answers were also given to the student at the end. These are given below.

*When the legal speed limit for the New York City Thruway was increased from 55km/h to 65km/h, how much time was saved by a motorist who drove the 700km between the Buffalo entrance and the New York City exit at the legal speed limit?*

Calculate time to travel 700km at 55km/h which is 12.7hours and calculate time to travel 700km at 65km/h which is 10.8hours. The driver has saved  $12.7 - 10.8 = 1.95$  hours (nearly 2hours) on their journey.

### Diagram question

*A man is trying to push a large stone block across a uniform drawbridge which has a mass of  $m = 200\text{kg}$ . The mass of the large stone block and the man combined is  $M = 300\text{kg}$ . The drawbridge is of length  $L = 5.00\text{m}$  and is held up by a wire. The wire can withstand a maximum tension of  $5000\text{N}$ . The vertical distance between the drawbridge and the point where the wire is attached to the wall is  $h = 4.00\text{m}$ . How far along from the wall can the man push the stone block before the bridge collapses? Assume  $g = 10\text{ ms}^{-2}$ .*

Treating the hinge as the pivot point this problem can be solved considering torques.

Working out the angle between the wire and the drawbridge with  $\tan \phi$  gives  $\phi = 38.65^\circ$ . Subtract  $38.65$  from  $180$  to give the angle used in the torque equation which is  $\theta = 141^\circ$ .

The torque acting holding the drawbridge up (and acting upwards) is calculated as  $T = F \times r = Fr \sin \theta = 15,617\text{N}$ . Both the centre of mass of the drawbridge and the centre of mass of the stone act downwards.

At breaking point the upwards torque of the wire is just balanced by the downwards torque due to gravity of the stone and the bridge. Thus assuming the centre of mass is in the middle of the plank:

$$2000\text{N} \times 2.5\text{m} + 3000 \times x = 15617\text{N}$$

Rearranging this gives  $x = 3.54\text{metres}$  from the pivot (wall).

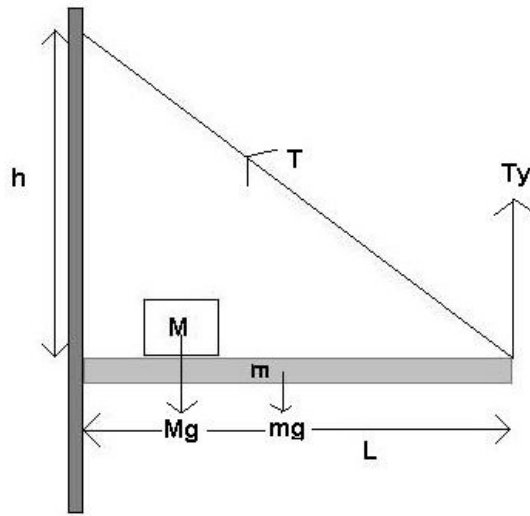


Figure A.2: Diagram of the drawbridge question

**Army Cadet Question:**

An army cadet uses a rope swing to cross a river. He starts on a high platform with his centre of mass 7.16 metres from the ground and grabs onto the rope which is 6m long, to swing. He lets go when the rope is vertical and his centre of mass is 1m off the ground. He needs to travel 5m horizontally to clear the river after letting go of the rope. If he successfully clears the river what is the tension in the rope just before he releases? The cadet weighs 80kg.

The approach is to find the tangential velocity, which because he lets go of the rope when vertical, is entirely horizontal. Once this velocity is found from projectile motion equations it can be entered into the equation for tension on a rope. The tension on the rope is determined by a force diagram. The sum of the forces acting provides the centripetal force which is the tension minus gravity when at the bottom of the circle.

If given the height of the platform the velocity can also be found from converting potential to kinetic energy, given  $mgh = (1/2)mv^2$

Vertically the cadet's velocity starts at 0m/s.  $s = (1/2)gt^2$ . Using  $s = 1m$  gives a time of  $t = 0.45s$  till he hits the ground - ignoring friction.

To find his initial velocity use horizontal component equations:  $s = ut$ , using  $s = 5\text{m}$  and  $t$  found above:  $v = 11.18\text{m/s}$ .

The tension on the rope is given by gravitational and centripetal forces  $T = mg + (mv^2)/r$ , putting the velocity in from above with mass = 80kg and  $r = 6\text{m}$ :  
 $T = 2466\text{N}$

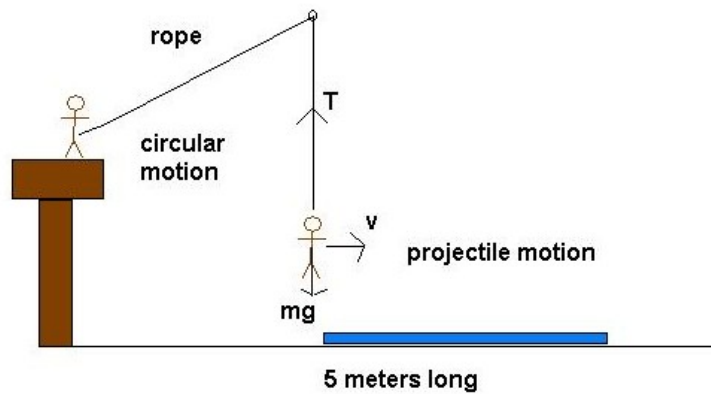


Figure A.3: Diagram of the army cadet question

If students used both projectile motion and energy conservation to find the velocity they were marked for one of the approaches and not awarded double marks.

## A.3 Allocation of Marks

### Allocation of Marks in the Drawbridge and Army Cadet questions

On the drawbridge question the breakdown of marks were as follows.

- Recognising that torque applies - 1 mark
- Stating  $\tau = r \times F$  - 1 mark
- Calculating the angle between the wire and the drawbridge = 38.6 degrees - 1 mark
- Using the angle to find the maximum tension exerted by the wire upwards  $T = f \times r = Fr \sin\theta = 15,617N$  - 1mark for symbolic form, 1mark for numerical answer (1mark if they just calculate the tension upwards with  $F_{up} = F \sin\theta = 3146N$ ) 2marks in total.
- Recognising that the centre of mass of the drawbridge acts through its centre - 1 mark
- Balancing torques correctly:  $T_{up} = W_{plank} \times d + W_{man/block} \times x$  - 1 mark for each part of the equation. 3 marks in total.
- Calculating x to be 3.5meters - 1 mark

For the army cadet question:

- Recognising that centripetal force applies - 1 mark
- Stating  $F_c = \frac{mv^2}{r}$  - 1mark
- Recognising one or both of projectile motion and energy conservation (either one of these approaches is fine) - 1 mark
- **If using the projectile motion** to find the release velocity:  $s = \frac{1}{2}st^2$  to find t - 1 mark,  $s = ut$  to find u - 1 mark,  $u=11.8m/s$  - 1 mark. 3 marks in total.
- **If using energy conservation** to find the release velocity:  $E_p = mgh$  and  $E_k = \frac{1}{2}mv^2$  - 1 mark,  $E_p = E_k$  - 1 mark,  $v=11.8m/s$  - 1 mark. 3 marks in total.

- A correct diagram and realising that  $v$  is horizontal at the point of release and that there are no external forces at the point of release - 1 mark
- Stating  $T = mg + \frac{mv^2}{r}$  - 2 marks in total (1 mark if they state that  $T = mg$  or  $T = \frac{mv^2}{r}$  and 1 mark off if all three parts in equation, but wrong signs.)
- Finding Tension=2466N - 1 mark

If students used both projectile motion and energy conservation to find the velocity they were marked for one of the approaches and not awarded double marks.

## A.4 Master Sheet for Researcher - 1B Students

**Master Sheet** - Experimental Protocol. To use whilst conducting think aloud

### Logistics

- Book room and bring booking times, room number and booking reference number
- Pen and paper ready - check memory, batteries etc.
- Tell students when to arrive and where to go
- Have the questions to the student ready all on paper (as diagram question is on paper, so keep all the same) and all the questions they will be asked
- Have answers to the questions
- Bring a calculator and layout ready for use (so they all use the same one, and do not have to start searching for one half way through the problem)
- Bring name, matric number and time of meeting to the interview

### A.4.1 Overall Plan

The plan of the interview: Approx 45 minutes to 1 hour

Warm up question - short: 2-5 minutes

First question - diagram question: 15 minutes

Second question - army man swinging on rope: 15-20 minutes

Retrospection on both, but on army question first - they can look at notes to remind them: 5-10minutes

## A.4.2 Welcome and warm up question

Explanation:

*Welcome. I am studying how students solve physics problems. Thank you for helping. Your name will not appear in any form of write up. Also there are no hidden agendas here I simply want to know how you problem solve. There will be a warm up question where I will explain to you how to do the procedure and gives you a chance to gets used to what you will be doing and then two main questions. You will do the warm up question with me in the room, then I will leave you by yourself with the recorder for the main questions. After this I will ask you another few questions at the end.. The aim of this is for you to talk aloud what you are thinking*

(Introducing idea in the introduction to get them used to it. Also don't say that I am going to be asking questions about What they did in case they think about it more because they know they have to explain it later.)

*This pen records what you say and write, if that is okay? So here is the warm up question. Please solve it as you would normally and think aloud while solving it. It is important that you say everything that you think or do. What ever is going through your mind say. Answer the question any way you want and do not worry about sounding stilted.*

Hand out Warm up question:

When the legal speed limit for the New York City Thruway was increased from 55km/h to 65km/h, how much time was saved by a motorist who drove the 700km between the Buffalo entrance and the New York City exit at the legal speed limit?

Prompts:

*Keep on talking*

*Just say what you are doing (stop them analysing and interpreting their problem solving techniques, but keep them talking- including the 'chatter' in the heads.)*

### A.4.3 First Main Diagram Question

*I will give you the first physics problem now for you to solve on the paper given. Please solve it as you would normally. Same as before, you had the right idea/ it is important to keep talking even more so than the warm up question, just whatever you are thinking. It is important that you say everything that you think or do. I will leave the room and come back in 15 minutes. I will be just outside in the Cafe/big study room.*

Protocol:

Leave the room for 15 minutes timed on a stopwatch.

Upon return *How are you getting on? Have you had enough or do you want another minute or so?* They are allowed an extra 2-5 minutes, but only if they ask for it. Judge by how far on and their attitude as to whether to move on immediately or ask them if they want to keep going. (*It was noted afterwards that students seemed to be honest that if they had enough they would say.*) After that stop this question and move onto the next, so they do not get too tired. No hints are given and no retrospection done so as to not influence the results from the next question, but also because the main aim of the diagram question is to study the students' use of the diagram, but we are also interested if they solved the problem or not.

Question:

A man is trying to push a large stone block across a uniform drawbridge which has a mass of  $m = 200\text{kg}$ . The mass of the large stone block and the man combined is  $M = 300\text{kg}$ . The drawbridge is of length  $L = 5.00\text{m}$  and is held up by a wire. The wire can withstand a maximum tension of  $5000\text{N}$ . The vertical distance between the drawbridge and the point where the wire is attached to the wall is  $h = 4.00\text{m}$ . How far along from the wall can the man push the stone block before the bridge collapses? Assume  $g = 10\text{ ms}^{-2}$ .



### A.4.4 Second Main Question

*I will give you the second physics problem now for you to solve. Again please solve it as you would normally, saying everything that you think or do. I will leave the room and come back in 15 minutes.*

Protocol:

Give second question - repeat protocol by leaving the room for 15 minutes. They are allowed an extra 2-5 minutes, time is a bit more flexible, because it is the last question and the more they try the more problem solving strategies in different circumstances i.e. having had plenty time to think about the problem, can be observed.

Second Main Question (20 mins)

An army cadet uses a rope swing to cross a river. He starts on a high platform with his centre of mass 7.16 metres from the ground and grabs onto the rope which is 6m long, to swing. He lets go when the rope is vertical and his centre of mass is 1m off the ground. He needs to travel 5m horizontally to clear the river after letting go of the rope. If he successfully clears the river what is the tension in the rope just before he releases? The cadet weighs 80kg.

*How are you getting on.. Would you like a hint?*

*Have you completed the question? Are you nearly finished or feel you have nearly got the answer? If the answer to both of these questions is no then ask: Would you like a hint?*

Once back in 15 minutes if they are not successful they can have the hints below, given one at a time in the order stated below.

Hints to give after they have had an attempt at the solution for 15minutes:

- Before he lets go of the rope the cadet is performing circular motion
- Centripetal force = net force into the centre =  $T - mg$
- In equation form:  $F_c = T - mg = (mv^2)/r$  and find  $v$
- Find  $v$  using projectile motion

### A.4.5 Retrospection

*Please explain to me the problem solving strategies you used to solve the problem.*

Emphasis on problem solving strategies - what was your overall approach at this question, what did you do generally to solve the problem?

Turn to the page of working where they did this to aid memory. Ask the relevant questions from those given below:

*Thinking back to the first main question you did with the diagram Can you remember why you added to/ didn't add to the diagram given? Why did you draw another diagram? How did you use this diagram to help you - or did you not think about that consciously?*

### A.4.6 Validity of think aloud

Were you ever thinking something that you feel wasn't captured by the written or spoken materials?

End of recording

Explain answer to questions and answer any questions about the problems given that the student may have.

*Thank you for participating*

Volunteer to give them feedforward/feedback on their strategies via email or another informal meeting.

*Would you like to receive a short summary of feedforward relating to what you did - There is a lot of literature on how experts problem solve, so I can give you a bit of feedback about what you did compared to this literature. Your Amazon*

*voucher will be sent soon, once it has been organised.*

*Especially for the first few people - please do not talk to anyone about the problems or solutions until the end of next week, in case they are asked to do this.*

Thank you

## A.5 Consent Form

**Problem Solving**

**CONSENT FORM**

If you are happy to participate please complete and sign the consent form below.

**Please  
Initial Box**

1. I confirm that I have read the attached information sheet on the above project and have had the opportunity to consider the information and ask questions and had these answered satisfactorily.

