



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.



Mapping the transition - content and pedagogy from school through university

Katherine Alice Slaughter

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy
to the
University of Edinburgh
May 2012

Abstract

A study has been carried out at the University of Edinburgh in order to examine how physics students' abilities and attitudes towards study change during their time at university. This is a large topic with numerous possible avenues of research, as a result the field has been narrowed for this thesis in order to focus on three main subject areas; how students adapt during the transition from school to university, how students attitudes towards studying physics change during an undergraduate degree and, finally, student data handling skills in the undergraduate laboratory with links to whether student perceptions of their data handling skills are consistent with their ability.

It has been found that students may face difficulties going from school to university study. Students potentially face gaps in their prior learning due to differences in school leaving qualification syllabi, which is compounded by instructors having expectations of student ability that are higher than student actual ability. It has been seen that students become less positive in their attitudes towards study over the course of their first year of instruction, potentially due to a drop in confidence.

In the subject area of attitudes towards study, longitudinal studies have been carried out in order to examine the expert-like thinking of students. Results gathered are suggestive of a selection effect with the most expert-like thinkers possessing levels of expert thinking similar to those of physics instructors, even when initially entering the degree program.

Investigation of student laboratory work has shown that there is a large gap between student estimations of their own ability and the reality of such skills. This has been demonstrated by contrasting the results of surveys examining student perceptions towards practical work with data gathered from a data handling diagnostic test that has been designed and implemented as part of this thesis.

Declaration

Except where otherwise stated, the research undertaken in this thesis was the unaided work of the author. Where the work was done in collaboration with others, a significant contribution was made by the author.

K.A. Slaughter
May 2012

Acknowledgements

Firstly, I would like to acknowledge the help and support of everybody that has contributed to this thesis, there are too many to name but I hope that anyone I have not mentioned here will know how grateful I am.

I would like to thank my parents, Martin and Pamela Slaughter, who have always offered support and encouragement and without whom I would certainly not be where I am today.

I would also like to thank my fiance Gareth Davies and my family and friends who have all patiently listened to me talk about my work and helped me in many ways.

My supervisors Professor Simon Bates and Dr Marialuisa Aliotta gave me the opportunity to join the Physics Education Research group at its inception and have offered help and advice throughout my studies, as has Dr Ross Galloway who has offered invaluable support over the last three years.

The work in this thesis would not have been possible without funding from the University of Edinburgh and I gratefully acknowledge both the Principal's Teaching Award and the School of Physics and Astronomy for their financial support.

Contents

Abstract	i
Declaration	ii
Acknowledgements	iii
Contents	iv
List of figures	vi
List of tables	xiii
1 Introduction and Motivation	1
1.1 The school to university transition	2
1.2 Attitudes towards physics study	5
1.3 Experimental work	10
1.4 Structure of thesis	15
2 Methodology	16
2.1 Educational context of the study	16
2.2 The Colorado Learning Attitudes About Science Survey (CLASS)	19
2.2.1 The survey instrument	20
2.2.2 Design and validation of the instrument	20
2.2.3 Useage of the CLASS instrument at the University of Edinburgh	24
2.3 Statistical tests	28
2.3.1 T-Tests	28
2.3.2 ANOVA	30
2.3.3 Chi-Squared	30
2.3.4 Instrument reliability tests	31
2.4 Pre and post event testing	35

3	The School to University Transition	37
3.1	Academic situation	38
3.1.1	Differing academic backgrounds	38
3.1.2	Changes in course content	39
3.1.3	A-Levels and SQA award exams	41
3.1.4	Physics A-Level examinations	42
3.1.5	Physics Higher and Advanced Higher qualifications	43
3.1.6	Mathematics A-Level examinations	44
3.1.7	Mathematics Higher and Advanced Higher qualifications	45
3.1.8	Discussion	45
3.2	Mathematics skills	46
3.2.1	The mathematics diagnostic	46
3.2.2	Staff expectations survey	47
3.2.3	Results	48
3.2.4	Discussion of results	52
3.3	Scientific reasoning skills	53
3.3.1	Investigation of scientific reasoning using the Lawson Test of Scientific Reasoning	55
3.3.2	Test questions and results	57
3.3.3	Discussion of results	71
3.4	1st year physics student attitudes towards study	72
3.4.1	Implementation of the CLASS survey	73
3.4.2	Results of CLASS survey	74
3.5	Chapter summary	76
4	Attitudes Towards Physics Study	79
4.1	1st year attitudes and beliefs towards study	80
4.1.1	Results of whole class	82
4.1.2	Results of physics majors and non-majors	84
4.1.3	Comparison of male and female attitudes towards study	90
4.2	Pseudo-Longitudinal (cross-sectional) study	92
4.2.1	Validity of the pseudo longitudinal approach	94
4.2.2	Overall favourable and unfavourable attitudes	95
4.2.3	Category results	99
4.2.4	Selection effect	100
4.2.5	CLASS in schools	102
4.3	Fully longitudinal study	104
4.3.1	Overall favourable and unfavourable attitudes	105
4.3.2	Category results	106
4.3.3	Completion of study	108
4.4	Validity of CLASS for individuals	109
4.5	Comparison of longitudinal methods	111
4.6	Chapter summary and discussion	112

5 Student Laboratory Work	117
5.1 Attitudes to experimental work	118
5.1.1 The attitudes of schools students towards experimental work	119
5.1.2 The attitudes of first year university students towards experimental work	120
5.1.3 Comparison of attitudes	121
5.1.4 Discussion of results	134
5.2 The Data Handling Diagnostic	137
5.2.1 Design and validation	138
5.2.2 Results	143
5.2.3 Section summary	147
5.3 Chapter summary	148
6 Conclusions, Discussion and Future Work	150
A Colorado Learning Attitudes about Science Survey	162
B Glossary of Terms	167
C School Leaving Qualifications	170
D Mathematics Diagnostic	173
E Staff Expectations Survey	176
F Lawson Test of Scientific Reasoning	177
G Focus Group on expert-like thinking	188
G.0.1 Focus Group 26th April 2010	188
G.0.2 Focus Group 28th April 2010	192
H Perceptions of School Practical Work Survey	197
I Perceptions of University Practical Work Survey	200
Bibliography	203
Publications	211

List of Figures

1.1	An agree-disagree plot showing a decrease in expert-like thinking for students in four universities in the USA. For further details see reference [43]	8
1.2	Graph showing the decline in uptake of A-Level science subjects in school over a period of 10 years from 1990-2000. For further details see [38] from which the image is used.	14
2.1	Figure and caption from Adams <i>et al</i> 2006. “Women’s. (a)N = 88 and Men’s (b) N = 62 responses to “What would a physicist say?” and “What do YOU think?” ” For further details see reference [34].	24
2.2	Figure and caption from Heckler and Sayer 2010 [94]. The change in student performance, by week, when asked to answer a question related to the motion of a charge. There is a large variation in student score on a week by week basis. For further details see reference.	36
3.1	Comparison of student performance on a mathematics diagnostic test administered by the University of Coventry. The different shaped markers represent different cohorts taking the diagnostic test, from those in 1991 who received a grade N at A-Level (a fail grade) to those who received a grade C in 1997. Although axis labels are not provided on the figure it is believed that the y-axis represents test score. For further details see reference [25].	40
3.2	Representation of staff responses to Question 1 of Staff Expectations survey and actual student performance on the test. Staff N=19, Student N=125. Mean student score is 29%	48
3.3	Representation of staff responses to Question 2 of Staff Expectations survey. Staff responses are shown in blue with a dark red line corresponding the student standard deviation (16%) Staff N=19, Student N=125	49

3.4	Representation of staff responses to Question 4 of Staff Expectations survey. Staff responses are shown in blue with a red line corresponding to student performance on the question. Staff N=18, Student N=125	50
3.5	Representation of staff responses to Question 5 of Staff Expectations survey. Staff responses are shown in blue with the bar corresponding to three in which the student performance was highest marked by numbers 1,2 and 3.Staff N=18, Student N=125. Numbers 1-3 represent the three questions on which the students scored most highly.	51
3.6	Representation of staff responses to Question 6 of Staff Expectations survey. Staff N=18, Student N=125. Numbers 1-3 represent the three questions on which the students scored the lowest.	52
3.7	Profile of scores in Lawson Test of Scientific Reasoning for Chinese and American high school students. For details see Bao <i>et al</i> (Reference [101])	54
3.8	Image originally accompanying Questions 3 and 4 of the Lawson test of Scientific Reasoning (See Appendix F)	58
3.9	Graphical representation of results for secondary school (N=193) and university students (N=111) answering Question 1 of the abridged Lawson Test of Scientific Reasoning.	59
3.10	Image originally accompanying Questions 9 and 10 of the Lawson test of Scientific Reasoning (See Appendix F)	61
3.11	Graphical representation of results for secondary school (N=192) and university students (N=105) answering Question 2 of the abridged Lawson Test of Scientific Reasoning.	62
3.12	Image originally accompanying Questions 17 and 18 of the Lawson test of Scientific Reasoning (See Appendix F)	63
3.13	Graphical representation of results for secondary school (N=192) and university students (N=105) answering Question 3 of the edited Lawson Test of Scientific Reasoning.	64
3.14	Image originally accompanying Questions 19 and 20 of the Lawson test of Scientific Reasoning (See Appendix F)	66
3.15	Graphical representation of results for secondary school (N=194) and university students (N=101) answering Question 4 of the abridged Lawson Test of Scientific Reasoning.	67
3.16	Image originally accompanying Questions 11 and 12 of the Lawson test of Scientific Reasoning (See Appendix F)	69
3.17	Graphical representation of results for secondary school (N=194) and university students (N=101) answering Question 5 of the edited Lawson Test of Scientific Reasoning.	70

3.18	Percentage of expert-like thinking for secondary school students and 1st year university students. Each pair of bars is twinned with the solid lines representing favourable expert-like thinking and the hatched lines unfavourable expert-like thinking. The favourable thinking corresponds to the left hand axis and the unfavourable to the right. The first two x-axis categories represent the students in secondary schools (those in school physics classes (N=92) and those on pre university visit days (N=65)). The second two categories represent the average score of the university students, pre and post first year instruction (N=385). Error bars represent the standard error on the mean.	75
4.1	Percentage of favourable expert-like thinking pre- and post- first year instruction for three consecutive academic years and all the years combined. The hatched bars indicates pre-instruction data and filled bars post-instruction. Error bars represent the standard error on the mean. N=147 for the 2008-09 data, N= 134 for 2009-10 and N=104 for 2010-11. Combined N=385 for both pre and post instruction data	81
4.2	Percentage of favourable expert-like thinking pre- and post- first year instruction. The hatched bar indicates pre-instruction and the filled post-instruction. Error bars represent the standard error on the mean. N=385 for both pre and post instruction data . . .	83
4.3	Percentage of unfavourable expert-like thinking pre- and post-first year instruction. The hatched bar indicates pre-instruction and the filled post-instruction. Error bars represent the standard error on the mean. N=385 for both pre and post instruction data	84
4.4	Percentage of favourable expert-like thinking for physics majors (N=257) and non-majors (N=127) both pre- and post- first year instruction. The hatched bars indicates pre-instruction data and the filled bars post-instruction. In all cases error bars represent the standard error on the mean.	85
4.5	Percentage of unfavourable expert-like thinking for physics majors (N=257) and non-majors (N=127) both pre- and post- first year instruction. The hatched bars indicates pre-instruction data and the filled bars post-instruction. In all cases error bars represent the standard error on the mean.	86
4.6	Percentage of favourable expert-like thinking for physics majors (N=N=257) and non-majors (N=127) both pre- and post- first year instruction and biological sciences students mid instruction (N=82). The hatched bars indicates pre-instruction, the filled post-instruction and the cross hatched mid instruction. In all cases error bars represent the standard error on the mean.	87

4.7	Percentage of unfavourable expert-like thinking for physics majors (N=257) and non-majors (N=127) both pre- and post- first year instruction and biological sciences students mid instruction (N=82). The hatched bars indicates pre-instruction, the filled post-instruction and the cross hatched mid instruction. In all cases error bars represent the standard error on the mean	88
4.8	Comparison of number of UCAS points earned at school, with overall favourable CLASS score. 60 UCAS points represents a Grade B at Higher or a Grade D at A-Level. 120 UCAS points represents a Grade A at either A-Level or Advanced Higher (Values correct for university entry up to and including September 2012). The dark line on each box represents the median for each grouping.	89
4.9	Percentage of favourable expert-like thinking for male physics students (N=222 for 2008-10) (N=72 for 2010-11) and female physics students (N=61 for 2008-10) (N=32 for 2010-11) both pre- and post- first year instruction. The hatched bars indicates pre-instruction and the filled post-instruction. In all cases error bars represent the standard error on the mean.	90
4.10	Percentage of unfavourable expert-like thinking for male physics students (N=222 for 2008-10) (N=72 for 2010-11) and female physics students (N=61 for 2008-10) (N=32 for 2010-11) both pre- and post- first year instruction. The hatched bars indicates pre-instruction and the filled post-instruction. In all cases error bars represent the standard error on the mean.	91
4.11	Percentage of favourable and unfavourable expert-like thinking for school pupils, students in an undergraduate degree, postgraduate students and staff. The green bars show the percentage of favourable expert-like thinking and are plotted against the left hand axis while the blue bars show the percentage of unfavourable expert-like thinking and are plotted against the right hand axis. The errors bars represent the standard error on the mean.	96
4.12	Favourable CLASS score for individual fourth and fifth year students compared to career destination. The purple line represent the average CLASS score of the whole fourth and fifth year class	102
4.13	Percentage of favourable and unfavourable expert-like thinking for school pupils in Year 4 (N=80), Year 5 (N=59) and Year 6 (N=18) of secondary school. The green bars show the percentage of favourable expert-like thinking and are plotted against the left hand axis while the blue bars show the percentage of unfavourable expert-like thinking and are plotted against the right hand axis. The errors bars represent the standard error on the mean.	103

4.14	Percentage favorable and unfavorable CLASS scores over three consecutive years for the same group of students (N=35). The blue bars show the percentage of favourable expert-like thinking and are plotted against the left hand axis while the green bars show the percentage of unfavourable expert-like thinking and are plotted against the right hand axis. The error bars in each case represent the standard error on the mean.	106
4.15	Percentage changes in CLASS category favourable scores between 1st and 3rd year.	107
4.16	Individual student change in class position in CLASS score between Year 1 and Year 3. The individual students are numbered between 1 and 35 in order to maintain the confidentiality of the students, the change in rank position in the class is shown on the x-axis. . .	110
4.17	Percentage scores of favorable expert-like thinking collected from both a fully longitudinal study and through a cross-sectional study. First year data: N=35 (Fully Longitudinal) and N=127 (Cross-sectional). Second year data: N=35 (Fully Longitudinal) and N=105 (Cross-sectional). Third year data: N=35 (Fully Longitudinal) and N=61 (Cross-sectional).	112
5.1	Question 1 from the Attitudes to Practical Work at University Survey. See Appendix I for the full survey and Reference [110] for further details	121
5.2	Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Answers to question asking where they would describe the practical work encountered as Understandable	122
5.3	Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Answers to question asking where they would describe the practical work encountered as Interesting	123
5.4	Answer profile for students answer two surveys asking them to consider their experiences of school and university practical work. Answers to question asking where they would describe the practical work encountered as Organised	124
5.5	Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Answers to question asking where they would describe the practical work encountered as the best part of physics	124
5.6	Answer profile for students answer two surveys asking them to consider their experiences of school and university practical work. Answers to question asking where they would describe the practical work encountered as enjoyable	125

5.7	Question 2 from the Attitudes to Practical Work at University Survey. See Appendix I for the full survey and Reference [110] for further details	126
5.8	Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Responses to the statement “ <i>I prefer to have written instructions for experiments</i> ”	127
5.9	Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Responses to the statement “ <i>Discussions in the lab enhance my understanding of the subject.</i> ”	128
5.10	Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Responses to the statement “ <i>I had few opportunities to plan my experiments during the lab work</i> ”	129
5.11	Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Responses to the statement “ <i>The experimental procedure was clearly explained in the instructions given</i> ”	130
5.12	Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Responses to the statement “ <i>There was good linkage between experiments and the relevant theory</i> ”	131
5.13	Question 3 from the Attitudes to Practical Work at University Survey. See Appendix I for the full survey and Reference [110] for further details.	132
5.14	The item difficulty index for 23 questions of the prototype Data Handling Diagnostic. The solid lines at 0.3 and 0.9 show the lower and upper acceptable criterion for score. The dotted line shows the mean score for the test.	140
5.15	The item discrimination index for 23 questions of the prototype Data Handling Diagnostic. The solid line at 0.3 shows the lower acceptable criterion for score. The dotted line shows the mean score for the test.	141
5.16	The point biserial coefficient for 23 questions of the prototype Data Handling Diagnostic. The solid line at 0.2 shows the lower acceptable criterion for score. The dotted line shows the mean score for the test.	142

5.17	The mean fraction of students answering each of the 23 test items correctly, averaged over all trial classes. The error bars show the standard deviation between classes, and the columns are colour-coded according to the expert view of each questions anticipated difficulty. The solid line is an unweighted linear fit to the answer profile. Figure reproduced with permission. For full details see Reference [112]	145
5.18	Mean test scores for each class in the physics program at institution A. The error bars show the standard error on the mean score. Figure and caption reproduced with permission. For full details see Reference [112]	146
C.1	Grid showing the structure of each of the seven physics A-Level syllabi and the two SQA physics qualifications.	172

List of Tables

3.1	Raw numbers of secondary school students choosing each answer option for Question 1 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer A/Reason E) is shown in bold. Total N=193	59
3.2	Raw numbers of first year university students choosing each answer option for Question 1 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer A/Reason E) is shown in bold. Total N=111	60
3.3	Raw numbers of secondary school students choosing each answer option for Question 2 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer E/Reason C) is shown in bold. Total N=192	62
3.4	Raw numbers of first year university students students choosing each answer option for Question 2 of the abridged Lawson test of Scientific Reasoning. Correct answer as for Table 3.3. N=105	62
3.5	Raw numbers of secondary school students choosing each answer option for Question 3 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer B/Reason E) is shown in bold. Total N=189	65
3.6	Raw numbers of first year university students students choosing each answer option for Question 3 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer B/Reason E) is shown in bold. Total N=103	65

3.7	Raw numbers of secondary school students choosing each answer option for Question 4 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer A/Reason D) is shown in bold. Total N=194	68
3.8	Raw numbers of first year university students students choosing each answer option for Question 4 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer E/Reason C) is shown in bold. Total N=101	68
3.9	Raw numbers of secondary school students choosing each answer option for Question 5 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer B/Reason A) is shown in bold. Total N=180	71
3.10	Raw numbers of first year university students students choosing each answer option for Question 5 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer B/Reason A) is shown in bold. Total N=100	71
3.11	Number of completed student surveys from secondary school students and first year undergraduate	74
4.1	Number of completed student surveys for the 2008-09, 2009-10 and 2010-11 academic years, broken down by degree studied and gender. Where totals of sub-cohorts do not equal the overall cohort total this is due to either degree or gender information being unavailable for a few students.	82
4.2	Number of completed surveys included in the pseudo-longitudinal study	93
4.3	Pre and post instruction CLASS scores for the first year physics class over three consecutive years. Both favourable and unfavourable scores are shown. Bracketed figures represent the standard error on the mean.	95
4.4	Overall percentages of favourable CLASS responses, for all categories and cohorts surveyed. The bracketed figure represent the standard error on the mean.	99

4.5	Overall percentages of unfavourable CLASS responses, for all categories and cohorts surveyed. The bracketed figure represent the standard error on the mean.	100
4.6	Raw numbers of student surveys collected at each stage of the study, the final numbers used throughout the study and the percentage of surveys retained for students who completed all three surveys.	105
4.7	Favorable scores collected through fully longitudinal and cross-sectional methods. The abbreviation ‘Long.’ refers to data collected in the fully longitudinal study and ‘Pseudo.’ refers to data collected in a cross-sectional study. The numbers in brackets represent the standard error on the mean.	113
4.8	Unfavorable scores collected through fully longitudinal and cross-sectional methods. The abbreviation ‘Long.’ refers to data collected in the fully longitudinal study and ‘Pseudo.’ refers to data collected in a cross-sectional study. The numbers in brackets represent the standard error on the mean.	113
5.1	Results of test deployment with 10 trial institutions. Table shows year of study, degree discipline, number of students in each class, mean score and standard deviation of mean score. Note that the years have been renumbered so that Scottish institutions begin with a year zero: this is so that comparisons can be made between institutions from around the UK and Ireland and reflects the fact that Scottish degrees are one year longer than those in the rest of the UK.	144

Chapter 1

Introduction and Motivation

The process of completing an undergraduate degree can be viewed, in itself, as a transition where students enter a degree, most commonly directly from school, and leave to join the workforce or continue on to further study. During that time students will have studied new subjects and met people from a variety of different backgrounds and cultures. For most this is seen as an enriching experience with many people suggesting that either the chosen subject of study [1] or the university experience as a whole [2],[3],[4] was one of the most formative episodes of their life. Claims such as these can, by their very nature, only ever be anecdotal as they are virtually impossible to quantify. In this thesis, while the focus is on the transitions faced and changes experienced while passing from secondary school to degree graduation, the emphasis is on specific, measurable changes that students experience during an undergraduate degree.

This project originally began as a one year Masters by Research looking at the transitional period faced by physical science students going from secondary school to university. This research remains a key part of the project and is discussed in Chapter 3. However, some of the results produced from the initial research presented searching questions as to the longer term changes in undergraduates; these areas are discussed in Chapters 4 and 5.

The over arching theme of this thesis, if the content was to be described by only one sentence, could perhaps be thought of as *'investigating what students gain during an undergraduate degree'*, a topic which has been studied from many angles such as graduate attribute studies [5],[6], studies into the employment prospects of young people [7],[8] and socio-economic considerations [9]. In addition, studies

have also considered what communities and countries gain from a university educated workforce [10]. None of these topics will be considered in this thesis and it is, therefore, perhaps more accurate to instead consider it as a body of work which looks at how students **change** during the course of a degree rather than what they gain. This thesis does not seek to examine graduate attributes nor to present an exhaustive list of the changes experience by students during a degree. It should be considered as a map of the evolution of specific qualities in students during their time at university.

The research questions in this thesis, at their most general, are concerned with looking at the changes that occur in students over the course of an undergraduate degree. This is such a large and general subject area that to carry out a literature review of all related subject material would be a task that could well last the length of the research project itself. As such the decision has been made to focus on work that relates to three main areas of research; the school to university transition, student laboratory work, and finally, attitudes and beliefs towards study. The research into the school-university transition consists of two additional areas (scientific reasoning and content knowledge) which were chosen as part of the scope of the project when it was originally conceived as a one year research project but are now less central to the issues addressed in this thesis as a whole. In this introductory chapter the background of the project and reasoning behind the decision to study each area will be discussed in the sections below. As the topics addressed in each of the three results chapters can be read as three distinct studies, there is additional information included at the start of each chapter explaining the motivation behind the research, with comparisons to previously conducted work included alongside the research itself.

1.1 The school to university transition

Whilst this thesis now looks at changes through all years of an undergraduate degree it is known that the changes students experience going from secondary school to university (and over their first year of study) will be some of the largest that they are faced with during the course of their education. The level of upheaval students experience, finding themselves confronted with far larger class sizes, less teacher to student contact time, and often the need to adopt new

learning styles to cope with the higher level and greater volume of the material presented to them, can mean that students feel overwhelmed and struggle to cope in tertiary education. It is well documented that this is a key point in students' academic careers [11],[12], and one where student retention can be a large issue. The physical sciences (with physics especially effected) experience a "leaky pipe" effect [13] with students failing to complete their degrees and leaving higher education at various stages of their studies [14]. The retention of these non-completing students is of concern to the management of universities as the departure of students will impact academic institutions financially as well as there being a moral responsibility to ensure that all recruited students have an equally fair chance of completing the degree [15] and there exists a potential reputation risk to the university if students are often seen to fail.

In order to help ease the transition from secondary school it is important to first identify the potential problems that students may face. There has been extensive research into the first year experience [16] with issues such as retention, student support, learning and teaching and student perseverance all well represented. Retention may be impacted in the first year of study by issues with student academic ability and student attendance. In these cases, as might intuitively be expected, research has shown that those students who are academically 'weaker' [17] or do not attend lectures and tutorials [18] have a lower chance of succeeding in their chosen courses and as a result are more likely to exit higher education.

In this research project it was decided to expand upon the potential difficulties experienced by students in their first year and attempt to broaden the research into the first year experience that had already been carried out in the UK. Four different topics were identified as potential problem areas to be investigated and constitute the original scope of this project when it began as an MSc: Students' ability to reason scientifically; student laboratory work; subject content knowledge and, finally, attitudes and beliefs towards the subject. Both the areas of attitudes towards study and student laboratory work have grown considerably since the project was extended to form a doctoral thesis, and the results of investigations carried out on these topics can be found in Chapters 4 and 5 respectively. Background and introductory material for the areas of student attitudes and laboratory work can be found in the sections 1.2 and 1.3

respectively.

As will be discussed in Chapter 3, the issue of student preparedness for university is an emotive issue. At the time of writing this thesis, the most recent figures for A-Level pass grades (the prerequisite qualification of university study for students in England, Wales and Northern Ireland) showed that the pass rate had risen for the 29th year in a row [19]. This continued increase leads to an inevitable yearly debate in the media as to whether the examinations are getting easier and if so, are they are still fit for purpose (examples can be seen in [20],[21]). Presented within Chapter 3 is a review of the most recent assessments of the school leaving qualifications in physics and mathematics alongside a comparison of the different options offered by the numerous examination awarding bodies in England, Wales and Northern Ireland, which highlight the differences in course content the students in the first year cohort may be facing. The number of different examination boards offering A-Level physics and the variation in the course content within the syllabi, means that it is a difficult task for universities to be fully aware of which topics will be familiar to all students and which will be new material for some of the class. This exercise of comparing course content to provide a concise picture of student prior knowledge could be viewed as a first step towards enacting Ausubel's dictum "*Ascertain what the student knows, and teach accordingly*" [22].

Teacher expectations of students can have a profound effect on the way that students are taught (and in turn how students learn), both in terms of the content covered with a class and the way in which the teachers interact with students [23],[24]. This issue is also explored, albeit briefly, in Chapter 3 in the context of university entry level mathematics. Studies have shown that the mathematics content of the A-level examination has been decreasing in the period from 1991 to 1997 (the point at which the particular study ends), and as a direct result students entering university were seen to have poorer mathematics skills than the generation before them [25],[26]. In contrast the studies included in this thesis show that teaching staff have high expectations of student mathematics skills, far higher in fact than the actual ability of these students, a finding which is discussed in greater detail in Chapters 3 and 6.

The final area discussed within the context of the school-university transition is that of scientific reasoning. Sometimes known as scientific enquiry, scientific

reasoning relates to the skills needed to carry out science by applying the scientific method [27]. This itself is an enormous area of study, with scholars studying the scientific method going as far back as Aristotle [27]. The presence of the scientific reasoning component of this thesis is small, not due to a lack of inspiration for research questions but rather due to the opposite; the field represents so many potential avenues of research that it would not be possible to focus on this topic and to do justice to the other research presented within this thesis. As a result the concept of scientific reasoning is used within this thesis as an assessment tool, another way to measure the changes experienced by students within their first year of university study rather than another area of research in itself. A conscious choice was made to use an assessment instrument that is widely used in science education research, The Lawson Test of Scientific Reasoning [28] which has been used in many different trials ([29] and [30] are just two examples of many). This meant that options were available to expand this area of study, if required, through comparison to previously conducted studies. In actual fact due to the expansion of other areas of this thesis this option was not explored. However, the results of using the Lawson Test with students on either side of the school-university boundary are discussed within Chapter 3, with further work on scientific reasoning and problem solving continuing in the form of another PhD thesis within the University of Edinburgh Physics Education Research group [31].

1.2 Attitudes towards physics study

In Chapters 3 and 4 the issue of attitudes towards study are discussed. Over and above their content knowledge, there is evidence that the way students think *about* a subject has a profound effect on performance in that subject. It has been suggested that from as early as primary school children develop misconceptions which can shape their attitudes and beliefs of how they view science in general [32]. It has long been established that external social factors and stereotypes create a series of beliefs in students that can affect performance [33]. In fact it has been suggested in previous research that attitudes and beliefs may be a better indicator of student ability than school examination results [34]. This has led to increased research into students' attitudes and beliefs in a variety of subject disciplines (just a few examples of a great number of studies can be seen in [35],

[36], [37]).

The study of student attitudes towards learning and the impact that these can have on student performance have been a subject of science education research for several decades both for discipline based research and for those researchers based in Education Departments (see [38] for an extremely thorough review of non-discipline based literature on the subject).

One of the reasons the literature is so rich in attitude-based studies is that it is such a broad subject area. Gardner suggests that one of the first points for a researcher of student attitudes to consider is whether they are interested in studying “*attitudes towards science*” or “*scientific attitudes*”[39], the former of these two covering pupil “*feeling, beliefs and values held about an object that may be the enterprise of science, school science, the impact of science on society or scientists themselves*” and the latter relating to “*features that might be said to characterize scientific thinking and are cognitive in nature*” [38].

In this thesis the focus will be only on attitudes towards science, specifically how such beliefs formulate and change over the course of an undergraduate degree. While this is an area that perhaps is traditionally thought of as being far more qualitative, a variety of survey instruments have been established to assess quantitatively the attitudes and beliefs of physics students. There are many of these instruments that have been developed, including the Views of Nature of Science questionnaire (VNOS)[40], the Views About Science Survey (VASS)[41], the Maryland Physics Expectations Survey (MPEX)[42] and the Colorado Learning Attitudes about Science Survey (CLASS)[34]. Studies undertaken with these instruments have allowed better understanding of the attitudes of science students, the evolution of these attitudes over time and the differences in attitudes between physics students and physics academics and practitioners.

In the area of physics, the instrument that has been most widely used to examine student attitudes is the Colorado Learning Attitudes about Science Survey (CLASS)[34]. This measures the attitudes and beliefs of university students through a series of attitudinal questions and calculates their level of expert-like thinking by comparing the answers to those of physics professors (further details of the development and use of the CLASS survey can be found in Chapter 2). The survey has been used at many institutions across North America but has been far less widely adopted in the UK. The survey is usually

administered both before and after a specific course is taught (known as a period of instruction) thus allowing the effect of the course on students attitudes to be assessed through the shift in the students answers.

The survey was developed by the Physics Education Research group at the University of Colorado building on a previous survey, MPEX, produced by the Redish Group at the University of Maryland [43]. One perhaps surprising feature of the MPEX survey was that a downward trend was observed in the attitudes and beliefs of university undergraduates after their first year of study, with student levels of expert-like thinking seen to significantly decline after a period of instruction. Students were surveyed at many institutions, with both traditional and more innovative teaching styles represented, and it was found that at all the institutions included in the study student levels of expert-like thinking declined significantly after a year of university teaching. Redish also pioneered the use of the ‘agree-disagree’ plot in order to represent the findings of such studies. The percentage of favourable scores (the extent to which answers agree with the expert responses) are plotted on the y-axis, with the percentage of unfavourable (the answers which are at odds with the expert opinion) on the x-axis. An answer that was in perfect agreement with the expert response would therefore be plotted in the top left hand corner, conversely if responses are becoming less expert over time the move would be towards the bottom right hand corner. An example ‘agree-disagree’ plot on which the results for four institutions (surveyed pre and post instruction) is shown in Figure 1.1. It should be noted that the expert response in Figure 1.1 is not at the (0,100) position, this is due to the fact that not all the answers on the survey were found by the experts to have an overall consensus of agreement.

This decline in expert-like thinking is also observed in using the CLASS survey. It has been consistently shown that student attitudes are seen to decline after a period of instruction [34],[44], unless student epistemologies are specifically targeted in the course design [45],[46],[47]. This decline, which has been widely reproduced in academic institutions around North America and the rest of the world¹, has led to CLASS increasingly being used as a measure of the success of course redesigns, with students taking a specific course being surveyed before and

¹Although there are few published studies which report this decline, perhaps due to the fact it has been established in Physics Education Research as the expected result when looking at the changes in attitudes over time.

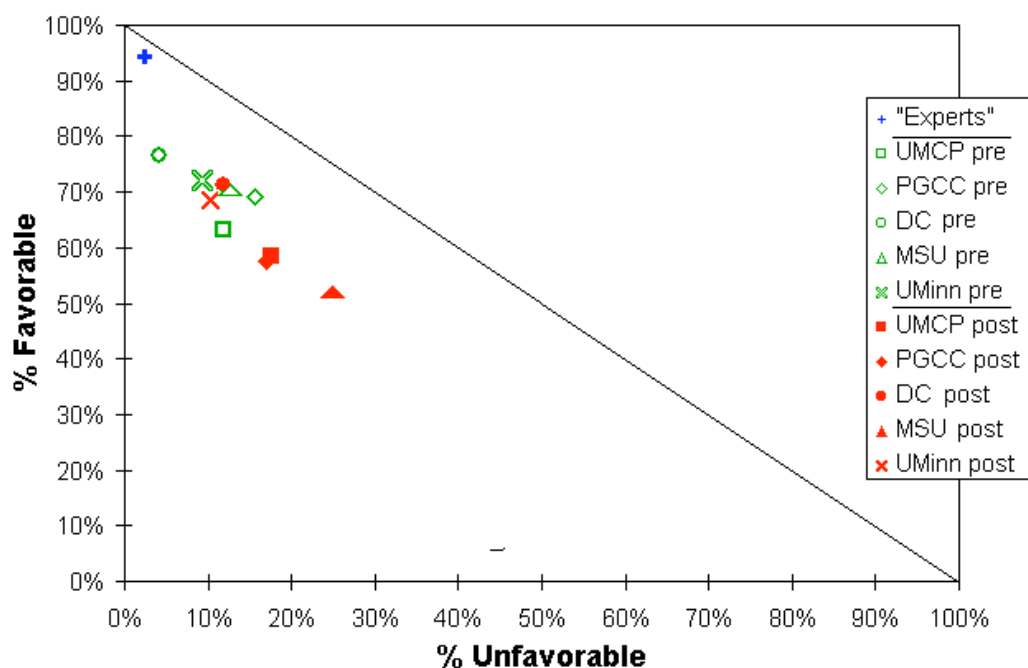


Figure 1.1: An agree-disagree plot showing a decrease in expert-like thinking for students in four universities in the USA. For further details see reference [43]

after the changes in order to assess the benefit of the redesign [45],[47],[48],[49]. Another attitudinal survey with science students to show a decline in positive student attitudes is that of Anders and Berg at Umeå Universitet, Sweden who examined shifts in attitudes in chemistry students [50] using a survey based on the works of Perry [51]. The study reported a negative shift in student attitudes, (although this was not the focus of the study). In addition the study was used to identify the students displaying the biggest positive and negative shifts, six of these students were then interviewed at length to try and determine the root of these changes. Following the interviews the responses were categorised into “choice”, “activity” and “persistence”, all of which could be classed as either positive or negative. Overall they found that those with negative attitude shifts were lacking in positive motivational choices. They concluded that students can lose or gain motivation in the subject for varying reasons including teacher attitude and class atmosphere.

A literature search reveals that published papers that report a decline in student attitudes after study are relatively uncommon, perhaps due to the

fact course designers and instructors are unwilling to publish such statistics as they may be seen as a reflection on the teaching of the instructor rather than symptomatic of a change in the way of thinking in the student cohort. Even in the case of the CLASS survey where the move towards more negative attitudes after one year of has been established as a standard result in most academic institutions, there are few examples of these results being published, although in this case this may be due to the fact this decline is now considered a ‘known result’.

Given that most physics instructors would be surprised and disappointed to be told that a course which is considered successful by every other measure is likely to produce students that are less positive than when they commenced their studies it is remarkable that the reason behind this decline has not been directly examined by researchers. Studies have looked at course reforms that claim improved CLASS scores [45],[46],[47] but do not pinpoint why in particular the scores have been improved or what causes the scores to decline in the first place. One idea that may be connected to this decline in expert-like thinking is a loss of confidence with students feeling less confident in their abilities as a result of the transitions taking place in their way of learning during their first year of study. One model that describes such a lack of confidence is Perry’s Model [51], a model of learning which suggests a theory of intellectual and ethical development of university students. Perry describes students as being within a development process of learning rather than having static personality traits which affect the way they learn. A drop in confidence is detailed in Perry’s model which contains nine stages of intellectual development of university students. These nine stages can be grouped into 4 categories known as ‘dualism’, ‘multiplicity’, ‘relativism’ and ‘commitment’. Most students will begin university in the first stage of intellectual development, dualism, which relates to the idea that there is only one correct answer and it is the job of the learner to gain the right answer from the instructor. Multiplicity is an extension of dualism where students extend their horizons to incorporate the idea that an answer can be not yet known, as well as right or wrong. The step to the third stage, relativism, where students realise that knowledge is relative and context dependent is seen as a very large step of progress in development. It is at this stage that students can experience a loss of confidence and step back to previous levels, with a cyclical process established

where some students may never reach the uppermost categories. It could be that the drop in expert-like thinking seen at the end of first year instruction relates to a step back to a previous learning stage.