2. I understand that my participation in the study is voluntary and that I am free to withdraw at any time without giving a reason and without detriment to any treatment/service

3. I understand that the interviews will be audio-recorded

4. I agree to the use of anonymous quotes

5. I agree that any data collected may be passed to other researchers

I agree to take part in the above project

\_\_\_\_\_  
Name of participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name of person taking  
consent

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

## **A.6 Temporal Maps**

# 1B STUDENTS: DRAWBRIDGE QUESTION

Mark: 2/10

Student 1 Drawbridge Q

FDPEE	Time(minutes)
Code	2 4 6 8 10 12 14
Focus on the problem	[Bar chart showing activity blocks]
Describing the physics	[Bar chart showing activity blocks]
Planning	[Bar chart showing activity blocks]
Execution	[Bar chart showing activity blocks]
Evaluation	[Bar chart showing activity blocks]
Silence and no writing	[Bar chart showing activity blocks]

ACTIVITY	Time(minutes)
Code	2 4 6 8 10 12 14
Reading	[Bar chart showing activity blocks]
Translation	[Bar chart showing activity blocks]
Analysis	[Bar chart showing activity blocks]
Exploration	[Bar chart showing activity blocks]
Planning	[Bar chart showing activity blocks]
Implementation	[Bar chart showing activity blocks]
Verification	[Bar chart showing activity blocks]

REPRESENTATIONS	Time(minutes)
Code	2 4 6 8 10 12 14
Equation symbolic	[Bar chart showing activity blocks]
Equation numerical	[Bar chart showing activity blocks]
Free numbers	[Bar chart showing activity blocks]
Diagram	[Bar chart showing activity blocks]
Words	[Bar chart showing activity blocks]

Mark: 2/10

Student 2 Drawbridge Question

FDPEE	Time(minutes)
Code	2 4 6 8 10 12 14 16 18 20 22 24 26
Focus on the problem	[Bar chart showing activity blocks]
Describing the physics	[Bar chart showing activity blocks]
Planning	[Bar chart showing activity blocks]
Execution	[Bar chart showing activity blocks]
Evaluation	[Bar chart showing activity blocks]
Silence and no writing	[Bar chart showing activity blocks]

ACTIVITY	Time(minutes)
Code	2 4 6 8 10 12 14 16 18 20 22 24 26
Reading	[Bar chart showing activity blocks]
Translation	[Bar chart showing activity blocks]
Analysis	[Bar chart showing activity blocks]
Exploration	[Bar chart showing activity blocks]
Planning	[Bar chart showing activity blocks]
Implementation	[Bar chart showing activity blocks]
Verification	[Bar chart showing activity blocks]

REPRESENTATIONS	Time(minutes)
Code	2 4 6 8 10 12 14 16 18 20 22 24 26
Equation symbolic	[Bar chart showing activity blocks]
Equation numerical	[Bar chart showing activity blocks]
Free numbers	[Bar chart showing activity blocks]
Diagram	[Bar chart showing activity blocks]
Words	[Bar chart showing activity blocks]





Mark: 9/10

Student 5 Drawbridge Question

FDPEE	Time (minutes)
Code	2 4 6 8 10 12
Focus on the problem	0-1.5, 1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
Describing the physics	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
Planning	
Execution	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
Evaluation	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
Silence and no writing	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12

ACTIVITY	Time (minutes)
Code	2 4 6 8 10 12
Reading	
Translation	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
Analysis	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
Exploration	
Planning	
Implementation	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
Verification	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12

REPRESENTATIONS	Time (minutes)
Code	2 4 6 8 10 12
Equation symbolic	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
Equation numerical	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
Free numbers	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
diagram	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12
Words	1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, 4.5-5, 5-5.5, 5.5-6, 6-6.5, 6.5-7, 7-7.5, 7.5-8, 8-8.5, 8.5-9, 9-9.5, 9.5-10, 10-10.5, 10.5-11, 11-11.5, 11.5-12

Mark: 10/10

Student 6 Drawbridge question

FDPEE	Time (minutes)
Code	2 4 6 8 10
Focus on the problem	0-1, 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Describing the physics	1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Planning	
Execution	5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Evaluation	5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Silence and no writing	0-1, 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12

ACTIVITY	Time (minutes)
Code	2 4 6 8 10
Reading	0-1, 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Translation	
Analysis	0-1, 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Exploration	
Planning	
Implementation	5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Verification	5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Silence	0-1, 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12

REPRESENTATIONS	Time (minutes)
Code	2 4 6 8 10
Equation symbolic	4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Equation numerical	5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Free numbers	1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
diagram	0-1, 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12
Words	0-1, 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 9-10, 10-11, 11-12





Student 5, Drawbridge Question

Mark: 0/10

FDPEE	Time (minutes)																			
Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Focus on the problem	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Describing the physics																				
Planning																				
Execution																				
Evaluation																				
Silence and no writing																				

ACTIVITY	Time (minutes)																			
Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Reading	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Translation																				
Analysis																				
Exploration																				
Planning																				
Implementation																				
Verification																				
Silence																				

REPRESENTATIONS	Time (minutes)																			
Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Equation symbolic																				
Equation numerical																				
Free numbers																				
diagram																				
Words																				

Student 6, Drawbridge Question

Mark: 10/10

FDPEE	Time (minutes)					
Code	1	2	3	4	5	6
Focus on the problem	█	█	█	█	█	█
Describing the physics						
Planning						
Execution						
Evaluation						
Silence and no writing						

ACTIVITY	Time (minutes)					
Code	1	2	3	4	5	6
Reading	█	█	█	█	█	█
Translation						
Analysis						
Exploration						
Planning						
Implementation						
Verification						
Silence						

REPRESENTATIONS	Time (minutes)					
Code	1	2	3	4	5	6
Equation symbolic						
Equation numerical						
Free numbers						
diagram						
Words						





Mark: 2/10

Student 3 Army Cadet Question

FDPEE		Time(minutes)							
Code		2	4	6	8	10	12	14	16
Focus on the problem		█	█	█	█	█	█	█	█
Describing the physics		█	█	█	█	█	█	█	█
Planning									
Execution									
Evaluation									
Silence and no writing									

ACTIVITY		Time(minutes)							
Code		2	4	6	8	10	12	14	16
Reading									
Translation		█	█	█	█	█	█	█	█
Analysis									
Exploration									
Planning									
Implementation									
Verification									

REPRESENTATIONS		Time(minutes)							
Code		2	4	6	8	10	12	14	16
Equation symbolic									
Equation numerical									
Free numbers									
Diagram		█	█	█	█	█	█	█	█
Words									

Mark: 5/10

Student 4 Army Cadet Question

FDPEE		Time(minutes)							
Code		2	4	6	8	10	12	14	16
Focus on the problem		█	█	█	█	█	█	█	█
Describing the physics		█	█	█	█	█	█	█	█
Planning									
Execution									
Evaluation									
Silence and no writing									

ACTIVITY		Time(minutes)							
Code		2	4	6	8	10	12	14	16
Reading		█	█						
Translation		█	█	█	█	█	█	█	█
Analysis									
Exploration									
Planning									
Implementation									
Verification									

REPRESENTATIONS		Time(minutes)							
Code		2	4	6	8	10	12	14	16
Equation symbolic									
Equation numerical									
Free numbers									
Diagram		█	█	█	█	█	█	█	█
Words									

Mark: 10/10

Student 5 Army Cadet Question

FDPEE		Time (minutes)													
Code		2	4	6	8	10	12	14							
Focus on the problem		[Bar chart showing activity distribution]													
Describing the physics		[Bar chart showing activity distribution]													
Planning		[Bar chart showing activity distribution]													
Execution		[Bar chart showing activity distribution]													
Evaluation		[Bar chart showing activity distribution]													
Silence and no writing		[Bar chart showing activity distribution]													

ACTIVITY		Time (minutes)													
Code		2	4	6	8	10	12	14							
Reading		[Bar chart showing activity distribution]													
Translation		[Bar chart showing activity distribution]													
Analysis		[Bar chart showing activity distribution]													
Exploration		[Bar chart showing activity distribution]													
Planning		[Bar chart showing activity distribution]													
Implementation		[Bar chart showing activity distribution]													
Verification		[Bar chart showing activity distribution]													

REPRESENTATIONS		Time (minutes)													
Code		2	4	6	8	10	12	14							
Equation symbolic		[Bar chart showing activity distribution]													
Equation numerical		[Bar chart showing activity distribution]													
Free numbers		[Bar chart showing activity distribution]													
diagram		[Bar chart showing activity distribution]													
Words		[Bar chart showing activity distribution]													

Mark: 8/10

STUDENT 6: ARMY QUESTION

FDPEE		Time(minutes)													
Code		2	4	6	8	10	12	14							
Focus on the problem		[Bar chart showing activity distribution]													
Describing the physics		[Bar chart showing activity distribution]													
Planning		[Bar chart showing activity distribution]													
Execution		[Bar chart showing activity distribution]													
Evaluation		[Bar chart showing activity distribution]													
Silence and no writing		[Bar chart showing activity distribution]													

ACTIVITY		Time(minutes)													
Code		2	4	6	8	10	12	14							
Reading		[Bar chart showing activity distribution]													
Translation		[Bar chart showing activity distribution]													
Analysis		[Bar chart showing activity distribution]													
Exploration		[Bar chart showing activity distribution]													
Planning		[Bar chart showing activity distribution]													
Implementation		[Bar chart showing activity distribution]													
Verification		[Bar chart showing activity distribution]													
Silence and no writing		[Bar chart showing activity distribution]													

REPRESENTATIONS		Time(minutes)													
Code		2	4	6	8	10	12	14							
Equation symbolic		[Bar chart showing activity distribution]													
Equation numerical		[Bar chart showing activity distribution]													
Free numbers		[Bar chart showing activity distribution]													
diagram		[Bar chart showing activity distribution]													
Words		[Bar chart showing activity distribution]													







Student 5, Army Cadet Question

Mark: 2/10

FDPEE	Time (minutes)																	
	Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Focus on the problem		█	█	█	█	█												
Describing the physics					█	█	█	█		█		█	█	█				
Planning																		
Execution																		
Evaluation																		
Silence and no writing		█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█

ACTIVITY	Time (minutes)																	
	Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Reading																		
Translation		█	█	█	█													
Analysis					█	█	█	█		█								
Exploration				█														
Planning																		
Implementation																		
Verification																		
Silence		█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█

REPRESENTATIONS	Time (minutes)																	
	Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Equation symbolic																		
Equation numerical																		
Free numbers																		
diagram		█	█	█	█													
Words																		

Student 6, Army Cadet Question

Mark: 10/10

FDPEE	Time (minutes)							
	Code	1	2	3	4	5	6	7
Focus on the problem		█	█	█	█	█	█	
Describing the physics				█		█	█	█
Planning								
Execution								
Evaluation								
Silence and no writing		█		█	█	█	█	█

ACTIVITY	Time (minutes)							
	Code	1	2	3	4	5	6	7
Reading		█		█				
Translation		█	█	█	█			
Analysis			█	█	█	█	█	█
Exploration								
Planning								
Implementation								
Verification								
Silence		█		█	█	█	█	█

REPRESENTATIONS	Time (minutes)							
	Code	1	2	3	4	5	6	7
Equation symbolic								
Equation numerical								
Free numbers								
diagram		█	█	█	█			
Words								

FDPEE		Time (minutes)										
Code		1	2	3	4	5	6	7	8	9	10	11
Focus on the problem		█	█	█	█	█	█	█	█	█	█	█
Describing the physics				█	█	█	█	█	█	█	█	█
Planning			█									
Execution										█	█	█
Evaluation												
Silence and no writing			█		█		█					

ACTIVITY		Time (minutes)										
Code		1	2	3	4	5	6	7	8	9	10	11
Reading											█	█
Translation		█	█	█	█	█	█	█	█	█	█	█
Analysis				█	█	█	█	█	█	█	█	█
Exploration												
Planning			█									
Implementation											█	█
Verification												
Silence			█		█		█					

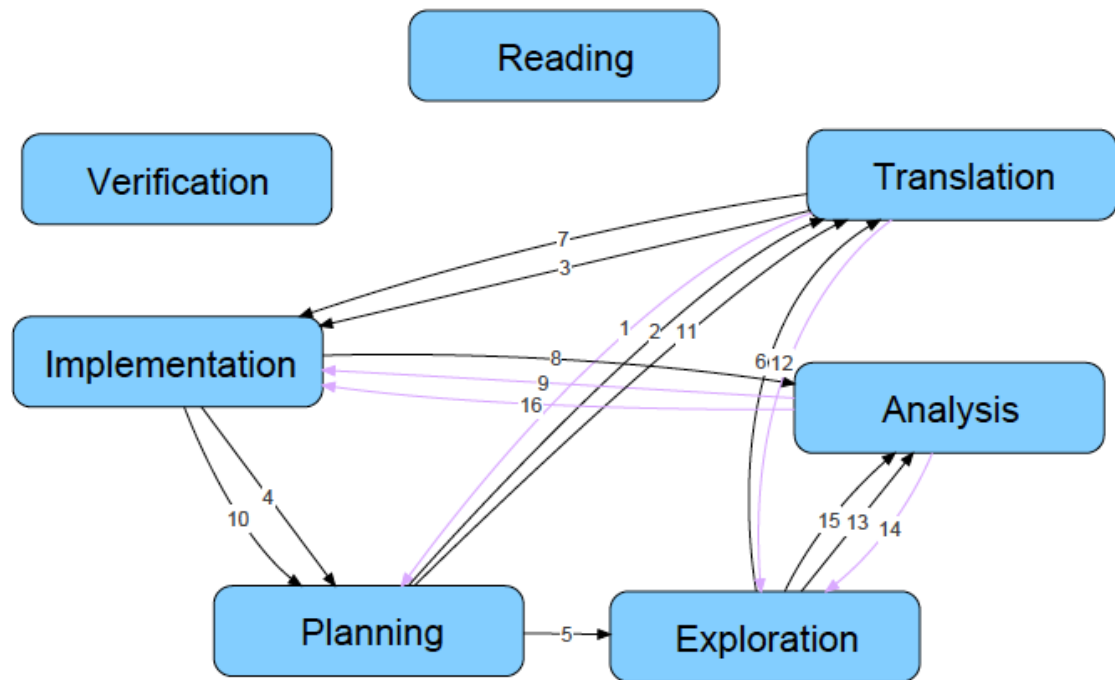
REPRESENTATIONS		Time (minutes)										
Code		1	2	3	4	5	6	7	8	9	10	11
Equation symbolic								█				
Equation numerical											█	█
Free numbers				█					█	█	█	█
diagram		█	█	█	█	█	█	█				
Words									█			