In the six years between the publishing of the CLASS survey and the completion of this thesis, use of the CLASS survey has grown considerably and it is increasingly used, not just to look at a class as a whole but to study the expert-like thinking of different demographics of a physics class. The instrument has been used to investigate gender differences in expert-like thinking [44],[52] with male students found to generally hold more expert-like views and a larger decline observed in the expert-like thinking of female students post instruction. Differences in attitudes between physics majors and non-majors are often commonly investigated [34],[53](perhaps unsurprisingly non-majors are seen to be less expert-like in their views than physics majors). This trend has been further investigated with the Colorado Physics Education Research Group going further in this area to explore the differences in attitudes seen between those who intend to major in physics and those who actually do major in physics [54]. The results of these studies will be discussed in further detail in Chapters 3 and 4 in relation to the work carried out using the CLASS instrument at the University of Edinburgh.

Within Chapter 3 the CLASS instrument is used in order to investigate the changes in student attitudes on either side of the school-university boundary, as well as during the first year of study (as the instrument was originally designed to be used). In Chapter 4 longitudinal studies have been carried out in order to attempt to establish what happens after the drop in expert-like thinking reported in previous studies and when (if ever) students go on to become fully expert-like thinkers, a question that is discussed both in terms of data collected as part of this thesis and comparatively to other studies which have emerged on similar topics.

1.3 Experimental work

Laboratory based experimental work is a key component of an undergraduate physics degree, with even those students enrolled on a theoretical physics degree programme required by UK guidelines from the Institute of Physics to carry out

some experimental work in the first years of the degree program [55]. The reasons for compelling students to carry out such work are not just as preparation for future employment prospects (in fact less than 20% of UK graduates from physics degree programs entered employment in scientific services in 2008 [56]) but as an opportunity to learn and practice the numerous transferable skills that can be gained through laboratory based work.

The skills potentially gained in laboratories are extensive. A thorough review of the learning objectives of university laboratories has been carried out by Richardson *et al* at the University of Sydney [57]. While this review was carried out in the context of the Australian educational system many of the learning objectives highlighted are in clear agreement with those outlined by the Institute of Physics in the UK [55], such as the planning of experiments, data analysis, interpretation of uncertainties and presentation of data, to name just a few of these themes. These ideas are also echoed in guidelines from the American Association of Physics Teachers which state five learning objectives for student laboratory work “*the art of experimentation*”, “*experimental and analytical skills*”, “*conceptual learning*”, “*understanding basic knowledge of physics*”, and finally “*developing collaborative learning skills*” [58]. These categories suggest that there should be great potential for learning opportunities in the undergraduate laboratory, not just through the linking of experiment and theory but also through the skills (both analytical and collaborative) learnt.

Experimental lab work is one of the areas that students tend to have the strongest feelings, both positive and negative, about when starting university. These views tend to be so varied as the experience of practical work in secondary education can be extremely different from school to school, due to different interpretations of the syllabus [59], teacher preference of areas to teach as well as financial and timetable constraints of individual schools.

The study of various aspects of university experimental laboratories is popular in science education research fields. Herrington states that the study of laboratory instruction is so important because “*the laboratory is a different instructional context from class room instruction*” [60].

As has already been eluded to in Section 1.1 and will be further discussed in Chapter 3, due to the diverse nature of school qualifications (both between Scottish and English Qualifications and between the English qualifications

themselves) there is no absolute definition of what students entering tertiary education will have experienced in their secondary education. Within many university physics departments there is a general feeling that school experimental work is leaving students underprepared for conducting science research [59],[61] and that the students who have obtained the skills needed to do well in school science are not necessarily being given the skills needed to be researchers “*Our current system of science education and assessment favours the syllabus-bound, diligent pupils who are happy to learn the facts, often by rote, and answer questions almost parrot-fashion. Whilst such pupils may well gain the highest grades at GCSE and often also at advanced level, are they really the sort of people who will thrive in the research laboratory where the safety of the syllabus has gone and innovation is the key to success?*” [61].

Johnstone [62] describes practice in laboratory classes in schools as “*recipe following*”, with students engaged in the mindless following of instruction manuals. He suggests while it is partly due to the nature of the teaching it is also due to the naturally limiting nature of working memory capacity. It is possible to overload the mind, causing students to revert to recipe following.

One of the reasons that the undergraduate laboratory is so often stated as being of vital importance is the opportunity it offers for learning by inquiry, a method of learning that is highly prized by many educators. In 1996 the American National Research Council ([63]) described inquiry as “*a multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating the results. Inquiry requires identification of assumptions, use of critical and logical thinking, and consideration of alternative explanations.*” This is a description which is still very relevant to the learning objectives outlined by the Institute of Physics today [55]. Although it should be stated that while laboratory based work may provide the opportunity to learn such skills, there are also large numbers of papers that lament the perceived lack of room for scientific inquiry in the school and university laboratory ([64] and [65] are just a two examples).

While it is clear that lab based work provides the opportunity for students

to explore and develop many different skills it is not apparent that a universal measure for such skills is available. Unlike for subjects such as Newtonian Mechanics and Electricity and Magnetism ([66] and [67] respectively) there is not, in the field of physics education, a widely used test of student experimental skills for undergraduate students. In Chapter 5 the development and testing of such a tool will be discussed, with reference to work carried out with 1200 undergraduate students in the UK and Ireland.

The decline in the number of students studying science, especially physics, is well documented and is the focus of many outreach programs designed to encourage students into STEM subjects. A lack of scientists raises problems for industry and research as well as the education of future generations (a topic which has been the focus of many research papers and will not be covered in this thesis). Figure 1.2 shows the decrease in the number of students choosing to study physics between 1990 and 2000. One of the reasons suggested for this decline is student attitudes towards and perceptions of science, with researchers suggesting that this decline in interest can be seen as early as primary school [68],[69]. In primary schools students often experience science through hands on work rather than the more traditional bookwork often experienced by students in secondary schools. It has been suggested that student attitudes towards science are generally positive in primary schools, although a decline in attitudes can be seen from as young age as ten due to decreasing practical work in the classroom as well as repetitive practice assessment for national tests [68]. This suggests that students have generally positive attitudes towards experimental work whilst in school (albeit which decline with increasing age) and should in theory be arriving at university with a positive view towards the experimental work element of their degree.

The attitudes of students in the undergraduate laboratory is an issue that has been on the periphery of science education research since at least the 1970s, when Shulman and Tamir stated “*We are entering an era when we will be asked to acknowledge the importance of affect, imagination, intuition and attitude as outcomes of science instruction as at least as important as their cognitive counterparts*”[70]. However, a review of the literature shows that there have been few studies which consider attitudes towards practical work, especially in the the undergraduate laboratory. One such study by Hanif and Sneddon [71]

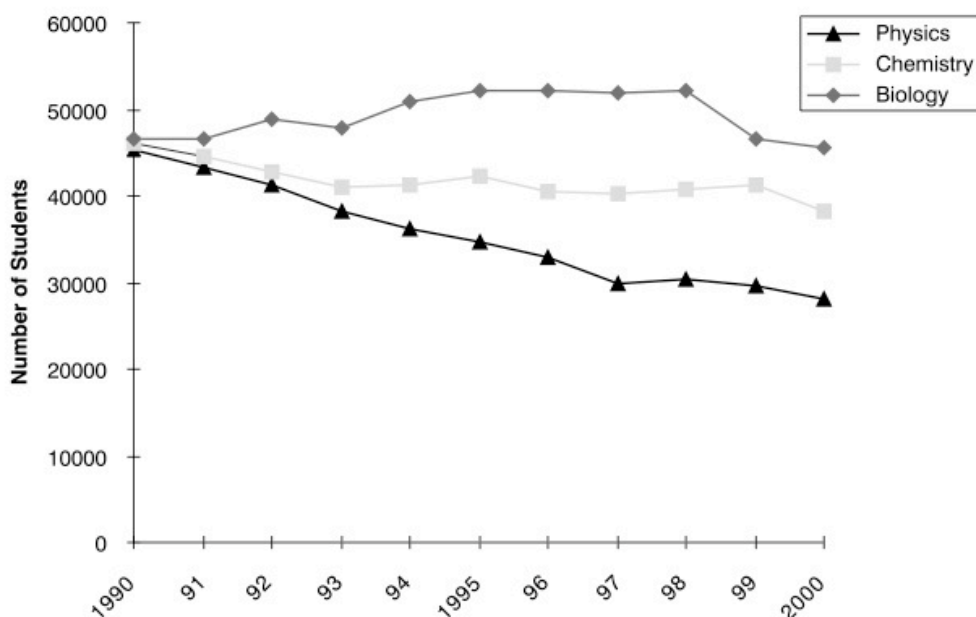


Figure 1.2: Graph showing the decline in uptake of A-Level science subjects in school over a period of 10 years from 1990-2000. For further details see [38] from which the image is used.

assesses physics undergraduates at the University of Glasgow at various stages of their degree and surveys them to gain a picture of students perceived values of laboratory work. They found that students thought that university laboratories were beneficial in illustrating theory as well as teaching practical skills. The second element of the work presented in Chapter 5 is a study of the attitudes of physics students on either side of the school university boundary, in work carried out with students at the Universities of Edinburgh, Glasgow and St. Andrews, aiming to establish not only the changes that students experience during their first year of study, but also a baseline of their attitudes and perceptions on leaving school. Combining the two themes of this chapter, it will be established how much the thinking of the students changes towards experimental work and if this, in turn, is reflected in the experimental skills gained by the students.

1.4 Structure of thesis

Contained within this thesis are, in addition to this introduction, five chapters which examine the transitions of students from school through university. Chapter 2 contains the methodology explaining the methods and tools used to carry out the work presented, Chapters 3-5 contain the results obtained through investigation of each of the topics discussed previously alongside comparison to and discussion of previous work in the field and, finally, Chapter 6 contains the conclusions of the thesis research and suggestions for future work.

Chapter 2

Methodology

The nature of this study means that there is not one single method that is appropriate to all the research carried out during the project. Each of the following results chapters presents work which has been implemented using different instruments and techniques in order to investigate the changes seen in students during the course of an undergraduate degree, with the appropriate tools needed to analyse the data collected being dictated by the nature of the data itself. There are, however, tools that have been used on numerous occasions as well as background information relating to the students that is relevant to all the studies. The role of this chapter is to serve as a reference chapter for the reader, illustrating the most pertinent contextual information for the study and describing the instruments and tests most commonly used.

2.1 Educational context of the study

As all studies carried out as part of this doctoral thesis have been carried out at the University of Edinburgh, or in secondary schools in local proximity to the University, it is beneficial to provide background information about the University as a institution, as well as the physics degree program, to aid reader understanding of the educational context in which this project has been implemented. The University of Edinburgh is one of the oldest universities in the UK, originally founded in 1583. The University is a member of the Russell Group, a group of some of the leading universities in the UK, all of whom focus heavily on academic research [72] and is consistently ranked in the top 50 universities in the world

[73],[74]. As a result of its reputation the University attracts high achieving students with incoming students typically possessing school leaving qualifications at the highest grades. The high level of competition for places means that most courses select rather than recruit prospective students. The University consists of 28,974 students split between postgraduate and undergraduate degrees and 3,283 members of academic staff across all disciplines (correct January 2011) [75].

The School of Physics and Astronomy, within the college of Science and Engineering, consists of 60 academic staff all of whom work with a research group specialising in an area of physics, in one of four Institutes; The Institute for Astronomy, The Institute for Condensed Matter and Complex Systems, The Institute for Particle and Nuclear Physics and The Edinburgh Parallel Computing Centre (which specialises in high performance computing). The variety of specialisms of the staff is reflected in the variety of undergraduate degree programs available to students, the school offers ten different physics degrees in total, at an undergraduate level. Five of these degrees are offered by the School of Physics and Astronomy alone: Physics, Mathematical Physics, Astrophysics, Computational Physics and Theoretical Physics. An additional five degree programs are offered jointly between physics and other schools within the university, these are: Physics with Meteorology, Physics and Music, Physics and Mathematics, Physics and Computer Science and Chemical Physics. Degrees programs in Scotland are typically one year longer than those in the rest of the UK, with the first year designed to act as a foundation year ensuring all students regardless of educational background (as is discussed in greater detail in Chapter 3) have the same content knowledge by the end of the first year of study. In the initial year of study, physics students are required to study one third physics, one third mathematics and the remaining one third is chosen by the individual student from another subject area. As this system is used in most of the first year degree courses across the university, many of those in the first year physics class will be from different degree programs, having chosen to take physics as their elective option. No separate physics class is offered for non-majors and all those choosing to take the first year physics class must be as qualified to do so as the majors taking the course. A typical first year class comprises approximately 200-300 students, of which around one half will be physics majors; it is very unusual for any non-majors to take any higher level courses after the first year, and impossible for

students to do so from third year upwards.

All the degrees offered in the School of Physics and Astronomy are typically studied over the course of four years, leading to a Bachelor of Science degree, or can (with the exceptions of Physics with Meteorology, Physics and Music, Physics and Mathematics and Physics and Computer Science) be studied as a five year Master of Physics degree. The entrance requirements for the Masters degree program are the same as those for the Bachelors degree but students are required to achieve higher grades in examinations in their third year of study in order to be allowed to continue onto the masters degree. Not all students who are qualified to stay on for the optional fifth year choose to do so, and the majority that do stay into the fifth year will then go on to study for a PhD.

The first year of the degree program is taught in a reformed workshop style [76] implementing many teaching methods advocated by Physics Education Researchers [66],[77],[78]. As students progress through the degree program classes are taught using increasingly traditional teaching styles (although some higher level courses have replaced traditional tutorials with workshop style classes). Students must take all the required physics courses in the proscribed year and must complete them before moving onto the next year of study. Students are offered little choice in the courses provided within their set degree program until the beginning of the fourth year of study.

Many of the investigations conducted with students reported in this study have taken place during the first year of the undergraduate degree. There are a variety of reasons for this, partly it is a natural consequence of the original scope of this project which was to investigate the transition from secondary school to the first year of undergraduate study. Another consideration is that in order to establish the changes that occur during an undergraduate degree it is necessary to establish the starting point of the attributes studied. A final contributing factor is that the learning environment established in the first year physics class lends itself very easily to the inclusion of Physics Education Research instruments and diagnostic tests. Two of the three instructors on the course are also Physics Education researchers and the continuing ambition to establish best teaching practice in the course means that new teaching methodologies are often trialled. In some cases, the implementation of new methods of teaching or measuring student skills has offered an opportunity for some of the data presented in this

thesis to be collected.

2.2 The Colorado Learning Attitudes About Science Survey (CLASS)

As outlined in Chapter 1, the Colorado Learning Attitudes about Science Survey (CLASS) is now one of the most commonly used measures of student attitudes in Physics Education Research. The survey allows a quantitative measure of how students think about physics by comparing the answers of students to those of expert physicists. The instrument has been extensively tested and validated by the survey designers [34] and this overview is not intended to recreate or replace the detailed papers written by the instrument designers, but to act as an overview to the workings of the CLASS instrument.

The survey consists of 42 attitudinal questions, each of which is answered on a five-point Likert scale, from strongly agree to strongly disagree. A copy of the survey is included in Appendix A. At the point of design, the survey authors carried out exploratory factor analysis in order to determine possible categories that the questions of the survey could be placed in. Eight categories in total were found which are as follows; Personal Interest, Real World Connection, Problem Solving General, Problem Solving Confidence, Problem Solving Sophistication, Sense Making and Effort, Conceptual Understanding and Applied Conceptual Understanding. The creation of these categories means that student survey scores can be analysed over all the questions or by individual categories.

One of the key identifying features of CLASS and one of its precursors, MPEX (Maryland Physics Expectations Survey) [42],[43], is that the student responses are compared to a previously collected expert opinion in order to give a percentage of expert-like thinking of the survey cohort. Not only does this provide a quantitative measure of the differences in epistemologies that exist between students and academic staff but it also allows changes in student attitudes over time to be measured by comparing to the same reference point.

2.2.1 The survey instrument

Developed in the early 2000s as an extended and more extensive version of the Maryland Physics Expectations Survey (MPEX) [42],[43] the survey was first presented to the Physics Education Research community at the 2004 PERC Conference [79]. The survey looks to explore the way students think about physics rather than examining content knowledge. The 42 questions on the survey are designed to be subjective and answered on a five-point Likert scale.

After the initial design and refinement of the survey (as detailed in the next subsection), the survey was given to physics faculty at the University of Colorado who were asked to give their responses to the questions. In total 12 physics faculty completed survey responses were collected and used to come up with an ‘expert’ response to the survey items [34], where there was not a general consensus between the staff responses discussion was used between the academic staff in order to try and agree on an expert response [80]. This ‘expert’ set of answers to the survey can then be used to compare student responses to the expert response to give the percentage to which the students surveyed are thinking like experts.

The nature of the studies most usually carried out using the survey instrument (where the changes in attitudes over a period of instruction are examined) means that the survey cannot be completed anonymously as it must be possible to match the pre and post instruction surveys to each other for individual students. Some students can be resistant to filling in a questionnaire that probes how they think and can be identified to them [81], however, the fact that the surveys are identifiable to specific students does allow for possibilities in the analysis of the data. If information is available on the background of the students it is possible to conduct studies into the differences seen in student epistemologies of, for example, male and female students [52],[82] and differences between physics majors and non majors [53],[54],[82].

2.2.2 Design and validation of the instrument

The original paper which accompanied the release of the CLASS instrument was an extensive overview of the design and validation of the survey [34]. Contained in this thesis is a brief overview that is intended to highlight the key features of the design and validation process, those requiring more in depth information

should consult the original paper.

The survey was based on original statements from the MPEX [42] and VASS (The Views about Science Survey) [41] surveys with some statements modified or discarded and others added. Five main design features are identified by the original survey authors as distinguishing it from previous attitudinal surveys, the most prominent of which is that the wording of the survey should not only be clear, concise and without possible alternate implications of the statement but also should be easily adapted for use in other science subjects (as has subsequently been done for chemistry and biology) [83],[84].

The modification and creation of new statements was initially carried out by listening to students talking about science to establish the sort of ideas and vocabulary that students use. Technical words were avoided so that the survey could be used by students of many levels of study. The survey designers specifically state that the survey is designed to be used in an academic context but that it should not be specific to any particular course or level.

There is much emphasis in the original paper on the scoring of the instrument. As has been previously mentioned in this thesis, the results of the survey can give either a percentage favourable (the percentage to which the student responses agree with the expert opinion) or a percentage unfavourable (the percentage to which the student answer is directly at odds with that of the expert). There is also a percentage score for each of eight categories, which the statements have been sorted into (as will be discussed in greater detail in the following sections). Of the total 42 statements, 27 fit into the eight categories and 9 more statements are included in the overall favourable and unfavourable scores but are not featured in any of the categories. The remaining six statements either have no expert score or, according to the survey authors, “*are not useful in their current form*”, no explanation is made in the paper as to why they were left in the survey or in what sense they were deemed to be “*not useful*”.

Using a five-point Likert scale, it is possible to score the responses using an ordinal scale or an interval scale. The survey authors argue that in interview the students presented arguments for choosing various points on the five-point scale which showed that they did not consider there to be equal weight between each different response and therefore an ordinal scale is used in scoring the survey, with a percentage of agreement quoted. The use of a five-point scale (as opposed

to perhaps a three- or seven-point scale) is an intentional part of the design with interviewed students stating that they would have been more likely to select the neutral option on many occasions if there was not a distinction between agree and strongly agree, for example. Despite the use of a five-point scale, during analysis the results are collapsed down to a three point scale with students either agreeing, disagreeing or choosing the neutral option to each statement. The survey authors assert that while this may have made the results less detailed they have not changed their accuracy “*we may have lost some definition but have no reason to believe we have distorted the results*” [34].

Built into the design of the instrument is that a minimum number of statements in the survey must be answered or the survey responses are discarded. The number needed to preserve the results is that 32 of the 36 scored statements must be answered. If statements are skipped but the minimum number of statements is still answered the analysis is conducted as if the statement does not exist for that students. The authors state that statistical analysis shows that skipping the statements provides a more accurate result than scoring the statements as neutral, although no evidence of this is presented in the paper.

At the time of publishing the paper (2006) the survey had been given to 7000 students as both a pre- and post-instruction test. The administration of the instrument is slightly different to the way it was used at the University of Edinburgh (detailed in the next subsection), as it was originally given in online format and the students were allowed between three and seven days to complete the survey. Course credit was given for taking the survey, including if the students only gave their name and student ID. A timer was included in the survey and all surveys that were completed in three minutes or less were discarded.

The validity of the instrument was tested using four different aspects of validity; face validity, construct validity, predictive validity and concurrent validity. Face validity involved survey responses and conducting interviews, both with the experts and students. The expert response against which student answers are compared was composed from three interviews with physics teaching staff at the University of Colorado, after which any ambiguous statements were removed. The survey was given to a further 12 experts to create the expert view used in scoring. Student interviews were used to discuss student interpretation of the statements alongside questions posed to the student to give an idea of the

character and study habits of the student being interviewed. In total, 34 students were questioned with a representative mix of students with respect to gender, degree intention and ethnic background. The construct validity refers to the many trials the survey has undergone and the statistical analysis of the responses. The predictive validity is described in the paper as the correlation of the survey responses with student incoming beliefs and course performance, although this is not further elaborated on in the paper and no results are presented. Finally, the concurrent validity involves analysis of the results to show that expected results are found, for example, that majors are more expert than non-majors.

Some of the statements of the survey have been placed into one or more of eight categories, as described previously. The process used to do so has been described extensively in the original paper, and will not be described in this thesis as they are not directly relevant to the analysis carried out during this project. The authors state that the majority of attitude surveys use what they term a ‘*predeterminism*’ method where the statements are group into categories at the point of survey design based on the beliefs and views of the survey designer [34]. An alternate method is to not include any categories until the data has been collected and then use exploratory factor analysis [85], to place items into categories based on student responses. The CLASS survey authors used a combination of the two methods where exploratory factor analysis was carried out on the statements placed into predetermined categories and then the results of the factor analysis were used for form new categories.

A key issue to do with the validity of the test is addressed by the authors in asking whether the students are answering what **they** think or what they think they **should** say. This is addressed in part in one of the trials carried out by the authors where students were asked to respond to each statement answering first what they think and then what they think a physicist would say, the results of this study show that the students can distinguish between the two, as illustrated in Figure 2.1 , with subsequent analysis of previously collected CLASS surveys showing the student responses more closely resemble the responses classed as “*What do **YOU** think*” this is further elaborated on in subsequent papers [86].

The reliability of the results is tested by looking at the similarities in the results given by the instrument with different sets of students in the same class (for example sampling the introductory physics class over several years). Analysis

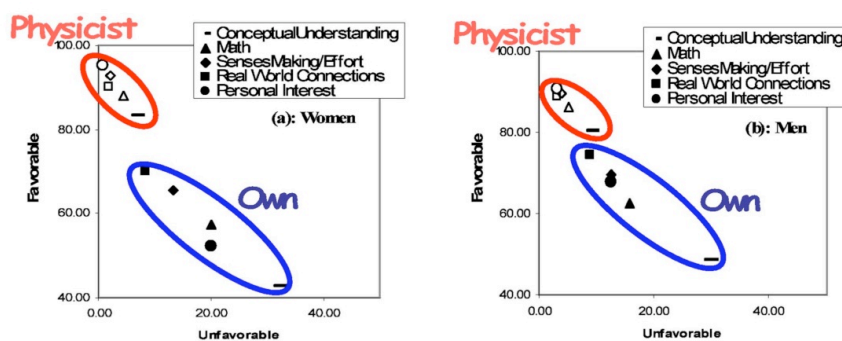


FIG. 3. (Color) Women's (a) ($N=88$) and men's (b) ($N=62$) responses to "What would a physicist say?" and "What do YOU think?"

Figure 2.1: Figure and caption from Adams *et al* 2006. "Women's. (a) $N = 88$ and Men's (b) $N = 62$ responses to "What would a physicist say?" and "What do YOU think?" " For further details see reference [34].

of the results shows that the student responses are highly similar over several points of sampling, suggesting the instrument has high test/re-test reliability. No consideration is made in the paper as to whether the results are valid for individual students, it is outlined in the future work section of the paper that this issue has not been addressed and will be covered in future work by the group, although no papers have appeared in the literature that address this point at the time of writing this thesis¹.

One final area addressed by the authors that is relevant to this thesis is the issue of interpretation of the results, with attention drawn to issues, other than instruction, which may affect student attitudes. One possible area is that of the timing of the survey with notable changes in student attitudes seen after university vacation periods (which highlight the necessity of sampling students at the same point in time if comparisons are to be drawn between different classes).

2.2.3 Useage of the CLASS instrument at the University of Edinburgh

The CLASS survey has been an instrumental tool for the research presented within this thesis, having been used in several different studies designed to

¹Data collected as part of this thesis suggests that the test may not be valid on an individual basis, as will be discussed in Chapter 4.

investigate the effect of an undergraduate degree on student attitudes and beliefs towards study (and conversely the effect of student epistemologies on degree performance).

The survey has principally been used in three different ways, to examine the first year undergraduate experience, to give a ‘snapshot’ of student attitudes at a specific point in time and to look at the evolution of attitudes over the course of the whole undergraduate degree. The details of how it has been used are outlined in the following section, alongside discussion of previous studies that have used the same methodologies.

Studies of first year attitudes have been carried out at many academic institutions in North America, as has been extensively discussed in Chapter 1 and briefly in previous sections of the current chapter. The method of the investigations is simple, students are asked to fill in the survey before any instruction at the university and then are asked to complete the same survey again at the end of the period of instruction (either one semester or one year depending on the specifics of the individual study).

In the case of the first year experience studies presented in Chapters 3 and 4 of this thesis a period of one academic year (two semesters) was used as the period of instruction. Students were surveyed in their first week of tutorials (week 2 of the academic year) and again in the final week of instruction prior to students departing on study leave before the end of first year examinations. The surveys were completed in paper format and no course credit was offered as an incentive to complete the survey. Prior to being given a copy of the survey the students were given a brief introduction to the survey which explained the history of the instrument and that it was used to investigate how students think about science. Emphasis was given to the fact that there were no correct answers to the questions but that it was important to answer honestly what they think the answers should be rather than trying to guess what their physics instructors might want them to think (it has been shown that students are able to draw a distinction between the two [86]). It was also explained to the students that the survey answers formed a very important part of the doctoral research of one of the Teaching Assistants on the first year course (the thesis author) and that they should answer seriously. Although it is not possible for the surveys to be completed anonymously as the responses of individual students need to be matched to allow comparison of pre-

and post-instruction attitudes, the students were reassured that their responses would not be shared with anyone prior to anonymization.

Due to the nature of the study, only matched student surveys were used (meaning the same student had filled in both pre- and post-teaching surveys) for analysis so that changes in the attitudes of individual students could be traced. If a student had filled in either a pre- or post-instruction survey but not both their survey answers were discarded; this led to approximately 40% of the collected surveys being discarded².

The survey has been administered pre and post first year instruction three times (with a fourth time being implemented at the time of writing this thesis) commencing in the 2008-09 academic year. On each occasion the survey has been implemented as described above with no deviations. The students take on average between 5 and 15 minutes to complete the survey, with the majority completing in around ten minutes. In order to allow the quality of student responses to be monitored the survey designers included a fail-safe question, which states

“We use this statement to discard the survey of people who are not reading the questions. Please select agree-option 4 (not strongly agree) for this question to preserve your answers”.

Approximately 4% (averaged over several years) of students who completed the survey during this study failed to choose the correct option and their survey responses were discarded. This is lower than seen in other published studies [34] and while it does not necessarily exclude all students who are not taking the survey seriously, it does allow the elimination of obviously unreliable data.

The three years worth of collected data have been subjected to statistical tests (a single factor ANOVA test) to examine any potential differences in the data collected in each year. This was done for the overall favourable and unfavourable scores as well as on a per category basis. No differences were detected and as a result the data presented in Chapters 3 and 4 is aggregated data from all three years.

Another use of the CLASS survey is seen in Chapters 3 and 4 of this thesis, where a one time use of the survey gives a ‘snapshot’ of student attitudes at a

²The number of discarded surveys appears large but is affected by a number of visiting students taking only one of the semesters of the course, as well as those who do not attend the final week of teaching in the second semester or miss the first survey through joining the course late.

certain point in time. This method is used in Chapter 3 where the attitudes of secondary school pupils are examined and extensively used in Chapter 4 where the results of the pseudo-longitudinal or cross-sectional study are presented. In this method the surveys can be completed anonymously as there are no subsequent surveys for the results to be compared to. This also presents the advantage that no survey responses have to be discarded if they are not part of a matching pair. In most instances that use this method the survey was implemented in paper format as described above. The exception to this was the snap shot data collected for years four and five of the undergraduate degree which was collected through an online survey tool as part of a Year 4 (Senior Honours) undergraduate project.

The final use of the instrument is in the fully longitudinal study, which is also presented in Chapter 4. This is an extension of the first year study with the surveying of the same students continuing into second and third year of the undergraduate degree. Unlike in the study of first year student attitudes, the students are not surveyed pre and post instruction in any of the subsequent years but at the end of teaching in each year, forming effectively successive post tests for the initial pre test administered at the beginning of the first year. The second year data was collected in paper format during the final two weeks of teaching before students departed on exam leave. The data from third year students was collected earlier in the semester (week six of ten) due to the difficulties of finding a timetabled slot when all the students would be in the same class. The number of students who failed to correctly answer the fail safe question had decreased to only 1% by the third year of data collection. The surveys were met with some resistance from the students in the second year of data collection, with students reluctant to fill in the same survey for a third time (as they had already completed it pre- and post- first year instruction). However, once it was explained to the students why it was necessary to use the same survey every time (rather than an isomorphic survey) and that the study the results would form an important part of the Physics Education Research group's research, the response rate was approximately 60%, for the whole year and approximately 95% of those who had attended the class in which the surveys were administered. In the third year of data collection another approach was used, with the introduction and administration of the surveys carried out by a member of staff not associated

with the Physics Education Research group and in a designated section of class time. The take up of the survey was higher on this occasion with 100% of those who attended the class completing the survey.

2.3 Statistical tests

All quantitative results, including those presented in this thesis, need to be subjected to statistical tests to establish the significance of the findings thus determining whether observed changes are the result of actual change of the measured quantity or just numerical fluctuations. These tests may either be used to compare one data set to another or to test the validity of the data collected or even the instrument used to collect the data. All the statistical tests that have been used as part of the analysis of any of the data included in the results chapters will be presented in detail in the following sections, alongside descriptions of which tests are appropriate to use in different circumstances. The data analysis that makes up this project has been carried out in Microsoft Excel either using the data analysis features built into the software or using the raw form of the statistical formulas. In the case of the data analysis tool pack being used this chapter will discuss the uses of the test and variants of the tests. Where statistical test have been carried out without the use of Excel's data analysis tool pack the equations used are specifically included and discussed. In order aid the reader's understanding of the uses of the tests, example uses have been provided for each of the tests discussed.

2.3.1 T-Tests

T-tests are used to assess whether the means of two distributions are statistically different to each other. It is one of the most commonly used statistical tests with the null hypothesis being that the two distributions are not different to each other. There are different types of t-test which are used depending on the characteristics of the distributions being tested.

Paired t-test

Paired t-tests are used when a variable is being measured for the same group of students before and after an intervention [87]. For example if a medical study wanted to examine the effect of a new drug on patient blood pressure, the patients' blood pressure would be measured before treatment, and measured after in order to look for any differences in blood pressure after administering the drug. In the case of this thesis paired t-tests are used most commonly to look at changes in some variable pre- and post- a period of instruction.

Independent t-test

An independent t-test is used when a fixed variable is being compared for two different groups [87]. Using again to use a medical analogy, an independent t-test would be used to assess whether there was a difference in the average blood pressure of male and female patients. When using an independent t-test there are two options available for analysis: either equal variance is assumed between the two samples or unequal variances. There are considered to be both positive and negative aspects of each of the two methods [88], with the unequal variance test being considered more reliable but an equal variance test being considered more sensitive. In this thesis, when an independent t-test is used, unequal variances have been assumed.

One and two-tailed t-tests

Another factor to consider is the 'tails' of the t-test, which addressed the concept of the predicted direction of any differences between two cohorts. If, before data is collected, it is predicted that one set of results will differ from another in a set direction, for example patients taking blood pressure reducing medication will have lower blood pressure than those in a control group, then a one tailed t-test is used when looking at the differences between groups. If no predictions were made prior to data collection then a two-tailed test is used, predictions can not be made after the data is collected. In this thesis no assumptions about the collected results were made and in consequence a two-tailed t-test has been used in all instances.

2.3.2 ANOVA

The ANOVA (**A**nalysis of **V**ariance) test is used when multiple t-tests need to be carried out on the same data set. The nature of t-tests means that there is a small possibility of the use of many t-tests providing a ‘false positive’ thereby finding statistically significant differences between groups which do not actually exist [87]. As is the case for t-tests, different versions of ANOVA are used depending on the type of data to be analysed. For paired data such as described for a paired t-test a two factor ANOVA is used, whereas for independent groups single factor ANOVA is used. The major difference between these two techniques is that two-factor can look at the effect of multiple treatments on the same group of patients, or in the case of this thesis the effect of multiple years of instruction on the same group of students. Two-factor ANOVA can be classed as either ‘with replication’ or ‘without replication’. Replication refers to the idea that two different measures have been recorded and the effects of each can be looked at either separately or together. Without replication assumes that multiple treatments can have occurred to the group, but looks for only differences in the groups overall.

2.3.3 Chi-Squared

While there are several version of the chi-squared test, Pearson’s test of goodness of fit is most commonly referred to as a Chi-Squared test and is the version that has been used in this thesis. It is shown in equation 2.1.

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} \quad (2.1)$$

Where o_i represents the observed frequency and e_i represents the calculated expected frequency. The resultant χ^2 value is then used in look up tables to determine find the corresponding p -value of significance.

The chi-squared test compares two distributions in order to asses whether they could both be drawn from the same sample. The test is most commonly used in this thesis to look at test score profiles for different cohorts of students in order to assess if performance is different for the two groups. Unlike a t-test which only look at the mean of the distribution, a chi-squared test will examine the whole range of scores.

2.3.4 Instrument reliability tests

As a tool, multiple choice testing is used extensively in Physics Education Research either in diagnostic instruments or to assess the learning that has occurred during a course or over a period of time. There are many widely used multiple choice instruments designed to test different subject areas including Newtonian Mechanics [66], Quantum Mechanics [89], Electricity and Magnetism [90], to name only a few. In order to assess the reliability and usefulness of these instruments standardised statistical tests can be carried out. Reliability can be thought of as a measure as to whether the test is “*consistent in itself and consistent across time*” [67] both of which are factors which can be assessed quantitatively. There are two areas of any multiple choice instrument to assess; the reliability of the individual test items and the reliability of the test as a whole. The reliability of the test in itself is important as if the instrument is known to be reliable confidence can be placed in the test retest reliability for individual students and the validity of comparing one group of students to another group to examine differences in ability of the two cohorts.

In this section, three tests will be outlined which examine the reliability of individual test items and a further one is included which measure the reliability of the test as a whole. These tests were carried out in relation to the work presented in Chapter 5, where the design and implementation of a multiple choice test to examine student data handling and laboratory skills is described. The reliability tests carried out on the Data Handling Diagnostic followed the procedure of those outlined in the reliability testing of the Brief Electricity and Magnetism Assessment (BEMA). An overview of the tests is presented here but readers should consult the original paper for further details if more information is required [67]. An additional whole test reliability statistic is often used to measure the test-retest reliability of an instrument known as the Kuder-Richardson Test [67]. It is not included in this overview as it was not used to test the Data Handling Diagnostic, due to the fact that a requirement of the Kuder-Richardson test statistic is that the test covers only one subject topic and all the questions are of similar levels of difficulty. The Data Handling Diagnostic was specifically designed to be on several subject areas and uses questions of varying difficulty making the test unsuitable for use on this occasion.

The Item Difficulty Index

The item difficulty index (P) is a measure of how difficult the cohort taking a test found an individual test item and is shown by the proportion of correct responses. It is the ratio of the number of students that correctly answered a specific test item, divided by the number of students attempting the item, as shown in Equation 2.2, where N_1 represents the number of correct responses and N represents the total number of responses.

$$P = \frac{N_1}{N} \quad (2.2)$$

The range of the difficulty index is from $[0,1]$ with 1 representing a situation where all students correctly answered the question and 0 a case where no students answered the item correctly. The widely agreed acceptable values for the Item Difficulty Index is between 0.3 and 0.9, with an ideal value of 0.5 for test items [67],[91]. The Item Difficulty Index is sensitive to the population taking the test, as a test that is found to be too easy or too hard of a level for one group of students will not necessarily be the same for students at a different level of study. It is possible to calculate an average Item Difficulty Index for all K items on the test, as shown in equation 2.3, where again the final value of \bar{P} must be compliant with the acceptable range of values.

$$\bar{P} = \frac{1}{K} \sum_{i=1}^K P_i \quad (2.3)$$

Item Discrimination Index

The Item Discrimination Index (D) measures the ability of test items to discriminate between the strongest and weakest of the students. If the discrimination index is high then the item is consistently correctly answered by those students with more robust knowledge of the subject matter than those with weaker knowledge. If the discrimination index is low this shows that the weaker students are more often correctly answering the question than the stronger students. There are two ways of calculating the the discrimination index, both of which have advantages and disadvantages, as are discussed below.

The first method is known as the 50%-50% method and involves the cohort being split into two groups (Low and High) decided by whether their overall total

score on the test is higher or lower than the median. Once the data is split into the two halves the discrimination index is calculated for each individual item on the test, as shown in Equation 2.4, where N_H is the number of correct responses for the item in the high scoring group, N_L is the number of correct responses in the low scoring group and N is the total number of students who took the test.

$$D = \frac{(N_H - N_L)}{N/2} \quad (2.4)$$

The second method is known as the 25%-25% method and uses only the top and bottom quartiles of the whole class, although apart from this adaption the method of calculation is identical, as seen in Equation 2.5

$$D = \frac{(N_H(\text{top}25\%) - N_L(\text{bottom}25\%))}{N/4} \quad (2.5)$$

Which method to use is the choice of the individual carrying out the research and will depend on the data available. The 50%-50% method can underestimate the discriminatory power as the inclusion of the two middle quartiles (who are likely to have more unstable knowledge) can make the results less clear. The 25%-25% uses only the most consistent quartiles, making it more unlikely that the discrimination index will be underestimated, however it does require discarding 50% of the data which may not be viable depending on the amount of data collected.

The possible range of scores for the Item Discrimination Index is from $[-1,1]$ where $+1$ corresponds to a very highly discriminating test and -1 a case where everyone in the low group correctly answers the question and no one in the high group does so. An item is considered to be of good discriminating power if it has a value of $D=0.3$ or higher [67],[91]. As for the Item Difficulty Index, it is possible to calculate an average discriminatory index for all K items on the test, as seen in Equation 2.6

$$\bar{D} = \frac{1}{K} \sum_{i=1}^K D_i \quad (2.6)$$

Point Biserial Coefficient

The Point Biserial Coefficient (r) measures the consistency of an individual test item to that of the whole test and is a measure of the correlation between student

scores on a specific item and their score of the whole test. An item with a high point biserial coefficient indicates that if students answered correctly on this question they are likely to score highly on the test overall. To calculate the coefficient the difficulty of the individual item needs to be considered (as calculated in Equation 2.2) and compared to the average score of those answering the question correctly, as is explained algebraically in equation 2.7. \bar{X}_i represents the average total overall score for those who answered the item correctly, \bar{X} is the average total score of the whole sample, S_x is the standard deviation of the total score of the sample and P is the item difficulty index.

$$r_i = \frac{\bar{X}_i - \bar{X}}{S_x} \sqrt{\frac{P}{1-P}} \quad (2.7)$$

The value of the point biserial coefficient can be in the range of $[-1,1]$ with $+1$ representing an item which is highly correlated with the test. The recognised value of a consistent item is that $r \geq 0.2$ [67],[91]. A calculation of the average point biserial coefficient for the whole test is shown in Equation 2.8 , the average value must also meet the criteria outlined for individual test items.

$$\bar{r} = \frac{1}{K} \sum_{i=1}^K r_i \quad (2.8)$$

Ferguson's Delta

Ferguson's Delta (δ) looks at the test as a whole rather than individual test items. It aims to give a value of how discriminatory the test is as a whole by investigating the distribution of all the student scores, with the idea that a well designed test will yield a wide distribution of student scores. This is calculated by comparing individual student scores to each other to look at the frequency of occurrence of each different score. Ferguson's delta is the ratio of unequal pairs of scores to the maximum number of unequal pairs the test could hypothetically produce. The calculation of Ferguson's delta is represented in Equation 2.9 where f_i is the number of occurrences of each score, N is the number of students in the sample and K is the number of test items. The range of values possible for Ferguson's Delta is $[0,1]$ with values higher than 0.9 considered to show high discrimination [67],[92].

$$\delta = \frac{N^2 - \sum f_i^2}{N^2 - N^2/(K + 1)} \quad (2.9)$$

2.4 Pre and post event testing

Within the scope of this project extensive use has been made of the concept of pre and post event testing, where some defined quality of student learning is measured before and after a period of instruction. The period of instruction used can vary in length from long term projects, such as those seen in the longitudinal studies in Chapter 4 to only a semester or even less. The role of the pre-instruction test is to establish a baseline measure of the student attribute to be studied, after the test some intervention takes place (the intervention may be a period of ‘normal’ instruction), following which a post instruction test is used to measure any changes in the measured attribute.

The concept of pre and post event testing is widely used in Physics Education Research with examples including assessing learning gains on the Force Concept Inventory [66],[93] and changes in student attitudes [34]. One criticism that has been raised against the technique is the random nature of the dates chosen to survey the students, with suggestions that the time chosen to survey students can make a difference to test scores; ‘the winter break effect’ [34]. At a finer level of detail it has been proposed that in some cases only a few days of difference in timing the test can make large differences to the results gained [94], as shown in Figure 2.2.

The nature of the post-instruction test can vary depending on the preference of the experiment designers. In the field of Physics Education Research it is most common to see the same test being used pre and post instruction, although some tests favour using a different but homogeneous test post-instruction. An isomorphic test can avoid concerns related to the idea that any improvements in student attributes could be due to students having already attempted the test, however difficulties arise in ensuring a test is truly homogeneous. In this project there are no examples of students being given isomorphic tests pre- and post-instruction with the same test instruments always being used for both tests. The only example of true pre and post event testing in this thesis is connected with the use of the CLASS instrument, where the changes in attitudes of the **same**

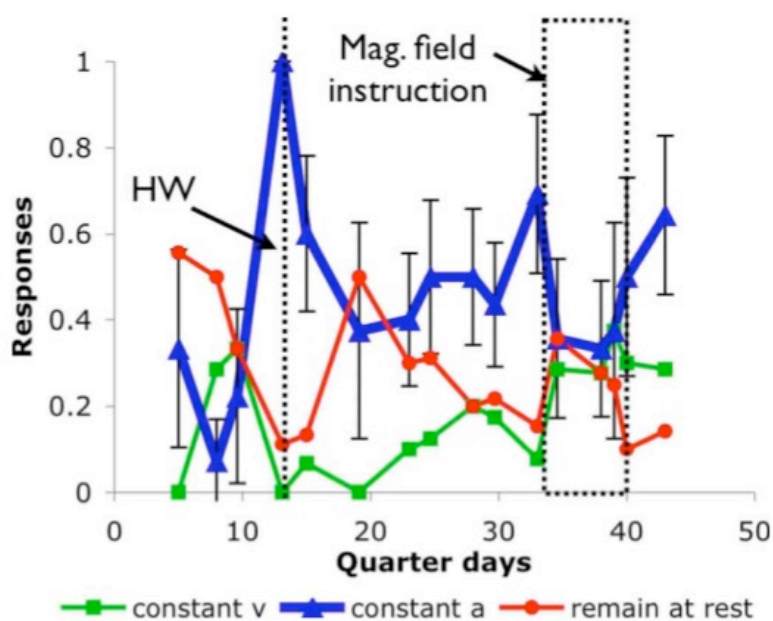


FIG. 5. (Color online) Responses by day on a question about the subsequent motion of a charge placed at rest in a uniform field. The correct answer (blue and bold) increases dramatically and falls rapidly. Later, during the unit on magnetic forces and fields, there may be an interference dip. Some minor responses have been omitted.

Figure 2.2: Figure and caption from Heckler and Sayer 2010 [94]. The change in student performance, by week, when asked to answer a question related to the motion of a charge. There is a large variation in student score on a week by week basis. For further details see reference.

students have been measured at different time points. Other examples in this thesis of comparisons of student attributes at different time points are not truly pre- and post-event testing as different groups of students have been compared, using the assumption that the students in one year group will not differ largely to the students in the year groups surrounding them.

Chapter 3

The School to University Transition

In the introductory chapter of this thesis, it was explained that this project was originally conceived as a one year Masters by Research that would investigate the secondary school to university transition, with emphasis on three key areas; student attitudes towards study, student laboratory work and student scientific reasoning. The expansion of this project to a doctoral thesis has meant that the work in the areas of attitudes towards study and laboratory work has grown considerably and can be found in Chapters 4 (as well as within this Chapter) and 5 respectively.

The expansion of the project to look at the whole undergraduate experience should not make the work on the school-university transition any less important. This remains a key time in a student's life where the changes students are faced with will probably be some of the largest they face during their education. This is a key point for the retention of students to the degree program [11],[12], making it an important issue both for the well-being of the students but also the financial prosperity of the university.

With so many factors having an influence on a student's transition from school to university it would be impossible to consider all those that contribute to whether students remain in the degree program. As such in this Chapter only four factors are considered; student prior learning, the expectations of teaching staff on the degree program, student scientific reasoning and, finally, student attitudes towards study. Both the areas of student scientific reasoning and

attitudes towards study have been considered by looking at groups of students on either side of the school-university boundary through the testing of students in local secondary schools as well as in the first year physics class at the University of Edinburgh.

3.1 Academic situation

In order to understand the changes that occur in students during the course of an undergraduate degree it is necessary to first establish the starting point for these students. In all the areas covered in this thesis students have been given pre-tests at the beginning of their degrees to establish base-lines against which any changes can be measured.

In terms of subject content, it is not the intention of this thesis to cover what students are expected to learn over the course of a degree, however given the different entry points and educational backgrounds of the students in the first year class at the University of Edinburgh, it is prudent to summarise what physics undergraduate students will have covered at school and the expectations, in terms of prior knowledge, placed on them when they enter university. In this chapter both mathematics and physics qualifications will be considered as both are prerequisite qualifications for entrance to the degree program.

In the following sections terms and acronyms will be used that are specific to the UK educational system, they have been defined where used but readers may also wish to refer to the Glossary of Terms included in Appendix B.

3.1.1 Differing academic backgrounds

Of the students in the first year class at the University of Edinburgh approximately 45% are from Scottish schools, 45% are from schools in the rest of the UK and the final 10% are international students. Students educated in Scotland will have studied typically five subjects after the age of 16: the qualifications gained as a result are known as Highers, which are awarded by the Scottish Qualifications Agency (SQA). These qualifications are considered sufficient to go on to tertiary education; however many students choose to stay in school to complete further study in more depth in (usually) three of these five subjects, pursuing qualifications known as Advanced Highers. Some independent schools

in Scotland offer students to the opportunity to study A-Levels (the system used in the rest of the UK) rather than the qualifications offered by the SQA. All three qualifications are accepted by Scottish universities for entrance to university meaning that there is already variation in the amount of prior learning the Scottish students may have.

The rest of the UK has final leaving qualifications known as Advanced Levels (A-Levels) and Advanced Subsidiary Levels (AS-Levels). Four or five AS-Levels will be studied as the initial education post 16 years of age, after which students will typically choose three of the five subjects to continue to full A-Level. Students must normally have completed A-Levels in order to study at university, although a few are admitted through foundation level schemes aimed at widening participation. A-Levels are taught through a modular system with six modules in a full A-Level, the first three modules comprise the AS-Level qualification and the final three are known as the A2 qualification.

3.1.2 Changes in course content

Much has been written over recent years about the perceived decline in A-Level standards: the continuing rise in the number of students receiving the highest grade at A-Level is reported every year in the media [95]. In 1999 the London Mathematical Society commissioned a report into basic mathematics skills of students and the levels of preparation for mathematics-based degree courses [96]. They reported that *“Mathematics, Science and Engineering Departments appear unanimous in their perception of a qualitative change in the mathematical preparedness of incoming students, even among the very best”*.

A study has been carried out by the University of Coventry over several decades, looking at the performance of students in a mathematics diagnostic test and comparing them to the grades they received at A-Level [25],[26]. The test measures student ability in seven areas of mathematics, including arithmetic, algebra, trigonometry and calculus. It was found that, on average, students with approximately the same score on the test compared to the scores of their peers in previous years possess markedly different A-Level grades. One such profile is seen in Figure 3.1 - the lines on the graph are used as a guide to the eye and are not significant to the data. Figure 3.1 shows that the scores of students who received a grade N at A-Level in 1991 (a fail grade) had a very similar results

profile on the test to those students who received a Grade C in 1997. A similar study carried out at the University of York between 1979 and 1999 found similar results, with student performance on a diagnostic test static from year to year, until 1990, when a decline in student scores began to be observed [26]. The nature of this apparent decline can be attributed to various factors, including the changing focus of the syllabus and decline in teacher standards [26]. However the reasons behind the decline are not the focus of this section, rather they are included to highlight the changing subject knowledge students can be expected to have on entering university. It is the case that the depth of the academic content students will have covered prior to entering university has decreased with time, thus meaning that students are embarking on physics degrees far less well prepared than their predecessors - leading to problems with student performance and retention.

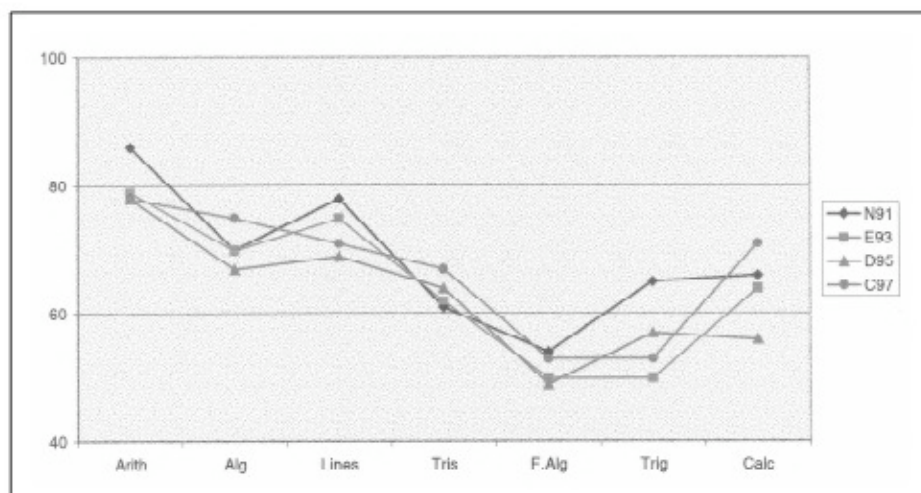


Figure 3.1: Comparison of student performance on a mathematics diagnostic test administered by the University of Coventry. The different shaped markers represent different cohorts taking the diagnostic test, from those in 1991 who received a grade N at A-Level (a fail grade) to those who received a grade C in 1997. Although axis labels are not provided on the figure it is believed that the y-axis represents test score. For further details see reference [25].

The problems associated with A-Level examinations are by no means confined to the mathematics A-Level, with similar problems also reported for physics. A

review of GCSE and GCE (A-Level) physics carried out in 2009 [59] shows that there has been a significant reduction in content covered in both GCSE and A-Level since 2002. The report also outlined the emotive trend at GCSE to look at the social implications of technology as opposed to looking at physical concepts. For example in one examination paper, students were asked to discuss the advantages and disadvantages of CCTV, the internet and mobile phones. Overall GCSEs were found to have a reduced level of quantity and depth of explanation expected of the students than had been expected five years previously. This is illustrated by the fact that several exam boards increased the overall time of examination by adding more multiple-choice exams, but reduced the written exams, allowing students far less opportunity to show any deeper understanding of the subject and were not “*effective at assessing certain higher order skills, such as the ability to describe, explain or evaluate*” [59].

In schools teachers may be under pressure to ensure students perform well, so that the school places highly in school league tables. This can lead to teachers being forced to ‘teach to the test’, ensuring pupils know enough facts to pass the exams but with little time to explore concepts further or focus on valuable problem solving and reasoning skills [97],[98].

For A-Level qualifications one notable change between 2001 and 2007 was a decline in the time allotted for examining practical skills and investigative work from 20% in 2001 (except for OCR which allowed 16.7%) to 17.5% for the highest exam board and 12.5% for the lowest.

Between 2001 and 2007 a decline in standards of students receiving Grade A at A-Level has been noted [59], however this was judged to be difficult to measure due to the differing nature of the examinations, compounded by the varying content of the various different A-Level syllabi.

3.1.3 A-Levels and SQA award exams

As mentioned previously the students in the first year class at the University of Edinburgh are from differing backgrounds, with at least five different school leaving qualifications (Advanced Highers, Highers, A-Levels, Irish Leaving Certificate and the International Baccalaureate) represented each year in the

qualifications of the incoming class¹. To add a further level of complexity to the situation there are five different exam boards in the UK offering seven different A-Level Physics qualifications (two of the examination boards, AQA and OCR, offer two different qualifications). Whilst there is core material that all A-Level syllabi must cover the specific course content and the order in which the topics are covered is left up to the discretion of the individual examination boards.

The different A-Level qualifications available from the various examination boards has led to some controversy, with teachers under increasing pressure to make sure schools perform well in national examinations there have been suggestions that some teachers intentionally choose the easiest examination boards [99]. This would appear to be a natural response to being presented with a variety of examinations with varying course content and examination styles, however the effect for students is that they may well be leaving school with qualifications that make them less well prepared for tertiary education than some of their peers.

In the following sections the similarities and differences between the A-Level qualifications themselves, as well the differences between A-Levels and the Scottish examinations, will be outlined to give the reader a greater understanding of the varying course content students entering the university physics degree will have been exposed to. In order to qualify to study the physics course at the University of Edinburgh all students must have both Physics and Mathematics qualifications to at least Higher Level (or A-Level), to that end, both subjects will be considered in this chapter.

3.1.4 Physics A-Level examinations

There are five different examination boards offering A-Levels in the UK (AQA, CCEA, EdExcel, OCR and WJEC), however there are actually seven different physics A-Level syllabi available in the UK as both AQA and OCR offer two different A-Level specifications.

All seven physics A-Levels consist of six independent modules, the first three of which are combined to give an AS-Level qualification and the latter three form the A2 component of the qualification (which combine with the AS modules to

¹Although there are relatively small numbers of students with the Irish Leaving Certificate and the the International Baccalaureate

give a complete A-Level qualification). Of these six modules, two consist of some form of practical, experimental or laboratory work, although the exact nature of the practical work varies from board to board.

Of the seven variations available the most obvious difference between boards is the way the information is presented to the students. Three of the seven use a context based approach with the physics in each module presented using real life contexts and examples in order to aid understanding. The titles of the modules reflect this method of display using ideas such as “Harmony and Structure in the universe” (AQA-B) and “Physics at Work” (Edexcel). These context based courses are generally modern, reformed courses that have recently been created, such as Edexcel’s Salter’s physics which was created by the Science Education Research group at the University of York [100]. Two of the three exam boards which offer such A-Levels as an option leave the decision up to the individual schools by offering both a traditional style and a context based course.

Another key difference between the examinations available is the differing level of experimental work offered as part of each different A-Level. What is consistent is that students will have studied some form of practical work in two out of the six modules involved in a full A-Level qualification, however this is where the only concrete similarities end. Some of the A-Levels insist that practical work is carried out in the form of a practical exam (WJEC and CCEA), others give individual schools the option between examination and course work investigations, whilst others have only small elements of practical work with more emphasis on preparing reports and presentations (OCR-B and Edexcel).

A summary of the modules involved for each A-Level and details of the practical work involved in each module can be found in Appendix C.

3.1.5 Physics Higher and Advanced Higher qualifications

The Scottish Qualification Authority (SQA) oversees the secondary education qualifications for all of Scotland, meaning that all Scottish students study from the same syllabus and sit the same examinations. As a result the picture is far less complicated than for students in England and Wales.

The Higher qualification consists of three modules ‘*Mechanics and Properties of Matter*’, ‘*Electricity and Electronics*’ and finally ‘*Radiation and Matter*’. There is no practical module to the course (with practical work included alongside

bookwork in some modules), and is undertaken in one year of study. The lack of experimental work means that those students who choose to study at university immediately after studying Highers will potentially have little experimental experience when faced with university laboratory work.

A new Higher qualification has been developed and has been gradually released alongside the former qualification from June 2010. None of the students included in this doctoral thesis would have had the opportunity to study the new program and as a result it is not included in this overview. However any reader using this thesis as a starting point for future work into the content of the qualifications should be aware of the changes that can gradually be implemented for students beginning Higher study from September 2010.

The Advanced Higher consists of four modules; '*Mechanics*', '*Electrical Phenomena*', '*Wave Phenomena*' and '*Physics Investigation*'. The physics investigation is weighted at one sixth of the total Advanced Higher qualification and requires students to design and implement a physics experiment to be carried out alone by the student, over one full school term.

3.1.6 Mathematics A-Level examinations

As is the case for the Physics A-Levels, there are five examination boards that offer a Mathematics A-Level qualification. The picture is slightly less complicated than for Physics in that none of the examination boards offer more than one qualification. There is, however, still a lack of continuity amongst the A-Levels, partly in the modules offered and partially arising from the choices made available to the individual schools as to what modules of the course to teach. The A-levels all consist of six modules and all five of the examination boards have four modules of Pure/Core mathematics which make up a compulsory part of the qualification. The remaining two units are left to the discretion of the individual schools and are usually selected from either statistics or mechanics modules (or a combination of the two). In addition three of the five examination boards also offer modules in decision maths which can be taken in the place of the mechanics or statistics modules. This level of choice means that students may be entering a physics degree having taken between 0 and 2 mechanics modules. The entrance requirements at the University of Edinburgh do not stipulate which modules must have been studied, meaning that the incoming students will, in all

likelihood, have a wide variety of mathematics backgrounds.

One further difference that may arise is the method of assessment, with one of the examination boards (AQA) offering schools the choice between a coursework or written examination as assessment for two of its modules. This decision, again, would be taken by the individual schools.

3.1.7 Mathematics Higher and Advanced Higher qualifications

The Scottish Higher examination does not have any choice of units and all candidates must take three modules, (Mathematics 1-3) which cover the areas of algebra, geometry and trigonometry. Students must pass internal unit assessments in addition to an externally set examination. A statistics module was previously available as an optional unit but this level of choice has now been removed. This is very similar to the Advanced Higher qualification which also consists of three units assessed through unit assessments and an external exam. The units cover the areas of calculus, algebra and geometry. Optional modules in Mechanics and Statistics have been removed from the course, meaning that all students taking SQA Advanced Highers will have covered the same material.

3.1.8 Discussion

It is difficult to say with any certainty what a student can be expected to know on entering university. It is not a case of being able to read one syllabus and assume all students will have covered the same material, looking only at the A-Level qualifications we see that there are significant differences in terms of content covered, the way information is presented, the way practical skills are taught (in the case of physics) and the way assessment of the learning objectives is carried out.

The two-tiered leaving point for Scottish students (with students starting university with either Highers or Advanced Highers) means that even amongst the Scottish students there is not uniformity of content knowledge. This means that it is almost inevitable that students in their first year will find themselves faced with physics and mathematics that they have never encountered before. Whilst this exposure to new knowledge is one of the central aims of university

learning, it may be, never the less, unsettling for new students, especially if the perception is that your peers are not in the same situation. There is no obvious remedy to this situation while the A-Levels (in particular) are so varied, but educators in higher education can perhaps help to ease the anxiety of students by being aware of the differences in content knowledge, taking steps to diagnose gaps and emphasising the importance of self study.

3.2 Mathematics skills

For an individual undertaking a physics degree, the study of physics and mathematics are inherently linked. Mathematics is often viewed as the language of physics, without it many physical concepts are virtually impossible to explain. The previous sections discussed, in part, the changes in the content of secondary school Mathematics and Physics leaving qualifications and the differing grade profiles over time. A side effect of this is that often those teaching incoming students gauge what students should be able to do on entering university by considering their own school experiences, however the considerable changes in the structure and content of the qualifications means that these expectations may be misaligned with what the students have actually covered in their previous education.

As has already been discussed in Chapter 1, this thesis does not seek to explore the evolution of mathematics ability over time. However, it is valuable to be able to establish a baseline of student mathematics ability for students on entry to the degree. If there is a difference between student ability and staff expectation of student ability this can have an impact of the usefulness of the teaching for the student. In the following sections the results of a mathematics diagnostic test given to first year physics students at the University of Edinburgh during week one of the 2009-10 academic year will be discussed, along with results of a survey of physics teaching staff at the university that examines staff expectations of the performance of the first year students on the diagnostic test.

3.2.1 The mathematics diagnostic

The diagnostic test was designed by a member of teaching staff within the School of Physics and Astronomy. It was not possible to be involved in the design

process of the test, with the results being made available to the Physics Education Research Group after the test had been administered. A copy of the test can be found in Appendix D. It comprised 20 questions that students were asked to answer in 30 minutes. Students were given no prior warning that they were going to be examined, although all the questions were on topics the students should, in theory, have encountered during secondary school.

The test was graded using a binary marking system where students were awarded one mark if they had correctly answered the question and zero if they had answered incorrectly or only partially answered the question. Questions where students had made no attempt to answer were recorded as blank.

It appears that the test was overly ambitious in the number of questions it asked the students to attempt, as only two students out of a total of 125 (1.6%) attempted all 20 questions, with 12 out of 20 being the average number of questions attempted. In addition the test had not been subjected to any tests in order to ascertain the validity of the instrument such as those outlined in Chapter 2. The decision to use binary marking, where students score one mark if the answer is fully correct and zero in all other cases, may partially account for the low scores recorded as it is more typical within the university to give students partial credit for work which is correctly attempted but incorrect in the final numerical answer.

The overall performance of the class on the test was very low, with students scoring, on average, 29% (this compares with a university recognised standard pass rate of 40%).

3.2.2 Staff expectations survey

In addition to administering the mathematics diagnostic to the the incoming first year students, a survey was designed to assess the expectations of the teaching staff as to how they thought the students would perform on the test. All teaching staff were invited via email to take part in the online survey where they were shown a copy of the test and asked to predict student performance, a full copy of the survey can be found in Appendix E. There are 60 academic staff in the School of Physics and Astronomy, of whom 19 (32%) completed the survey.

3.2.3 Results

In the following sections the results gained from both instruments will be presented and student performance will be contrasted to staff expectation in each of the areas addressed in the expectations survey. Each question will be addressed in turn, with the staff responses displayed alongside the results for the students. Question 3, which asked if a link would be seen between secondary school grades and diagnostic test performance, is not presented here as the relevant secondary school grades were not available in order to complete the analysis.

Question 1 “To the nearest 10%, what do you think the mean test score will be when all our first year students take this test?”

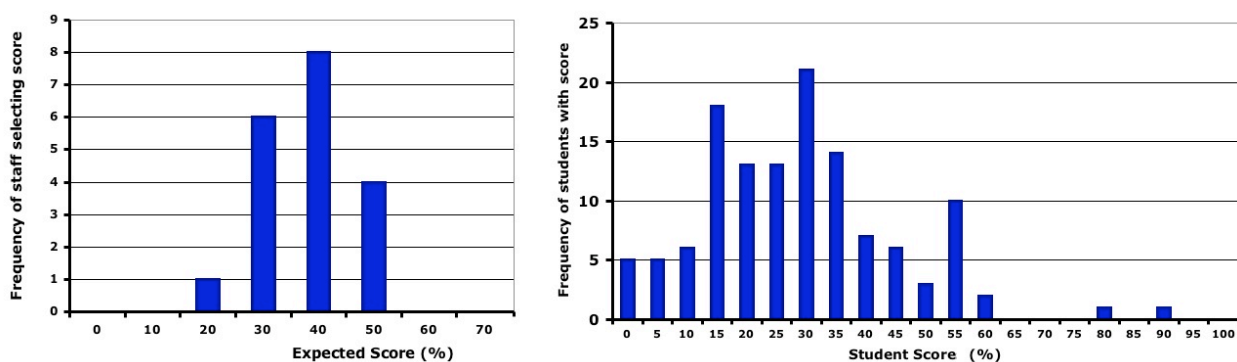


Figure 3.2: Representation of staff responses to Question 1 of Staff Expectations survey and actual student performance on the test. Staff N=19, Student N=125. Mean student score is 29%

As can be seen in Figure 3.2 none of the teaching staff thought that the average mark on the test would be greater than 50%, however 12 out of the 19 staff surveyed (63%) believed the average mark would be above the pass mark (the university recognised standard pass rate is 40%). The overall performance of the class on the test actually was very low, with students scoring, on average, 29%, and few students (24% of the class) scoring a pass grade or higher.

Question 2 “To the nearest 5%, what do you think that the standard deviation of scores will be when all our first year students take this test?”

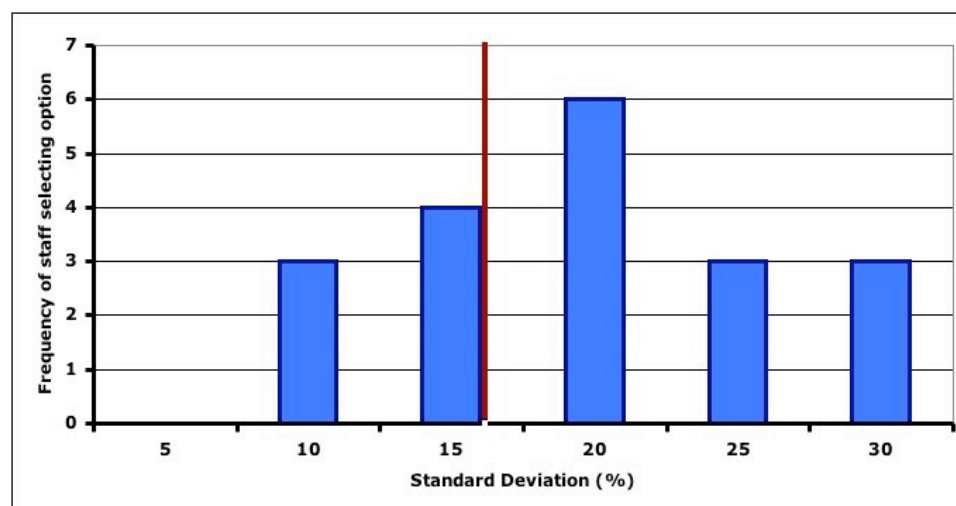


Figure 3.3: Representation of staff responses to Question 2 of Staff Expectations survey. Staff responses are shown in blue with a dark red line corresponding the student standard deviation (16%) Staff N=19, Student N=125

The standard deviation in the student scores was 16% (15% to the nearest 5% as asked in the staff survey). Figure 3.3 shows that there was a large range of responses to the question, with the majority (63%) of staff predicting a larger standard deviation than was recorded, suggesting that staff anticipate a larger range of student abilities within the first year cohort than actually exists.

Question 4 “To the nearest 10%, what percentage of the first year class do you think will get the first question correct?”

The first question of the test asked students to rearrange the equation shown in Equation 3.1 to find an equation in terms of h .

$$F = G \frac{Mm}{(R + h)^2} \quad (3.1)$$

As shown in Figure 3.4 the most popular staff response to the survey was also the response that agreed with student performance, with 72% of the students

correctly answering the question. There are a wide range of staff responses to this question with a 70% range between the lowest predicted score and the highest.

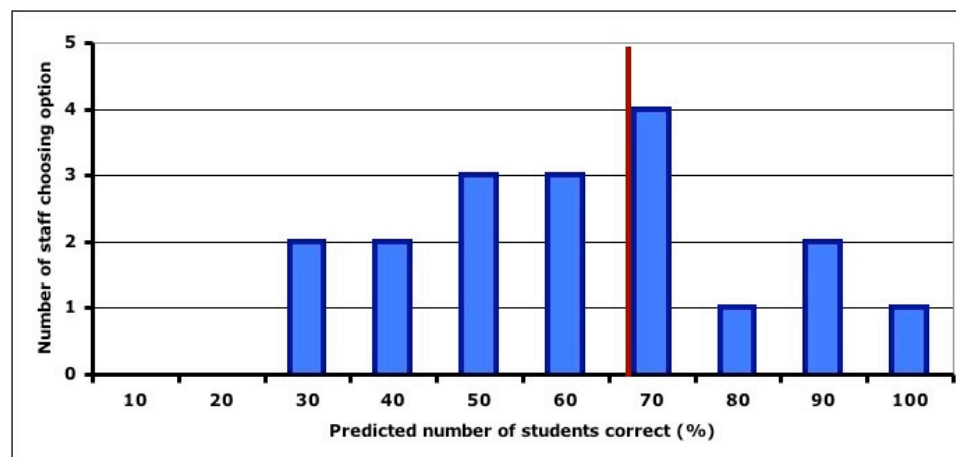


Figure 3.4: Representation of staff responses to Question 4 of Staff Expectations survey. Staff responses are shown in blue with a red line corresponding to student performance on the question. Staff N=18, Student N=125

Question 5 “Please identify what you think are the three questions that they will find the easiest”

In this question staff were asked to identify the three questions which they thought the students would find easiest, they were not, however, asked to rank their choices. As a result the responses to this question have all been accorded equal weight and summed with the question with the highest total number of votes deemed to be the question staff thought would be the easiest.

As only raw test scores are available for analysis the definition of ‘easiest’ in this case has been taken as the questions with the highest percentage score for the cohort and does not include considerations such as time spent on questions.

The question with the most votes from the staff survey was Question 1 which was also the question the student cohort scored the mostly highly on with 72% answering correctly. The questions the staff ranked as the 2nd and 3rd easiest (Questions 7 and 15) were not displayed in the student scores, although the second and third highest scoring questions for students were identified in the top five ranked responses by staff.

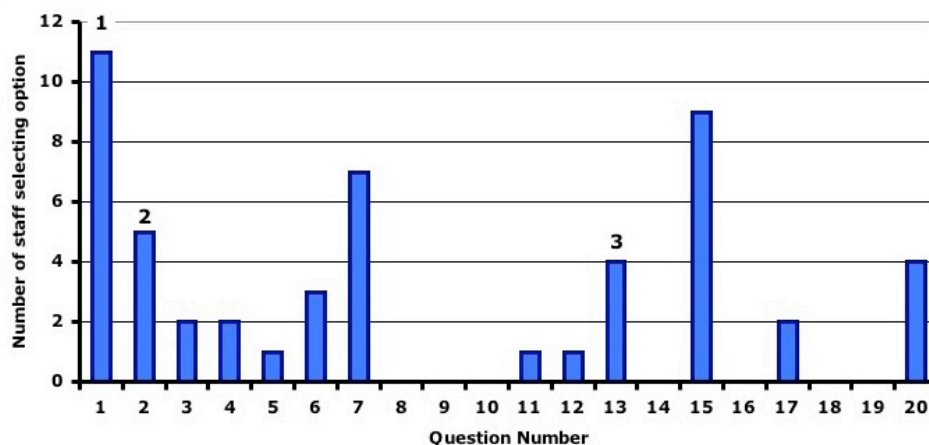


Figure 3.5: Representation of staff responses to Question 5 of Staff Expectations survey. Staff responses are shown in blue with the bar corresponding to three in which the student performance was highest marked by numbers 1,2 and 3. Staff N=18, Student N=125. Numbers 1-3 represent the three questions on which the students scored most highly.

Question 6 “Please identify what you think are the three questions that they will find the hardest”

As in the previous question the staff responses were all weighted equally for this survey item thus identifying the questions students would find the most difficult is complicated by the fact that so few students completed the test. What is not known from the raw scores is if students worked tactically through the test paper, attempting the questions they found easy and choosing to leave blank those they found difficult or if they simply worked through the test and attempted them in order. The only option with this data is to count the total number of incorrect responses for each question (excluding blanks) and assume that the pattern displayed by the students who attempted all questions can be further extrapolated to include responses for the whole class.

None of the three questions identified by the staff as ‘difficult’ (Questions 16, 18 and 19) were reflected in the student test scores. The three questions with the most incorrect responses were all in the first ten questions of the test, this may be a consequence of the fact that more students attempted the first half of the test and left many subsequent questions blank.

If the blank responses are included in the analysis as incorrect answers then

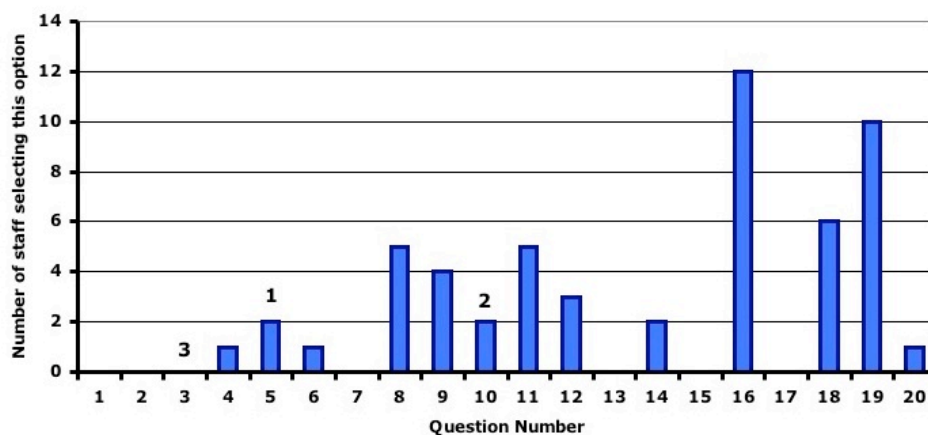


Figure 3.6: Representation of staff responses to Question 6 of Staff Expectations survey. Staff N=18, Student N=125. Numbers 1-3 represent the three questions on which the students scored the lowest.

the results differ with only one of the three identified questions still featuring in the most poorly completed student questions. Using this approach the staff correctly identified one of the three questions that students found most difficult, question 16.