## **A.7 Transition Maps**

1B STUDENT – DRAWBRIDGE QUESTION

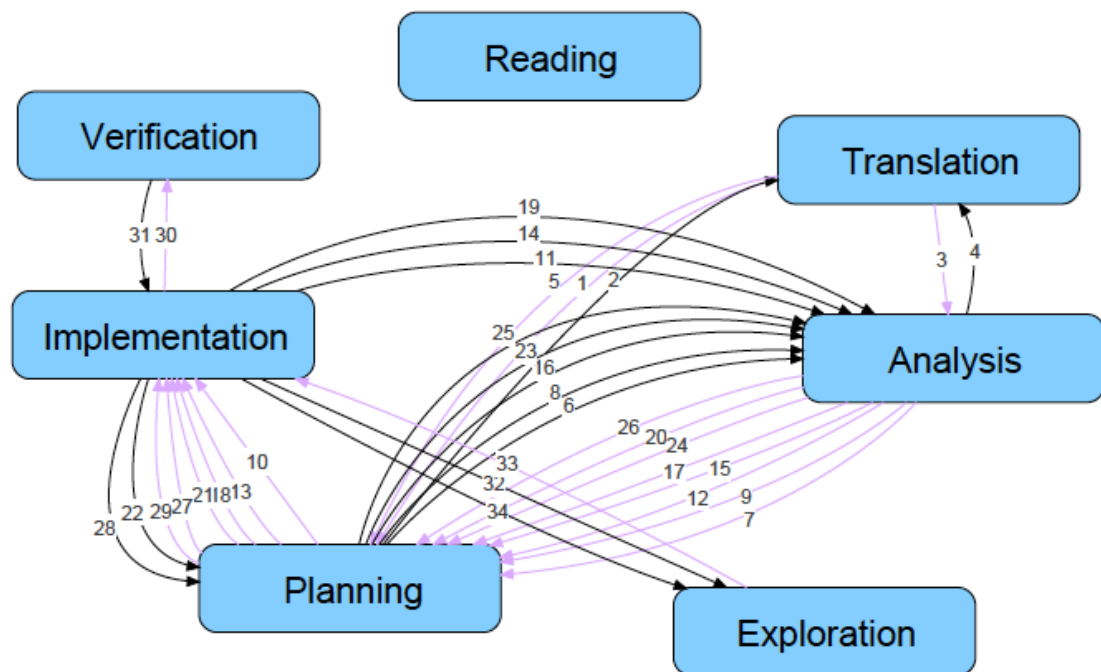
Student 1

Mark: 2/10



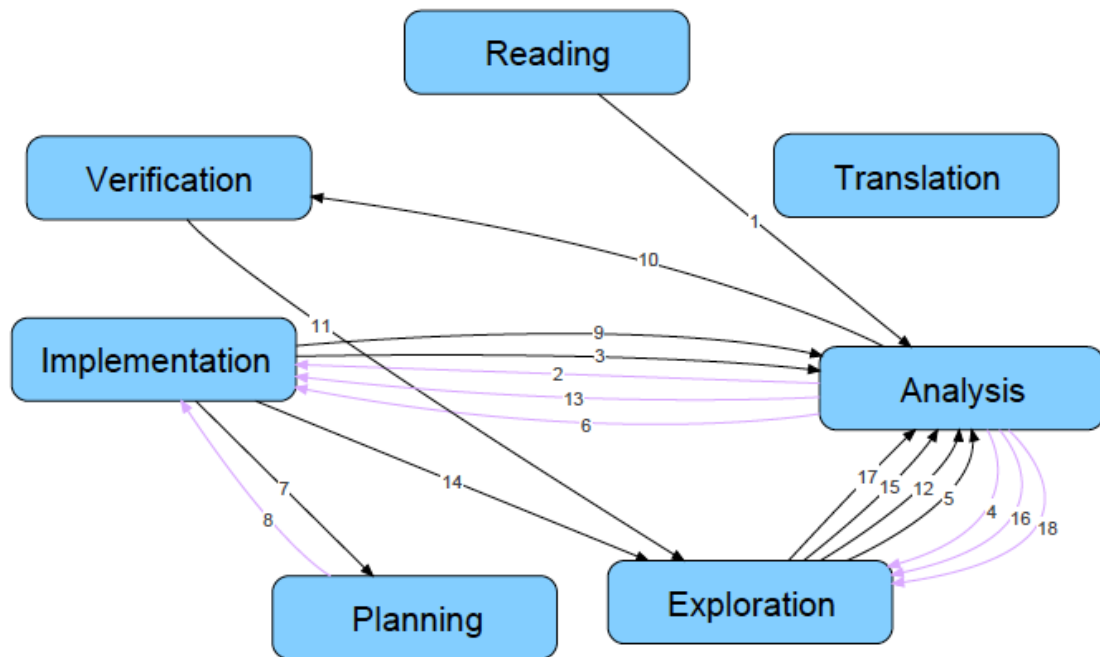
Student 2

Mark: 2/10



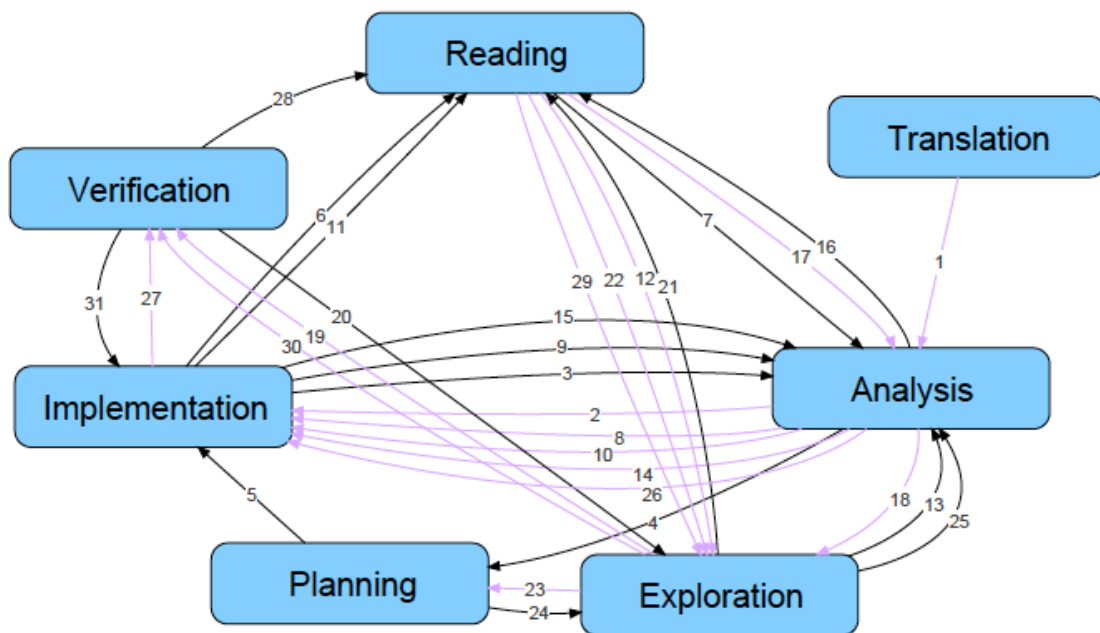
Student 3

Mark: 3/10



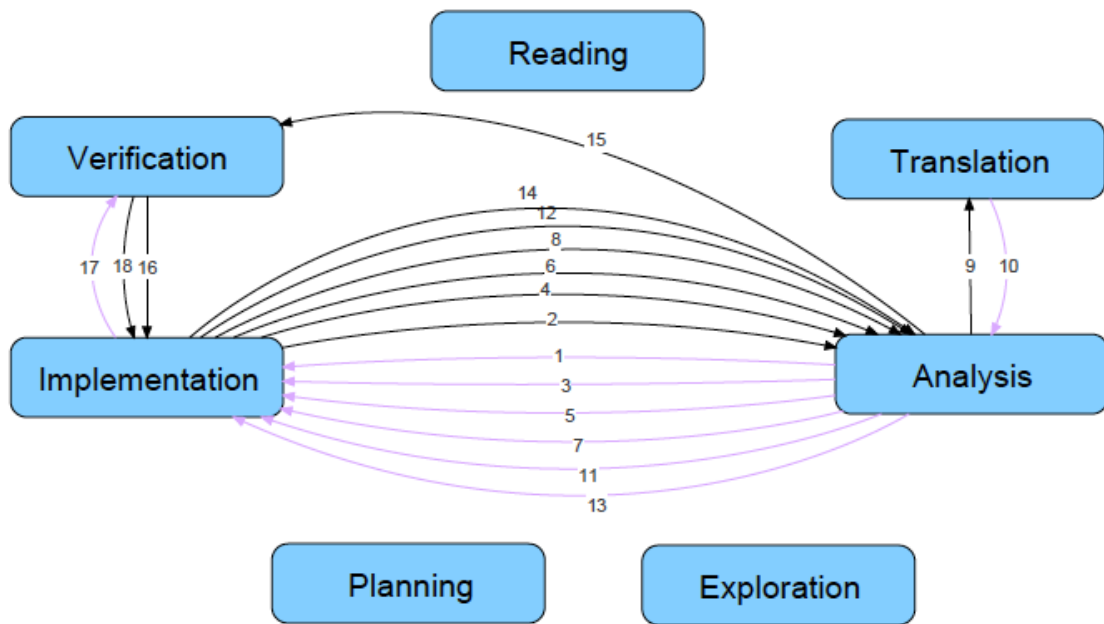
Student 4

Mark: 1/10



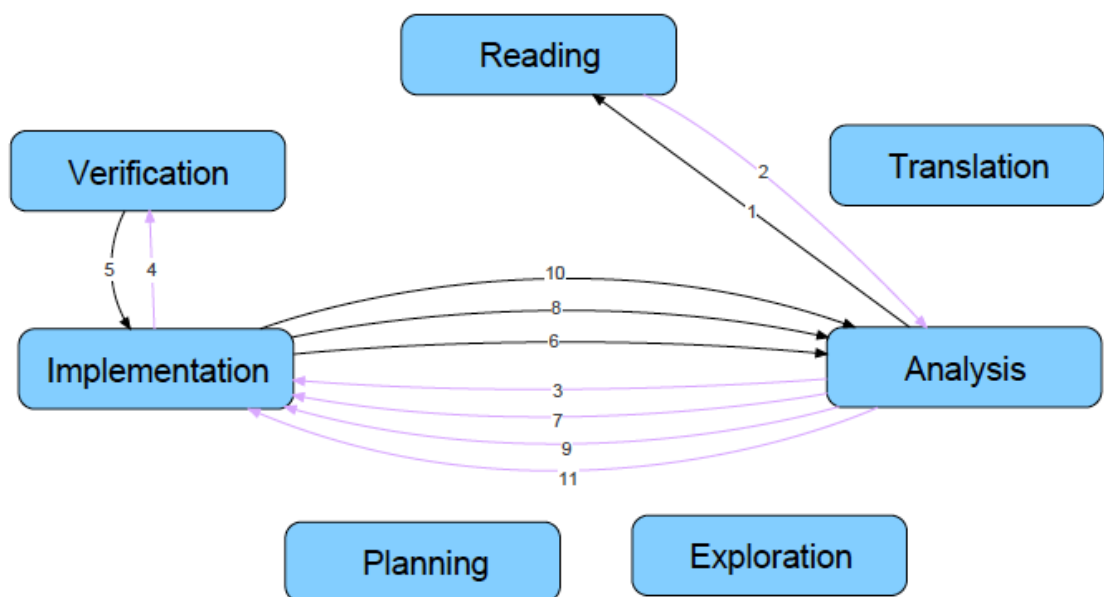
Student 5

Mark: 9/10



Student 6

Mark: 10/10

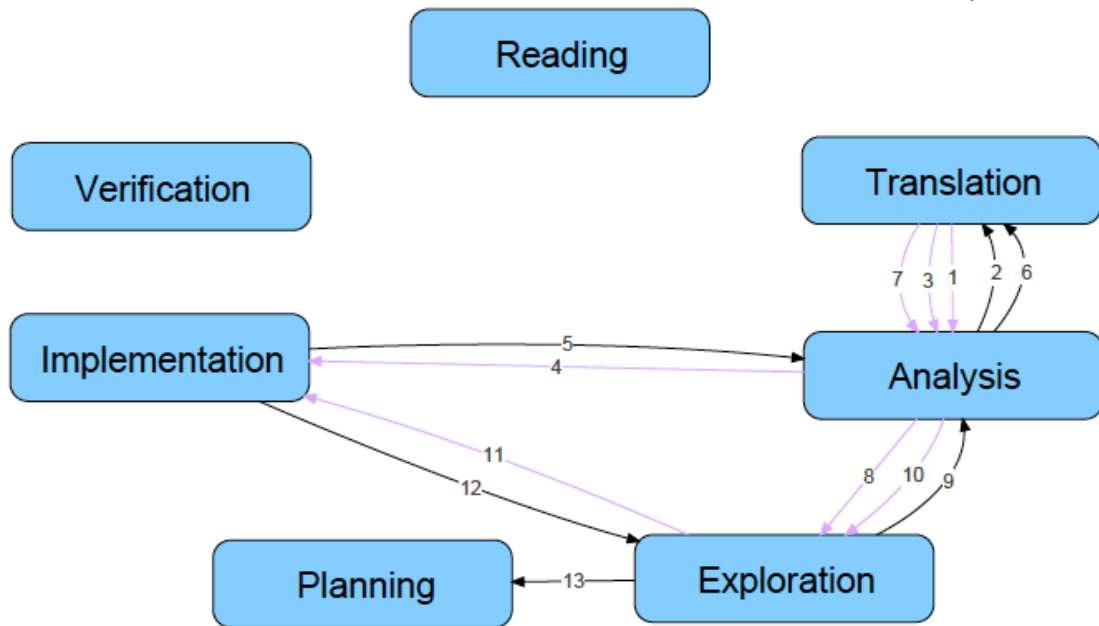




1B STUDY ARMY CADET QUESTION

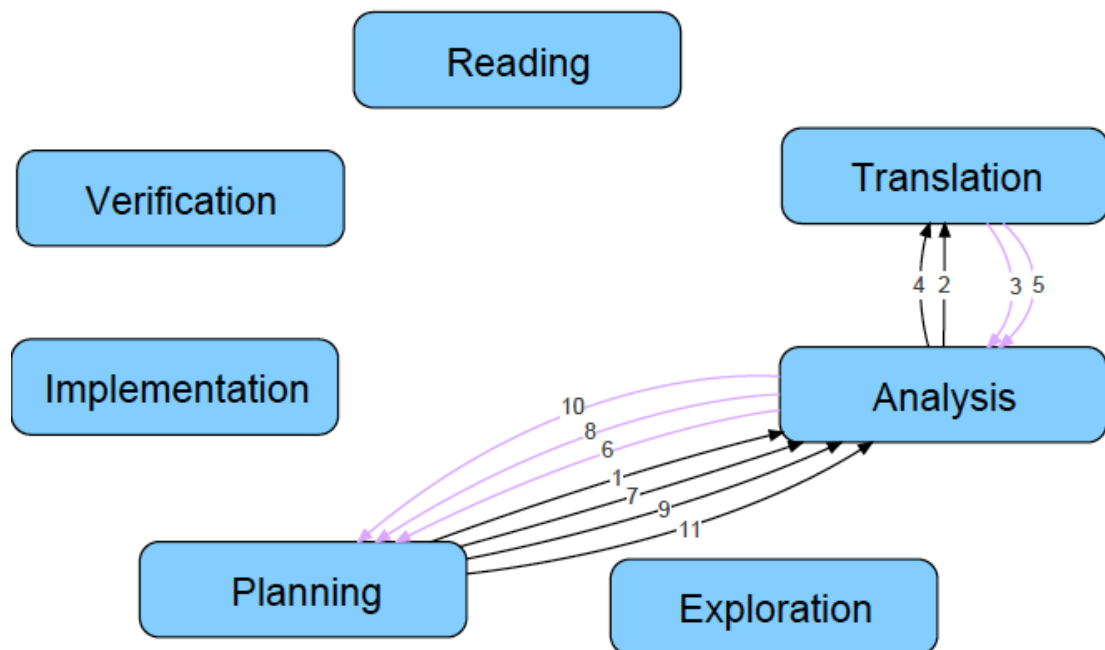
Student 1

Mark: 1/10



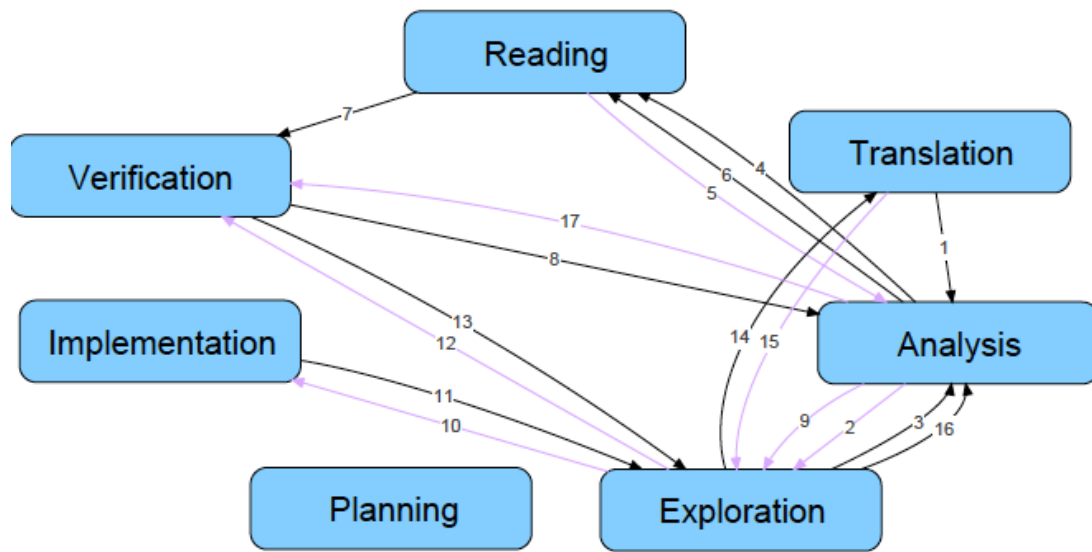
Student 2

Mark: 5/10



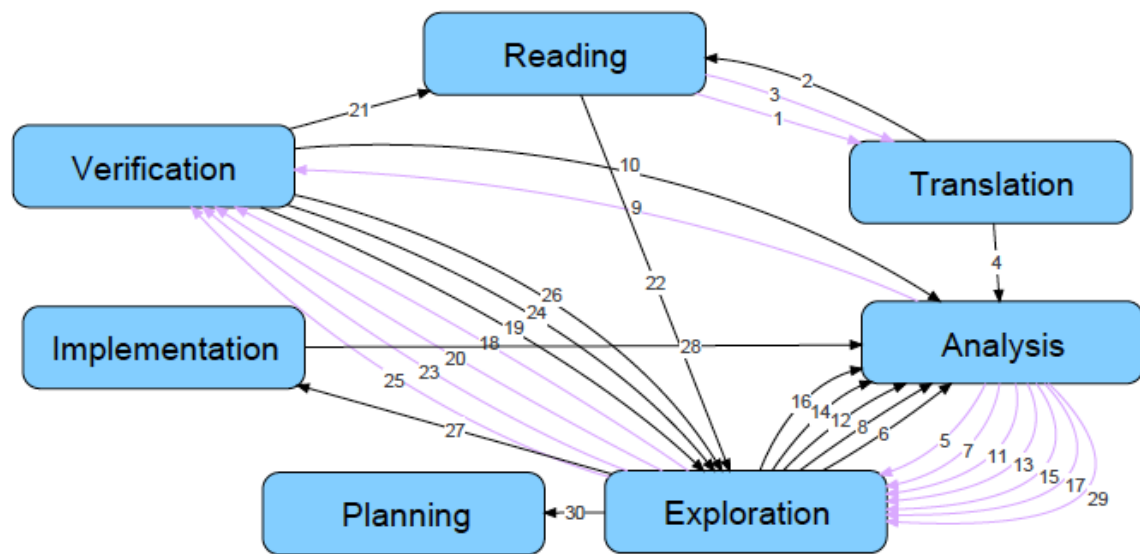
Student 3

Mark: 2/10



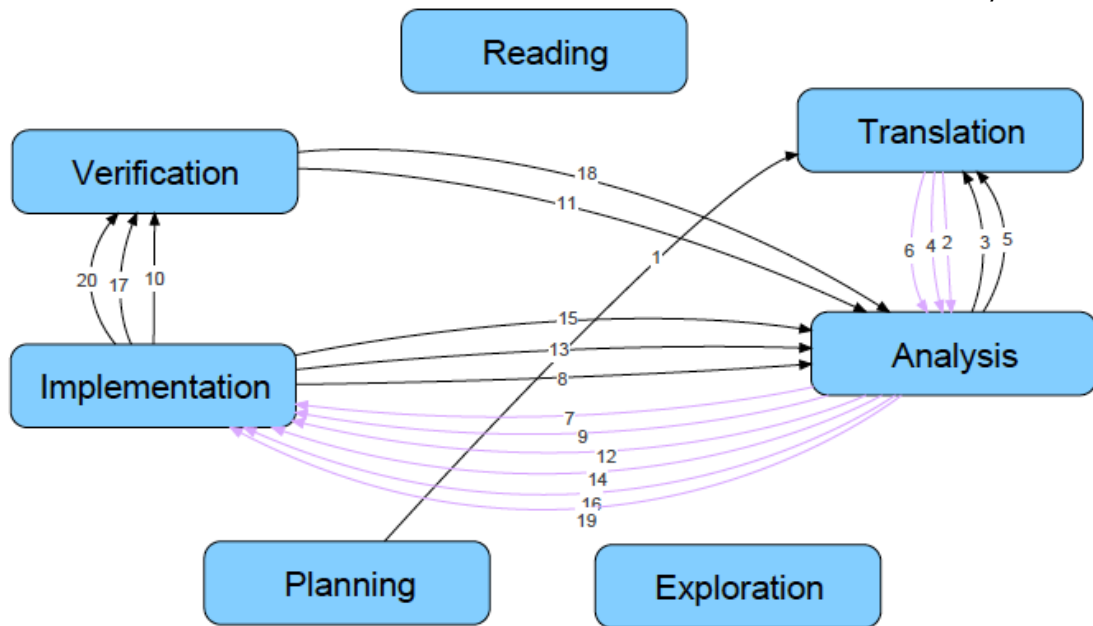
Student 4

Mark: 5/10



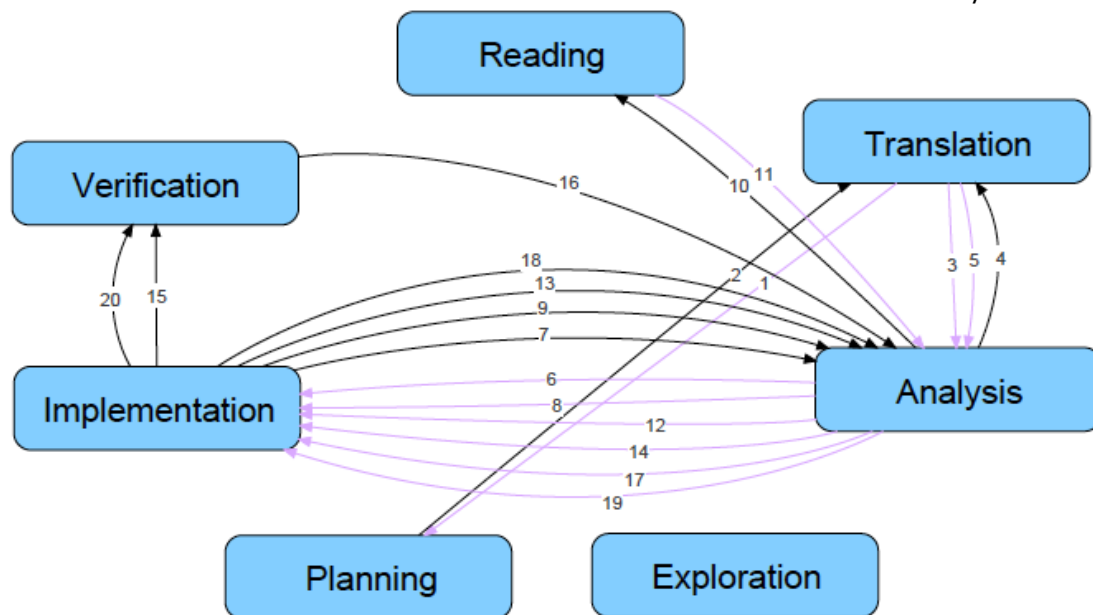
Student 5

Mark: 10/10



Student 6

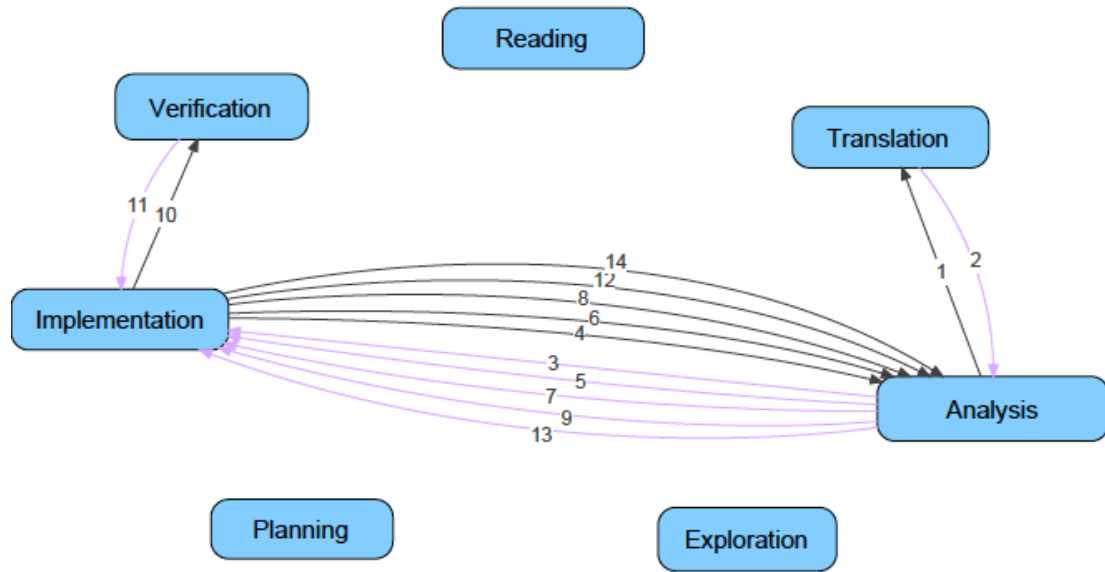
Mark: 8/10



2B STUDY – DRAWBRIDGE QUESTION

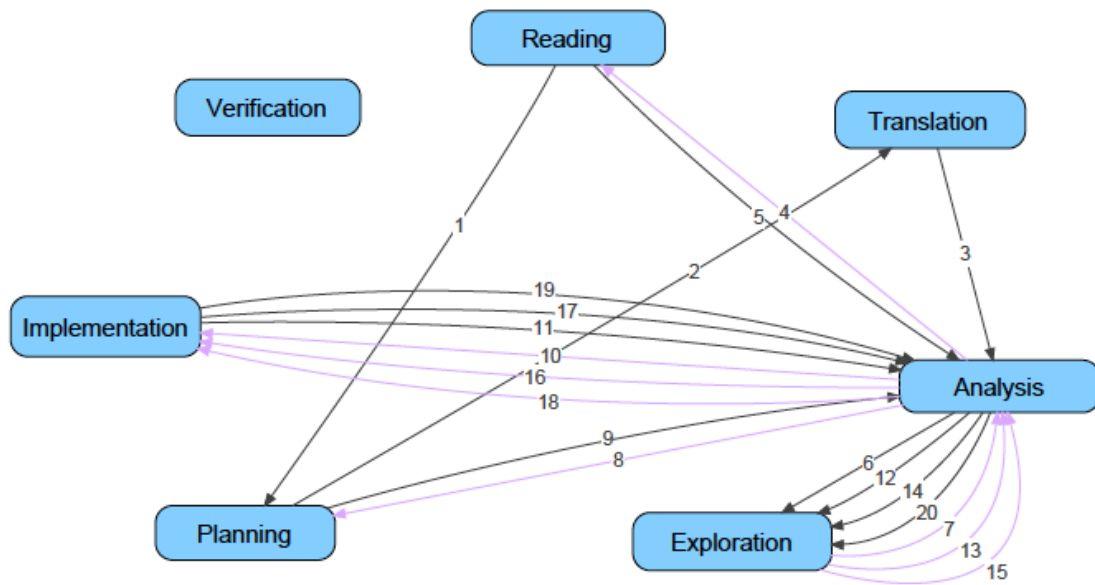
Student 1

Mark: 6/10



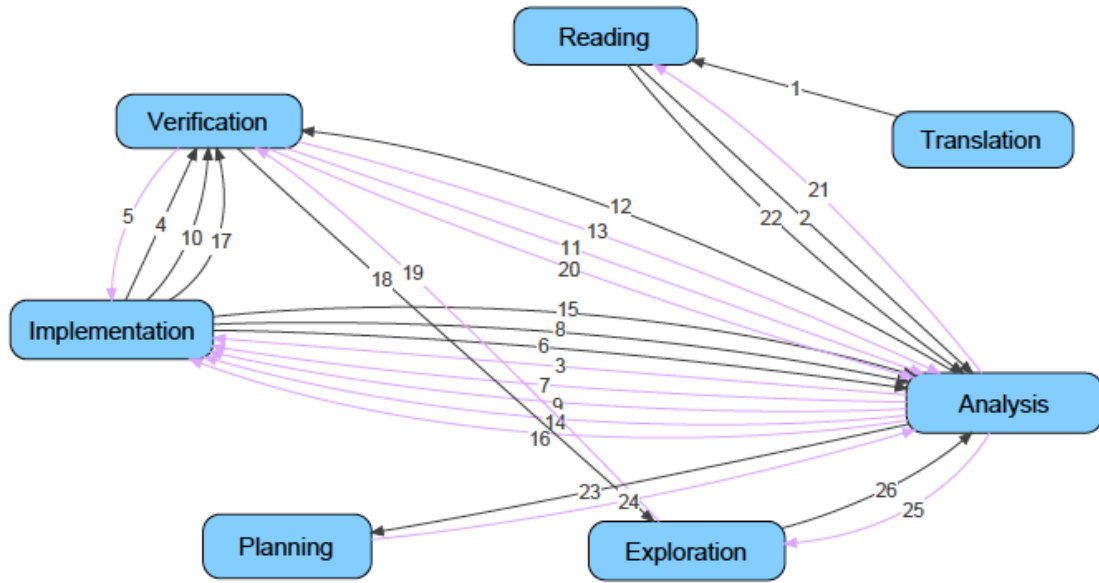
Student 2

Mark 2/10



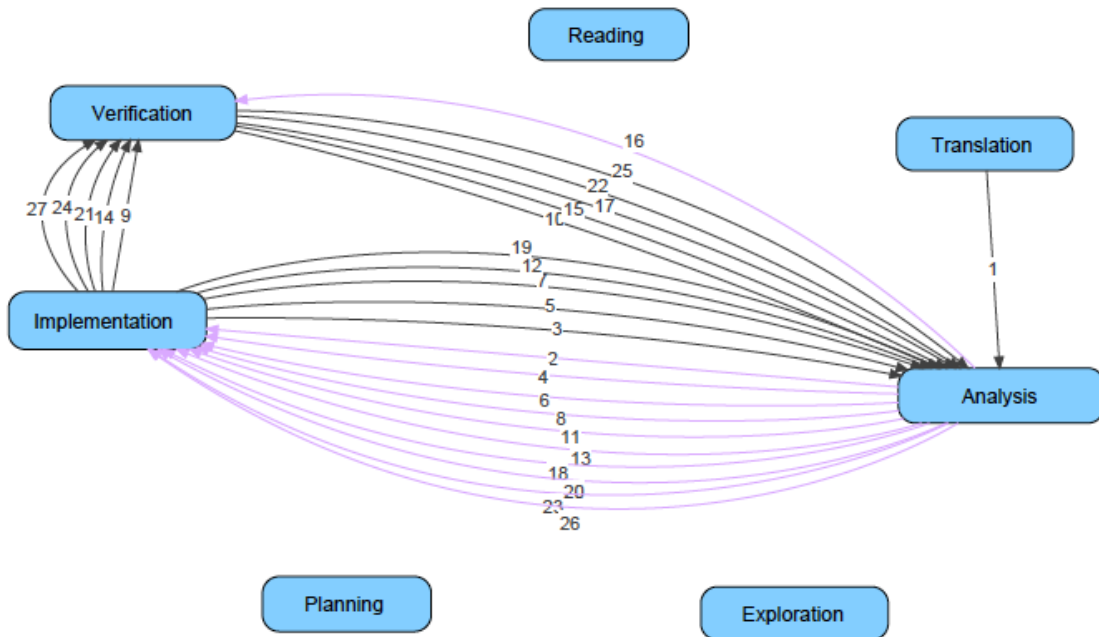
Student 3

Mark: 0/10



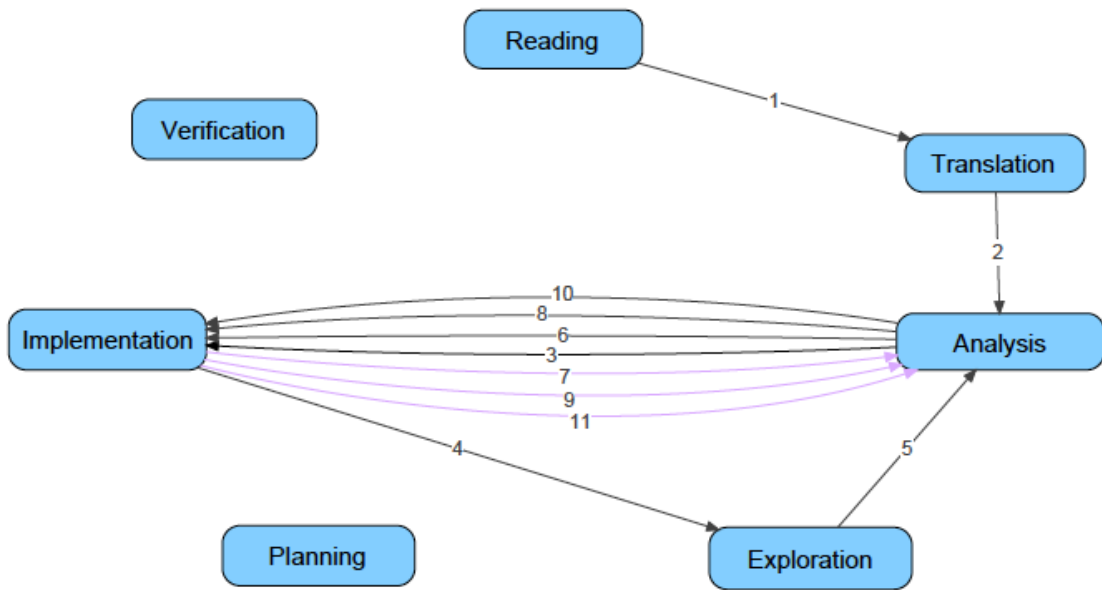
Student 4

Mark: 10/10



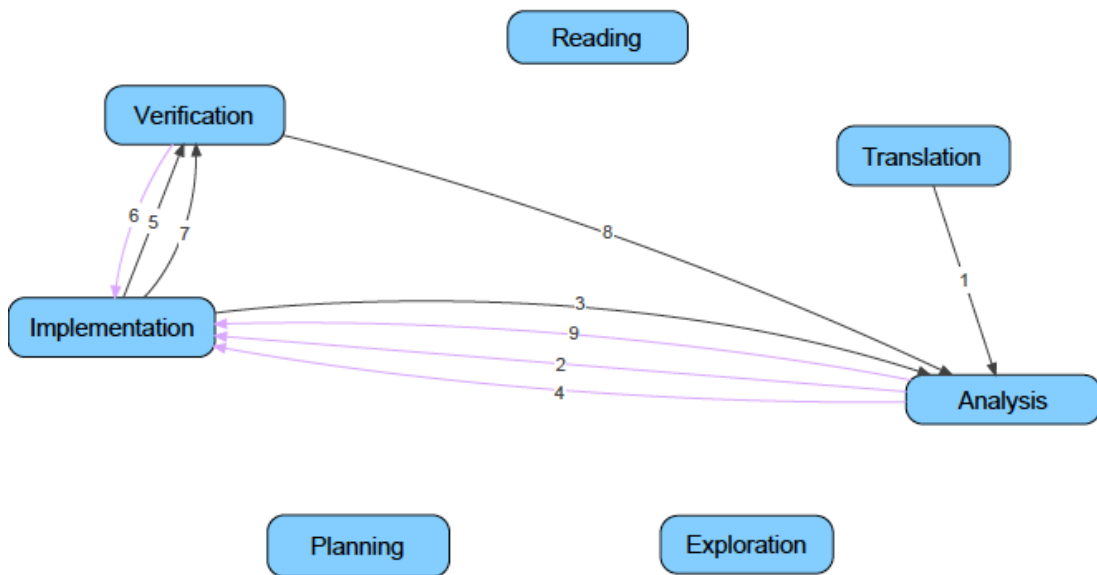
Student 5

Mark: 0/10



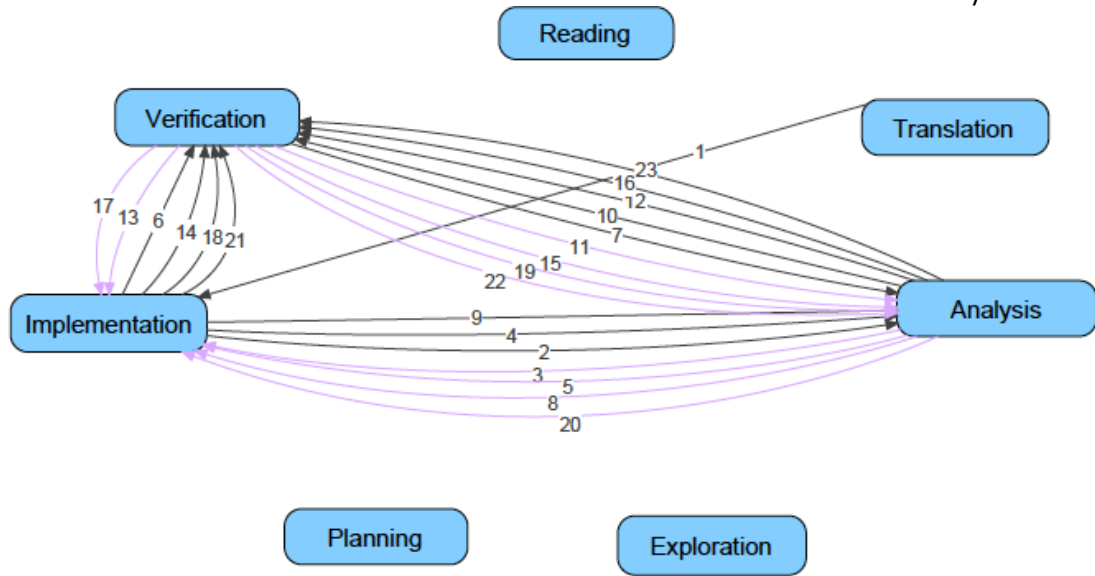
Student 6

Mark: 10/10



Student 7

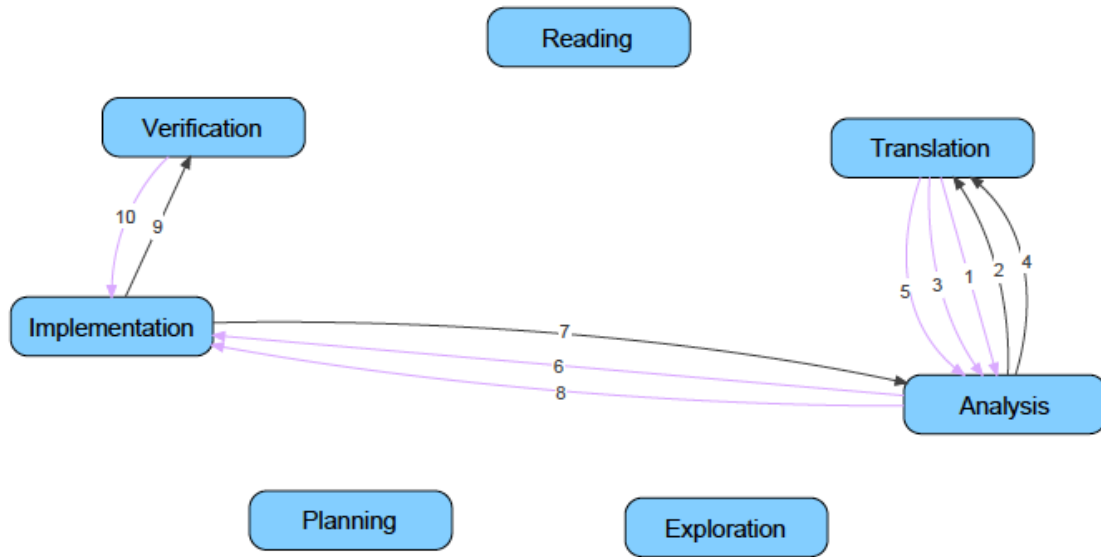
Mark: 3/10



2B STUDY – ARMY CADET QUESTION

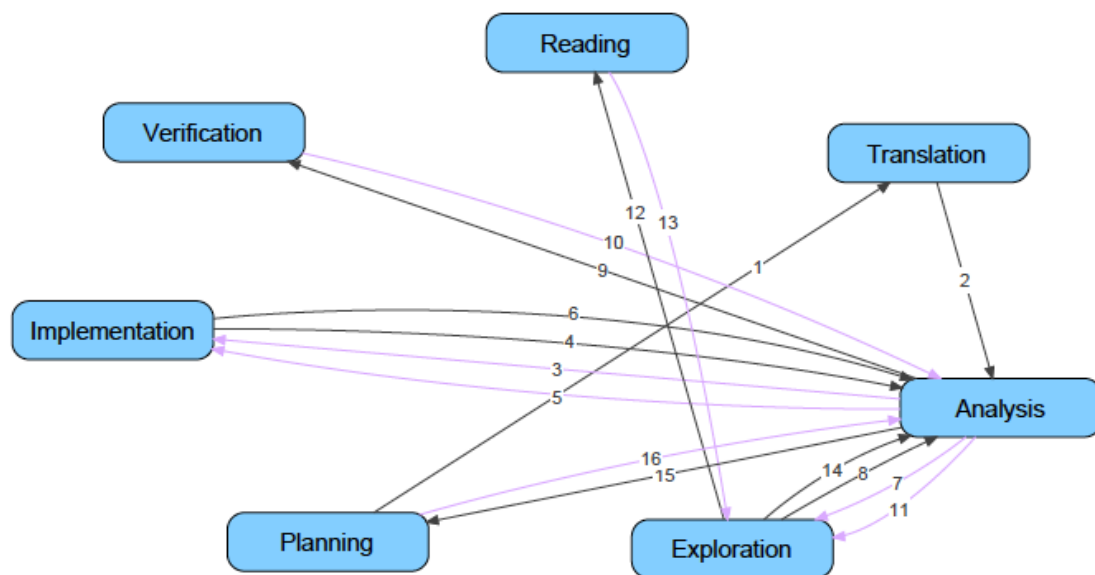
Student 1