3.2.4 Discussion of results

The fact that the diagnostic test employed to assess student mathematical ability was potentially flawed, in terms of the length and level of the test, will have some impact on the findings of this study. It is the case that questions in the second half of the test were far less widely attempted than those in the first half, which is probably due to the short time the students had to complete the test. This makes it difficult to be sure which questions students found the most easy and difficult rather than which they had time to complete and which they didn't. However while the results of the test may not be of great use to judge the student mathematical ability they are still of interest when comparing student performance and teaching staff expectations of that performance.

Staff expected a higher pass rate to the exam than was actually achieved by the students, given that the time allocated for the test had been made clear to the staff it is reasonable to compare the expectations of the staff and the performance

of the students. The teaching staff also predicted a larger standard deviation than was recorded suggesting that staff anticipated a larger range of student abilities than was found using the test.

The staff predicted the questions that students would find the easiest accurately with the top three scoring questions all within the top five ranked by the staff. The staff were less successful in identifying the questions students found difficult but this may be due to the deficiencies in the test meaning that it is hard to identify if students simply didn't attempt the ones they thought would be the most challenging.

Despite the identified flaws with the test instrument it is still possible to take information from the findings which add to the picture of the school-university transition. It would appear that staff tend to over estimate the the mathematics ability of the incoming students, whether that is in terms of the mathematics they have learnt or the speed at which they can solve problems (a skill that is linked to the amount of practice students have had). This over estimation of mathematics skills may lead to students being presented with problems that they find overwhelming or impossible with their current levels of knowledge, denting student confidence and leaving teaching staff frustrated.

3.3 Scientific reasoning skills

In order for students to become successful scientists they must first acquire a skill set which will enable them to carry out scientific procedures. These transferable skills are used in conjunction with subject knowledge in the process of solving science problems and are highly prized in science teaching (and in fact are sometimes regarded just as highly as content knowledge) [101].

One such skill is scientific reasoning, a skill that must be developed and is not necessarily part of our everyday reasoning [102]. Results of studies into levels of scientific reasoning may show very different results for similar groups of people depending on whether they have been forced to build upon their reasoning skills [102].

One test in this area that has been used extensively in science education is Lawson's Classroom Test of Scientific Reasoning [28]. It seeks to examine common categories of reasoning such as proportional reasoning, deductive and inductive

reasoning, control of variables, probability reasoning, correlation reasoning, and hypothesis evaluation [101]. Each of the questions in the test asks the student to consider a problem and then select an answer from a multiple choice list of answers. The next question to students is then to think about **why** the answer they picked is the correct answer and to pick a reason, again from a list of multiple choice options.

The Lawson Test of Scientific Reasoning has been used to investigate the differences between students in different educational systems (Chinese students and students in the USA (N=5760)) at the end of their high school education [101]. Chinese students have studied far more high school physics than their American counterparts. While it was seen that the Chinese students scored much more highly in tests examining their content knowledge (results consistent with the much greater level of physics education the students have had) the profiles of the two cohorts were nearly identical for scientific reasoning (see Figure 3.7), suggesting increased levels of high school physics content does not increase scientific reasoning skills. In this case the authors conclude that “students ideally need to develop both content knowledge and transferable reasoning skills, researchers and educators must invest more in the development of a balanced method of education, such as incorporating more inquiry-based learning that targets both goals.” [101].

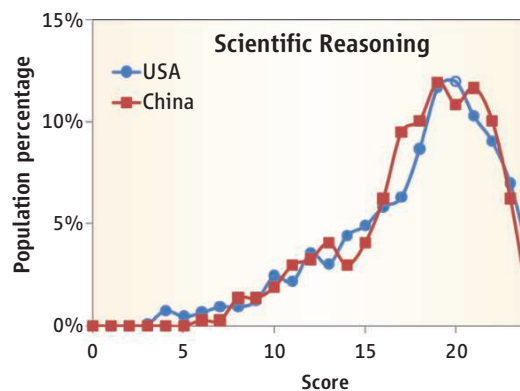


Figure 3.7: Profile of scores in Lawson Test of Scientific Reasoning for Chinese and American high school students. For details see Bao *et al* (Reference [101])

These results show that increased content knowledge is not an indicator of increased scientific reasoning, and yet good scientific reasoning is a necessary

skill of a practicing scientist. It is assumed that, over time, students studying science will increase in their ability to reason scientifically but this is seldom explicitly tested. In the following sections a brief investigation into the differences in reasoning ability on either side of the school university transition gap will be outlined, along with the results found.

3.3.1 Investigation of scientific reasoning using the Lawson Test of Scientific Reasoning

In order to look for any differences in scientific reasoning between secondary school students and those in their first year of undergraduate study, a sample of students from both cohorts were asked to attempt selected questions from the Lawson Test of Scientific Reasoning.

Due to the time scales of this study (which was carried out before the conversion of this project from a masters to a doctoral thesis) it was not possible to track the changes in scientific reasoning ability in a specific group of students over a period of time. As a result testing took place with secondary school students concurrently to testing with students in the first year of a physics degree. In order to justify comparing the two groups attempts were made to make them as similar as possible. The students in secondary schools were all studying physics at the highest level and were from schools with a long track record of sending students to the University of Edinburgh. Whilst this does not guarantee that the students will attend the university or even study physics degrees it suggests that the students sampled will not be too dissimilar to those who will study at the university. In addition, more data was collected on university pre-offer visit (UCAS) days where prospective physics students spend a day visiting the university. Again, visiting is not a guarantee of taking up a place but it does suggest that those taking the test are the *types* of students who will go on to study physics at the university.

The amount of time available to spend with the secondary school students was limited (a 20 minute time slot) thus meaning that there was not sufficient time to carry out the whole test, although a full copy of the test can be seen in Appendix F. The questions used were selected as covering a wide range of areas within the test and a variety of question styles. In total five pairs of questions were used, each of which will be presented below as they appeared in the test (including

relevant diagrams), along with the responses for both cohorts. As mentioned in the previous section the test questions require the students to not only select a multiple choice answer for each question but following that to select the **reason** that they chose their previous answer, therefore a student must correctly give the answer and reason for their answer to be fully correct. For all the questions the results are displayed in a matrix so that the total class break down for both cohorts can be easily seen. The results are then further plotted to show the percentages of each cohort that answered both the question and reason correctly.

A different method of data collection was used for each of the two cohorts; this was not an intentional part of the design of the study but was necessitated by the difficulties in gaining face to face test time with the first year undergraduate students. All data for the secondary school students was collected using electronic voting hand sets or ‘clickers’. Access to the students was gained either through visits to secondary schools within a 50 mile radius of Edinburgh or from visits of prospective students to the University of Edinburgh on UCAS visit days. There was not a specific time limit on the test and, using the voting system software supplied with the clickers, it was possible to check that all students had answered the questions before moving on. However as only a 20 minute time slot was available for the data collection there was an element of time pressure for the students. This was not the case for the university student data which was collected through an online survey with no time limit. In this case there was an incentive of a prize draw for book vouchers used to encourage the participants to take part and the test could be taken at any time over a week long period: 38% of the class (N=111) completed the test. The online test was not compulsory, as result there is likely to be an element of self-selection bias in the group that chose to take the test.

Another subtle difference between the two methods of data collection is the inclusion of a “don’t know” option in the online version of the test. When the secondary school students were given the test they were given verbal instructions that if they did not know how to answer they should make an “educated guess” and if they felt they could not do so they should not vote, which accounts for the blank responses presented in the results below. As the design of the online survey required all students to answer the question before they could move onto the next one a “don’t know” option was added to each question in order to

encourage students to move on to the next question as opposed to giving up on the test entirely, however only 1% of the total responses used this option.

In order to analyse the data collected the results for each question were separated into three categories: “completely correct answer”, “correct answer but incorrect reason” and finally “incorrect answer”, which covered all answers that did not match the correct response, including those that had given an incorrect answer but a correct reason as these were assumed to be those student that were guessing the answers. By categorising the answers this way the results profiles for each of the two cohorts could be easily compared and contrasted and a chi squared test could be used to establish the significance of any differences between the two groups.

3.3.2 Test questions and results

Ten questions were asked in total consisting of five multiple choice questions followed in each case by a multiple choice list of reasons for the given answer. The questions are presented below in the order they were given to the students and as they originally appeared. After each question the results for the two groups are presented and discussed.

Question 1 (Questions 3 and 4 on original test)

“To the right are drawings of two cylinders filled to the same level with water. The cylinders are identical in size and shape. Also shown at the right are two marbles, one glass and one steel. The marbles are the same size but the steel one is much heavier than the glass one. When the glass marble is put into Cylinder 1 it sinks to the bottom and the water level rises to the 6th mark. If we put the steel marble into Cylinder 2, the water will rise:”

- A** to the same level as it did in Cylinder 1
- B** to a higher level than it did in Cylinder 1
- C** to a lower level than it did in Cylinder 1

Because:

- A** the steel marble will sink faster.

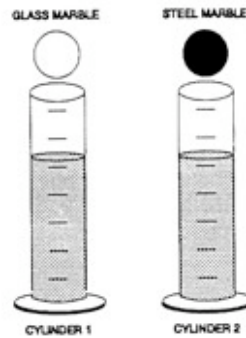


Figure 3.8: Image originally accompanying Questions 3 and 4 of the Lawson test of Scientific Reasoning (See Appendix F)

- B** the marbles are made of different materials.
- C** the steel marble is heavier than the glass marble.
- D** the glass marble creates less pressure.
- E** the marbles are the same size.

The first question given to the students was intentionally chosen to be relatively easy and, hopefully, help the students to gain confidence on the test. As can be seen in Figure 3.9 the question did not pose a significant problem for either cohort, with 76% of the secondary school students giving the correct answer and reasoning and 97% of the university students also doing so. There are only a very small number of secondary school students who gave the correct answer and the incorrect reasoning (and no university students that did), with the majority of incorrect responses from secondary school students choosing answer B *“to a higher level than it did in Cylinder 1”* with reasoning option C *“the steel marble is heavier than the glass marble”*. The full results for both cohorts are shown in Tables 3.1 and 3.2 below. In order to establish if the two cohorts’ answer profiles were statistically different to each other a chi-squared test was carried out on the results using three bins “completely correct answer”, “correct answer but incorrect reason” and finally “incorrect answer”. The results of the test showed the two cohorts were statistically different ($\chi^2=24.141$, $N=294$, $p=0.0001$).

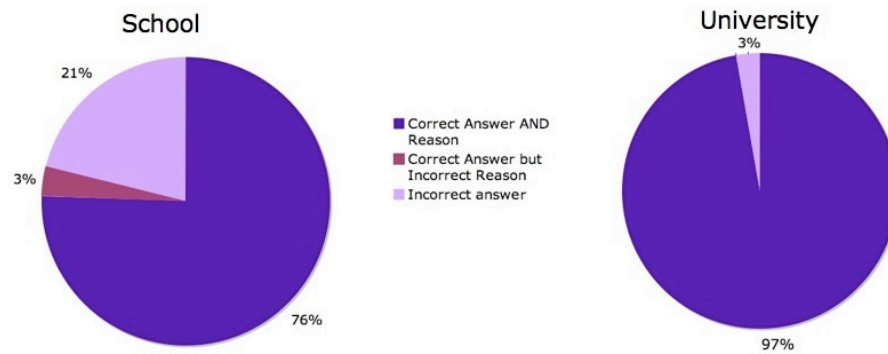


Figure 3.9: Graphical representation of results for secondary school (N=193) and university students (N=111) answering Question 1 of the abridged Lawson Test of Scientific Reasoning.

Table 3.1: Raw numbers of secondary school students choosing each answer option for Question 1 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer A/Reason E) is shown in bold. Total N=193

Reason/Answer	A	B	C	Blank
A	0	0	0	0
B	1	5	1	0
C	1	21	0	1
D	0	5	0	0
E	146	3	1	4
Blank	4	0	0	0

Table 3.2: Raw numbers of first year university students choosing each answer option for Question 1 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer A/Reason E) is shown in bold. Total N=111

Reason/Answer	A	B	C	Don't Know
A	0	0	0	0
B	0	0	0	0
C	0	1	1	0
D	0	0	1	0
E	108	0	0	0
Don't Know	0	0	0	0

Question 2 (Questions 9 and 10 on original test)

At the right are drawings of three strings hanging from a bar. The three strings have metal weights attached to their ends. String 1 and String 3 are the same length. String 2 is shorter. A 10-unit weight is attached to the end of String 1. A 10-unit weight is also attached to the end of String 2. A 5-unit weight is attached to the end of String 3. The strings (and attached weights) can be swung back and forth and the time it takes to make a swing can be timed. Suppose you want to find out whether the length of the string has an effect on the time it takes to swing back and forth. Which strings would you use to find out?

- A only one string
- B all three strings
- C 2 and 3
- D 1 and 3
- E 1 and 2

Because:

- A you must use the longest strings.
- B you must compare strings with both light and heavy weights.

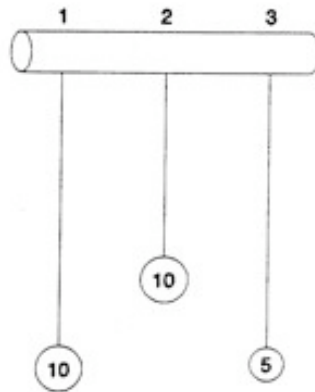


Figure 3.10: Image originally accompanying Questions 9 and 10 of the Lawson test of Scientific Reasoning (See Appendix F)

C only the lengths differ.

D to make all possible comparisons.

E the weights differ.

The results profile for this question is remarkably similar to that of the previous question. 75% of the secondary schools students gave both the correct answer and reason, “*Strings 1 and 2*” because “*only the lengths differ*”. 3% of students gave only the correct answer with incorrect reasoning and 22% gave an entirely incorrect answer. Unlike in the previous question there is some variation in the results of the university students with 89% giving the correct answer and reasoning, 2% the correct answer without a correct reason and finally 9% giving an incorrect answer overall. The full results for both cohorts are shown in Tables 3.3 and 3.4 below.

A chi squared test between the two cohorts established that the two results profiles are statistically different to each other ($\chi^2=9.34$, $N=297$, $p=0.0094$).

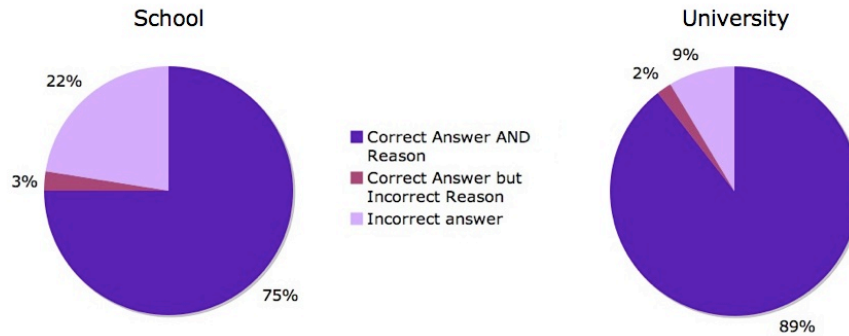


Figure 3.11: Graphical representation of results for secondary school (N=192) and university students (N=105) answering Question 2 of the abridged Lawson Test of Scientific Reasoning.

Table 3.3: Raw numbers of secondary school students choosing each answer option for Question 2 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer E/Reason C) is shown in bold. Total N=192

Reason/Answer	A	B	C	D	E	Blank
A	0	1	0	1	0	0
B	0	0	1	7	2	3
C	2	1	1	3	144	2
D	0	13	0	0	2	0
E	0	0	0	4	1	2
Blank	1	1	0	0	0	0

Table 3.4: Raw numbers of first year university students students choosing each answer option for Question 2 of the abridged Lawson test of Scientific Reasoning. Correct answer as for Table 3.3. N=105

Reason/Answer	A	B	C	D	E	Don't Know
A	0	0	0	0	0	0
B	0	1	0	0	0	0
C	1	0	0	0	94	0
D	0	3	0	0	2	0
E	0	0	0	2	0	0
Don't know	0	0	0	0	0	2

Question 3 (Questions 17 and 18 on original test)

Three red square pieces of wood, four yellow square pieces, and five blue square pieces are put into a cloth bag. Four red round pieces, two yellow round pieces, and three blue round pieces are also put into the bag. All the pieces are then mixed about. Suppose someone reaches into the bag (without looking and without feeling for a particular shape piece) and pulls out one piece. What are the chances that the piece is a red round or blue round piece?

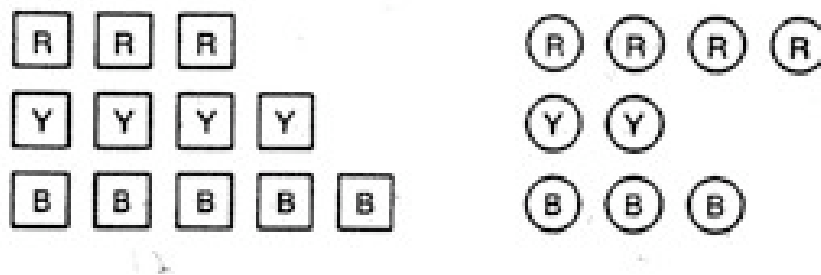


Figure 3.12: Image originally accompanying Questions 17 and 18 of the Lawson test of Scientific Reasoning (See Appendix F)

- A cannot be determined
- B 1 chance out of 3
- C 1 chance out of 21
- D 15 chances out of 21
- E 1 chance out of 2

Because:

- A 1 of the 2 shapes is round.
- B 15 of the 21 pieces are red or blue.
- C there is no way to tell which piece will be picked.
- D only 1 of the 21 pieces is picked out of the bag.

E 1 of every 3 pieces is a red or blue round piece.

The third question of the test, which looked at the students' ability to use probability, produced a results profile very similar to that of the previous two questions. The correct answer and reason was "1 chance out of 3" because "1 of every 3 pieces is a red or blue round piece." Looking first at the secondary school students' results it can be seen that 77% of the students managed to give the correct answer and reasoning, 7% gave the correct answer without the correct reasoning and 16% gave an entirely incorrect answer. The university students once again performed better with 91% giving the correct answer and reasoning, 2% giving the correct answer with an incorrect reason and 7% answering incorrectly.

The chi squared test established a significant difference between the two cohorts ($\chi^2=9.62$, $N=297$, $p=0.0079$).

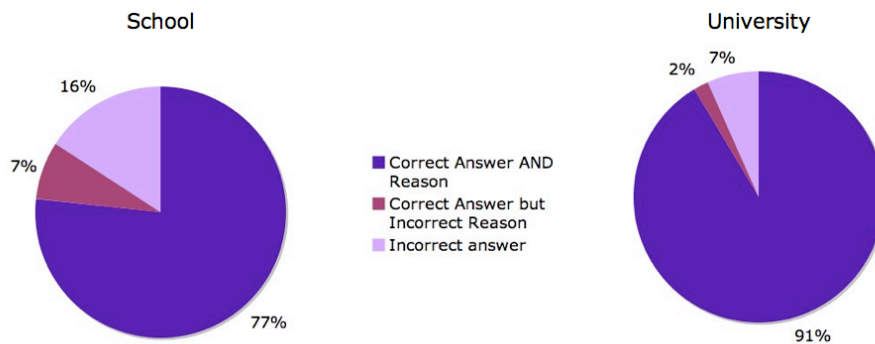


Figure 3.13: Graphical representation of results for secondary school (N=192) and university students (N=105) answering Question 3 of the edited Lawson Test of Scientific Reasoning.

Table 3.5: Raw numbers of secondary school students choosing each answer option for Question 3 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer B/Reason E) is shown in bold. Total N=189

Reason/Answer	A	B	C	D	E	Blank
A	0	1	0	0	0	0
B	0	1	0	5	0	0
C	1	1	0	0	0	0
D	0	4	3	0	4	1
E	1	145	1	9	0	5
Blank	0	7	0	0	0	0

Table 3.6: Raw numbers of first year university students choosing each answer option for Question 3 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer B/Reason E) is shown in bold. Total N=103

Reason/Answer	A	B	C	D	E	Don't Know
A	0	0	0	0	0	0
B	0	0	0	2	0	0
C	0	0	0	0	0	0
D	0	2	1	1	0	0
E	0	94	1	2	0	0
Don't know	1	0	0	0	0	1

Question 4 (Questions 19 and 20 on original test)

Farmer Brown was observing the mice that live in his field. He discovered that all of the mice were either fat or thin. Also, all of them had either black tails or white tails. This made him wonder if there might be a link between the size of the mice and the color of their tails. So he captured all of the mice in one part of his field and observed them. Below are the mice that he captured. Do you think there is a link between the size of the mice and the color of their tails?



Figure 3.14: Image originally accompanying Questions 19 and 20 of the Lawson test of Scientific Reasoning (See Appendix F)

- A appears to be a link
- B appears not to be a link
- C cannot make a reasonable guess

Because:

- A there are some of each kind of mouse.
- B there may be a genetic link between mouse size and tail color.
- C there were not enough mice captured.
- D most of the fat mice have black tails while most of the thin mice have white tails.

E as the mice grew fatter, their tails became darker

It is in this question that we first see a divergence from the answer profiles established in Questions 1, 2 and 3. While it is still the case that the university students outperformed the secondary school students, the percentages of students giving the correct answer and reason has reduced by a large amount. However, comments given by the university students, via the online survey, revealed that in this question it is not as straight forward as students answering incorrectly but instead may be down to differing interpretations of the question.

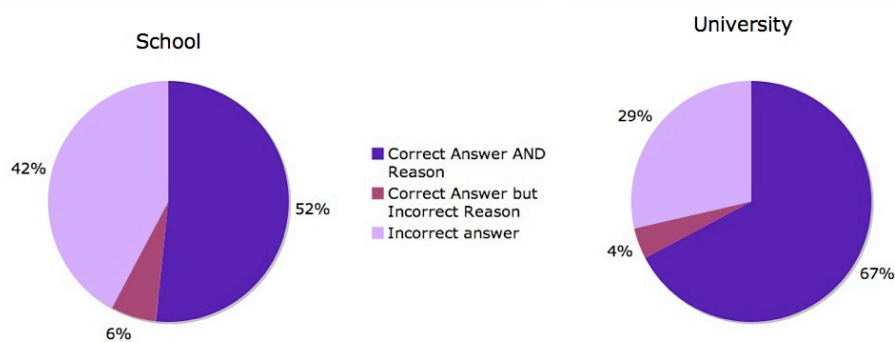


Figure 3.15: Graphical representation of results for secondary school (N=194) and university students (N=101) answering Question 4 of the abridged Lawson Test of Scientific Reasoning.

The correct answer, as given by Lawson in the original test is option A “*appears to be a link*” because “*most of the fat mice have black tails while most of the thin mice have white tails*”, however in the free text response box provided to all the university students taking the test in the online format many argued that this question was misleading as they felt with such a small sample size a firm conclusion could not be drawn, and as a result chose options C for both parts “*cannot make a reasonable guess*” because “*there were not enough mice captured*”. The full results for the secondary school students showed that 51% of the cohort gave the correct answer with reasoning, 6% gave the correct answer without correct reasoning and 42% gave an incorrect answer. It is worth noting that included in the quoted figure for entirely incorrect answer are 11% of students who answered that they couldn’t make a reasonable guess as the sample size was too small. For the university students there is a similar picture with 67% giving

the correct answer with reason, 4% correct without reason and 29% entirely incorrect, however in this case the number of students who answered that the sample size was too small to make a judgement was even higher with 16% of the cohort giving this answer. These raw numbers are displayed for both the cohorts in Tables 3.7 and 3.8.

Despite both groups seeming to similarly interpret the given question there is still a statistically significant difference between the two groups, at the 95% confidence level, ($\chi^2=6.75$, $N=295$, $p=0.034$).

Table 3.7: Raw numbers of secondary school students choosing each answer option for Question 4 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer A/Reason D) is shown in bold. Total N=194

Reason/Answer	A	B	C	Blank
A	0	11	11	1
B	6	3	3	2
C	0	2	22	9
D	100	4	5	5
E	5	2	0	0
Blank	1	0	2	0

Table 3.8: Raw numbers of first year university students choosing each answer option for Question 4 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer E/Reason C) is shown in bold. Total N=101

Reason/Answer	A	B	C	Don't Know
A	0	4	1	0
B	3	0	1	0
C	0	3	16	0
D	68	1	1	0
E	1	0	0	0
Don't Know	0	0	1	1

Question 5 (Questions 11 and 12 on original test)

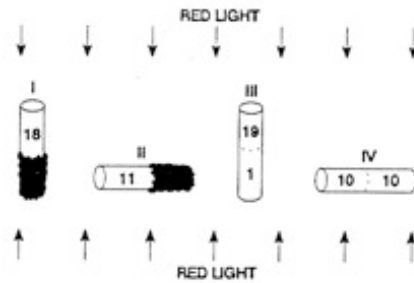


Figure 3.16: Image originally accompanying Questions 11 and 12 of the Lawson test of Scientific Reasoning (See Appendix F)

Twenty fruit flies are placed in each of four glass tubes. The tubes are sealed. Tubes I and II are partially covered with black paper; Tubes III and IV are not covered. The tubes are placed as shown. Then they are exposed to red light for five minutes. The number of flies in the uncovered part of each tube is shown in the drawing. This experiment shows that flies respond to (respond means move to or away from):

- A red light but not gravity
- B gravity but not red light
- C both red light and gravity
- D neither red light nor gravity

Because:

- A most flies are in the upper end of Tube III but spread about evenly in Tube II.
- B most flies did not go to the bottom of Tubes I and III.
- C the flies need light to see and must fly against gravity.
- D the majority of flies are in the upper ends and in the lighted ends of the tubes.
- E some flies are in both ends of each tube.

In the final question of the abridged test we again see a different answer pattern in the student responses. Unlike in Question 4 where this may be due to student interpretation of the question, in Question 5 there is no evidence to suggest that students were choosing to interpret the question any differently to as it was intended.

Looking at the charts in Figure 3.17 we can see that, at 21% correct for school students and 43% for university, this is the question which had the fewest students giving the correct answer “*gravity but not red light*” because “*most flies are in the upper end of Tube III but spread about evenly in Tube II*”. It is also the case that, for both cohorts, this is the question which sees the highest percentage of students giving a correct answer but with incorrect reasoning with 26% of secondary school students and 25% of university students doing so. This could be an example of students showing incomplete scientific reasoning, where they are able to reason an answer but not able to fully articulate **why** it is the answer.

University students were more likely than secondary school students to be able to answer the question correctly with 53% of secondary school students giving an entirely incorrect answer compared to 32% of university students.

As in all previous questions the raw students numbers were used in a chi squared calculation between the two cohorts. It was found that the two answer profiles were statistically different with a value of ($\chi^2=17.76$, $N=295$, $p=0.0001$).

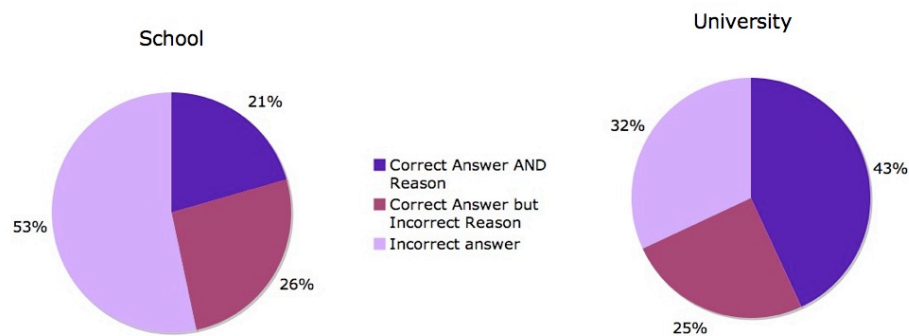


Figure 3.17: Graphical representation of results for secondary school (N=194) and university students (N=101) answering Question 5 of the edited Lawson Test of Scientific Reasoning.

Table 3.9: Raw numbers of secondary school students choosing each answer option for Question 5 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer B/Reason A) is shown in bold. Total N=180

Reason/Answer	A	B	C	D	Blank
A	7	37	8	1	2
B	2	39	3	1	2
C	4	2	3	0	1
D	17	4	30	3	2
E	1	1	4	2	0
Blank	1	1	1	1	0

Table 3.10: Raw numbers of first year university students students choosing each answer option for Question 5 of the abridged Lawson test of Scientific Reasoning. Initial answers to the question are displayed along the top of the table with reasons displayed vertically. The correct answer and reason (Answer B/Reason A) is shown in bold. Total N=100

Reason/Answer	A	B	C	D	Don't Know
A	2	43	2	1	0
B	1	19	0	0	0
C	1	1	3	2	0
D	3	3	8	1	0
E	0	1	0	6	0
Don't Know	0	0	0	0	2

3.3.3 Discussion of results

A study was carried out using approximately 300 students which has established striking differences in the scientific reasoning between the two cohorts. In all five of the questions asked the university students outperformed the secondary school students scoring more highly and more consistently as a group. In all the questions the differences in the results profiles for the secondary school pupils and the university students were statistically significant at the $p=0.05$ level. Given the high score of the university students in the first three questions it could be argued that the test was not of a high enough level to adequately probe the scientific

reasoning of these students. Whilst this may be true it remains a valuable tool in highlighting the differences in the two cohorts on either side of the tertiary education gap. The university students would not have learned specifically how to answer the questions (with the exception of Question 3 as probability may well have been taught in mathematics classes) however in the first year of university they have obviously developed more reasoning ability.

The fourth question of the test was the first to show more variation in the answer profiles, while part of this may be down to interpretation of the question there are still a reasonable number of students in both cohorts who failed to answer correctly without going for an alternate interpretation of the question.

In the final question, particularly for the secondary school students, we see an example of incomplete reasoning where approximately 25% of the students in both cohorts are able to answer the question correctly but not the reason why.

It is impossible from this study to draw any firm conclusions as to what has caused these differences in scientific reasoning, indeed even whether it is the nature of the teaching at university or a natural development with age. Perhaps what is being observed is a natural consequence of the ‘Hidden Curriculum’ [103] where students learn skills on a meta cognitive level without specifically being taught (an idea which will be explored further, with reference to data handling skills in Chapter 5). What we can say, without speculation, is that it has been shown that differences do exist meaning that when students first enter tertiary education educators should be aware that some students may have less developed scientific reasoning, manifesting itself in an inability to articulate **why** an answer is correct.

3.4 1st year physics student attitudes towards study

As described in Chapters 1 and 2, one of the most widely used instruments to examine student attitudes towards study is the Colorado Learning Attitudes about Science Survey (CLASS) [34] which has evolved from previous attitudinal surveys as an instrument that allows the calculation of a percentage of expert-like thinking for either individual students or a whole class by comparing student responses with expert opinions. Studies have shown that on entering tertiary

education students' levels of expert-like thinking will generally decline and become less expert after one semester of higher education [34], unless the curriculum design explicitly addresses these concerns [45]. While the survey has been used largely in North America there are now a few emerging examples of it being used in a wider context [44], with the pattern of a decline in student expert-like thinking, after the first year of instruction, being repeated. In this section the focus will be on student attitudes on either side of the tertiary education boundary and during the first year of teaching. Results of a longitudinal study using the survey as well as more detailed analysis of the data collected in the first year can be found in Chapter 4.

3.4.1 Implementation of the CLASS survey

The CLASS survey has been given to students in the first year physics class at the University of Edinburgh for three consecutive academic years, 2008-09, 2009-10 and finally 2010-11. The students were surveyed twice within each year, once in the very first week of semester one and again in the last week of teaching in the second semester, before the students depart on exam leave. The test was given in paper form with no explicit time limit for the students to fill it in, although they were asked to do so without talking and under 'examination conditions'. As the survey is designed to measure the changes in attitudes of the students it is necessary to track the identities of the students taking the test in order to match pre- and post-teaching surveys. All students were assured that their answers would be treated as confidential and would not be seen by any of the teaching staff. In total 385 surveys have been collected and have passed all data validation checks.

When surveying students in secondary schools pre- and post- instruction data was not collected, instead the students were surveyed once, in order to give an insight into their expert-like thinking at a particular time. The students were either surveyed during visits to local secondary schools (N=92) or were surveyed when visiting the University of Edinburgh on pre acceptance (UCAS) visit days (N=65). The two groups of secondary school students are considered to be distinct from each other as, although they are all of the same age and taking secondary school physics to the highest level, those in the pre university visit days have expressed an intention to study physics at university while those collected

during visits to schools may go on to study physics at university but this will not necessarily be the case. Unlike the data collected during first year university teaching the data collection with school pupils was done so anonymously as only one data collection point was used.

The total number of surveys collected for all the cohorts is shown in Table 3.11. As no statistically significant differences are detected between the data collected for any of the three consecutive years of first year data, the data has been combined to give one data set for first year attitudes.

Table 3.11: Number of completed student surveys from secondary school students and first year undergraduate

Cohort	Number
Secondary School	92
Pre University Visits	65
1st Year (2008-09)	147
1st Year (2009-10)	134
1st Year (2010-11)	104

3.4.2 Results of CLASS survey

The overall percentages of favourable and unfavourable expert-like thinking for each of the cohorts are shown in Figure 3.18 (the students surveyed on university visit days have been labelled ‘intending physics’).

Looking first at the favourable expert-like thinking, the secondary school students surveyed during physics classes scored, on average, 64(2)% expert-like thinking while the the secondary schools students who intended to study towards a physics degree have an average score of 72(2)%. This difference is statistically significant ($p=0.002$). Here, as in all subsequent cases, the bracketed number represents the standard error on the mean². There is no statistically significant difference between the secondary school students who intend to study physics and those beginning the first year physics course who have an average of 71(1)% ($p=0.68$).

²Calculated by dividing the standard deviation of the sample by the square root of the number of responses in the sample.

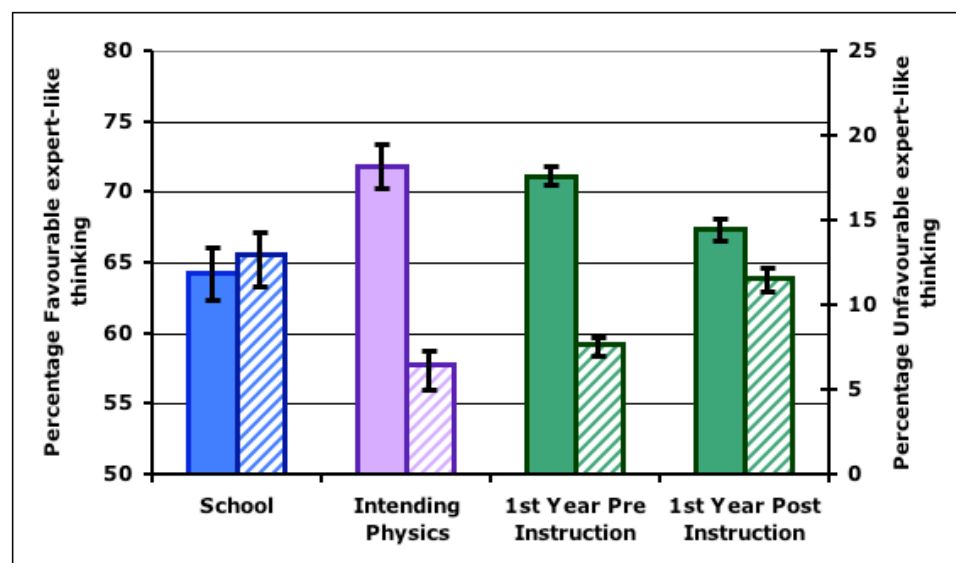


Figure 3.18: Percentage of expert-like thinking for secondary school students and 1st year university students. Each pair of bars is twinned with the solid lines representing **favourable** expert-like thinking and the hatched lines **unfavourable** expert-like thinking. The favourable thinking corresponds to the left hand axis and the unfavourable to the right. The first two x-axis categories represent the students in secondary schools (those in school physics classes (N=92) and those on pre university visit days (N=65)). The second two categories represent the average score of the university students, pre and post first year instruction (N=385). Error bars represent the standard error on the mean.

After two semesters of instruction, the students exhibit a reduction in expert-like thinking, with a class average of 67(1)% expert-like thinking; this decrease is significant ($p \leq 0.001$). The drop is not unexpected and is in line with previous studies, which, as previously stated suggest that unless a curriculum is designed to specifically address student epistemologies, a drop in expert like thinking will be seen [46].

The unfavourable scores follow the same pattern as the favourable, the students surveyed on school visits show the highest percentage of unfavourable expert-like thinking with an average of 19(1)%. This score is significantly different to that of the school students surveyed on pre university visits who recorded an average of 11(1)% ($p \leq 0.001$). Again there is no significant difference between those students intending to take physics and those starting a physics degree ($p=0.26$). After the first year of university instruction the students increase in

their unfavourable expert-like thinking, changing from 12(0)% to 15(1)%, this difference is significant ($p \leq 0.001$).

The difference in expert-like thinking between the two cohorts of secondary school students surveyed is perhaps suggestive of a selection effect with the students with the most expert-like thinking about physics being those who have already chosen to go on and study physics at university. The students were surveyed at the same point in the academic year and were taking the final school physics classes available before university, the only known difference between the two groups was that those surveyed on pre-university visit days have all applied for places on physics degree programs. The idea of a selection effect across transitions is explored in greater detail in Chapter 4.

3.5 Chapter summary

In this chapter an attempt has been made to better understand the changes that a student undergoes in their first year of undergraduate study, a time that is notoriously full of upheaval, distractions and new experiences. This is the time that most students are likely to leave a degree program, either to study a different course, or in some cases, leave higher education all together. Four distinct areas have been covered, the preparedness of students for academic content, the difference between the mathematical preparedness of the students and staff expectations, the evolution of scientific reasoning in the first year and the changes in student attitudes and expertise.

Study of school leaving qualification syllabi have uncovered that students entering the University of Edinburgh (although the results could be applied to most higher education institutions in the UK, in varying percentages of accuracy) will be coming from a breadth of different backgrounds with potentially very large differences in the subject knowledge that they have covered. The existence of seven different physics A-level syllabi result in differences in content covered but also (perhaps more importantly) experience gained, especially in practical work and long written answer style examinations. Even if the intake were entirely from Scottish students (who at present account for 45% of the first year undergraduate class) there would still not be continuity. The situation is less complicated as the Scottish Qualifications Agency (SQA) oversees all the qualifications with only one

syllabus for each. However due to the fact students can enter the undergraduate program with either Highers or Advanced Highers there is still the potential for some students to have had one year less physics experience on entering tertiary education.

It appears that teaching staff within the School of Physics and Astronomy have higher expectations of student mathematical ability than is achieved by the students. While the test used to measure the student mathematics skills has potential flaws, the most obvious being the lack of required time to complete the test, the staff were estimating the performance of students on that specific test, aware of the potential pitfalls for students. This difference between teaching staff expectations and student performance may be due to the changing course content of school leaving qualifications, meaning that staff assume material will be covered which, in fact, has not.

The development of scientific reasoning is known to continue into early adulthood and is also known to develop further as humans are forced to use their reasoning ability [102]. For this reason it is not surprising that the results found using the Lawson Test of Scientific Reasoning show that those who have had one more year of science education perform better than those who have not. However these results do offer a base line of the level of reasoning students are capable of on entering university, offering perhaps explanations for situations where instructors feel frustration at a student lack of willingness to *'think the problem through'*. As first year undergraduate students performed so highly on the test it may be that the test was too straight forward to reveal gaps in the cohorts' scientific reasoning. Due to the development of other areas of this thesis the decision was made not to develop a new testing instrument but these results have helped to contribute to another research project on Student Problem Solving [31].

It has been widely documented using several survey instruments [43],[34] that student levels of expert-like thinking decrease after one year of instruction. The results presented in this chapter suggest that this pattern is also seen in a university physics class in the UK and in addition provide insight into the expert-like thinking of students in school. It has been shown that students in a secondary school (highest level) physics classes display lower levels of expert-like thinking than those students, at the same level of study, who have already elected to study

physics at university. In contrast there is no difference in expert-like thinking between those students in school intending to study physics and those students who begin study on the physics degree program. The drop in expert-like thinking during the first year is examined further and discussed in Chapter 4. If this drop in expertise is due to a drop in confidence by the students then it may be valuable information for tertiary educators in attempting to increase retention as it may be possible to take steps to balance this drop, studies into which are beginning to emerge from North America [104].

The findings of this chapter present a mixed picture of the school to university transition for physics students. Depending on the school systems studied under and the level of qualifications taken students may feel under or over prepared for the academic workload, relative to their peers. Both scenarios can potentially have an effect on student confidence, and study habits. On entry to tertiary education we are seeing that the students have high levels of expert-like thinking when asked to think about how they study the subject (as do those in schools who have elected to study physics at the tertiary level), however after one year of study we see a decline in favourable attitudes. Initial results indicate that student mathematics ability does not reflect the expectations of those teaching them, which may have implications for the success of students on the course. In terms of student ability to reason scientifically it has been shown that there is a large gap in reasoning ability between secondary school and university, showing that students may initially feel overwhelmed by the tasks they face in the physics classroom. An awareness of these four areas, combined with relevant steps to reassure students, may help to ease the school-university transition.

Chapter 4

Attitudes Towards Physics Study

In the previous chapter, results were presented where the Colorado Learning Attitudes about Science Survey (CLASS) had been used to examine student attitudes towards study on either side of the school-university transition. In this chapter the usage of the instrument is presented in greater detail and with a far wider range of people surveyed. Firstly, a more in depth study of the changes in attitudes of the first year physics class will be discussed, followed by the presentation of data showing the evolution of student attitudes and beliefs towards study during the whole of the undergraduate degree.

When looking at the attitudes of the students in the first year physics class at the University of Edinburgh, in Section 3.5 of Chapter 3, it was shown that after one year of instruction, a decline was seen in student ability to think like an expert physicist. The observed decline was not unexpected, with previously published studies showing a decline after the first year of instruction[34],[44], as has been extensively discussed in Chapter 1¹. What has been far less widely investigated is what happens to these student attitudes **after** the first year of study, with no published work in this area available at the beginning of this PhD research project. Although two studies have been published since which partially address this issue ([53] and [105]), there still have not been studies making use of CLASS that examine the attitudes of physics students at higher levels in as greater level of detail or breadth of survey cohort as those covered in this thesis. One of

¹It should be noted that few studies are actually published showing this decline, this is believed to be due to the fact that it is now considered the accepted result when using this instrument and therefore it is more commonly seen that studies are published when they specifically do not find this decline.

the factors which may have deterred researchers from investigating the changes in student attitudes and beliefs over the course of a degree program is the large time scale involved in such a study. In order to investigate the expert-like thinking of the same group of students over the course of an undergraduate degree would require four years of data collection, longer than the time period of most doctoral research projects (and a considerable time commitment for any research group). Presented in this thesis is a more detailed examination of first year undergraduate student attitudes towards study, followed by two further sections which address potential ways to carry out longitudinal studies looking at the changes in student attitudes over time. Firstly a cross-sectional or pseudo longitudinal study is discussed in which multiple cohorts are sampled in the same academic year and inter-year comparisons are drawn and secondly an incomplete fully longitudinal study where the attitudes of the same students have been measured over three consecutive years.

4.1 1st year attitudes and beliefs towards study

In the following sections the change in the attitudes and beliefs of the first year students at the University of Edinburgh will be discussed, with reference to the data that has been collected over three consecutive years, along with a discussion of the differences seen in the expert-like thinking of various sub cohorts of the class as a whole, specifically, the differences between physics majors and non-majors and the differences seen when separating the individual results by gender.

As has been explained in Chapter 2, the first year class is made up of approximate 300 students per year, of which half are intending to study for physics degrees and the other half are taking physics as a complimentary subject to another degree. In this thesis the two halves have been labelled physics ‘majors’ and ‘non-majors’ but it is important to distinguish between our non-majors and those at a North American higher education institutions. All the students who take first year physics must be qualified to take the course and these entrance requirements are no different for the majors and non-majors meaning that they only thing that distinguishes the two cohorts is that the majors have chosen to take a physics degree while the non-majors have chosen to take a degree in a different subject.

Statistical tests (single factor ANOVA) were carried out on the data to look for any statistically significant differences between the three sets of first year data collected over three successive academic years. No differences were found in inter year comparisons (either in the overall data or any of the sub groups of the cohort) for the any of the years of data. Given that no statistically significant differences exist the data was combined to give a larger sample size and higher confidence in the statistical tests carried out. The level of similarity in the results from year to year is illustrated in Figure 4.1 which shows the data from each academic year separately as well as combined.

The one exception to the combining of data is in the gender sub cohort for the 2010-11 data where the female data was not included in the combined data, this was because while all the other sub groups of the cohort remained statistically consistent with the previous two years, the female student data was not consistent. It is presented in this thesis, along side the data for the previous two years, along with discussion as to why the results may be different.

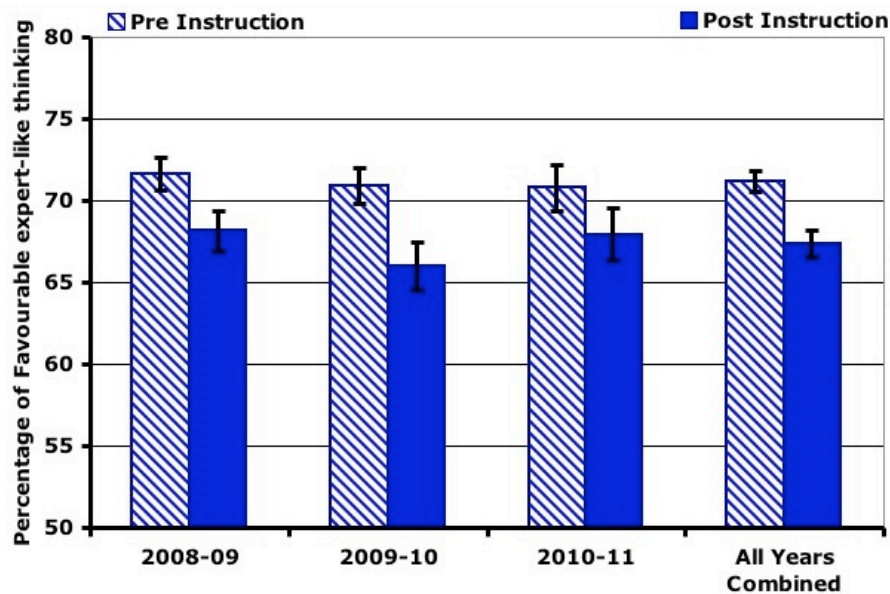


Figure 4.1: Percentage of **favourable** expert-like thinking pre- and post- first year instruction for three consecutive academic years and all the years combined. The hatched bars indicates pre-instruction data and filled bars post-instruction. Error bars represent the standard error on the mean. N=147 for the 2008-09 data, N= 134 for 2009-10 and N=104 for 2010-11. Combined N=385 for both pre and post instruction data

Due to the fact that only matched survey answers (where students have filled in both pre- and post teaching surveys) are of use in this study approximately 40% of the collected survey responses had to be discarded. Presented in Table 4.1 are the numbers of completed matched surveys for each academic year. The reason the 2010-11 academic year is smaller is due to the fact the class size had reduced by approximately a third due to a change in recruitment policy.

Table 4.1: Number of completed student surveys for the 2008-09, 2009-10 and 2010-11 academic years, broken down by degree studied and gender. Where totals of sub-cohorts do not equal the overall cohort total this is due to either degree or gender information being unavailable for a few students.

	2008-09	2009-10	2010-11
All	147	134	104
Physics Majors	110	93	54
Non Majors	37	41	49
Male	112	108	72
Female	34	27	32

4.1.1 Results of whole class

Looking first at the combined data for all three years of the study (2008-2011), as presented in Figure 4.2, we can see that the students are entering the degree program with high levels of expert-like thinking, with 71(1)% agreement with the expert responses; here, as in all subsequently quoted values, the bracketed figure represents the standard error on the mean.

After two semesters of instruction, the students exhibit a reduction in expert-like thinking, with a class average of 67(1)% expert-like thinking; this decrease is significant ($p=0.009$). The drop is not unexpected and is in line with previous studies, which, as previously stated suggest that unless a curriculum is designed to specifically address student epistemologies, a drop in expert like thinking will be seen [46]. It is, however, worth noting that the students are coming in with relatively high levels of expert-like thinking compared to physics students in North America, with their level of expert-like thinking approximately 6% higher pre-instruction than those seen by the CLASS survey authors [34] and while there is a significant drop, the students are still remaining at high levels and record

approximately an 8% higher percentage than those in the same paper. This pattern is repeated in a study from outside the US where students in Saudi Arabia also show a drop in expert-like thinking post-instruction [44]. Therefore while the students at the University of Edinburgh begin and end their first year instruction with higher levels of expert-like thinking, the overall drop is of approximately the same magnitude as those seen in international studies.

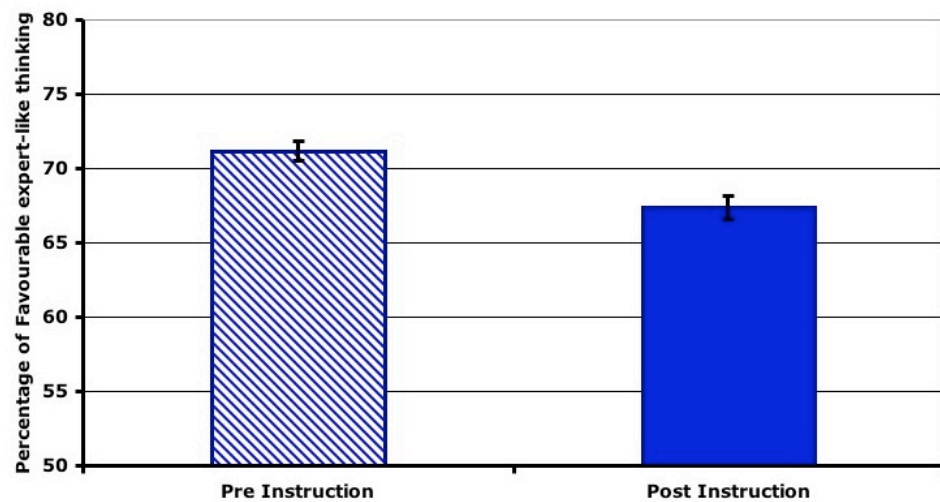


Figure 4.2: Percentage of **favourable** expert-like thinking pre- and post- first year instruction. The hatched bar indicates pre-instruction and the filled post-instruction. Error bars represent the standard error on the mean. $N=385$ for both pre and post instruction data

The inverse pattern is repeated when looking at the unfavourable data, as shown in Figure 4.3, the students commence the program with low levels of expert-like thinking, with 12(0)% disagreement with the expert opinion, after one year of instruction an increase in disagreement with the expert opinion is seen with the class scoring 15(1)% disagreement. Due to the neutral option available when answering all questions it is not necessarily the case that the unfavourable scores when combined with the favourable, will total 100%. Whether the unfavourable scores seen here are in line with those seen in other academic institutions is not known, due to the fact that only the favourable scores are reported in most published literature (including those by the survey designers).

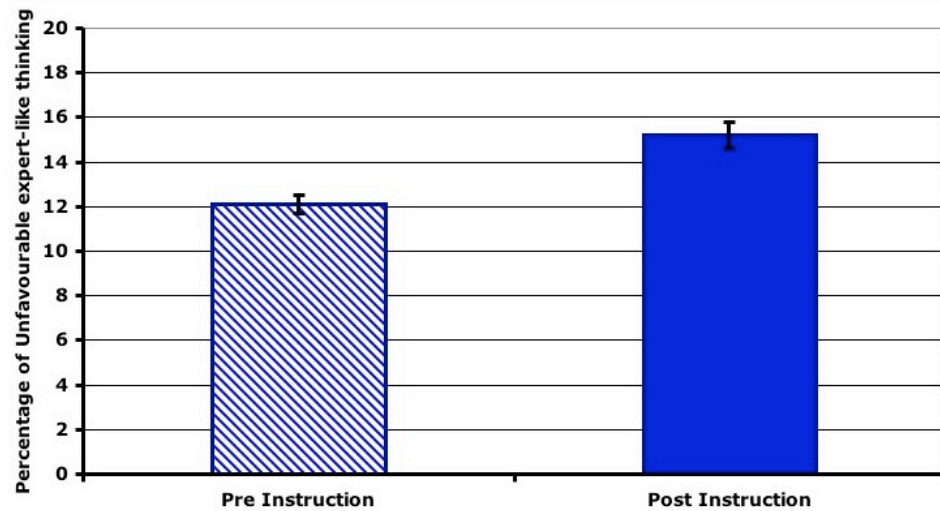


Figure 4.3: Percentage of **unfavourable** expert-like thinking pre- and post- first year instruction. The hatched bar indicates pre-instruction and the filled post-instruction. Error bars represent the standard error on the mean. $N=385$ for both pre and post instruction data

4.1.2 Results of physics majors and non-majors

One way to split the class into different cohorts is by looking at those who intend to study physics as a degree (majors) and those who are taking physics as an outside subject (non-majors). Figures 4.4 and 4.5 shows the pre- and post-instruction percentages of expert-like thinking for both majors and non-majors, in terms of favourable and unfavourable expert-like thinking.

There are several points of interest emerging from the results in figures 4.4 and 4.5: firstly, the majors and non-majors are starting from different levels of expert-like thinking; for favourable expert-like thinking this difference is 5% pre-instruction and is statistically significant ($p=0.004$), although in terms of unfavourable expert-like thinking there is not a significant difference ($p=0.10$). Recalling that all students must meet the same minimum entrance requirements and have carried out the same level of high-school physics as the physics majors, the only difference between the two groups is the intended degree course of study. This difference in background distinguishes the students in this study from those in other studies that have looked at the differences between majors and non-majors. However the results found remain in line with those seen in

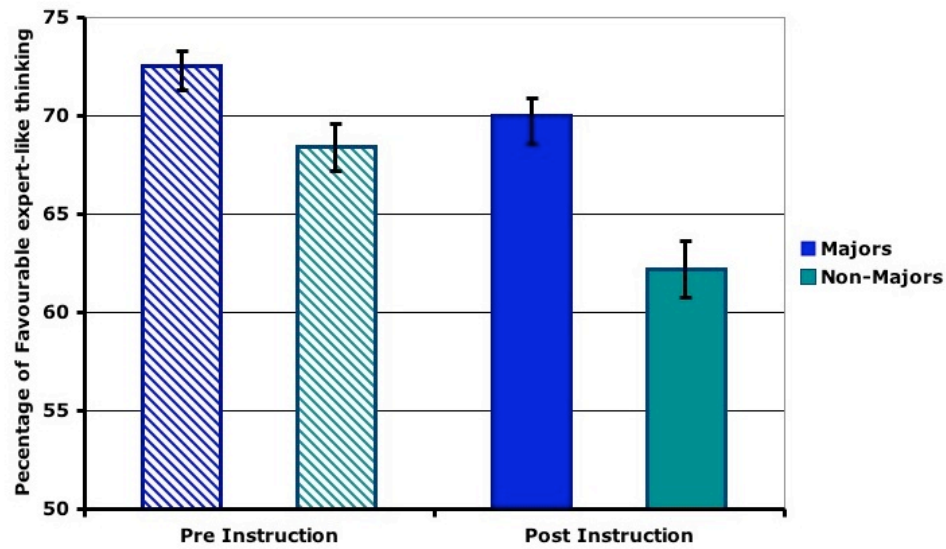


Figure 4.4: Percentage of **favourable** expert-like thinking for physics majors (N=257) and non-majors (N=127) both pre- and post- first year instruction. The hatched bars indicates pre-instruction data and the filled bars post-instruction. In all cases error bars represent the standard error on the mean.

previous studies where, for example, engineering students have been shown to have significantly less expert-like views than physics majors [53].

The other point of interest is the magnitude of the decrease in expert-like thinking. The majors show a relatively small drop in agreement from the expert opinion going from 73(1)% pre instruction to 70(1)% post teaching. In contrast the non-majors show a change of 6%, going from 68(1)% pre-instruction to 62(1)% post. The change for the majors is significant ($p=0.001$), as is the change for non-majors ($p \leq 0.001$). Similar changes are seen in the unfavourable expert-like thinking with a significant difference now existing between the two groups ($p \leq 0.001$).

On observing that differences exist in the attitudes and beliefs of physics majors and non-majors, even before instruction has taken place, further investigations were conducted into the differences between these two specific sub groups. The unique structure of the Scottish degree system provides a sub-cohort of students that are different to those represented as non-majors in the previously published literature, as has been extensively discussed in previous chapters. In an attempt to contrast the attitudes of the students in the first year physics class

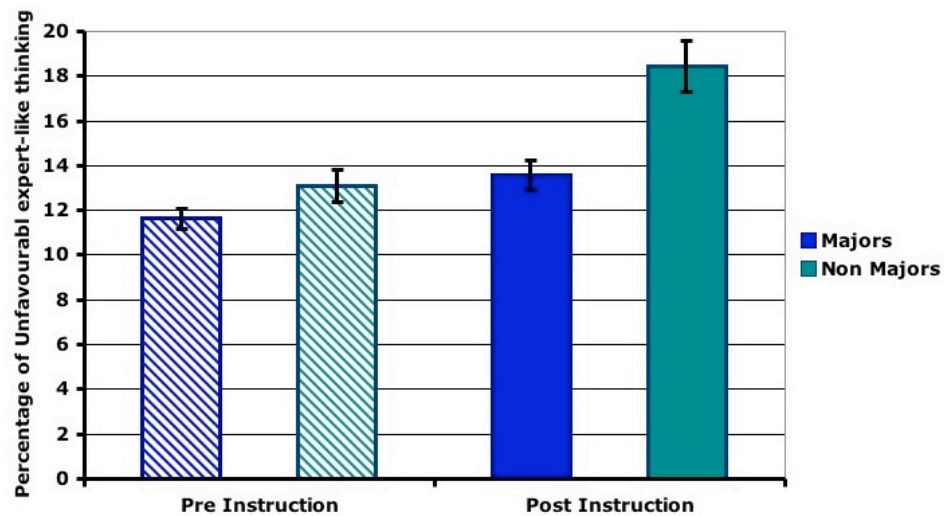


Figure 4.5: Percentage of **unfavourable** expert-like thinking for physics majors (N=257) and non-majors (N=127) both pre- and post- first year instruction. The hatched bars indicates pre-instruction data and the filled bars post-instruction. In all cases error bars represent the standard error on the mean.

to those more traditionally thought of as ‘non-majors’ the CLASS survey was given to students taking a first year physics module as part of a course offered to biological science students wishing to explore quantitative methods (Quantification in the Life Sciences 1 (QUILS))[106]. The biological science students were surveyed only once, either in week 5 or week 6 of Semester 1 of the 2010-11 academic year. As discussed in Chapter 2 the time at which students are tested can have an effect on the results found; as the biological sciences students were sampled at a point between the time of the pre- and post-instruction tests of those in the physics class (and on only one occasion) it is not possible to determine whether the scores gained should be most accurately compared to those of the physics class before or after instruction. The students were asked to fill in the survey on paper during physics tutorials. As with dissemination of the survey to the physics class, the students were given a short introduction as to the nature of the survey and what the results would be used for. Students were asked to complete the survey anonymously as no follow up surveys were necessary. None of the 82 completed student surveys failed the fail-safe question to determine data quality.

The levels of expert-like thinking displayed by the biological sciences students

are depicted in Figures 4.6 and 4.7, along side the scores for the physics majors and non-majors as have previously been shown in Figures 4.4 and 4.5. The biological sciences students' average favourable score was 20% lower than the pre-instruction of the non-majors with a class average of 48(2)%, similarly the unfavourable score for the biological sciences students was 19% more in disagreement with the expert opinion than that of the non-majors. Post instruction the gap remains with a 14% difference between both the favourable and unfavourable scores. All the differences reported are statistically significant ($p \leq 0.001$ for the pre-instruction and post instruction scores scores (both favourable and unfavourable)).

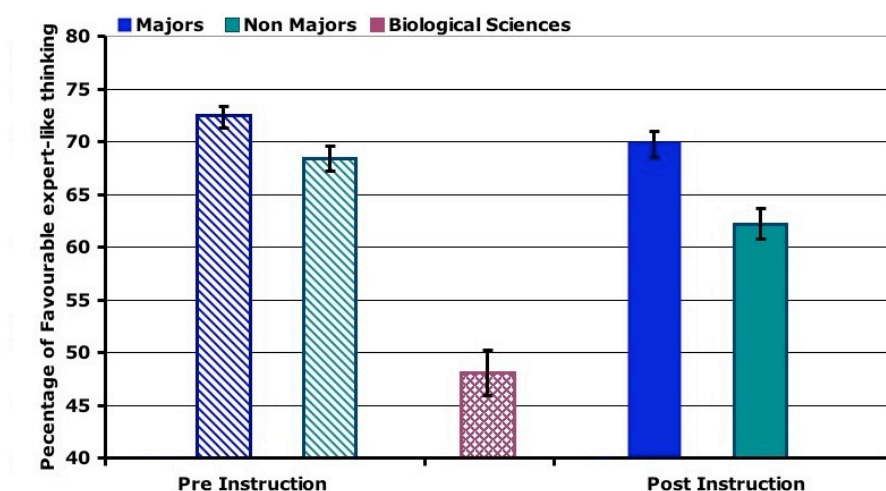


Figure 4.6: Percentage of **favourable** expert-like thinking for physics majors ($N=257$) and non-majors ($N=127$) both pre- and post- first year instruction and biological sciences students mid instruction ($N=82$). The hatched bars indicates pre-instruction, the filled post-instruction and the cross hatched mid instruction. In all cases error bars represent the standard error on the mean.

The far lower scores of the biological sciences students in terms of their attitudes towards physics are perhaps not surprising. The course title and description does not reveal the course content (one half physics, one half maths) and the students enrolled on the course need not have studied physics to the highest level at school. The results do, perhaps, give a more accurate measurement of the views of a typical non-major as they would be termed in a North American institution with the score of the biological sciences students in line with those seen for non majors at other institutions [34],[53]. This again suggests that the non-majors at the University of Edinburgh provide an insight

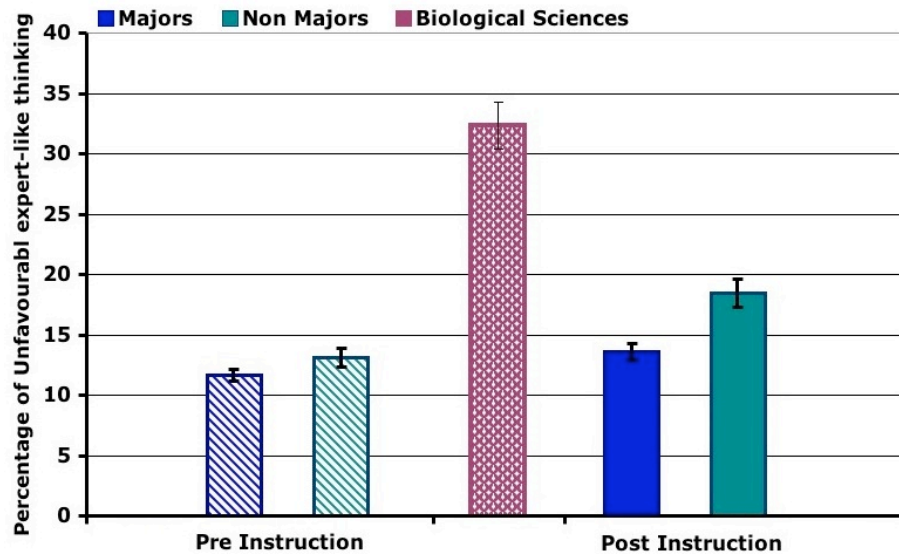


Figure 4.7: Percentage of **unfavourable** expert-like thinking for physics majors ($N=257$) and non-majors ($N=127$) both pre- and post- first year instruction and biological sciences students mid instruction ($N=82$). The hatched bars indicates pre-instruction, the filled post-instruction and the cross hatched mid instruction. In all cases error bars represent the standard error on the mean

into the differences in epistemologies that exist where the cohort is split only by choice of degree and no other factors.

Alongside collecting data into a different group of students studying a first year physics course, the difference in CLASS scores between the physics majors and non-majors within the first year physics class was further probed by looking into the educational background of the students in the class. An investigation was carried out to determine whether there was any correlation between the secondary school leaving qualifications and grades of the first year physics class and their CLASS score, with the hypothesis being that the non-majors may possess lower incoming scores or be less qualified than the majors.

All qualifications in the United Kingdom, regardless of whether they were gained in Scottish or English schools, are assigned to a national tariff of points by the Universities and Colleges Admissions Service (UCAS)[107]. By using this points system the qualifications of all the students can be compared without complication as to the awarding body (for example a Grade A at either A-Level or Advanced Higher would both be awarded 120 UCAS points). No statistically

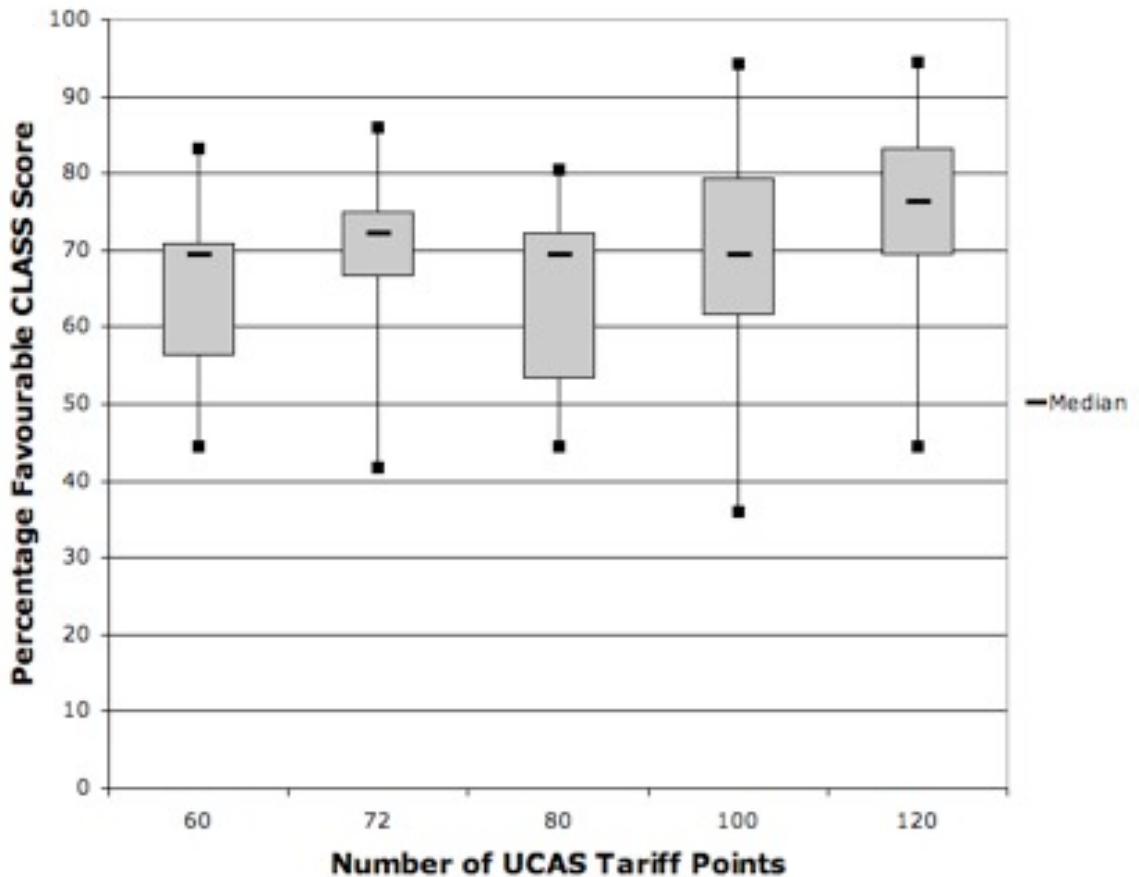


Figure 4.8: Comparison of number of UCAS points earned at school, with overall favourable CLASS score. 60 UCAS points represents a Grade B at Higher or a Grade D at A-Level. 120 UCAS points represents a Grade A at either A-Level or Advanced Higher (Values correct for university entry up to and including September 2012). The dark line on each box represents the median for each grouping.

significant difference is seen between the entry qualifications in physics and mathematics of the majors and non-majors ($p=0.20$). In addition the correlation between physics background of the students and CLASS score has been examined. It does not appear that previous physics background is a strong predictor of CLASS score, as illustrated in Figure 4.8.

4.1.3 Comparison of male and female attitudes towards study

As mentioned previously, differences were identified in the female cohort of the 2010-11 data when compared to other years. No differences were found in the data between 2008-09 and 2009-10 and for this reason the data has been combined for 2008-10 and the 2010-11 academic year data is presented separately. Splitting the whole data set to look at the differences between male and female students, again revealed differences. Figures 4.9 and 4.10 shows the percentage of favourable and unfavourable expert-like thinking for each of the two groups.

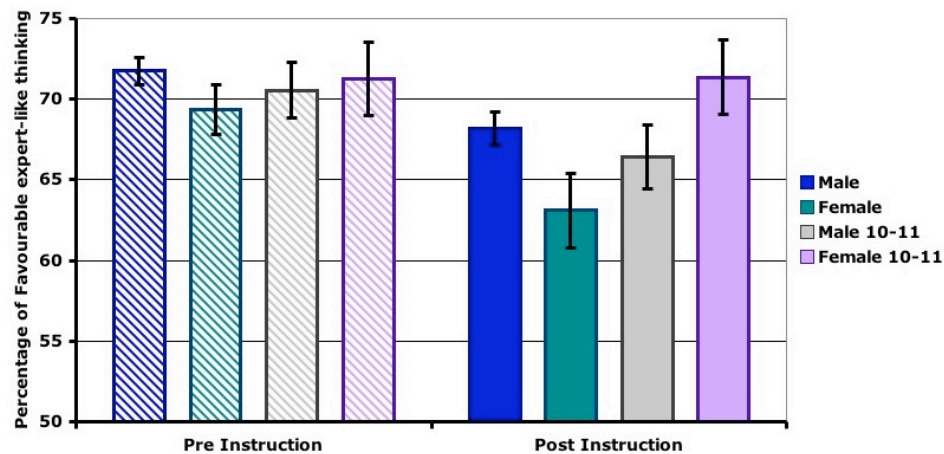


Figure 4.9: Percentage of **favourable** expert-like thinking for male physics students (N=222 for 2008-10) (N=72 for 2010-11) and female physics students (N=61 for 2008-10) (N=32 for 2010-11) both pre- and post- first year instruction. The hatched bars indicates pre-instruction and the filled post-instruction. In all cases error bars represent the standard error on the mean.

Looking first at the 2008-10 data there is a key difference between the subgroups shown in Figures 4.4 and 4.5 and those shown in 4.9 and 4.10. Unlike the majors and non majors there is no statistically significant difference between the male and female students before instruction, either in terms of favourable or unfavourable thinking ($p=0.18$ and $p=0.71$ respectively). The male students record 72(1)% expert-like thinking before instruction and the female students 69(2)%. After instruction a gap has opened between the scores of the two groups with the male students scoring 68(1)% and the female students 63(2)%. The two

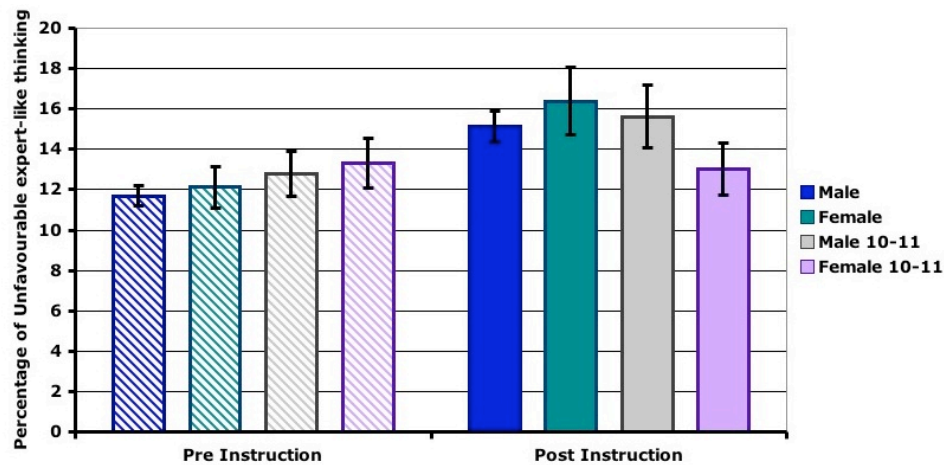


Figure 4.10: Percentage of **unfavourable** expert-like thinking for male physics students (N=222 for 2008-10) (N=72 for 2010-11) and female physics students (N=61 for 2008-10) (N=32 for 2010-11) both pre- and post- first year instruction. The hatched bars indicates pre-instruction and the filled post-instruction. In all cases error bars represent the standard error on the mean.

groups are now statistically different to each other ($p=0.049$), indicating that the first year experience has a greater impact on the attitudes of the female students.

It is when the data is split between male and female students that the 2010-11 academic year data does not follow the same pattern as previous years, more specifically in the performance of the female students, as can be seen in Figures 4.9 and 4.10. In the case of the male students there is no statistically significant difference when compared to the data from the two previous years. Looking at the pre-teaching data $p=0.53$ when comparing the 2010-11 academic year to the data from 2008-09 and $p=0.62$ when comparing to the 2009-10 data. There is also no statistically significant difference looking at the post instruction data for the male students ($p=0.21$ between the 2010-11 data and that from 2008-09 and $p=0.86$ between 2009-10 and 2010-11). This pattern is repeated for the unfavourable scores with $p=0.40$ and $p=0.39$ before instruction and $p=0.43$ and $p=0.82$ post instruction.

The data for female students, however, is not consistent with previous years. There is still consistency at the 95% confidence level when looking at the pre-instruction scores ($p=0.79$ between the 2008-09 and 2010-11 data and $p=0.32$ between 2009-10 and 2010-11). It also remains true that there is no gap between

the male and female students pre-instruction. It is in the post-instruction data that we see a move away from the pattern seen in previous years. Previously the female students have declined in their levels of expert-like thinking after one year of instruction, however in this year they have remained constant. In addition, where we have seen in previous years that, post-instruction, there is a statistically significant difference between the male and female students this is not the case for this data ($p=0.12$).