Mark: 7/10



Student 2

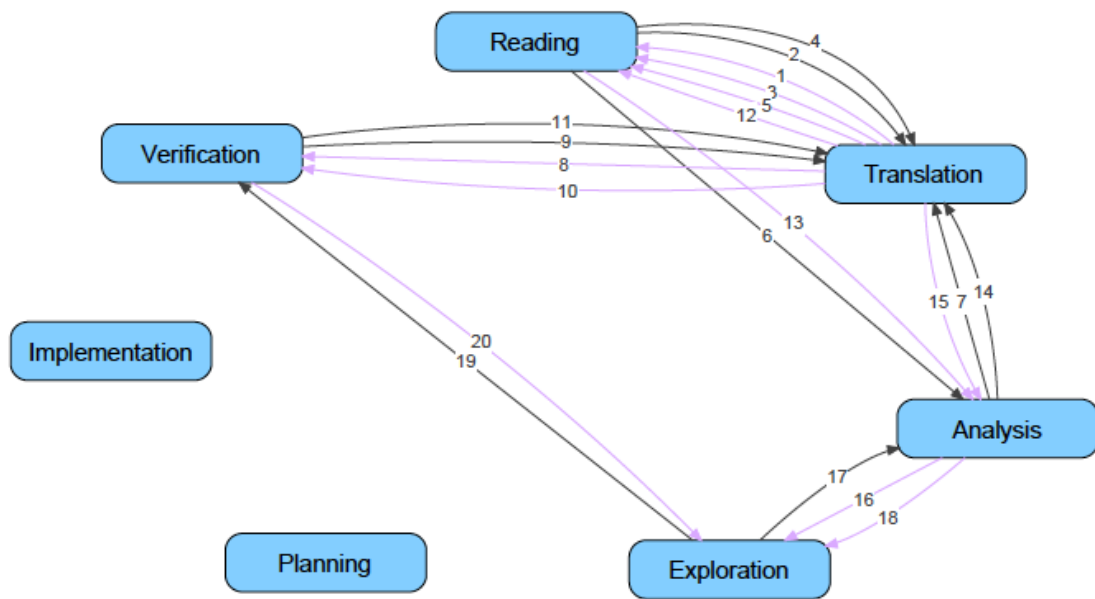
Mark: 4/10





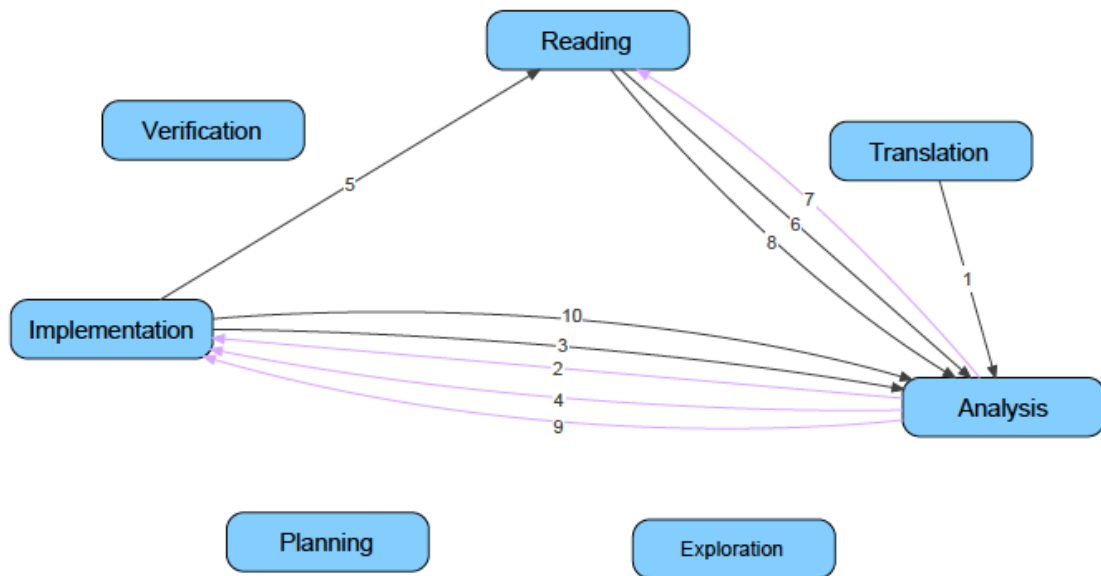
Student 3

Mark: 1/10



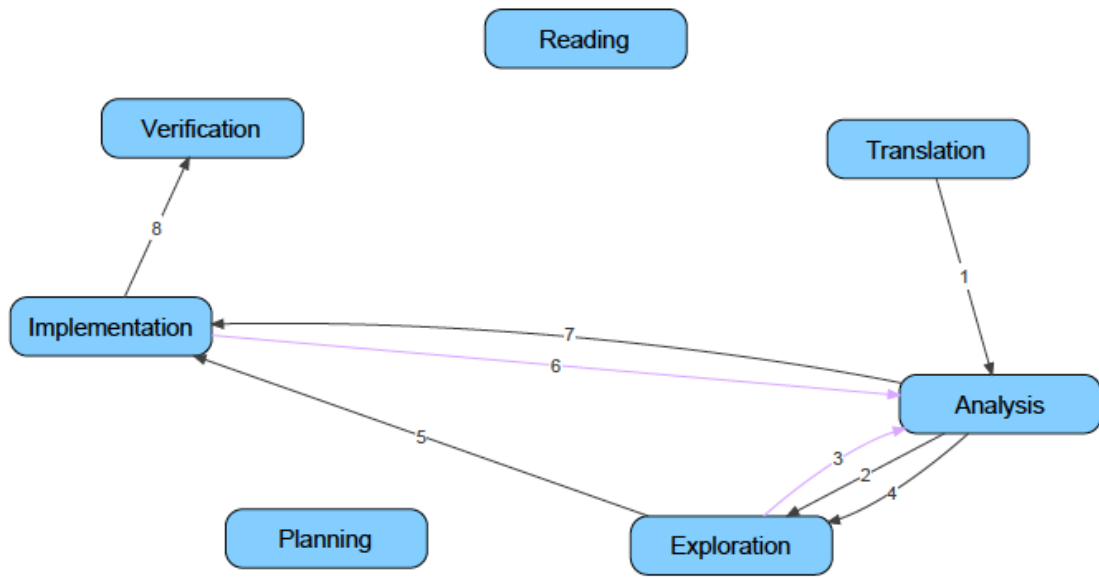
Student 4

Mark: 6/10



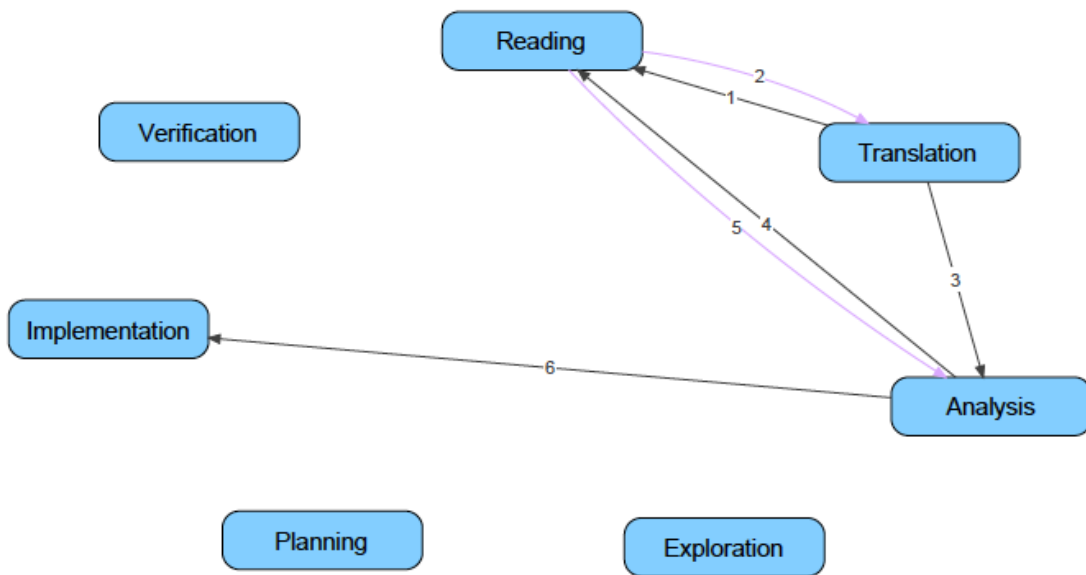
Student 5

Mark: 2/10



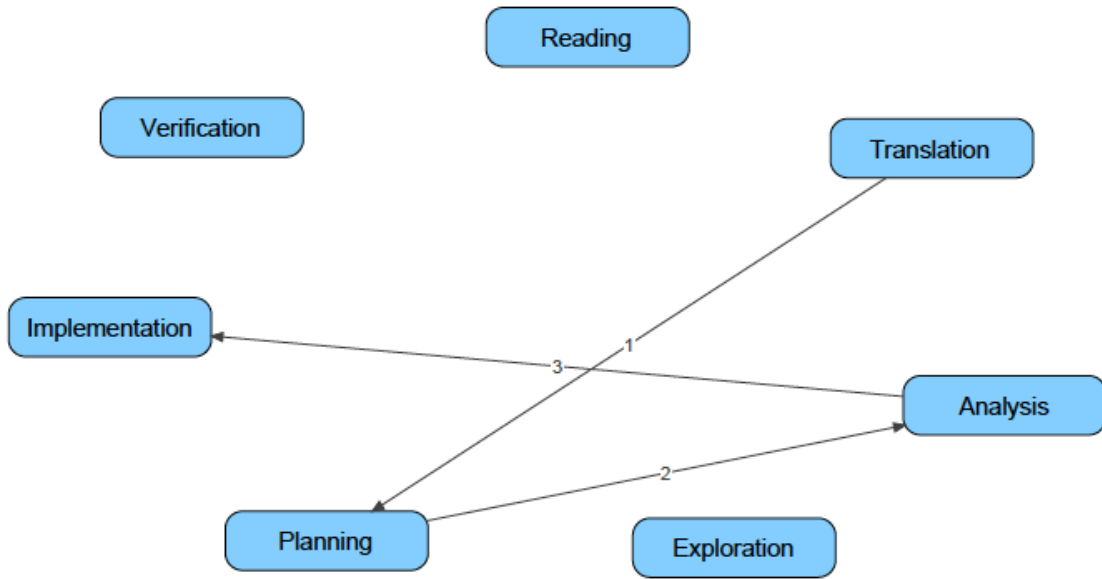
Student 6

Mark: 10/10



Student 7

Mark: 5/10



# Appendix B

## Smartpen Use: Code of Conduct



## Code of Conduct Please read all



Thank you for agreeing to take part in this study.

### Why are we doing this?

This is an ambitious study funded by a teaching development grant awarded by the Principal's Teaching Award Scheme of the University. We are looking at the ways students learn using the Livescribe smartpen technology. We are also interested in how students solve problems when using the pen.

Having a Livescribe smartpen will be beneficial to you, as you will have recordings and notes that you can listen back to from lectures, workshops and problems you solved at home. This can help with learning and later on with exam revision and deciding what information you want to take into the open-book exam.

The loan and your use of this pen involves certain responsibilities on your part. You must agree to perform the minimum usage and upload to a laptop as set out below. Below sets out what you can expect from us in terms of confidentiality and what we expect from you.

### Confidentiality

In order to keep track of the pens and ensure everyone returns them, your matriculation number will be linked to the pen number: the project officer (Marsali Wallace) will be the only person who has access to this data. For any outputs associated with the project (reports and any presentations, verbal or written), all the data will be anonymous and your name or matriculation number will not be linked to the data. Your performance on the pens will not in any way affect your course grade or mark on homework assignments.

We do not care if you get a particular question correct or not, we are looking at the process. There will be masses of data, so there will not be time to listen to everything in great detail. Just solve problems as you would normally and speaking if you want. See the example pencast for an example of a problem solved using the pen.

### How often to use your pen

If you want to stop using your pen for whatever reason, you don't have to tell us what the reason is, just hand the pen back to Marsali Wallace 1618 JCMB, Kings Buildings. However, if you do agree to use your pen we ask that you use it to record at least one physics thing a week. This thing could be a lecture, a problem you try, part of a workshop, making your Peerwise question, writing your revision notes, or any other physics related task.

You can use your pen for whatever you want, but please try to avoid uploading non-physics work onto the laptop in workshops. See the instructions or ask in workshops how to upload to your own computer and then delete the files off the pen.

#### When to back up your pen

We expect you to upload your data to a laptop ideally every week. There will be someone around in workshops to help you do this. Please do not leave it more than two weeks without uploading to the EdPER laptop, as your pen may get full up.

#### Attending workshops

If you do not attend a workshop where you should have backed up your pen, please do it at the next one. You can always arrange a time with Marsali Wallace (edper pens email) to upload out of workshop time, though this is only in particular circumstances and should not be used as a regular solution.

#### Keeping your pen, headphones and paper

Please keep your pen in good condition. A pencil case is provided for storing the pen, headphones, pen lid, spare ink and connector cable. Store the box and hand back everything you were given, except the paper. Hand back the box with pen, pen lid and connector cable, the pencil case and the headphones. There are different ear bud sizes for the headphones which should also be stored in the pencil case.

#### Returning your pen

Please return your pen in the last workshop to Marsali Wallace, or if you wish to keep it slightly longer please return to JCMB 1618 by

Please read the instructions provided before starting to use your pen and for more information see the Physics 1A course page on WebCT.

By signing this you agree to the above conditions of use.