Whether this change in the behavior of the female students is evidence of some change in the cohort or a statistical anomaly is impossible to determine from one data point and further testing of the cohort is needed to establish this. It is possible that the result does reflect changes in the physics class as the make up of the class is noticeably different to that of previous years; the class is only two thirds of the size of the proceeding year (due to a change in recruitment policy) meaning that students get more individual attention. The smaller number of places available on the course also impacted on the types of students on the course as there was far greater competition for places therefore meaning that only the strongest students were given places on the degree program. In addition this is the first year where non-majors outnumber the majors on the course meaning that the demographic of the classes is slightly different. The nature of the first year physics class means that new teaching methods are constantly being tested and evaluated with the students, in the 2010-11 academic year a new system to encourage student interactive learning and participation was being implemented [77] which may also have had an effect on the class.

4.2 Pseudo-Longitudinal (cross-sectional) study

The Physics Education Research group at the University of Edinburgh have made use of the term ‘pseudo-longitudinal’ to describe the first of the studies we have carried out which examine student attitudes towards science over a long period of time. A similar study which has been carried out in the US in parallel to our own study uses the alternate term of ‘cross-sectional’ to describe the same method [53]. The method involves giving the CLASS survey to cohorts in differing years of the degree program and comparing the expert-like thinking of one year group to the year groups around them. The method of data collection is the ‘snap

shot' method which was described previously where the surveys are completed anonymously as each cohort is surveyed only once and it is not necessary to compare the changes in any particular cohort over any period of instruction.

The study aims to examine the attitudes towards physics of a range of people currently involved in practicing physics, from school pupils and students (both undergraduate and postgraduate) through to academic staff in university physics departments. All data included in the study was collected in the first semester of the 2009-10 academic year², with the data in years four and five of the undergraduate degree and the academic staff responses collected as part of a fourth year undergraduate senior honours research project. In total 637 survey responses passed all the data validation checks for CLASS (as detailed in Chapter 2) and were included in the results presented here. A break down of the numbers involved in each cohort is shown in Table 4.2.

Table 4.2: Number of completed surveys included in the pseudo-longitudinal study

Level	Cohort	Number
Secondary School	School	92
Secondary School	School (Intending Physics)	65
Undergraduate	1st Year	127
Undergraduate	2nd Year	105
Undergraduate	3rd Year	61
Undergraduate	4th Year	57
Undergraduate	5th Year	23
Postgraduate	1st Year	33
Staff	Post docs or academic staff	74

There are two cohorts of data collected in schools, as has been discussed in Chapter 3, the data labelled 'school' was collected in secondary school physics classes in and around Edinburgh. The pupils were all taking the highest levels of physics classes available at secondary level (either Higher or Advanced Higher qualifications) but it is not known if any of the pupils surveyed intended to go on to study physics at university. Those students in the 'intending' cohort are also in school taking the highest level of physics, however the data was collected

²With the exception of the third year data which was collected at the start of the second semester and the data related to the school pupils (school and intending) which was collected in the previous academic year.

on pre university visit days from students who have decided to study physics at university. Years 1 through 5 of the undergraduate data were collected from students in the physics degree program at the University of Edinburgh and the postgraduate data was collected from surveys given to the incoming PhD students at the University. The data included for the first year undergraduate students includes only the physics majors, as they are the only students that would progress into the higher stages of the degree program.

The cohort labelled ‘staff’ are academic staff from the Universities of Edinburgh, Glasgow and St. Andrews, and a few academic staff who responded to an email asking for participants which was sent to a mailing list of physical sciences practitioners in the UK³. The school student surveys, as well as those for years one through three of the undergraduate degree and the postgraduate students were collected in paper format, while those for years four and five of the undergraduate degree and for the staff were collected through an online survey.

4.2.1 Validity of the pseudo longitudinal approach

For the pseudo-longitudinal method to be valid, one assumption must be made with regards to all the data in the study; that the students in each of the cohorts are the same *types* of students as those recruited in the years around them. This is a feature of the fact that as all data is collected in the same academic year, in order to make inter year comparisons we are required to accept that the students in one year are taken from the same sample as the students in any other year. This assumption can be tested, in part, through the repeated surveying of year cohorts to look for consistencies with in years. In Table 4.3 the overall favourable expert-like thinking scores are reported for repeated measurements for the first year cohort. As above only the majors are included here as only they are used in the pseudo-longitudinal study although similar consistency is seen when looking at the class overall or even just the non-majors.

ANOVA tests on the three data sets show that there are no significant differences between the year cohorts. When firstly looking at the pre instruction data it is found that $p=0.98$ for favourable expert-like thinking and $p=0.33$ for the unfavourable data. In the case of the post-instruction data $p=0.23$ when comparing favourable expert-like thinking and $p=0.20$ for the unfavourable data.

³Physical Sciences -Education JISC mailing list.

Table 4.3: Pre and post instruction CLASS scores for the first year physics class over three consecutive years. Both favourable and unfavourable scores are shown. Bracketed figures represent the standard error on the mean.

Year Cohort	Pre Fav.	Pre Unfav.	Post Fav.	Post Unfav
2008-2009	72(1)	11(1)	71(1)	13(1)
2000-2010	73(1)	11(1)	68(2)	15(1)
2008-2009	73(1)	13(1)	72(2)	13(1)

Where other repeated measurements have been made the same pattern is reproduced with consistency shown for repeated sampling of the same year cohort in consecutive academic years. The third year of the undergraduate degree has been sampled on two occasions (during the 2009-10 and 2010-11 academic years), with no differences seen in the cohort averages ($p=0.27$ for favourable data and $p=0.33$ for unfavourable). The incoming postgraduate cohort has also been surveyed twice (again during the 2009-10 and 2010-11 academic years), looking at the overall favourable scores no significant difference is found between the two ($p=0.17$) however for the unfavourable data there is a significant (although marginal) difference with $p=0.045$. The sample size of the 2010-2011 postgraduate data was relatively small ($N=21$) which can lead to uncertainty in the measurement, especially in the unfavourable sections where the percentages are so low that a change in one overall answer from negative to positive can have a large effect. For this reason the difference in unfavourable score between the two postgraduate measurements do not lead us to believe that the method is not valid.

Of the 16 measurements included in this comparison 15 of them show no significant difference which is strongly suggestive of the fact that the pseudo-longitudinal method is valid. The question of validity will be addressed again in Section 4.4 with reference to comparisons between a pseudo longitudinal and fully longitudinal method.

4.2.2 Overall favourable and unfavourable attitudes

The overall average expert-like thinking for each cohort has been calculated and compared to the neighbouring cohorts, the favourable and unfavourable scores for each of the groups are shown in Figure 4.11.

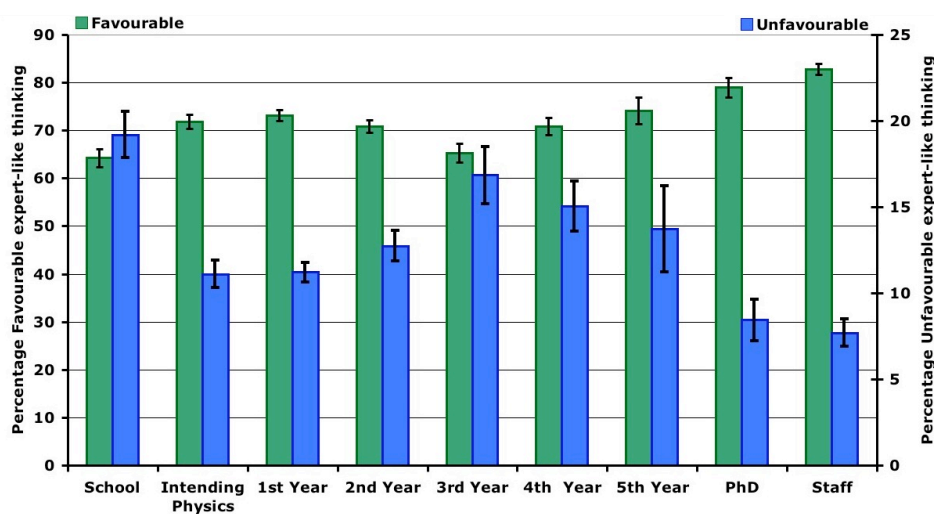


Figure 4.11: Percentage of favourable and unfavourable expert-like thinking for school pupils, students in an undergraduate degree, postgraduate students and staff. The green bars show the percentage of favourable expert-like thinking and are plotted against the left hand axis while the blue bars show the percentage of unfavourable expert-like thinking and are plotted against the right hand axis. The errors bars represent the standard error on the mean.

There are three areas of interest in the figure: The area on the left hand side of the plot, which looks at the attitudes of those in school and the first year of university, the central section which displays the scores for students during an undergraduate degree and, finally, the section on the right hand side which represent the attitudes of students in their masters year, first year of PhD and those working in physics. Each of the sections outlined will be discussed separately for ease of understanding.

The first area of interest highlighted is that of the attitudes of students in schools and in the first year of university, this has already been discussed in Chapter 3 as part of a study on the school-university transition and will only be briefly revisited here. A large difference is seen in the attitudes of those in school taking the highest physics qualification and those at the same stage in their schooling who have also elected to study physics at university, this difference is statistically significant ($p=0.0019$). No differences are seen in the expert-like thinking of students in school who intend to study physics and those students who begin a physics degree at the University of Edinburgh ($p=0.48$). Conversely looking at the unfavourable scores the same pattern is seen with students in school

classes showing significantly higher unfavourable expert-like thinking ($p=0.001$) and no difference existing between those intending to do a physics degree and those who begin the physics degree program ($p=0.90$).

The attitudes of students during the undergraduate degree is the next area of focus. The expert-like thinking the students displayed is predominately flat, with the exception of third year, with no statistically significant difference between the attitudes of students on entering the degree program in first year and those exiting their degrees at the end of a Masters degree. An ANOVA test shows that there is a difference in the undergraduate data over all ($p=0.0044$), however isolating the third year results and rerunning the ANOVA test without the third year data shows no significant differences in any of the other results ($p=0.42$). Looking at the unfavourable results a similar pattern exists with the third year cohort an obvious outlier (although it should be noted that there is no significant difference between the 3rd year and 4th year data $p=0.42$). Having established that the third year data are obviously different to that in the surrounding cohorts several steps have been taken in order to investigate this point further: Firstly focus groups were held with students in years four and five of the undergraduate degree program where a version of Figure 4.11 was shown to the group and students were asked to discuss the figure (timelines and transcripts of the focus groups are included as Appendix G). The students were not surprised by the drop in expert-like thinking seen in third year with one student saying;

“Maps perfectly how I felt about physics.....Maps perfectly how I thought the rest of the year felt about physics.”

with others stating that they felt they had progressively lost confidence in their abilities during the course of their degree

“1st year - you think you’re an expert because you don’t know enough about it.”

“it’s psychological. You realise even after studying it for hours and hours, you’re not going to do well in the exam. I lost confidence and enthusiasm in 3rd year.”

While the students feel that their expert-like thinking is linked to their confidence in the subject (an idea that will be explored further in Section 4.5), this is not necessarily the only explanation for the low score in the third year, it is also possible that the result is anomalous. To test this theory the third year

cohort in the 2010-11 academic year were also surveyed. As has been seen in Section 4.2.1 no significant differences were found between the data collected in the 2009-10 academic year and the following year, however, the 2010-11 average is numerically higher. If the 2010-11 data is used as the third year value of expert-like thinking instead of the original and compared to the other years as shown in Figure 4.11 an ANOVA test shows no significant difference between 1st and 5th year for the favourable scores ($p=0.19$) although a difference still remains for the unfavourable scores ($p=0.034$). This suggests that the low score for third year may produce a false positive effect where a difference is detected which does not necessarily exist, further data collection is needed to establish whether this is the case.

The final area of interest in Figure 4.11 is the transition from undergraduate to postgraduate, as seen on the right hand side of the plot. Here we see that there is a significant difference ($p=0.0033$) between the students in fourth year of the undergraduate degree and those embarking on PhD programs. The fifth year, however, appears to act as an intermediate point, with an average favourable (and unfavourable) score of expert-like thinking which lies between those of the year four and PhD cohorts. Year five is not significantly different to either of the years which neighbour it ($p=0.32$ and $p=0.17$ for favourable scores and $p=0.65$ and $p=0.067$ for unfavourable), again suggesting that this is an intermediate year. Recalling that year five is the optional masters year which involves a year long research project, it is perhaps not surprising that year five acts as the bridging year with students who are considering doing a PhD generally opting to stay on for the fifth year of study and conversely some students who choose to take the masters year deciding to stay on for a PhD after experiencing a year in research.

No difference is detected between the scores of those embarking on a PhD and the those of staff ($p=0.10$ for the favourable scores and $p=0.61$ for unfavourable scores). Interestingly the staff scores are 17% away from 100% expert-like thinking, although it should be noted that the survey authors have stated that a few of the questions included in the survey do not have an expert opinion and so it is not likely to be possible to score 100%. This does however raise questions as to the difference in epistemologies between American and British academic staff, a topic which will be discussed in Chapter 6.

4.2.3 Category results

Until this point, this chapter has focussed on the total scores of expert-like thinking over all questions in the CLASS survey, however the categories that 27 of the questions have been placed into at least one of (for details refer to Chapter 2) can offer additional insight into the thinking of each cohort. Interestingly, while the overall scores are largely flat over the course of the undergraduate degree, there is by no means consistency in the scores at the category level, as illustrated in Tables 4.4 and 4.5 .

Table 4.4: Overall percentages of favourable CLASS responses, for all categories and cohorts surveyed. The bracketed figure represent the standard error on the mean.

Categories	School		Undergraduate					Post	Staff
	Doing	Int.	1st	2nd	3rd	4th	5th		
Personal Interest	66 ₍₃₎	80 ₍₃₎	80 ₍₂₎	74 ₍₂₎	67 ₍₃₎	76 ₍₃₎	80 ₍₃₎	80 ₍₄₎	86 ₍₂₎
Real World									
Connection	67 ₍₃₎	75 ₍₃₎	77 ₍₂₎	75 ₍₃₎	68 ₍₃₎	74 ₍₃₎	67 ₍₅₎	74 ₍₄₎	85 ₍₂₎
PS General	74 ₍₃₎	85 ₍₂₎	81 ₍₂₎	75 ₍₂₎	67 ₍₃₎	74 ₍₂₎	73 ₍₄₎	82 ₍₃₎	85 ₍₂₎
PS Confidence	72 ₍₃₎	84 ₍₃₎	78 ₍₂₎	71 ₍₃₎	63 ₍₄₎	71 ₍₃₎	63 ₍₆₎	80 ₍₃₎	77 ₍₃₎
PS Sophistication	64 ₍₃₎	75 ₍₃₎	76 ₍₂₎	65 ₍₃₎	54 ₍₃₎	67 ₍₃₎	60 ₍₅₎	79 ₍₃₎	84 ₍₃₎
Sense making/ effort	74 ₍₂₎	82 ₍₂₎	82 ₍₂₎	77 ₍₂₎	75 ₍₃₎	75 ₍₃₎	88 ₍₂₎	85 ₍₃₎	93 ₍₁₎
Conceptual									
Understanding	66 ₍₃₎	74 ₍₂₎	74 ₍₂₎	68 ₍₂₎	58 ₍₃₎	69 ₍₃₎	71 ₍₅₎	78 ₍₄₎	81 ₍₂₎
Applied Conceptual									
Understanding	52 ₍₃₎	59 ₍₂₎	62 ₍₂₎	59 ₍₃₎	50 ₍₃₎	59 ₍₃₎	62 ₍₅₎	75 ₍₃₎	74 ₍₃₎

The variation in categories seen makes it perhaps more surprising that no overall change in scores is seen between the beginning and end of an undergraduate degree, while the student scores are increasing in some categories as students progress through a degree they are also decreasing in others. The system used by the survey designers to create the question categories means that not all questions fit into only one category, some into more than one category and some do not fit into any categories at all, as a result it is not possible to say that any changes in overall score will directly correspond to variations in the category scores per cohort.

Whilst the story reflected by the category changes is not immediately obvious, there exist changes which are apparent: As might be expected staff are more expert-like or at the same level as school pupils in all categories suggesting an

Table 4.5: Overall percentages of unfavourable CLASS responses, for all categories and cohorts surveyed. The bracketed figure represent the standard error on the mean.

Categories	School		Undergraduate					Post	Staff
	Doing	Int.	1st	2nd	3rd	4th	5th		
Personal Interest Real World Connection	20 ₍₃₎	5 ₍₁₎	6 ₍₁₎	8 ₍₁₎	16 ₍₂₎	11 ₍₂₎	6 ₍₂₎	6 ₍₂₎	5 ₍₁₎
PS General	18 ₍₂₎	10 ₍₂₎	7 ₍₁₎	10 ₍₁₎	18 ₍₃₎	12 ₍₂₎	21 ₍₅₎	8 ₍₂₎	5 ₍₁₎
PS Confidence	10 ₍₂₎	3 ₍₁₎	5 ₍₁₎	7 ₍₁₎	13 ₍₂₎	12 ₍₂₎	9 ₍₃₎	6 ₍₁₎	6 ₍₁₎
PS Sophistication	9 ₍₂₎	4 ₍₁₎	7 ₍₁₎	8 ₍₂₎	17 ₍₃₎	15 ₍₂₎	15 ₍₅₎	5 ₍₂₎	10 ₍₂₎
Sense making/ effort Conceptual Understanding	15 ₍₂₎	7 ₍₂₎	6 ₍₁₎	14 ₍₂₎	21 ₍₃₎	16 ₍₃₎	19 ₍₄₎	8 ₍₂₎	6 ₍₁₎
Applied Conceptual Understanding	14 ₍₂₎	5 ₍₁₎	7 ₍₁₎	9 ₍₁₎	12 ₍₂₎	11 ₍₂₎	6 ₍₂₎	5 ₍₁₎	2 ₍₁₎
	18 ₍₂₎	10 ₍₂₎	10 ₍₁₎	17 ₍₂₎	21 ₍₂₎	18 ₍₃₎	18 ₍₃₎	11 ₍₃₎	11 ₍₂₎
	27 ₍₂₎	19 ₍₂₎	17 ₍₁₎	20 ₍₂₎	24 ₍₂₎	24 ₍₃₎	22 ₍₄₎	12 ₍₃₎	14 ₍₂₎

upward trend in expert-like thinking over all the categories. However this is not a gradual increase with students becoming slowly more expert-like over time, in fact the 5th year are seen to be less expert-like than the first year students in some categories. The third year categories are all noticeably lower than those in the rest of the undergraduate degree, perhaps suggesting that the third year dip seen in the overall data is not merely a statistical anomaly. That such variation can be recorded in the categories while the overall score remains constant is a feature of the CLASS instrument that will be discussed in Chapter 6 as part of a larger discussion on the uses and limitations of the CLASS instrument.

4.2.4 Selection effect

The transitions seen at the entry and exit points of the undergraduate degree, as shown in Figure 4.11, show that expert-like thinking in physics increases at each time the cohort chooses to further specialise in the subject. There is a clear difference between those in secondary school physics classes and those in the same classes that have chosen to study physics at the tertiary level (and those in the first year of undergraduate study). Equally at the finishing point of the degree program, where students can either go on to conduct physics research and further study through a PhD or leave the program we see a marked difference with those embarking on PhD study clearly more expert-like than those in the undergraduate

degree. It is possible that these increases are due to students simply learning ‘how to think like a physicist’ during their studies, however if this were the case it might be expected that a gradual increase in expert-like thinking would be seen over the course of an undergraduate degree. Another possibility, and one that we favour, is that a selection effect is being seen with the most expert-like thinkers choosing to continue studying physics at each transition point.

In order to investigate the hypothesis of a selection effect a small study has been carried out where students in the 4th and 5th years of the degree program were asked after graduation to state their plans post finishing their degree. The students were given four options via an informal email by a member of teaching staff, with students being asked to select the option which mostly closely matched their future plans. The options were as follows:

- Job, real world activity
- Undecided, travel.
- Further study MSc etc
- Further study PhD, here or elsewhere.

It was hoped that enough students would respond to the email that it would be possible to look for any correlation between CLASS score and further study, and thus test the idea of a selection effect. Unfortunately only a small number of students responded to the survey (N=21), which is only 26% of those who filled in the CLASS survey in years 4 and 5, however they were a representative sample of the class both in terms of CLASS score (no significant difference between the responders and the whole class ($p=0.16$)) and in terms of degree classification awarded.

Figure 4.12 shows the results collected in the survey, with the purple line marking the average CLASS score of the whole of years four and five. The figure is suggestive of a selection effect with more of the students with above average CLASS scores going on to study in further education, however the small sample size means that no conclusive statistic can support this hypothesis, with the CLASS score of those students going on to study a masters or PhD shown to be no different to those (who responded to the survey) that are going on to jobs or are undecided ($p=0.99$). Further testing of the selection effect hypothesis will be addressed in sections 4.4 in relation to the fully-longitudinal study.

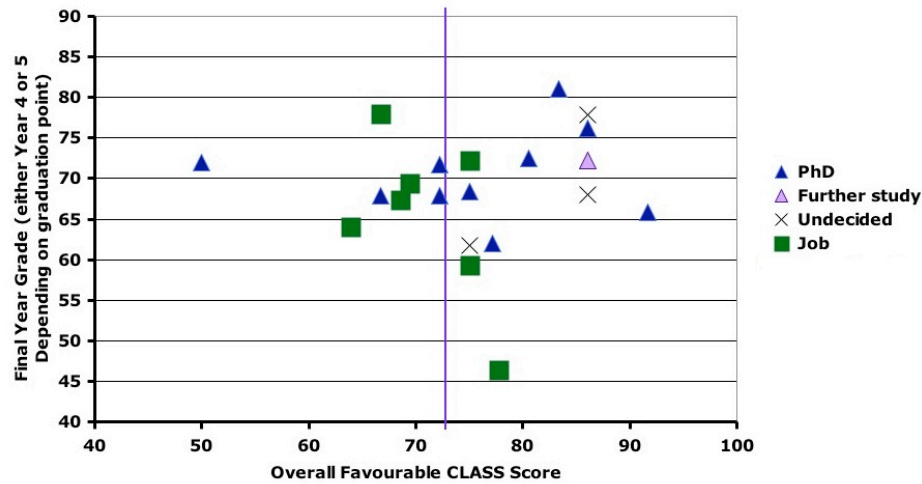


Figure 4.12: Favourable CLASS score for individual fourth and fifth year students compared to career destination. The purple line represent the average CLASS score of the whole fourth and fifth year class

4.2.5 CLASS in schools

Another insight into a possible selection effect is given from data collected as part of a 4th year undergraduate Senior Education Placement (SEP) Project where the CLASS survey was given to students in a local Edinburgh secondary school. The students were in the final three years of secondary education (Years 4, 5 and 6) and were all studying physics. The survey was given to students in class time and on paper, in total 157 surveys were collected and passed the data validation checks. As the data was not collected as part of this thesis the quality of the data has not been assessed to the level presented in the rest of this thesis, with no information available as to the number of surveys which had to be disregarded due to failure of validation checks. The raw data has been reanalysed for use in this thesis and it is the results as analysed by the author of this thesis which are presented here.

The overall favourable and unfavourable CLASS scores for the three years of school students are shown in Figure 4.13. Like that discussed in Section 4.2.2, this is pseudo-longitudinal data with all the data collected at the same point in time.

As in Figure 4.11, a transition is seen with a clear difference between years five and six. In year four the students display 47(2)% favourable expert-like

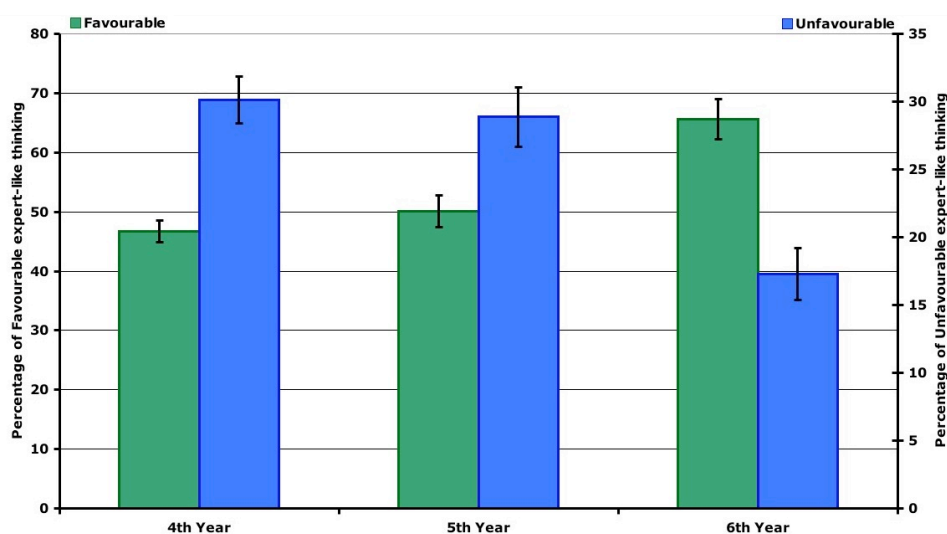


Figure 4.13: Percentage of favourable and unfavourable expert-like thinking for school pupils in Year 4 (N=80), Year 5 (N=59) and Year 6 (N=18) of secondary school. The green bars show the percentage of favourable expert-like thinking and are plotted against the left hand axis while the blue bars show the percentage of unfavourable expert-like thinking and are plotted against the right hand axis. The errors bars represent the standard error on the mean.

thinking, with the fifth year students recording 50(3)%, this difference is not seen to be statistically significant ($p=0.29$). There is however a large difference seen between those students in year five and those in year six, with the sixth year students displaying 66(3)% expert-like thinking, this difference is statistically significant ($p\leq 0.001$). The sixth year of study is the final year of school physics available to students and those in the class will be study Advanced Higher Physics, an option year which many students planning to study physics at university will take although it is not a prerequisite for university physics study. The percentage found for the sixth year students is not statistically different to that included in the pseudo-longitudinal study presented in Section 4.2.2 ($p=0.74$), which was collected with the same age of students at the same point of study.

The difference in the expert-like thinking between the fifth and sixth year students is suggestive that a selection effect is at work with the most expert-like of the students choosing to study physics at the highest school level. It is unfortunate that the surveys were completed anonymously meaning that no follow up work could be carried out as to whether the most expert-like students

do progress to higher levels of physics study, however that was beyond the scope of the Senior Education Placement project from which this data was taken.

4.3 Fully longitudinal study

Until this point, when considering the long term changes in student expert-like thinking, the only approach considered has been that of a pseudo-longitudinal methodology, where all data in the study is collected in the same academic year. Associated with this approach is the necessary assumption that it is valid to assume any year cohort is the same on a year to year basis. The methodology described in this section explores another approach to examining student expert-like thinking over the course of a degree program, that of a fully-longitudinal study. In this method the same students are surveyed as they progress through the degree program, thus eliminating the need to make assumptions about the similarities of the year cohorts.

This study commenced in September 2008, on the first day of undergraduate teaching of the 2008-09 academic year. The same cohort of students have been asked to complete the CLASS survey four times: pre and post first year instruction, post second year instruction and, finally, at the end of third year instruction. The dip in expert-like thinking seen in the first year of study has already been extensively discussed in section 4.1 and accordingly will not be included in this study. As only matched student responses are of interest in this particular study, some survey responses at each point of sampling have been disregarded (since the particular student failed to complete a corresponding survey at another point in time). Response rates can be seen in Table 4.6. The average results for each year group do not differ significantly by only using the fully matched responses as compared to the complete, non-matched data set. In total, 35 students have completed the initial pre-instruction survey as well as the post second and third year surveys while also passing all relevant checks for the quality of the data.

In all cases the survey was completed in paper format in a set-aside segment of class time. The students were asked to complete the survey without talking to other class members, and while the survey could not be completed anonymously (due to the need to track the changes of individual students), students were

	Collected	Retained	% Retained
1st Year	265	35	13%
2nd Year	105	35	33%
3rd Year	58	35	60%

Table 4.6: Raw numbers of student surveys collected at each stage of the study, the final numbers used throughout the study and the percentage of surveys retained for students who completed all three surveys.

asked only to provide their student ID number and, as in previous uses of the instrument, students were always told what the data would be used for and assured that responses would be treated confidentially. No course credit or other incentives were offered to the students to complete the survey.

Some students were reluctant to fill in the survey during the second and third year data collection periods often stating “*I’ve already done that survey*”. On these occasions it was explained to the students why it was necessary to use the same survey each time but no student was compelled to fill in the survey as this might affect the quality of the data.

The final data needed for this study, the attitudes of the 4th year students (when the majority of the students leave the degree program and graduate with a BSc) will not be collected until Spring 2012, which is outside the time constraints of this Doctoral thesis, however the data will be collected, along with destination data for the departing students as part of ongoing work stemming from this initial research.

4.3.1 Overall favourable and unfavourable attitudes

Figure 4.14 shows the overall percentage scores at each of the three sample points, with favourable scores plotted against the left hand axis and unfavourable scores on the right hand axis. The figure shows that the overall scores are remarkably constant with no statistically significant differences in any of the transitions shown. Looking first at the favourable data, the first year data (collected prior to any university teaching) records a student average of 70(2)%, which remains constant in second year and drops slightly to 68(3)% by the end of third year. However, an ANOVA test (for repeated measurements) carried out on all the data shows that there is no statistically significant difference between the years

($p=0.54$).

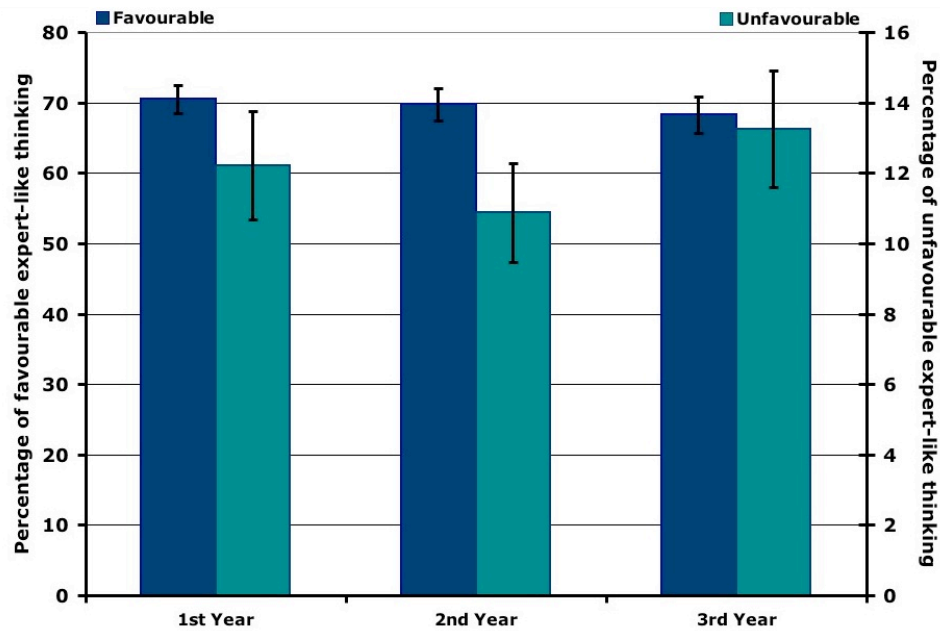


Figure 4.14: Percentage favorable and unfavorable CLASS scores over three consecutive years for the same group of students ($N=35$). The blue bars show the percentage of favourable expert-like thinking and are plotted against the left hand axis while the green bars show the percentage of unfavourable expert-like thinking and are plotted against the right hand axis. The error bars in each case represent the standard error on the mean.

Looking now at the unfavourable scores – the percentage to which the answers given are at odds with the expert opinion – again we see no significant changes. The score recorded at the start of the first year of 12(2)% decreases to 11(1)% by the end of second year and is then seen to increase at the end of third year to 13(2)%. Again using an ANOVA test no statistically significant differences between the years ($p=0.18$) are found.

4.3.2 Category results

Further investigation into the expert-like thinking of the students was carried out by splitting the results for each year group into the 8 scored categories of the CLASS survey. The percentage shifts for favourable scores in each category are displayed in Figure 4.15. In this case the changes have been calculated by looking

at the differences between the percentage scores in the first year and third year.

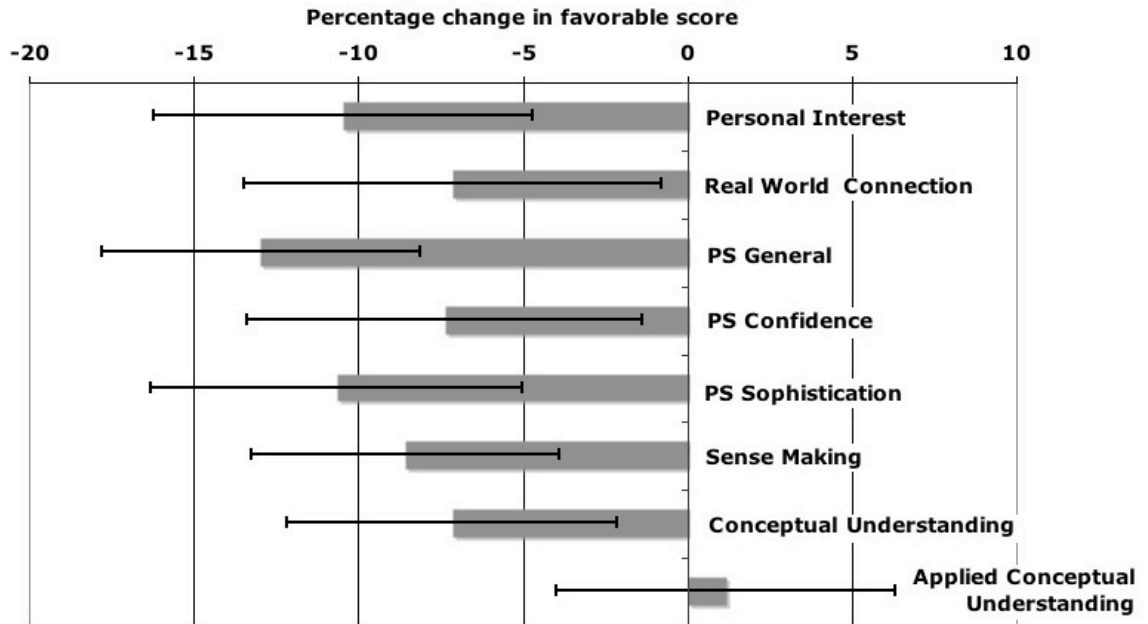


Figure 4.15: Percentage changes in CLASS category favourable scores between 1st and 3rd year.

Statistical tests investigating the differences in category score between all years indicate that 10 of the 16 possible categories (8 categories, which can be scored in both favourable and unfavourable percentages) remain constant between 1st and 3rd year. For the favourable data, significant shifts are seen in three categories. Firstly in ‘*General Problem Solving*’, where students have declined from 80.5(3)% expert-like thinking to 67.5(4)% ($p=0.0033$), also in ‘*Personal Interest*’ where students have changed from 76(4)% to 66(4)% ($p=0.04$) and, finally, in ‘*Problem Solving Sophistication*’ where students have decreased from 69(4)% to 59(4)% ($p=0.015$).

In the case of the unfavourable data three differences are statistically significant: these are in the ‘*Personal Interest*’ category with students changing from 7(2)% to 15(3)% ($p=0.00013$), the ‘*Real World Connection*’ category with a change from 7(2)% to 14(3)% ($p=0.025$) and lastly in the ‘*General Problem Solving Category*’ where students have increased in unfavourable expert-like thinking from 7(2)% to 10(2)% ($p=0.021$).

A previous study which incorporates the CLASS scores of freshmen and

seniors at MIT [105] also finds variation on a category basis, with significant shifts seen in three categories: in favourable scores for *'Real World Connection'* and an increase in unfavourable scores in the *'Sense Making'* category (neither of which we see) and the *'Personal Interest'* category which we also find, with a comparable magnitude of shift.

4.3.3 Completion of study

One of the disadvantages of the fully-longitudinal method of data collection is the large amount of time needed to complete the study. There is no way of increasing the speed at which data can be collected as it is a matter of waiting for the students to pass through the four or five years of an undergraduate degree; a time period which is longer than that set aside for a doctoral research project. Included in this thesis are the results from the first three years of a fully longitudinal study but no further, however the study will be completed by the thesis author as future work arising from this thesis. The data for the students in their fourth year of study will be collected in spring 2012 and a concerted effort will be made to ensure as many students as possible fill in the survey in order to have the largest sample size. Those students that have completed the CLASS survey for all previous three sample points will be the highest priority of data to collect, however all data collected will be useful. Unlike in the fully longitudinal survey, as discussed above, it is intended that as well as looking at the year to year changes in students, comparisons will also be made between students on entering the degree program in year one and those graduating in year four, which will hopefully lead to a larger sample size, as seen in studies conducted at MIT [105]. Looking only at the changes seen between the beginning and end of the degree will counteract the problems arising from attrition from the degree program with fewer students available to complete the survey in each year due to some leaving the degree program in each year.

Alongside the usual CLASS survey questions the students are also going to be asked additional demographic questions looking at their career plans after graduation, with the objective to test the selection effect hypothesis by looking for correlations between those who go on to postgraduate study and CLASS score. In addition it is planned that the CLASS scores from first year will be examined to investigate any correlation with low CLASS score and those who leave the

degree program with out graduating (either due to failure to pass examinations or due to personal choice).