Signed:

Print name:

Date:

# Appendix C

## The List of Technical Words

## Final List of Technical Words

Acceleration, accelerating  
Accurate  
Acting, act (in context e.g. gravity acts downwards)  
Amplitude  
angle  
Angular  
Anti-clockwise, clockwise  
Approximately  
Area  
At rest  
Average  
Axis  
Balance, balanced  
cancels, cancelled out  
Centre  
Centre of mass  
circle  
Charge, charged  
Coefficient  
Coefficient of friction  
Components  
Concentrated  
Conserved  
Constant  
Contact  
Cylinder  
Decimal places  
Decrease  
Density  
Direction  
Displaced, displacement  
Distance  
Distributed  
Divide, divided  
Downwards  
Electric  
Electrostatic  
Energy  
Equal  
Equally distributed  
Equilibrium  
Equation  
External  
Factor  
Final  
Finite  
Force/s, Reaction force  
Free body diagram  
Friction, frictional,  
Static Friction  
Frictionless  
Graph  
Gravitational

Gravity, -9.81  
Harmonic (motion)  
Hollow  
Horizontal, horizontally  
Hypotenuse  
Imprecise  
Inaccurate  
Increase  
Initially, initial  
Instantaneous  
Kilogram  
Kinetic  
Kinetic energy  
Launch  
Linear  
Magnitude  
Mass  
Maximum/minimum  
Minus  
Momentum, change in momentum  
Moment of inertia  
Multiplied, multiply  
Neglecting  
Newton's Law/s  
Normal contact force  
Observer  
Oppose movement  
Oscillation  
Over-precise  
Parabolic, parabola  
Particle  
Period  
Perpendicular  
Photon  
Positive, negative  
Potential  
Potential Energy  
Precise  
Projectile, projection  
Proportionality  
Pythagoras  
Radius  
Rate (only in context: constant rate or rate of rotation)  
Relativistic situation  
Revolution, Revolutions per second  
Resistance, air resistance  
Rotation, rotational, rotates  
Rotations per second  
scalar  
SHM, simple harmonic motion  
Sine, cosine  
Solid  
Slope

Speed  
Sphere  
Spiral  
Straight line  
Stationary  
Symbols, equations  
System  
Tension  
Terminal velocity  
Time  
Time of flight  
Torque  
Trajectory/ies  
Triangle  
Uniform  
Upwards  
Vector  
Velocity, change in velocity  
Vertically, vertical  
Weight  
Work, work done  
x times quicker  
Zero  
  
Symbols and equations  
each counted as one  
e.g.  $F=ma$ ,  $\mu_s$ .

If the word or phrase is on 1 line it counts as 1 technical word, over 2 lines are 2 technical words etc. So sometimes a phrase e.g. 'kinetic energy' counts as 1 technical word.



**Words NOT on list**

Either because not technical or not in the Course Hand book, but picked up in discussions coding:

Stopping

Times by

Change

Taut

Even

Point

Object

Relationship

Movement, Moving,

Motion

Operational space

Greater, Greatest,

Smallest

Less than

Perspective

Point of sliding

Size

Numbers, e.g. '3'

# Appendix D

## Expert Group Problem Solving

## D.1 Question and Answers from Expert Group Problem Solving

### Question One:

A small 50g block slides down a frictionless surface through height  $h = 10\text{cm}$  and then sticks to a uniform rod of mass 100 g and length 40 cm. The rod pivots about point O through angle  $\theta$  before momentarily stopping. Find  $\theta$ .

HINT: You can assume that there is no energy lost in the collision of the block with the rod.

### Answer:

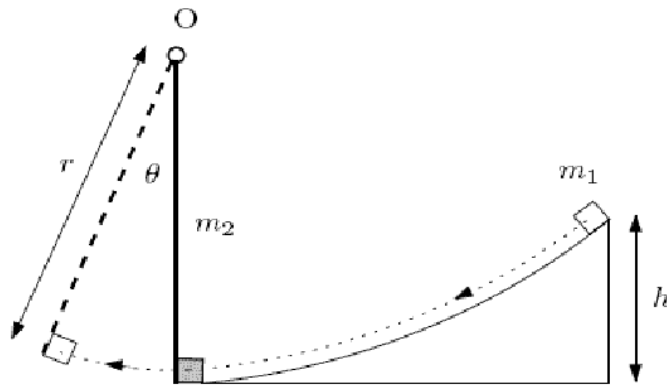


Figure D.1: Diagram for the block on a slope problem used in the expert group work

This problem uses conservation of energy. One way to solve this problem is as follows.

The potential energy of the system at the beginning of the problem (just before the block starts sliding), is equal to the potential energy of the system at the end (when it is momentarily stationary at the top of the motion).

$$m_1 = \text{block}$$

$$m_2 = \text{rod}$$

$r$  = length of the rod

In the initial position the potential energy of the block is  $m_1gh$ . When the rod pivots through angle  $\theta$ , the block reaches a height:

$$r - r\cos\theta = r(1 - \cos\theta)$$

And the centre of mass of the rod reaches a height:

$$\frac{r}{2} - \frac{r}{2}r\cos\theta = \frac{r}{2}(1 - \cos\theta)$$

Since the rod stops momentarily at that point, the kinetic energy is zero and so by using conservation of energy:

$$m_1gh = m_1gr(1 - \cos\theta) + \frac{m_2}{2}gr(1 - \cos\theta)$$

Cancel  $g$ 's and rearrange:

$$m_1h = r(1 - \cos\theta)(m_1 + \frac{m_2}{2})$$

$$\cos\theta = \frac{h}{r} \frac{m_1 + \frac{m_2}{2}}{m_1 + \frac{m_2}{2}}$$

Putting in the values  $h=0.1\text{m}$ ,  $r=0.4\text{m}$ ,  $m_1=0.05\text{kg}$ ,  $m_2=0.1\text{kg}$

$$\theta = 29 \text{ degrees}$$

### Question Two:

GRS 1915+105 (also known as V148 Aquilae) is a double-star system in the Milky Way Galaxy. One of the stars is a regular star but the other is a black hole surrounded by an accretion disk. The black hole is a microquasar, emitting intense X-rays and opposing jets of sub-atomic particles along the axis of rotation of the accretion disk. The speed of the ejected particles with respect to the black hole is 92.0% of the speed of light,  $c$ .

Determine the speed of one jet relative to the other.

HINTS: The speeds are a large fraction of the speed of light in a vacuum,  $c$ , so this is a relativistic problem.

The speed of light in a vacuum has the same value in all directions and in all inertial reference frames.

### Answer:

Consider two particles A and B ejected from the microquasar in opposite directions at a speed (as viewed from the quasar) of  $0.920c$ .

### Setting up the frames of reference:

$S$  is the frame of reference of the microquasar and  $S'$  is the frame of reference attached to particle A.

**Lorentz Transform Equations:**

If we take the sensible step of aligning the x axes,  $x$  and  $x'$  to be (anti)parallel to the motion of the particles, we can write,

$$x' = \gamma(x - vt) \quad (1)$$

where  $v$  is the speed of particle A as viewed from the quasar's frame of reference, S, and  $\gamma$  is the Lorentz factor,  $(1 - \frac{v^2}{c^2})^{-\frac{1}{2}}$ .

$$y' = y = \text{zero} \quad z' = z = \text{zero}$$

$$t' = \gamma(t - \frac{v}{c^2}x) \quad (2)$$

$u_{x'} = \frac{u_x - v}{1 - \frac{v}{c^2}u_x}$  is the Lorentz speed transformation.

$u_{x'}$  is the speed of B relative to A. This is the value we have been asked to find.  $v$  is the speed of A relative to S, which is  $0.920c$  and  $u_x$  is the speed of B relative to S, which is  $-0.920c$ . The minus sign denotes that B is travelling in the opposite direction to A.

Substituting these values into the Lorentz speed transformation:

$$u_{x'} = \frac{-1.840c}{1 + 0.0920^2}$$

$$u_{x'} = 0.997c = 298753316ms^{-1}$$

The value is less than  $c$  (check), but greater than  $0.920$ .

# Bibliography

- [1] IOP. The physics degree. Institute of Physics, June 2010.
- [2] P.B. Kohl and N.D. Finkelstein. Patterns of multiple representation use by experts and novices during physics problem solving. *Physical Review Special Topics - Physics Education Research*, 4(1), Jan-Jun 2008.
- [3] P. Heller, R. Keith, and S. Anderson. Teaching problem solving through cooperative grouping. part 1: Group versus individual problem solving. *Am. J. Phys.*, 60(7):627–636, July 1992.
- [4] P. Heller and M. Hollabaugh. Teaching problem solving through cooperative grouping. part 2: Designing problems and structuring groups. *Am. J. Phys.*, 60(7):637–644, July 1992.
- [5] W.K. Adams. *Development of a Problem Solving Evaluation Instrument; untangling of specific problem solving skills*. PhD thesis, University of Colorado, 2007.
- [6] K.A. Ericsson and H.A. Simon. *Protocol Analysis. Verbal reports as data*. Cambridge, Massachusetts, 1984.
- [7] I. S. Robertson. *Problem Solving*. Psychology Press, USA, 2001.
- [8] F. Reif. *Applying Cognitive Science to Education*. The MIT Press, 2008.
- [9] J. Larkin and F. Reif. Understanding and teaching problem-solving in physics. *Int. J. Sci. Educ.*, 1(2):191–203, 1979.
- [10] R.M. Garrett, D. Satterly, D.G. Perez, and J. Martineztorregrosa. Turning exercises into problems - An experimental study with teachers in training. *International Journal of Science Education*, 12(1):1–12, Jan-Mar 1990.
- [11] J.L. Docktor. *Development and Validation of a Physics Problem-Solving Assessment Rubric*. PhD thesis, University of Minnesota, 2009.
- [12] D.H. Jonassen. Toward a design theory of problem solving. *ETR&D-Educational Technology Research and Development*, 48(4):63–85, 2000.
- [13] D.H. Jonassen. *Learning to Solve Complex Scientific Problems*. Taylor and Francis Group, USA, 2007.

- [14] A. Johnstone. *Creative problem solving in chemistry: Solving problems through effective group work*, chapter Introduction. London: Royal Society of Chemistry, 1993.
- [15] C. Wood. The development of creative problem solving in chemistry. *Chemistry Education Research and Practice*, (7):96–113, 2006.
- [16] T.J. Nokes, C.D. Schunn, and M.T.H. Chi. Problem solving and human expertise. In Penelope Peterson, Eva Baker, and Barry McGaw, editors, *International Encyclopedia of Education*, pages 265 – 272. Elsevier, Oxford, 2010.
- [17] D. Fortus. The importance of learning to make assumptions. *Science Education*, 93(1):86–108, 2009.
- [18] P.C. Candy and R.G. Crebert. Ivory tower to concrete jungle - The difficult transition from the academy to the workplace as learning environments. *Journal of Higher Education*, 62(5):570–592, Sep-Oct 1991.
- [19] P.T. Hardiman, R. Dufresne, and J.P. Mestre. The relation between problem categorization and problem-solving among experts and novices. *Memory & Cognition*, 17(5):627–638, SEP 1989.
- [20] M.T.H. Chi, P.J. Feltovich, and R. Glaser. Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2):121–152, 1981.
- [21] W. Leonard, R. Dufresne, and J. Mestre. Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems. *Am. J. Phys.*, 64(12):1495–1503, December 1996.
- [22] A. Van Heuvelen. Overview, case study physics. *Am. J. Phys.*, 59(10):898–907, October 1991.
- [23] L. Ding, N. Reay, A. Lee, and L. Bao. Exploring the role of conceptual scaffolding in solving synthesis problems. *Physical Review Special Topics - Physics Education Research*, 7(2):020109 – 1–020109–11, 2011.
- [24] A.H. Schoenfeld. *Mathematical Problem Solving*. Academic Press, Florida, 1985.
- [25] A. Van Heuvelen. Learning to think like a physicist: A review of research-based instructional strategies. *Am. J. Phys.*, 59(10):891–897, October 1991.
- [26] D.H. Jonassen. Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *ETR&D-Educational Technology Research and Development*, 45(1):65–94, 1997.
- [27] M.S. Sabella and E.F. Redish. Knowledge organization and activation in physics problem solving. *American Journal of Physics*, 75(11):1017–1029, 2007.
- [28] R.E. Mayer. *Thinking, Problem Solving, Cognition*. W.H. Freeman and Company, 1983.

- 
- [29] C. Singh. When physical intuition fails. *American Journal of Physics*, 70(11):1103–1109, NOV 2002.
- [30] P.B. Kohl, D. Rosengrant, and N. Finkelstein. Strongly and weakly directed approaches to teaching multiple representation use in physics. *Phys. Rev. ST Phys. Educ. Res.*, 3(1):010108, June 2007.
- [31] P.B. Kohl and N.D. Finkelstein. Effect of instructional environment on physics students' representational skills. *Phys. Rev. ST Phys. Educ. Res.*, 2(1):010102, Jan 2006.
- [32] P.B. Kohl and N.D. Finkelstein. Student representational competence and self-assessment when solving physics problems. *Physical Review Special Topics - Physics Education Research*, 1(1), Jul-Dec 2005.
- [33] M. Suwa and B. Tversky. External Representations contribute to the dynamic construction of ideas. In Hegarty, M. and Meyer, B. and Narayanan, H, editor, *Diagrammatic Representation and Inference, Proc. Inference. 2002 Lecture Notes in Artificial Intelligence series*, pages 341–343. Springer, 2002.
- [34] E. Bagno, B.S. Eylon, and U. Ganiel. From fragmented knowledge to a knowledge structure: Linking the domains of mechanics and electromagnetism. *American Journal of Physics*, 68(7):S16 – S26, 2000.
- [35] F. Reif and J.I. Heller. Knowledge structure and problem solving in physics. *Educational Psychologist*, 17(2):102 – 27, 1982.
- [36] I.D. Beatty and W.J. Gerace. Probing physics students' conceptual knowledge structures through term association. *American Journal of Physics*, 70(7):750 – 758, 2002.
- [37] W. Gerace. Problem solving and conceptual understanding. In *Physics Education Research Conference 2001*, PER Conference Invited Paper, Rochester, New York, July 25-26 2001.
- [38] W.J. Leonard, W.J. Gerace, and R.J. Dufresne. Concept-based problem solving. *UMPERG Technical Report PERG-1999,12-Nov*, pages 3–18, 1999.
- [39] E. F. Redish. A theoretical framework for physics education research: Modeling student thinking. *Proceedings - International School of Physics Enrico Fermi*, 156:1 – 64, 2004.
- [40] A.A. diSessa. Toward an epistemology of physics. *Cognition and Instruction*, 10(2/3):pp. 105–225, 1993.
- [41] L. Vygotsky. *Mind and Society*, chapter Interaction between learning and development., pages 79–91. Harvard University Press, Cambridge, MA., 1978.
- [42] K. Heller and P. Heller. *The Competent Problem Solver*. University of Minnesota (A locally produced textbook), Minnesota, 1995.