4.4 Validity of CLASS for individuals

In Chapter 2, the original paper that accompanied the publication of the CLASS survey [34] was discussed, with particular emphasis on the design and validation of the instrument. One area of the validation which was considered was that of the validity of the survey for individual students, a topic which was not covered in the original paper but was signposted as an area of future work. Whilst carrying out the research presented in this chapter, the idea of individual validity was raised, with the view of testing the hypothesis that student attitudes are constant and the transitions seen are due to the the less expert-like students leaving the degree program. If the CLASS survey were valid for individuals it would be straight forward to prove that the student attitudes are remaining constant over set periods of time, however, this does not appear to be the case.

Whilst it would undoubtedly be unreasonable to expect that student responses are static over the course of three years, as factors such as individual attitudes, confidence and state of mind can vary from day to day, it does however appear that some individuals in the cohort undergo a large shift in their opinions over time. This is surprising given that there is no overall change in the expert-like thinking of the class between the first and third year of study. Of the 35 students included in the fully longitudinal study 21 (60%) remain within 1 quartile (8 places) of their original starting position, with the remaining 40% moving more than 8 rankings in relation to their peers)

Upon seeing the individual changes in CLASS score (as shown in Figure 4.16), in conjunction with the fact that the issue of individual validity is omitted from the CLASS design and validation paper [34], we suggest that the test is not valid for individuals. Private communication with Wendy Adams [80] ⁴ has revealed that the Physics Education Research group at the University of Colorado also have reason to believe that the survey is not valid for individuals, although they maintain this does not effect the validity of the test for whole classes, as they

⁴Conversation carried out between S.P. Bates, R.K. Galloway and W.K.Adams, PERC Conference, Omaha, Nebraska 2011

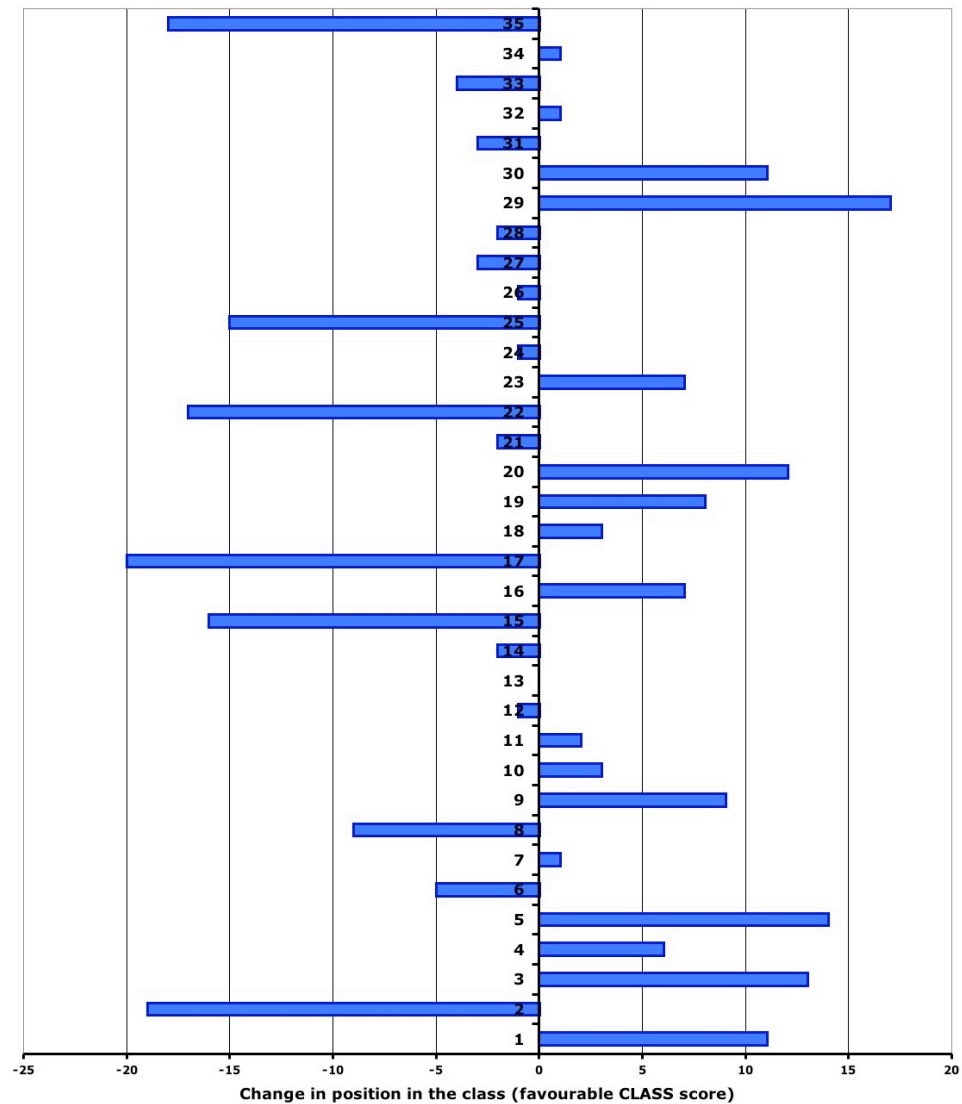


Figure 4.16: Individual student change in class position in CLASS score between Year 1 and Year 3. The individual students are numbered between 1 and 35 in order to maintain the confidentiality of the students, the change in rank position in the class is shown on the x-axis.

believe that individual day-to-day variation is largely offset by that of others meaning that the average score will remain unaffected.

The fact that the survey is likely not to be valid for individuals has potential implications for testing our selection effect hypothesis, as comparing the careers destinations of individual students to CLASS score on exit of the degree may well be meaningless. Instead it may be necessary to look at the average CLASS score

of those students who go on to postgraduate study and try to establish whether the mean of that group is any different to the mean score of their peers who do not choose to do so⁵.

4.5 Comparison of longitudinal methods

In Section 4.2.1, a possible way to validate the pseudo-longitudinal method of data collection were suggested, namely by comparing to data collected for the same year cohort in different academic years. The data acquired in the fully-longitudinal study offers an opportunity to further establish the reliability of this method by comparing the results of the fully longitudinal study to those collected in the pseudo-longitudinal or cross-sectional method. The fully longitudinal study appears to support the validity of the pseudo-longitudinal method as all the data collected is in agreement with the data obtained in the pseudo-longitudinal study. Comparisons between the two sets of data for favourable responses are shown in Figure 4.17.

As opposed to Figure 4.14, where the focus is on the changes in student performance between years, in Figure 4.17 information is gained by looking at bars of the same colour, which compare the data collected through the fully longitudinal method to that obtained in the pseudo-longitudinal study for three years of data collection: no significant variations are seen in any of the years considered.

Tables 4.7 and 4.8 contain the percentage scores (favorable and unfavorable, respectively) for each year group, as collected through both methods, and includes the p values found when comparing the data sets with t -tests. In all cases there are no significant differences between the data collected through either method.

The lack of any differences seen in the data collected using the two different methods appears to suggest that that the pseudo-longitudinal method is valid, at least for data collected during the course of the undergraduate degree. It is not possible to check the validity of the data collected in schools, although as has been discussed in Chapters 2 and 3, the students surveyed are the ‘type’ of

⁵In addition W.K. Adams did not have a conclusive answer as to the point at which the CLASS survey does become valid, asserting that large sample sizes are better for certainty of results, this being the case is of vital importance for testing the selection effect hypothesis that destination data is collected for as many of the 4th year class as possible.

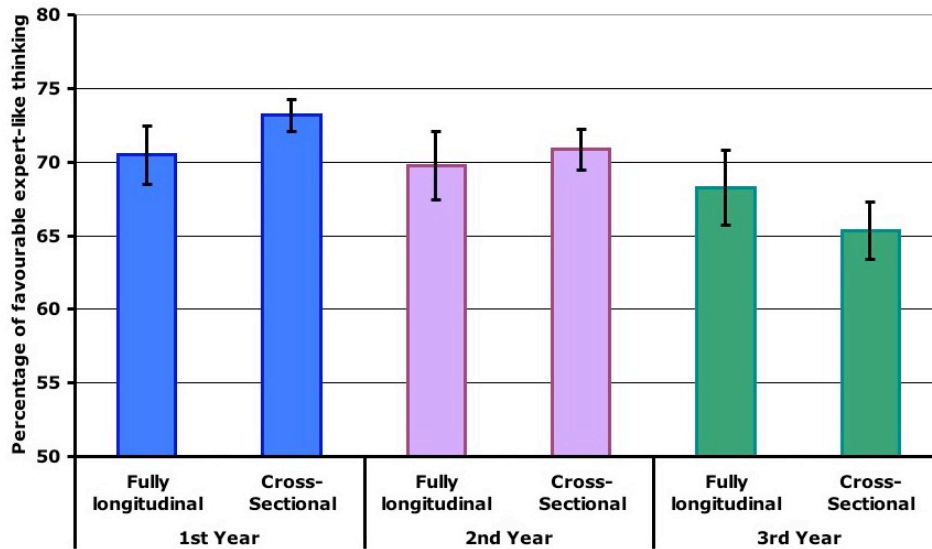


Figure 4.17: Percentage scores of favorable expert-like thinking collected from both a fully longitudinal study and through a cross-sectional study. First year data: N=35 (Fully Longitudinal) and N=127 (Cross-sectional). Second year data: N=35 (Fully Longitudinal) and N=105 (Cross-sectional). Third year data: N=35 (Fully Longitudinal) and N=61 (Cross-sectional).

student who enrolls on a degree at the University of Edinburgh.

4.6 Chapter summary and discussion

In this chapter extensive use has been made of the CLASS survey instrument in order to examine student attitudes and beliefs towards study, before, during and after an undergraduate degree. In all studies presented within the chapter there is one overarching theme, the change in student expert-like thinking over a period of time, with the period of time varying from study to study.

In the first investigation reported, student attitudes are explored during the first year of study of an undergraduate degree. All the students in this study become less expert-like in their thinking after one year of instruction. While this has been widely observed, in previous studies the reasons behind this drop are not yet known. One possible explanation may be related to issues with students' confidence. This drop in confidence is detailed in Perry's model of learning [51] which, as discussed in Chapter 1, details nine stages of intellectual development

Year	Long. (%)	Pseudo (%)	<i>p</i> value
1st Year	70(2)	73(1)	0.24
2nd Year	70(2)	70(1)	0.68
3rd Year	68(3)	65(2)	0.36

Table 4.7: Favorable scores collected through fully longitudinal and cross-sectional methods. The abbreviation ‘Long.’ refers to data collected in the fully longitudinal study and ‘Pseudo.’ refers to data collected in a cross-sectional study. The numbers in brackets represent the standard error on the mean.

Year	Long. (%)	Pseudo (%)	<i>p</i> value
1st Year	12(2)	11(1)	0.54
2nd Year	11(1)	13(1)	0.26
3rd Year	13(2)	17(2)	0.13

Table 4.8: Unfavorable scores collected through fully longitudinal and cross-sectional methods. The abbreviation ‘Long.’ refers to data collected in the fully longitudinal study and ‘Pseudo.’ refers to data collected in a cross-sectional study. The numbers in brackets represent the standard error on the mean.

of university students. These nine stages can be grouped into 4 categories known as ‘dualism’, ‘multiplicity’, ‘relativism’ and ‘commitment’. Most students will begin university in the first stage of intellectual development, dualism, which relates to the idea that there is only one correct answer and it is the job of the learner to gain the right answer from the instructor. Multiplicity is an extension of dualism, where students extend their horizons to incorporate the idea that an answer can be not yet known, as well as right or wrong. The step to the third stage, relativism, where students realise that knowledge is relative and context dependent is seen as a very large step of progress in development. It is at this stage that students can experience a loss of confidence and step back to previous levels.

Interesting differences are seen when looking at different populations of the first year cohort, while all the groups become significantly less expert like in their thinking, the starting points of the groups and the size of the decrease is by no means uniform for all the sub-cohorts.

Looking first at the physics majors and non-majors, the majors are observed to enroll on the physics degree program with significantly higher levels of expert-

like thinking than the non-majors, this is despite the fact that the non-majors must possess the same entry qualifications as the majors to be enrolled into the course. After one year of teaching both groups have decreased in their expert-like thinking, although the decrease is greater for the non-majors. The differences in the two cohorts cannot be attributed to the level of schooling as there is no difference seen in the entry grades of the two groups and in addition, investigation suggests that the grade received in secondary school physics is not strongly correlated with CLASS score.

The non-majors in this study are different to those in previously published, mainly North American studies, due to the fact their background is no different to that of the majors, with only university degree choice distinguishing the two groups. For further comparison the CLASS survey was administered to Biological Sciences students who were studying a brief physics course; the results gained show that the expert-like thinking (with regard to physics) of these biological sciences students was significantly lower than those of both the physics majors and non-majors. This result suggests that the non-majors at the University of Edinburgh represent an intermediate point compared to the two cohorts usually represented in studies, which investigate the differences between majors and non-majors.

Interesting differences are also seen when looking at the differences between male and female physics students, although unlike the physics majors and non-majors, no difference exists prior to university teaching. For the three years of data presented in this thesis in the first two years the female students are shown to be significantly less expert-like in their thinking after the period of instruction. In the final year of data collected the female students do no decrease in their expert-like thinking, the first (and only) time this is seen for any of the cohorts included in this thesis. Further data collection is needed to establish where this is an erroneous data point or the result of changes to the cohort and curriculum.

There are reasons to believe that the students in the 2010-11 physics class are different to their predecessors, which may explain the change seen in the pattern of expert-like thinking for the female students. Firstly the intake of the class is different, with a smaller overall class and, for the first time, there are more non-majors in the class than majors. There have also been changes in the recruitment method employed by the university, with a full selection process being in place,

rather than the active recruitment of students, which may have changed the ‘type’ of students in the class. Changes have been employed in the style of teaching used in the physics class with more student engagement activities such as the use of PeerWise [77]. In addition there is a higher percentage of female students in the class than in previous years, perhaps showing an example of the “critical mass” concept where students in minority groups perform better when there are a larger number of the same group in the class [108],[109]. It is impossible to do anything other than speculate as to what effect these differences may have had on the attitudes of the female students, but they are factors that should be considered before dismissing the 2010-11 CLASS data as erroneous.

The second and third studies reported in this chapter look at the changes in attitudes over a much longer period of time, before, during and beyond an undergraduate degree. Two different methods have been used to explore this area; the first using a pseudo-longitudinal or cross-sectional approach and the second using a fully longitudinal method. The advantages and disadvantages of both of these methods have been extensively discussed in this chapter, with both methods having advantages and disadvantages. Preliminary results from the fully longitudinal study appear to suggest that a pseudo-longitudinal method is valid, perhaps eliminating the need to track the same students as they progress through the degree program.

Results from the pseudo-longitudinal study suggest that student expert-like thinking is essentially static during an undergraduate degree, with transition stages at the entry and exit points of the degree program. A dip in the expert-like thinking may exist in the third year of the undergraduate degree program, with students interviewed attributing the dip to a drop in confidence, as also hypothesised to explain the decrease seen in first year. Taking into account the flat profile of expert-like thinking during the degree with steps up at either end, a selection effect hypothesis is suggested with the most expert-like students staying on to more physics study at each transition point. Further evidence of this selection effect is seen in data collected in a Scottish secondary school where the students become more expert-like in their thinking with each optional year of physics they study. The flat nature of the undergraduate expert-like thinking, (combined with data suggesting that secondary school grades are not a strong predictor of CLASS performance) suggest that it is unlikely that students grow

gradually more expert-like in their thinking with increased content knowledge.

The static nature of student expert-like thinking is also seen in the fully-longitudinal study, tracking student attitudes and beliefs towards physics during an undergraduate degree. Due to the time constraints of the PhD program only the first three years of this study are included in this thesis, however the results collected so far show that student expert-like thinking is not changing during the degree program.

Whilst completing the work presented in this chapter, questions have been raised as to the limitations of the CLASS instrument, with evidence appearing to suggest that the instrument is not valid for individuals or small number samples, a view which has been confirmed by one of the survey authors [80].

If the CLASS instrument is not valid for individuals it will make it more challenging to conclusively prove the selection effect hypothesis, however future data collection and analysis are scheduled for Spring 2012 to collect the final data for the fully longitudinal study with the necessary steps taken to collect enough data to be able to test the hypothesis. If the idea of a selection effect does prove to be correct, it raises interesting questions for future work, including addressing the issue of when these attitudes towards science study are formed and how we can influence student epistemologies if they are already rigidly set on entering tertiary education.

Chapter 5

Student Laboratory Work

The focus of the research in this chapter is on the experimental or laboratory aspects of an undergraduate physics degree. Laboratory work is a key component of any physics degree studied in the UK with all students, regardless of the ‘type’ of physics degree studied, required by Institute of Physics guidelines to undertake at least some experimental work [55]. It is one of the areas of a degree that students tend to have the strongest feelings about, both positive and negative, when starting university with opinions often founded in experience of experimental work at school. These views tend to be varied due to the fact that the experience of lab work in secondary education can be extremely different from school to school, due to different interpretations of the syllabus, teacher preference of areas to teach, as well as financial and timetable constraints of individual schools [59].

The origins of some of the work in this chapter lie in the study of the school-university transition, as discussed in Chapter 3. The topic of student attitudes towards laboratory work on either side of the tertiary education boundary was initially addressed in the first year of research for this project, when this body of work was intended as a masters by research project. In the following sections the results of attitude surveys given to students asking about views and perceptions of lab work at two different stages of study are presented. The first study was given to students at the beginning of their first term of university study asking them to retrospectively consider their experiences of experimental work during their time in secondary school, while the second survey was given at the end of the first year of university study asking students to reflect on their views of

university experimental work. The implementation of these studies, alongside the results gained from each and discussion of the differences between the two groups will be discussed in this chapter.

Another focus of the work within this chapter is that of the skills required by students to carry out experimental work, more specifically, ability to correctly handle data collected during laboratory classes. In keeping with the themes seen throughout this thesis, where changes are measured in students during the course of an undergraduate degree, it is the evolution of these skills which is the focus of this chapter, rather than attempting to teach such skills. In order to measure the data handling skills of students a diagnostic test has been developed and deployed with students enrolled in undergraduate degrees across the UK and Ireland. Details of the design and validation of the test, alongside results found from administering this test, are discussed in this chapter, followed by discussion as to the implications of the results found.

5.1 Attitudes to experimental work

All of the work in this section was carried out in conjunction with the Universities of Glasgow and St. Andrews, through a project led by the University of Glasgow. The aim of the research was firstly to retrospectively investigate the attitudes of students as to their perceptions towards school practical work and to probe their expectations of university experimental work on starting their degree. A second survey was then used to look at the attitudes of students towards labs after the first year of university teaching. The first half of this work has been reported in a study focusing on experimental work in schools, with an emphasis on differences in attitudes between students studying Highers, Advanced Highers and A-Levels [110].

In this thesis the data collected has been used for an alternate purpose; to look at the changes in attitudes towards practical work, between school and university. The work presented here uses only the data collected with students at the University of Edinburgh.

The work in these sections is not a true example of pre and post instruction testing, despite the apparent similarities between this method and those seen in Chapters 3 and 4. Unlike in the case of the CLASS survey it is not the same

survey that has been used in each instance, as some of the questions would not be appropriate given the context. It is also different to the situation seen in Chapter 3 where students are given selected questions from the Lawson test, as again the same test was used to compare groups, although in that situation the same students were not surveyed pre and post instruction.

Here the same cohort of students were surveyed on arrival at university and again post first year teaching, however the surveys were completed anonymously so it is only possible to track the trends in the attitudes of the cohort overall rather than the individual students. Six of the nine questions on the survey, are different between the two surveys and the three which are approximately the same are subtly different as the context has changed to ask about either attitudes towards labs in school or those in university. In the following sections details of the content and implementation of both the surveys will be discussed, firstly as the data has been presented by the collaboration, then followed by an alternate use of the raw data to look at the changes in attitudes over time. Copies of both surveys can be found in Appendix H and Appendix I.

5.1.1 The attitudes of schools students towards experimental work

A survey consisting of nine questions (five of which address opinions and perceptions and four which aim to establish demographic data) was given to 500 students (N for Edinburgh=243) in the November of semester one of the first year of the physics degree. The survey was designed by staff at the University of Glasgow, a copy of the survey can be found in Appendix H. The survey was deployed in three universities (Edinburgh, Glasgow and St. Andrews) with all universities surveying their students at the same time.

In order to allow additional information to be gained, after initial analysis of the data had been completed, focus groups were carried out with students in each of the universities, asking students to elaborate on the themes seen.

The findings of the project appear to suggest an extremely positive picture with students in school generally enjoying practical work. In general the students felt that lab work enhanced their understanding of physics and they were confident carrying out practical work on their own. When the data was split to look at the differences in attitudes between students with different school leaving

qualification (Highers, Advanced Highers and A Levels) differences were detected. For example, the project portion of the Advanced Higher meant that those students who had taken the qualification were more confident designing their own experiment. Conversely those students that had only studied Higher physics tended towards a preference for written instructions for practical work, most likely related to the fact that this is all they would have previously experienced. The final conclusions of the project state that with such a positive overall picture of school practical work, universities must work hard to ensure this enthusiasm continues.

5.1.2 The attitudes of first year university students towards experimental work

An adapted version of the survey that investigated attitudes of school practical work was given to students in the first year of undergraduate study at the end of the second semester, this time asking students to reflect on their experiences of undergraduate experimental work. On this occasion the number of students surveyed was increased by including two more universities from around the UK (Warwick and Hertfordshire), as well as the Universities of Edinburgh, Glasgow and St. Andrews.

As mentioned previously, the content of three of the questions was the same as those of the school practical work survey, although the context had been changed to look at university practical work rather than that in schools. Two of the questions were totally different to those on the first survey, generally focusing more specifically on individual aspects of the laboratory course, tailored to each institution¹. A full copy of the version of the survey given to students at the University of Edinburgh can be found in Appendix I.

The data collected using the survey for the University of Edinburgh has been made available for this thesis (N=136) and in the following sections the results will be discussed and compared to those collected in the school survey.

¹Acting in part as a course satisfaction survey.

5.1.3 Comparison of attitudes

Only three of the five attitude questions of the test can be directly compared between the two surveys, as the other questions differ between surveys and do not lend themselves to direct comparison. Of the three that can be compared, two focus on student experiences of laboratory work and the third asks students to consider why they think practical work is an important part of an undergraduate degree, therefore it is only the final question which actually looks at student opinion without dependence on context.

Question 1 “*What are your opinions of your practical work in physics in school / first year labs*”

Here is a way to describe a racing car.

quick <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> slow important <input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> unimportant safe <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> danger	The position of the cross in the box between the word pairs show that you consider it as very quick, slightly more important than unimportant and quite dangerous.
---	--

Use the same method to answer question 1.

Q1. What are your opinions about your present university laboratory experiences in physics?
(Cross ONE box on each line)

Useful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Useless
Not helpful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Helpful
Understandable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Not understandable
Satisfying	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Not satisfying
Boring	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Interesting
Well organised	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Not well organised
The best part of physics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The worst part of physics
Not enjoyable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Enjoyable

Figure 5.1: Question 1 from the Attitudes to Practical Work at University Survey. See Appendix I for the full survey and Reference [110] for further details

Figure 5.1 is a copy of Question 1 of the survey focusing on school practical work, included to illustrate the way the question was presented to the students. A six point scale was used, where it was not possible to select a neutral option, meaning that the student must select at least a slight preference to one opinion.

The answer profiles of the two cohorts were compared using a chi-squared test. Of the eight topics addressed in Question 1, three showed no statistically significant difference between the answers in relation to school laboratories

and those at university; the usefulness of the laboratories, the helpfulness of the laboratories and how satisfying the students found the laboratories. The remaining five questions showed a significant difference in the answers between the two surveys and they are illustrated in the figures below. In all cases the six point category on which the questions were scored has been renamed “strongly agree” through to “strongly disagree” despite the fact this is not how they were presented to the students, it is not believed any information is lost by renaming the categories for presentation in this thesis.

Understandable

Figure 5.2 shows the differing answer profiles for each cohort. While both sets of survey data show that the vast majority of students (91% for school practical work and 85% for university) find the laboratories understandable, there are more students who found school practical very understandable than who said the same for university. A Chi squared distribution reveals that there is a significant difference between the two ($\chi^2=8.51$, $N=379$, $df=2$, $p=0.037$).

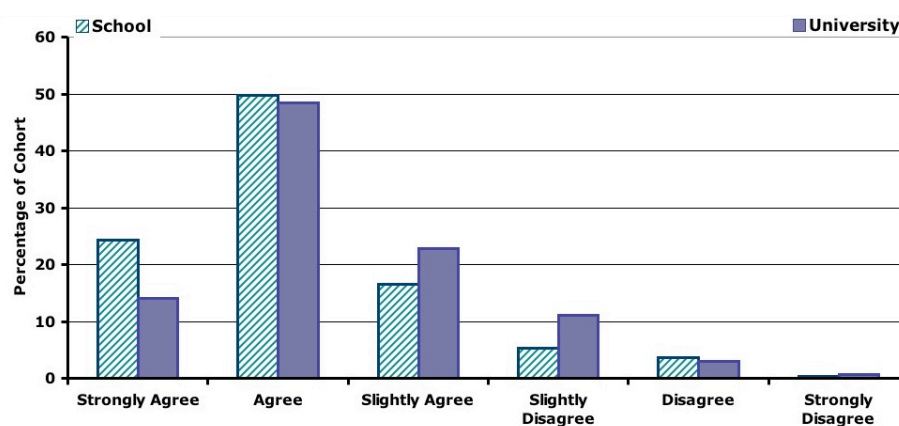


Figure 5.2: Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Answers to question asking where they would describe the practical work encountered as **Understandable**

Interesting

As was the case in Figure 5.2 it is also the case in Figure 5.3 that the majority of students find both the school and university practical work interesting, however with 75% of surveyed students finding school experimental work interesting and 57% saying the same for university, it can be seen that there is a distinct trend to find practical work less interesting at university. A Chi squared test confirms a difference between the two cohorts ($\chi^2=13.52$, $N=379$, $df =3$, $p=0.0036$).

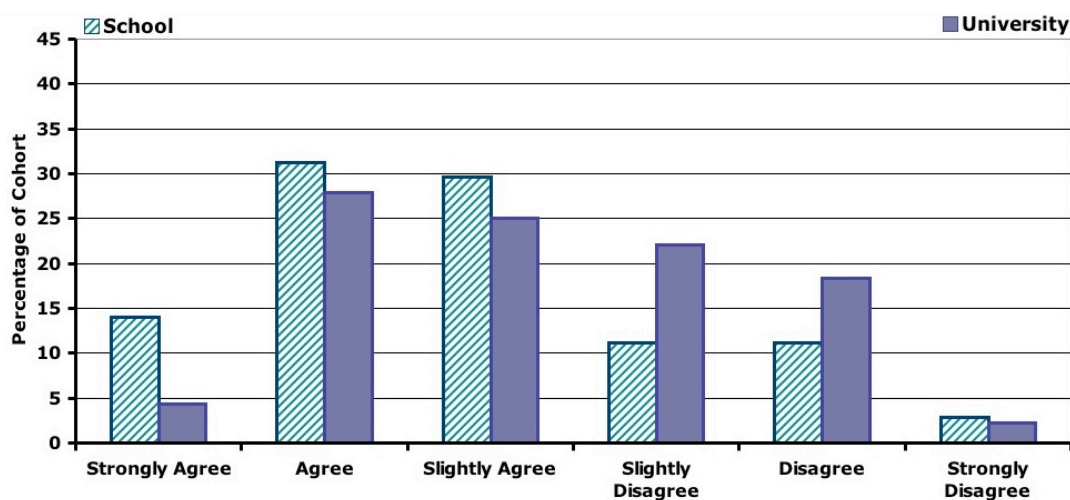


Figure 5.3: Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Answers to question asking where they would describe the practical work encountered as **Interesting**

Organised

Unlike in the previous two cases, it is now no longer the case that the two survey cohorts are in agreement in their responses. When asked if practical work sessions were well organised 62% of the students stated that those in school were, where as only 42% felt this was the case for university practical work. Again a Chi squared test shows the two distributions (as shown in Figure 5.4) are different ($\chi^2=25.15$, $N=379$, $df =3$, $p<0.001$).

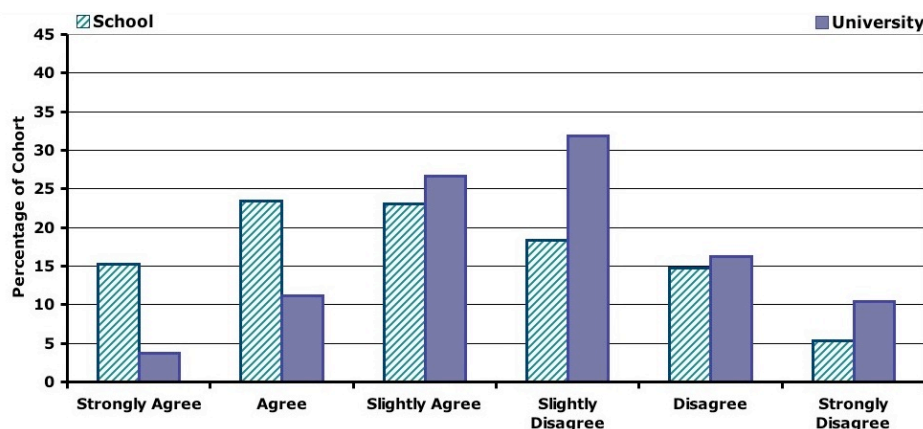


Figure 5.4: Answer profile for students answer two surveys asking them to consider their experiences of school and university practical work. Answers to question asking where they would describe the practical work encountered as **Organised**

Best Part of Physics

Here, as for the previous question, the students do not agree in an overall answer, as illustrated by Figure 5.5. 59% of students thought that school labs were the best part of physics whereas only 37% felt this about university practical work. The two distributions are different to each other with ($\chi^2=19.36$, $N=379$, $df=3$, $p<0.001$).

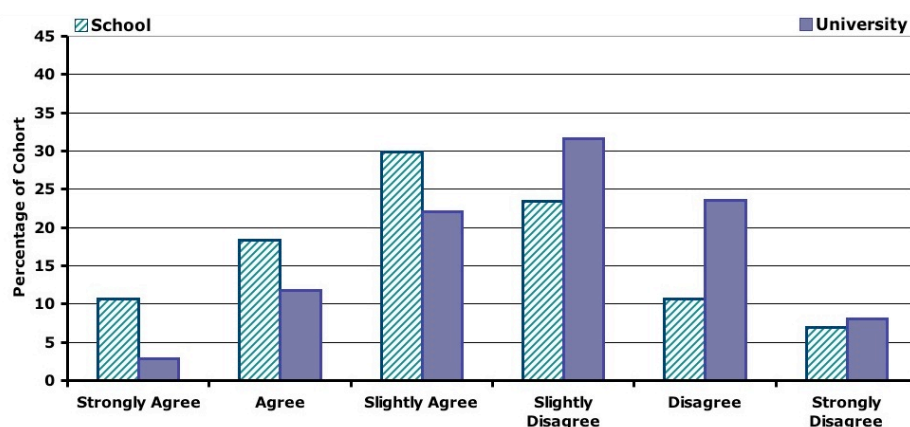


Figure 5.5: Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Answers to question asking where they would describe the practical work encountered as **the best part of physics**

Enjoyable

Here the most striking difference is observed between the two surveys, with the results seen very much at opposite ends of the spectrum to each other. School physics was deemed extremely enjoyable with 76% of the survey cohort stating that they enjoy practical work, in contrast only 18% of students said they found undergraduate lab work enjoyable. Unsurprisingly these distributions are statistically different to each other ($\chi^2=124.97$, $N=379$, $df=3$, $p<0.001$).

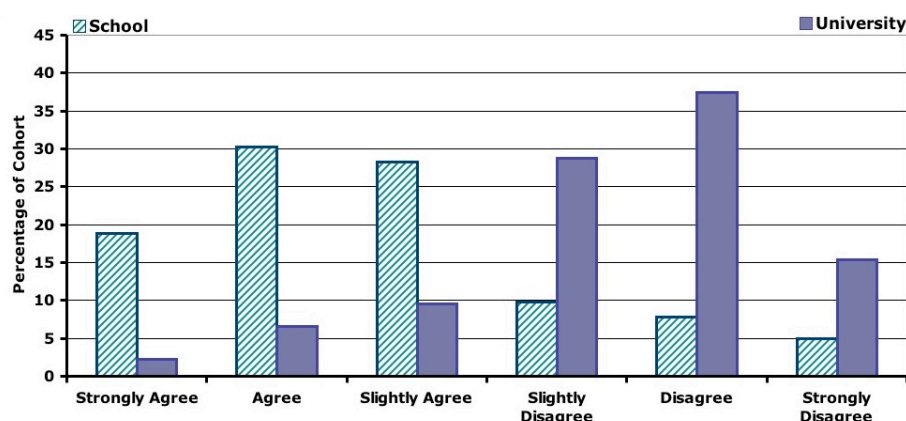


Figure 5.6: Answer profile for students answer two surveys asking them to consider their experiences of school and university practical work. Answers to question asking where they would describe the practical work encountered as **enjoyable**

A comparison of the two survey cohorts reveals that, in all five questions that display a significant difference between the two groups, the school experience of practical work is found to be more positive. Both cohorts would describe practical work as understandable and interesting but only the school cohort would describe their laboratory experiences as well organised, enjoyable and the best part of physics. The ‘enjoyable’ category sees the most marked difference between the groups with a complete swing from one end of the spectrum to the other, with 76% of school students describing practical as enjoyable compared to only 18% for university.

Question 2 “Think about your experiences in practical physics work (cross the box which best describes your opinion)”

Question 2 of the survey asked students to consider their experiences of practical work and answer eight statements on a five-point Likert scale from strongly agree to strongly disagree; a copy of the question is shown in Figure 5.7. Of the eight statements given to the students, five showed a significant difference between the experiences in school practical work and those at university. Those five statements will be discussed below, with histograms used to show the differences in the two distributions.

Q2. Think about your experiences in laboratory work in physics.
(Cross the box which best reflects your opinion).

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
(a) I prefer to have written instructions for experiments.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(b) Laboratory work helps my understanding of physics topics.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(c) Discussions in the laboratory enhance my understanding of the subject.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(d) I had few opportunities to plan my experiments during the laboratory work.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(e) I felt confident in carrying out the experiments in physics.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(f) The experimental procedure was clearly explained in the instructions given.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(g) I was so confused in the laboratory that I ended up following the instructions without understanding what I was doing.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(h) There was good linkage between experiments and the relevant theory.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 5.7: Question 2 from the Attitudes to Practical Work at University Survey. See Appendix I for the full survey and Reference [110] for further details

I prefer to have written instructions for experiments

Both of the surveys produced results indicating that students like to have written instructions for experiments, with 87% of students saying so after the first year of university study and 73% when considering school labs. The distributions (as illustrated in Figure 5.8) appear not too dissimilar to each other, however a chi squared test reveals they are statistically different ($\chi^2=11.2$, $N=379$, $df =1$, $p<0.001$), due to school students seeming to have less strong views on the subject, shown by more frequently selecting the neutral option.