- [43] T. Gok. The general assessment of problem solving processes and metacognition in physics education. *Eurasian J. Phys. Chem. Educ.*, 2(2):110–122, 2010.
- [44] D. Hestenes, M. Wells, and G. Swackhame. Force concept inventory. *Phys. Teach.*, 30(3):141–158, March 1992.
- [45] R. Lippmann. *Students' Understanding of Measurement and Uncertainty in the Physics Laboratory: Social Construction, Underlying Concepts and Quantitative Analysis*. PhD thesis, University of Maryland, 2003.
- [46] D. May and E. Etkina. College physics students' epistemological self-reflection and its relationship to conceptual learning. *Am. J. Phys.*, 70(12):1249–1258, December 2002.
- [47] A. Mason and C. Singh. Do advanced physics students learn from their mistakes without explicit intervention? *American Journal of Physics*, 78, 2010.
- [48] A. Mason and C. Singh. Using Reflection with Peers to Help Students Learn Effective Problem Solving Strategies. In *2010 Physics Education Research Conference*, volume 1289 of *AIP Conference Proceedings*, pages 41–44, Melville, NY, USA, 2010. Amer Assoc Phys Teachers, Amer Inst Physics.
- [49] A.L. Brown and A.S. Palinesar. *Knowing, learning, and instruction : essays in honor of Robert Glaser*, chapter Guided, Cooperative Learning and Individual Knowledge Acquisition, pages 393–451. Hillsdale, N.J. : L. Erlbaum Associates., 1989.
- [50] C. Howe, A. Tolmie, A. Thurston, K. Topping, D. Christie, K. Livingston, E. Jessiman, and C. Donaldson. Group work in elementary science: Towards organisational principles for supporting pupil learning. *Learning and Instruction*, 17(5):549 – 563, 2007.
- [51] Neville Bennett and Allyson Cass. The effects of group composition on group interactive processes and pupil understanding. *British Educational Research Journal*, 15(1):pp. 19–32, 1989.
- [52] E. Harskamp, N. Ding, and C. Suhre. Group composition and its effect on female and male problem-solving in science education. *Educational Research*, 50(4):307 – 318, 2008.
- [53] K.L. Nielsen, G. Hansen-Nygaard, and J.B. Stav. *Student Response Systems in Science and Engineering Education*, chapter D in Nielsen, K.L PhD Thesis: How the Initial Thinking Period Affects Student Argumentation during Peer Instruction: Students' Experiences versus Observations. 2012.
- [54] C. Turpen and N.D. Finkelstein. Not all interactive engagement is the same: Variations in physics professors' implementation of *Peer Instruction*. *Phys. Rev. ST Phys. Educ. Res.*, 5:020101, Aug 2009.

- [55] C. Turpen and N.D. Finkelstein. The construction of different classroom norms during peer instruction: Students perceive differences. *Physical Review Special Topics Physics Education Research*, 6(2):20123, 2010.
- [56] N. Mercer. *Words and Minds: how we use language to think together*. Routledge, London, 2000.
- [57] D. Barnes. *Exploring talk in schools : inspired by the work of Douglas Barnes / edited by Neil Mercer and Steve Hodgkinson.*, chapter Exploratory Talk for Learning, pages 1–15. London : SAGE, 2008., 2008.
- [58] J. Osborne. Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328(5977):463 – 466, 2010.
- [59] D. Christie, A. Tolmie, A. Thurston, C. Howe, and K. Topping. Supporting group work in scottish primary classrooms: improving the quality of collaborative dialogue. *Cambridge Journal of Education*, 39(1):141–156, 2009.
- [60] Jr. Perry, W. G. *Forms of Intellectual and Ethical Development in the College Years: A Scheme. Jossey-Bass Higher and Adult Education Series*. 1999.
- [61] J. J. Wellington and J. Osborne. *Language and literacy in science education / Jerry Wellington and Jonathan Osborne*. Buckingham : Open University Press, 2001., 2001.
- [62] M.P Farrell and F Ventura. Words and understanding in physics. *Language and Education*, 12:243–253, 1998.
- [63] S.F. Itza-Ortiz, N.S. Rebello, D.A. Zollman, and M. Rodriguez-Achach. The vocabulary of introductory physics and its implications for learning physics. *The Physics Teacher*, 41(6):330–336, 2003.
- [64] E. Mazur. *Peer Instruction: A User's Manual*. Series in Educational Innovation. Prentice Hall, 1997.
- [65] L.C. McDermott and P.S. Shaffer. *Tutorials in introductory physics / Lillian C. McDermott, Peter S. Shaffer, and the Physics Education Group, Department of Physics, University of Washington*. Prentice Hall series in educational innovation. Upper Saddle River, N.J. : Prentice Hall, c1998., 1998.
- [66] R.R. Hake. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1):64–74, JAN 1998.
- [67] M. Prince. Does active learning work? a review of the research. *Journal of Engineering Education*, 93(3):223 – 231, n.d.
- [68] J. Michael. Where's the evidence that active learning works?. *Advances in Physiology Education*, 30(4):159 – 167, 2006.

- [69] C.H. Crouch and E. Mazur. Peer instruction: Ten years of experience and results. *Am. J. Phys.*, 69:970–977, 2001.
- [70] C.H. Crouch, J. Watkins, A.P. Fagen, and E. Mazur. Peer instruction: Engaging students one-on-one, all at once. In E.F. Redish and P. Cooney, editors, *Reviews in Physics Education Research*, pages 1–1. American Association of Physics Teachers, 2007.
- [71] M.K Smith, W.B Wood, W.K Adams, C Wieman, J.K Knight, N Guild, and T.T Su. Why peer discussion improves student performance on in-class concept questions. *Science*, 323(5910):122 – 124, 2009.
- [72] P. Tynjälä. Towards expert knowledge? a comparison between a constructivist and a traditional learning environment in the university. *International Journal of Educational Research*, 31:357 – 442, n.d.
- [73] E. Mazur. Peer instruction: Getting students to think in class. In Edward F. Redish and John S. Rigden, editors, *The Changing Role of Physics Departments in Modern Universities, Part Two: Sample Classes*, AIP Conference Proceedings, pages 981–988. American Institute of Physics, 1997.
- [74] K.L. Nielsen and J.B. Stav. *Student Response Systems in Science and Engineering Education*, chapter F in Nielsen, K.L PhD Thesis: Student Response Systems in Physics Lectures: Does Voting Results Represent a Correct Image of Student Understanding? 2012.
- [75] I.D. Beatty, W.J. Gerace, W.J. Leonard, and R.J. Dufresne. Designing effective questions for classroom response system. *American Journal of Physics*, 74(1):31–39, 2006.
- [76] M.C. James and S. Willoughby. Listening to student conversations during clicker questions: What you have not heard might surprise you! *American Journal of Physics*, 79(1):123–132, 2011.
- [77] M.C. James, F. Barbieri, and P. Garcia. What are they talking about? Lessons learned from a study of peer instruction. *Astronomy Education Review*, 7(1):37–43, Feb-Dec 2008.
- [78] S. Kaartinen and K. Kumpulainen. Collaborative inquiry and the construction of explanations in the learning of science. *Learning and Instruction*, 12:189 – 212, n.d.
- [79] M.C. James. The effect of grading incentive on student discourse in peer instruction. *American Journal of Physics*, 74(8):689–691, 2006.
- [80] S.D. Willoughby and E. Gustafson. Technology talks: Clickers and grading incentive in the large lecture hall. *American Journal of Physics*, 77(2):180–183, 2009.

- [81] J. B. Williams and A. Wong. The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology*, 40(2):227–236, 2009.
- [82] F. Bacon. Open book examinations. *Education + Training*, 11(9):363–363, October 1969.
- [83] P. Race. Designing assessment to improve physical sciences learning. The Higher Education Academy - Physical Sciences Centre, March 2009.
- [84] R. Brightwell, J. Daniel, and A. Stewart. Evaluation: Is an open book examination easier?. *Bioscience Education e-Journal*, 3, 2004.
- [85] M.S. Gupta. Open-book examinations for assessing higher cognitive abilities [educator’s corner]. *IEEE Microwave Magazine*, 8(6):46, 2007.
- [86] J. Tuminaro and E.F. Redish. Elements of a cognitive model of physics problem solving: Epistemic games. *Physical Review Special Topics - Physics Education Research*, 3(2), Jul - Dec 2007.
- [87] W.K. Adams and C.E. Wieman. Problem solving skill evaluation instrument - Validation studies. In McCullough, L and Hsu, L and Heron, P, editor, *2006 Physics Education Research Conference*, volume 883 of *AIP Conference Proceedings*, pages 18–21, NY, USA, 2007. Amer Assoc Phys Teachers, Amer Inst Physics. Physics Education Research Conference, Syracuse, NY, JUL 26-27, 2006.
- [88] U. Flick. *An introduction to qualitative research*. SAGE, Los Angeles ; London, 2009.
- [89] M.W. Van Someren, Y.F. Barnard, and J.A.C. Sandberg. *The think aloud method*. Academic Press, London, University of Amsterdam, 1994.
- [90] J. Nielson, T. Clemmensen, and C. Yssing. Getting access to what goes on in people’s heads? - reflections on the think-aloud technique. *NordiCHI*, pages 19–23, October 2002.
- [91] D. Laurillard. *The Experience of Learning: Implications for teaching and studying in higher education*, chapter Styles and Approaches in Problem-Solving, pages 126–144. Number Chapter 8. University of Edinburgh, 3rd edition, 2005.
- [92] B. Kawulich. Participant observation as a data collection method. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2), 2005.
- [93] Livescribe. <http://www.livescribe.com/uk/>, [Accessed: 12 June 2013].
- [94] A. Schaack. Livescribe in k-12 education: Research support - a review of scientific evidence demonstrating the effectiveness of smartpen technologies for improving teaching and learning. online [http : //www.livescribe.com/uk/media/pdf/education/LivescribeK – 12ResearchSupport.pdf](http://www.livescribe.com/uk/media/pdf/education/LivescribeK-12ResearchSupport.pdf) [Accessed 24 June 2013], March 2009.

- [95] J.R. Boyle. Note-taking and secondary students with learning disabilities: Challenges and solutions. *Learning Disabilities Research and Practice*, 27(2):90 – 101, 2012.
- [96] N. Mercer. The analysis of classroom talk: Methods and methodologies. *British Journal of Educational Psychology*, 80(1):1–14, 2010.
- [97] B. Glaser and A. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Reserch*. Aldine, Chicago, Jun 1 1967.
- [98] A. Bryman. *Social Research Methods*. Oxford University Press, Oxford, 2008.
- [99] J. Potter and M. Wetherell. *Discourse and Social Psychology*. Sage Publications, London, 1987.
- [100] F. Pajares and D. M. Miller. Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, 86(2):193–203, 1994.
- [101] K. Hogan, B.K. Nastasi, and M. Pressley. Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction*, 17(4):pp. 379–432, 1999.
- [102] K. Charmaz. *Constructing grounded theory: a practical guide through qualitative analysis*. Sage Publications, London, 2006.
- [103] D. Halliday, R. Resnick, and J. Walker. *Fundamentals of Physics*. Wiley, USA, 7th edition, 2005.
- [104] D. Sands and T. Overton. Cognitive psychology and problem solving in the physical sciences. *New Directions*, (6):21–26, September 2010.
- [105] The School of Physics and The University of Edinburgh Astronomy. Physics 1a: Foundations, student course handbook, 2012/2013.
- [106] A. P. Field. *Discovering statistics using SPSS : (and sex and drugs and rock 'n' roll) / Andy Field*. Introducing statistical methods. Los Angeles, Calif. ; London : SAGE Publications, c2009., 2009.
- [107] Tufts University. Visual understanding environment. <http://vue.tufts.edu/> [16 April 2013].
- [108] J Meijer, M. V. J. Veenman, and B. H. A. M. van Hout-Wolters. Metacognitive activities in text-studying and problem-solving: Development of a taxonomy. *Educational Research and Evaluation*, 12(3):209 – 237, 2006.
- [109] A.K. Barbey and L.W. Barsalou. Reasoning and problem solving: Models. In Larry R. Squire, editor, *Encyclopedia of Neuroscience*, pages 35 – 43. Academic Press, Oxford, 2009.

- 
- [110] M. Prensky. Digital natives, digital immigrants part 1. *On the Horizon*, 9(5):1, 2001.
- [111] M. Prensky. Digital natives, digital immigrants part 2: Do they really think differently?. *On the Horizon*, 9(6):1, 2001.
- [112] S. Bennett and K. Maton. Beyond the "digital natives" debate: Towards a more nuanced understanding of students' technology experiences. *Journal of Computer Assisted Learning*, 26(5):321 – 331, 2010.
- [113] A.A. diSessa and B.L. Sherin. What changes in conceptual change?. *International Journal of Science Education*, 20(10):1155 – 91, 1998.
- [114] S. Bates and K. Brunton. Closing the feedback loop for clicker questions. *New Directions in the Teaching of Physical Sciences*, (6):30–33, September 2010.
- [115] B.J. Brooks and M.D. Koretsky. The influence of group discussion on students' responses and confidence during peer instruction. *Journal of Chemical Education*, 88 (11):1477–1484, Sep 2011.
- [116] R.R. Hake. Interactive engagement versus traditional methods: A six thousand student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1):64 – 74, 1998.
- [117] A.L. Strauss. *Constant Comparison: Qualitative Analysis for Social Scientists*. Cambridge University Press, New York, 1987.
- [118] A.H. Schoenfeld. *Handbook for Research on Mathematics Teaching and Learning*, chapter Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics, pages 334–370. MacMillan, New York, 1992.
- [119] J. E. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46):16569–16572, nov 2005.
- [120] G. Upton and I. Cook. *A Dictionary of Statistics*. Oxford paperback reference. Oxford ; New York : Oxford University Press, 2008., 2008.
- [121] H.T. Williams. Semantics in teaching introductory physics. *American Journal of Physics*, 67(8):670–680, 1999.
- [122] J. J. Wellington and G. Ireson. *Science learning, science teaching / Jerry Wellington and Gren Ireson*. London ; New York : Routledge, 2008., 2008.
- [123] K.L. Nielsen. *Student Response Systems in Science and Engineering Education*. PhD thesis, Norwegian University of Science and Technology, 2012.
- [124] J. N. Webb. *Game theory : decision, interaction, and evolution*. Springer undergraduate mathematics series. [London] : Springer., 2007.

- [125] K.L. Nielsen, G. Hansen-Nygaard, and J.B. Stav. Investigating peer instruction: How the initial voting session affects students' experiences of group discussion. *ISRN Education*, 2012.
- [126] P.B. Kohl and N.D. Finkelstein. Expert and novice use of multiple representations during physics problem solving. In *2007 Physics Education Research Conference*, volume 951 of *AIP Conference Proceedings*, pages 132–135, NY, USA, 2007. Amer Assoc Phys Teachers, Amer Inst Physics.
- [127] J. Larkin, J. McDermott, D.P. Simon, and H.A. Simon. Expert and Novice Performance in Solving Physics Problems. *Science*, 208(4450):1335–1342, 1980.

## Publications

M.Wallace and R. Galloway. 'Using Smartpen Technology to Observe Student Discussions in Physics Lectures', *Proceedings for HEA STEM Conference, Imperial College, London and The Royal Geographical Society*, 2012.