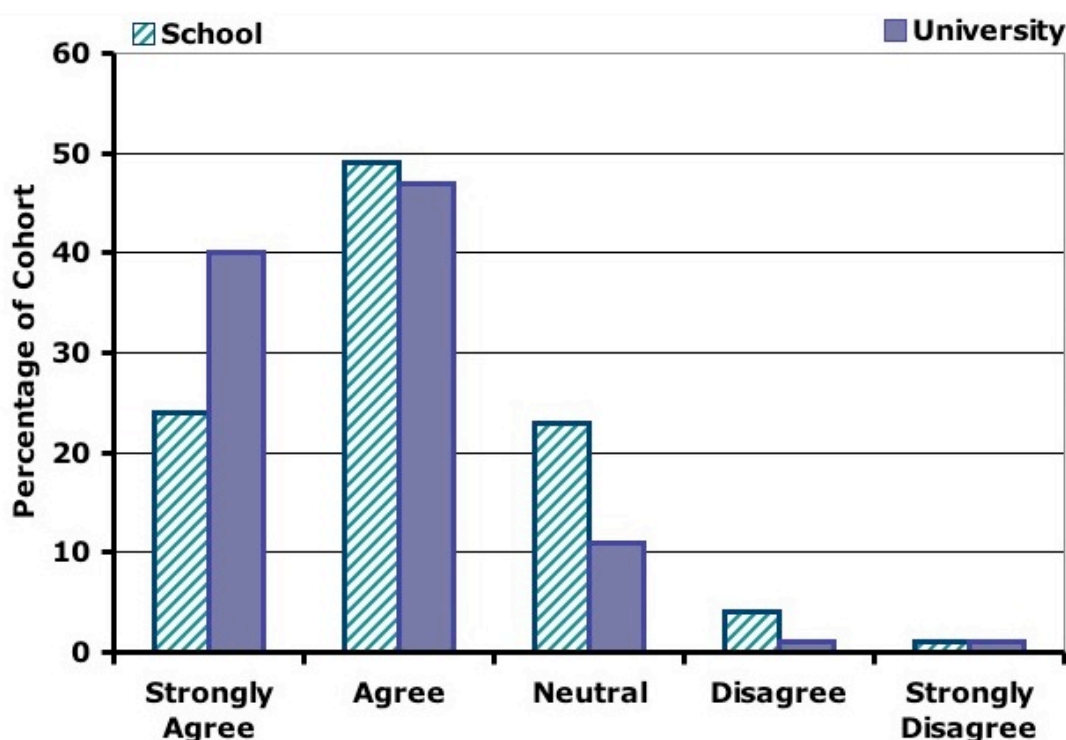


Figure 5.8: Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Responses to the statement “*I prefer to have written instructions for experiments*”

Discussions in the lab enhance my understanding of the subject

As with the previous statement a significant difference exists between the two survey cohorts ($\chi^2=11.2$, $N=379$, $df =2$, $p<0.001$), despite the fact they do not appear very different from each other (as seen in Figure 5.9). The school practical

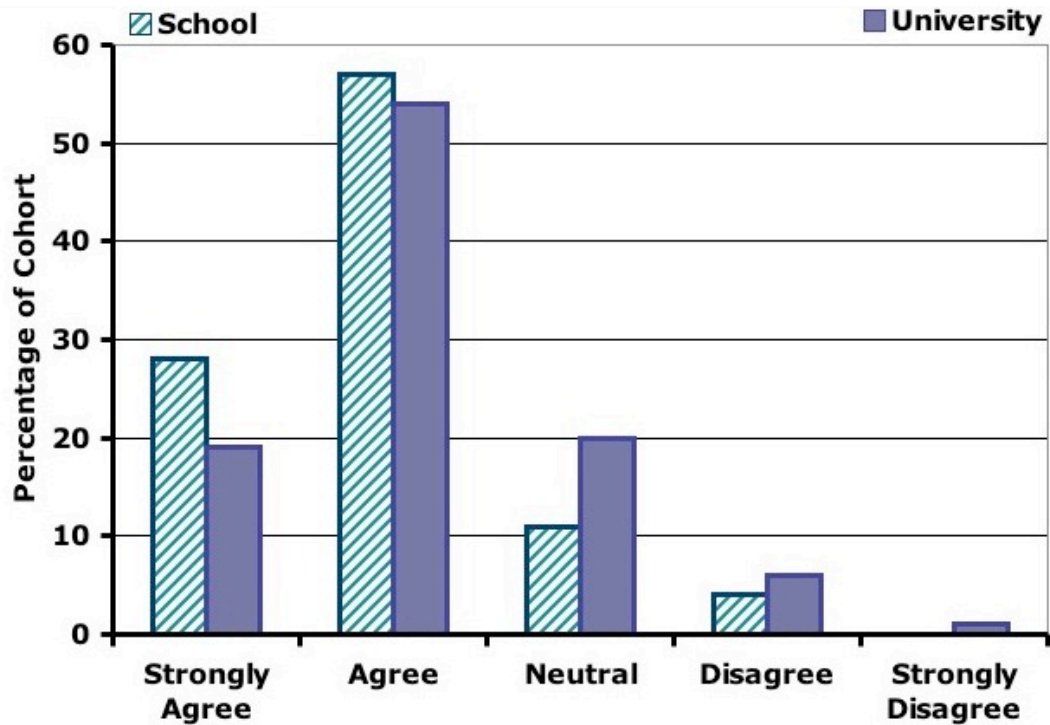


Figure 5.9: Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Responses to the statement *“Discussions in the lab enhance my understanding of the subject.”*

work survey reveals that 85% of students found discussion of practical work enhanced understanding, compared to 73% on the university survey suggesting that university laboratory classes are less likely to be enhanced by discussion, according to the survey results, than school practical work.

I had few opportunities to plan my experiments during the lab work

Figure 5.10 shows the distributions of the two cohorts when considering opportunities to plan practical work. The two distributions are similar with the difference between them lying in the fact university students were more likely to choose the neutral option for this statement. The distributions are significantly different to each other ($\chi^2=6.7$, $N=379$, $df=2$, $p=0.034$).

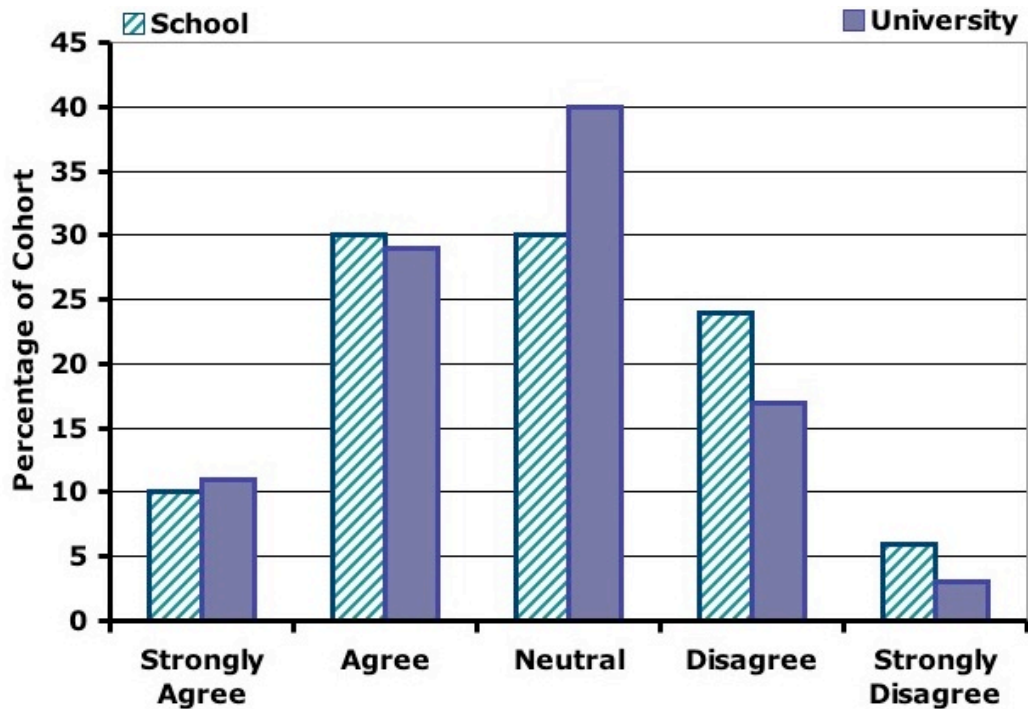


Figure 5.10: Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Responses to the statement *“I had few opportunities to plan my experiments during the lab work”*

The experimental procedure was clearly explained in the instructions given

The distributions for the two cohorts, on the subject of experimental procedure, are shown in Figure 5.11. In terms of agreement with the statement the two distributions are effectively the same with 51% for university labs and 50% for schools. The two distributions are, however, statistically different to each other ($\chi^2=7.4$, $N=379$, $df=2$, $p=0.025$) due to the fact that school students are more likely to be neutral towards the statement if they do not agree whereas university students tend towards disagreeing.

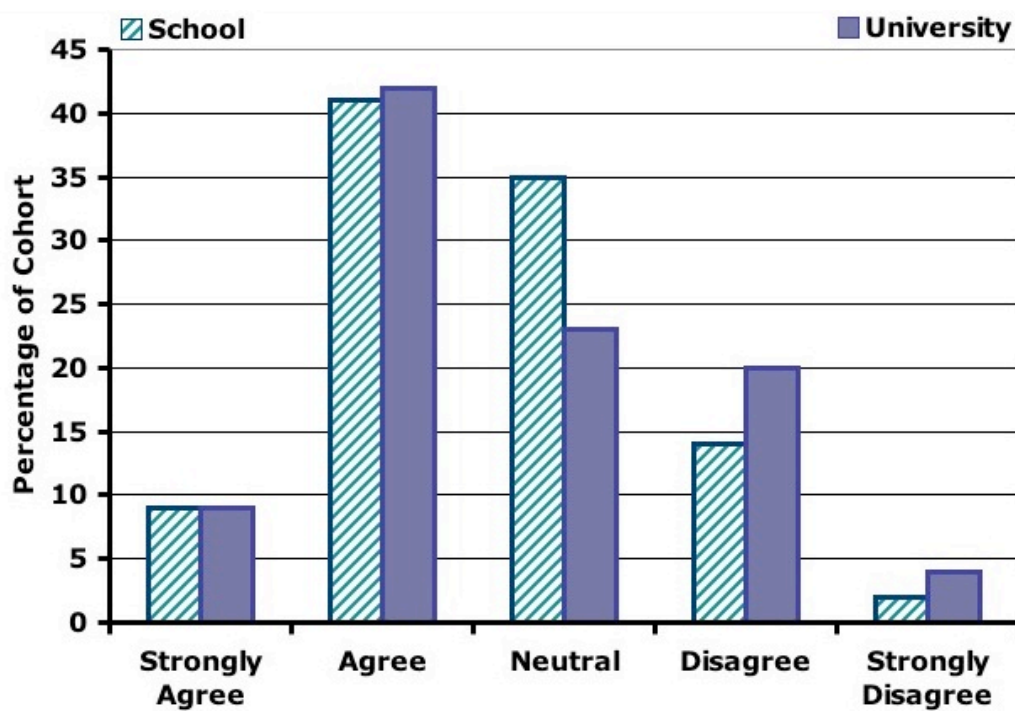


Figure 5.11: Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Responses to the statement *“The experimental procedure was clearly explained in the instructions given”*

There was good linkage between experiments and the relevant theory

As with the previous statement, the difference between the two cohorts when asked to consider the link between theory and practical work lies in student likelihood to select the neutral option. Figure 5.12 shows that, in terms of agreement with the statement the two distributions are very similar with 62% for university practical work and 68% for schools. The two distributions are, however, statistically different to each other ($\chi^2=7.3$, $N=379$, $df=2$, $p=0.026$). This difference may be due to the fact that after the agree options the school students tend to choose the neutral option whereas the university responses are more evenly split between neutral and disagree.

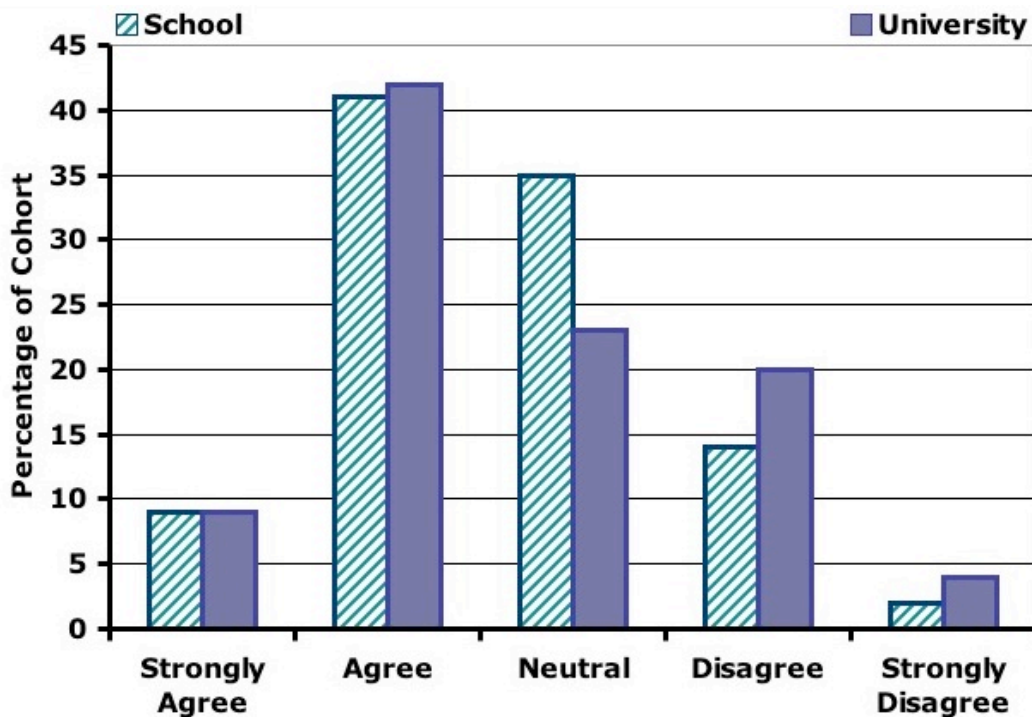


Figure 5.12: Answer profile for students answering two surveys asking them to consider their experiences of school and university practical work. Responses to the statement “*There was good linkage between experiments and the relevant theory*”

For the second question there were again five statements for which there was a statistically significant difference between the two cohorts, however few of these differences reveal any notable differences between the two in terms of

overall attitudes. Both university and school students prefer to have written instructions for experiments, with the number of university students stating a preference higher than that of school students. In addition, both groups also feel that discussion in the laboratory help their understanding. This is more the case for school students, however it is not clear if this is due to the fact that university students were less likely to have discussion about the work or whether it was the case that they found discussions less helpful. Both cohorts were in approximately 50% agreement that the experimental procedure was clearly explained, however the difference between the two groups was that those in school were more likely to select neutral for the statement (if not positive), while the university cohort had a higher frequency of students disagreeing with the statement. On the subject of opportunity to plan their own experiments the profiles were remarkably similar, however, one difference that was seen was that students in school were more likely to be neutral towards the statement where as those in university were more inclined to disagree. Finally, both cohorts felt that there was good linkage between experiments and theory, although, once again those in schools were more likely to agree with this statement.

Question 3 *“Here are several reasons why experimental work is an integral part of a physics course. Pick the three you consider to be integral and rank them in descending order of importance.”*

A copy of question 3 from the survey is shown in Figure 5.13. Eight statements were given to the students and they were asked to select and rank the three they thought were the most important.

Q3. Here are several reasons why laboratory work is an integral part of a physics course.

A. Physics is a practical subject	E. New discoveries are made by means of
B. Experiments illustrate theory for me	F. Experimental skills can be gained in the laboratory
C. Laboratory work allows me to test out ideas	G. Experimental work allows me to think about Physics
D. Experiments assist me to planning and organise	H. Experimental work makes Physics more enjoyable for me

Pick the THREE which you consider to be integral and rank them in descending order of importance in the boxes below.

Figure 5.13: Question 3 from the Attitudes to Practical Work at University Survey. See Appendix I for the full survey and Reference [110] for further details.

The ranked statements for the two survey cohorts in order of popularity were as follows:

School Cohort

- B, G, E, C, A, H, F, D

University Cohort

- B, E, F, G, A, C, H, D

These rankings were created by summing the percentage of students who selected each options as their first choice, with those who selected it as their second and those who selected it as their third, so that each ranking was accorded equal weight, as a result of this the total percentages over all questions will necessarily sum to 300% rather than 100%. This method was used when the results of the surveys were presented for publication [110], however in this thesis these percentages have been divided by three to give a clearer view for the reader, meaning that the total over all questions will sum to one hundred. The individual placings for each statement were also retained before they were combined and will be discussed along with selected results below.

Both cohorts selected option B “*Experiments illustrate theory for me*” as the most important aspect of practical physics with 22% of school students and 21% of university students selecting the statement as in their top three most important. Only the school pupils selected this option as their most important if only the first place rankings are considered (with 35% of the class doing so), with the university students selecting option E “*New discoveries are made by means of experiments*” as their most important option (with 26% selecting this option).

As well as being the option most often placed in first place by the university students, option E “*New discoveries are made by means of experiments*” was ranked highly overall for both cohorts, coming in second place for the university students (17% of the total cohort) and third place for school students (16% of the total rankings). It was also the second most highly placed statement for school students, if only the first place rankings were considered.

At the other end of the results both cohorts were in agreement as to the statement they considered the least important option D “*Experiments assist me to plan and organise*”, with 3% selecting this option from the school survey as well as 3% from the university survey. Similarly statement H “*Experimental work*

makes physics more enjoyable for me” received few votes, coming in 6th place for school students (with 10% of the total votes) and seventh place for university students with 6% of the vote.

One interesting difference between the two groups is the placing of option F *“Experimental skills can be gained in the laboratory”* which was placed as the third most important statement for the university cohort (with 17% of the votes) but was found in second to last place for the school students with 9% of the votes.

It would appear that the views of the two cohorts as to why practical work is important are largely similar with two of the eight statements appearing in the top three statements for both cohorts. Students feel the illustration of theory through practical work is the most important and that the progress of new discoveries through experimental work is also a key feature. They are also in agreement that the planning and organisational help gained through practical work is not a major contributing factor as to why experimental work is carried out, neither is the enjoyment of the work. One key recognised difference is the different weight the two cohorts place on the importance of experimental skills, a topic which will be discussed in the next section, with reference to the data handling diagnostic.

5.1.4 Discussion of results

The surveying of two different cohorts on the subject of practical work has revealed differences between the two groups. As a very broad summary it could be said that the overall experience of both cohorts tends to be positive towards practical work, although in all cases where differences do exist the school experience is seen to be more positive.

Both cohorts appear to be largely positive about their experiences of practical work, although the results from question one clearly show that those taking first year university experimental work do not find the process enjoyable. It is interesting to note that the results suggest both groups are confident in their ability, with 85% of university students and 91% of school students describing their practical work as understandable, a view that is not fully substantiated by the results seen when using a data handling diagnostic, as discussed in the next section.

Both cohorts feel that the primary role of practical work is to illustrate theory, however in both there is an emerging view point that the bigger picture is also

important, that practical work is important not just for building skills, but also for science as a whole as it leads to making new discoveries.

As stated in the introduction to this work, this cannot be considered an example of a true pre and post-instruction survey as the surveys are not identical and also, far more crucially the surveys are not measuring a baseline before and after teaching. The first survey asks students to retrospectively consider their experiences of school work while the second survey asks students to evaluate the experimental classes that they have just finished. The context here is crucial as the questions, with the exception being question 3, do not ask students to consider practical work as a subject but rather their individual experiences of the work. Whether practical work is rated as well organised or interesting will depend very much on the individual, but also the environment in which they have had their experiences of practical work. As a result, such questions may be of interest to the instructors in charge of running the practical courses but can tell us little about the changes in student attitudes towards practical work over their first year of study. In addition the retrospective element of the school survey means that students will be looking back to a year or more since they were involved in school practical work, a time period which may mean students are more vague about some of the organisational details of their school practical work, compared to answering the university survey in the last week of laboratory tuition, when it will be very much fresh in their minds.

Question three is the only question in the survey which is not context dependent and therefore offers an insight into how student attitudes towards practical work have changed over the course of their first experience of undergraduate laboratory teaching. While there is no change in the top and bottom ranked answers as to why practical work is an important part of a physics degree, interesting shifts in thinking are observed. Both cohorts agree that practical work is important for illustrating theory and making new scientific discoveries, however the move from seventh place to third place for the almost circular argument of “Experimental skills can be gained in the laboratory” shows that university students are beginning to think about the importance of the skills learned in the laboratory, as well as the results gained from experiment. The nature of the idea that practical work is important, because it makes you better at practical work, is perhaps a difficult one to comprehend, however perhaps this is an attempt to

convey a larger idea that because practical work can lead to new discoveries and is a large part of what it means to be a scientist, it is important to work towards gaining the skills set needed to be able to carry out practical work.

5.2 The Data Handling Diagnostic

In the previous section, the idea of a set of skills which are needed to carry out experimental work was briefly discussed. The discussion focused on an apparent student consciousness that these skills were needed in order to achieve success and that the best place to gain them was in the undergraduate laboratory. There are many skills that are required in order to be a successful experimental physicist, with the collection and analysis of data in the laboratory just one of them. Once raw data has been collected it can be transformed into quantitative outcomes through analysis and judgements as to the quality of the data and the suitability of the experimental method used. Investigation into the teaching of this topic was carried out by the Physics Education Research group at the University of Edinburgh² which revealed that these skills are rarely explicitly taught in undergraduate degrees, especially after the first year of instruction, however the absence of teaching of these skills does not mean that students are not expected to have acquired them, with some instructors stating that students are expected to pick up the skills intuitively as they work in the laboratory, again perhaps providing an example of the “hidden curriculum” [103].

As has been discussed in previous chapters, the use of diagnostic instruments to measure student ability over time is a widely used method in Physics Education Research, with diagnostics available to test many areas of an undergraduate degree such as Newtonian Mechanics [66], Electricity and Magnetism [67] and attitudes towards study [34] to name just a few. There was not, at the time of starting this project, a test which measured data handling skills, nor could one be found in related disciplines such as Chemistry or Engineering (such a tool has been developed in Canada at the same time as the development of the University of Edinburgh assessment tool, see Reference [111] for further details). The decision was made that in order to establish a quantitative measure of student data handling skills it was necessary to design such a test. This was carried out in the Physics Education Research group at the University of Edinburgh and in the following sections the design process will be discussed, followed by the results seen when trialled with 1208 students from academic institutions around the UK and Ireland. This has been written up for publication [112] and where the author of this thesis was not directly involved in the work the information included is

²As part of the initial work which lead to the creation of the Data Diagnostic Test

gained from the paper supplemented by interviews with the test designers³. In keeping with the themes of this thesis, the aspect of the data handling diagnostic that will be considered in this thesis is that of assessing a baseline of student skills and how those skills are changing over the course of an undergraduate degree. Design and background of the test is included to give context to the results.

5.2.1 Design and validation

The first step in the creation of the Data Handling Diagnostic test was to consult with ‘subject experts’ as to what key skills it was felt that students should be able to demonstrate. These experts were members of staff involved in the teaching of undergraduate experimental or data analysis classes from academic institutions around the UK. It was an objective of the design that the test should not be discipline specific and could therefore be used by any undergraduate course which involved experimental work; as a result experts were consulted from a broad range of physical sciences subjects, as opposed to just physics. Four main categories were devised by the experts as areas all undergraduate students should have expertise in:

- Averaging
- Precision, accuracy and sources of error
- Graphical methods (line fitting/outlier/extracting functional forms)
- Quantitative handling of experimental error

Once the subject categories had been decided upon, staff in the Physics Education Research group at the University of Edinburgh began to design a series of questions related to this topic. A bank of multiple choice questions was created, each with 4 answers⁴. The answers were designed based on common mistakes staff had seen students make so that the correct answers were not easy for the students to pick out by pattern matching. From the bank of questions a test was created with the questions selected chosen to give a wide coverage of the

³Thanks to R.K. Galloway, in particular, for the time he spent explaining the original design process.

⁴Three of the questions used were designed by Day and Bonn at the University of British Columbia, see Reference [111].

categories listed above, as well as a range of difficulty levels. This test was then given to the same group of subject experts for feedback as to the validity of the test as a whole and the clarity of the questions. Undergraduate students were also asked to comment on the meaning of the questions in order to assess the clarity from a user prospective. After validation with the experts and students, modifications were made to any questions that were deemed to be unclear and any that were considered to be unsuitable for modification were removed. The original test, after initial validation checks, consisted of 23 questions.

Reliability of the test

Having decided upon an initial set of questions to make up the data handling diagnostic test, the prototype test was given to students in years one through to five of an undergraduate degree in ten institutions in the UK and Ireland. In total 1164 students completed the test, with the level of participation from each of the ten institutions dependant on the way in which the students were exposed to the test. These test responses were then used to assess the validity and robustness of the test instrument, using statistical tests designed specifically for this purpose, as detailed in Chapter 2. The results of the tests will be discussed in this chapter, alongside the implications of the results for the instrument, however for further details of the tests themselves the reader should refer to Chapter 2. The validity and robustness tests were all carried out by the thesis author and mark the first personal involvement with the instrument.

Item Difficulty Index

The first test carried out on the instrument was to measure the Item Difficulty of the individual test items as well as of the test as a whole. Here, as for all subsequent quoted statistics, where the student did not answer the question their responses have been discounted rather than included as incorrect. The decision to remove blank answers from the total number of responses was taken so that the later questions were not erroneously found to be more difficult for the students due to the fact that time constraints meant that not all the students completed the test. On average each students left less than 1 out of the 23 questions blank.

Figure 5.14 shows the item difficulty index for all 23 items of the test. There are six questions that fall outside of the acceptable range (0.3-0.9) of which four

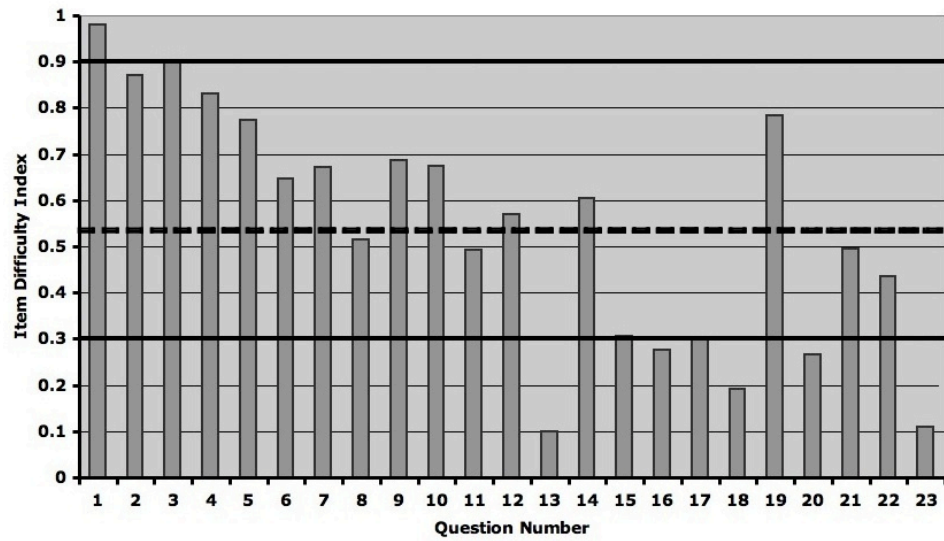


Figure 5.14: The item difficulty index for 23 questions of the prototype Data Handling Diagnostic. The solid lines at 0.3 and 0.9 show the lower and upper acceptable criterion for score. The dotted line shows the mean score for the test.

were substantially different from the ideal criterion and are therefore highlighted as needing possible changes. Of the four which did not fall within the desired range three fell below the lower criterion, meaning that few of the class managed to correctly answer the question, the final one came above the upper criterion, meaning that virtually all of the students answered it correctly. The item difficulty of the whole test was found to be 0.54, within the set criteria and close to the optimum figure of 0.50.

Item Discrimination Index

The result of subjecting the instrument to tests measuring the Item Discrimination Index, the ability of the individual test items (and the test as a whole) to distinguish between the students of high ability and those of low ability, is shown in Figure 5.15. Items which have a high discrimination index represent cases where those students that performed well on the test as a whole also scored highly on the item. Any item with a negative item discrimination index indicates that the question has been answered correctly more often by weaker students than stronger students and as such should be discarded.

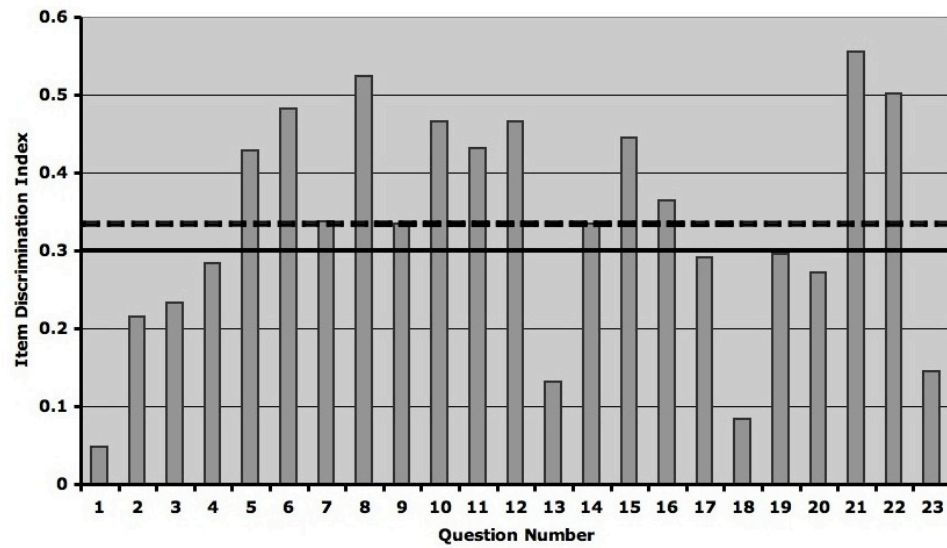


Figure 5.15: The item discrimination index for 23 questions of the prototype Data Handling Diagnostic. The solid line at 0.3 shows the lower acceptable criterion for score. The dotted line shows the mean score for the test.

As was discussed in Chapter 2, there are two versions of the item discrimination test and which to use is dependent on the individual circumstances. The 25-25 split test involves using only the upper and lower quartiles of the test and discarding the middle two quartiles which are considered to be less reliable. The 50-50 split uses all the data but can tend to under estimate the item discrimination due to the inclusion of the more unreliable middle quartiles; this test should be used when the data set is not large enough to justify discarding the data. With a sample size of over 1000 ($N=1164$) it was decided that the 25-25 split would be used for this instrument and the results shown in Figure 5.15 are for this test.

Figure 5.15 shows that the majority of items fall in a range of 0.2-0.6 with an average discrimination index for the whole test of 0.33, above the acceptable criterion of 0.3. It is not the case that all questions must be above the 0.3 range if the test as a whole has a score above the acceptable criterion and most questions are above or close to 0.3 [67]. Although no questions have a negative item discrimination index, four questions stand out as having lower discriminatory power than the others; 1, 13, 18 and 23. The low discriminatory power of these questions is due to their very low (or in the case of question 1) very high item difficulty index as it is not possible to judge the discriminatory power of a question

if the majority of the whole class are either giving the correct or incorrect answer.

Point Biserial Coefficient

The Point Biserial Coefficient measures the consistency of an item compared to the test as a whole, with the mean value enabling the measurement of the self consistency of the whole test. The acceptable range of values is a score greater than 0.2 with higher scores considered to show a higher level of consistency with the test.

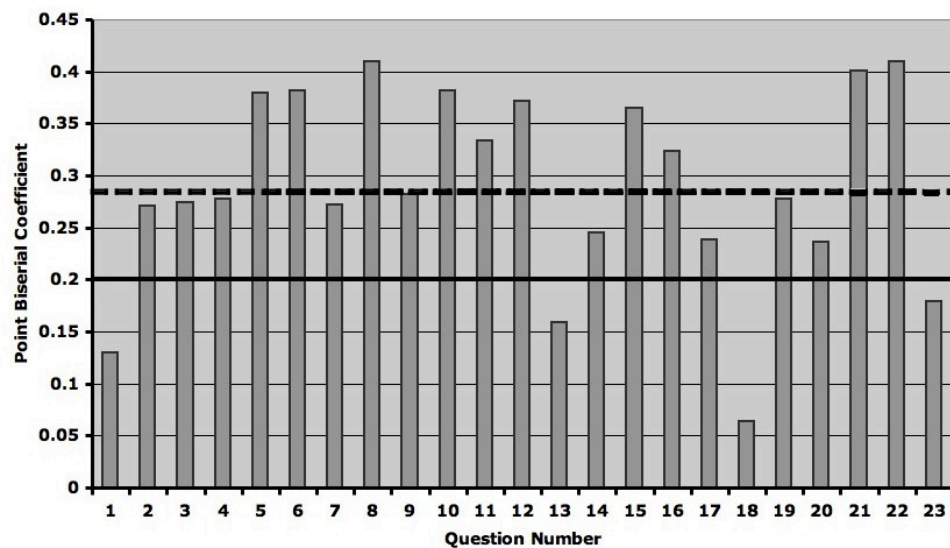


Figure 5.16: The point biserial coefficient for 23 questions of the prototype Data Handling Diagnostic. The solid line at 0.2 shows the lower acceptable criterion for score. The dotted line shows the mean score for the test.

Figure 5.16 shows the point biserial coefficient for all 23 items: 19 of the 23 items fall within the acceptable range, with the same four previously identified questions falling below the criteria (1,13,18 and 23). The average score for the whole test is 0.29, above the acceptable lower bound.

In addition to statistics that examine the reliability of individual items of the prototype test, a measurement known as Ferguson's delta was used to measure the reliability of the whole test. The test looks at the spread of student scores over the whole possible range of scores, with the view being that tests with high ability to discriminate between students will have a wide range of scores. A value

of greater than 0.9 is considered to show a test with high discriminatory power; the score for the Data Handling Diagnostic was found to be 0.95.

Another test which is commonly used to measure the reliability of an instrument as a whole is the Kuder-Richardson 21 (KR-21) test [67]. It has not been used when considering the reliability of the Data Handling Diagnostic as one of the key assumptions of the Kuder Richardson test is that the questions are all of the same difficulty level, a feature that is intentionally not the case for the Data Handling Diagnostic.

As can be seen from the results discussed above, the prototype Data Handling diagnostic was found to be reliable, valid and robust for all relevant statistical measures. Exceptions to this were found in questions 1, 13, 18 and 23, to which revisions were made before the publication of the test (for further details of these revisions the reader is referred to Reference [112]).

5.2.2 Results

Having found the prototype instrument to be fit for purpose, it is now possible to use the data collected in the initial assessment stage to investigate the performance of the whole cohort on the test, as well as looking for differences in cohorts within an institution where students at varying levels of study have taken the test.

In total 1164 students completed the test, with the cohort made up of students from ten different institutions in the UK and Ireland and with data collected from at least three different year cohorts at three of the ten institutions. Table 5.1 shows the breakdown of respondents by institution, as well as the mean score for each class.

The average score for the test overall was 12 correct answers out of 23 (52%), when looking at all data collected, without filtering by institution or year of study. Figure 5.17 shows the scores per question for the whole cohort; the questions have been colour coded by the survey designers to indicate the intended difficulty level of the question.

The data collected and trends seen are discussed as a whole, as well as on a per institution basis in works by other members of the design team and interested readers should refer to the original paper for details [112]. In keeping with the themes of this thesis the focus here will not be on the whole class statistics or

Table 5.1: Results of test deployment with 10 trial institutions. Table shows year of study, degree discipline, number of students in each class, mean score and standard deviation of mean score. Note that the years have been renumbered so that Scottish institutions begin with a year zero: this is so that comparisons can be made between institutions from around the UK and Ireland and reflects the fact that Scottish degrees are one year longer than those in the rest of the UK.

Institution	Degree Discipline	Year	N	Mean Score (%)	St. Dev. (%)
Institution A	Physics	0	256	50	11
Institution A	Physics	1	121	61	12
Institution A	Physics	2	64	62	11
Institution A	Physics	3	40	62	12
Institution A	Physics	4	27	63	13
Institution A	Chemistry	2	55	54	13
Institution B	Physics	0	93	47	11
Institution B	Physics	1	21	48	14
Institution B	Physics	2	22	46	11
Institution C	Physics	2	50	54	13
Institution C	Physics	3	54	56	13
Institution C	Physics	4	27	59	10
Institution D	Physics	1	71	59	13
Institution D	Physics	2	51	57	11
Institution E	Chemistry	1	56	44	9
Institution E	Chemistry	2	29	53	9
Institution F	Chemistry	1	74	49	11
Institution G	Chemistry	1	39	47	13
Institution H	Physics	2	30	39	9
Institution I	Physics	2	24	53	11
Institution J	Physics	2	5	62	12

variation between institutions but on the changes that are seen in students on a year by year basis in the **same** academic institutions.

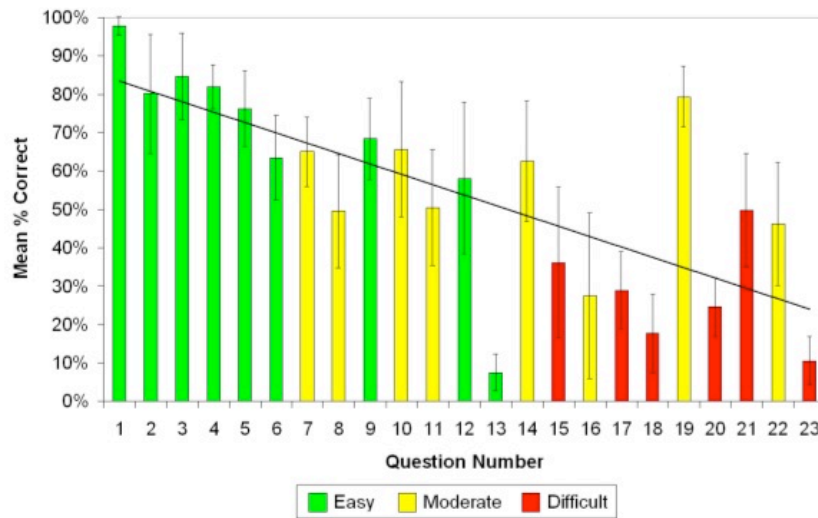


Figure 5.17: The mean fraction of students answering each of the 23 test items correctly, averaged over all trial classes. The error bars show the standard deviation between classes, and the columns are colour-coded according to the expert view of each questions anticipated difficulty. The solid line is an unweighted linear fit to the answer profile. Figure reproduced with permission. For full details see Reference [112]

Results by year cohort

Five of the ten institutions which deployed the Data Handling Diagnostic with their students did so for multiple years groups of the same degree program. All data collection was carried out at the same point in the same academic year meaning that the data collected is not a fully longitudinal study with the same students being followed through their degrees, but rather a ‘snapshot’ of the answers of students in all year groups at one point in time. In this sense it resembles the method used in the pseudo-longitudinal CLASS study described in Chapters 2 and 4, and has associated with it the same implicit assumption, that the intake criteria for students has not changed so the same types of students are likely to be found in each consecutive year.

The most complete set of data available for comparison purposes are those collected at Institution A where all five years of an undergraduate degree program

have been surveyed⁵.

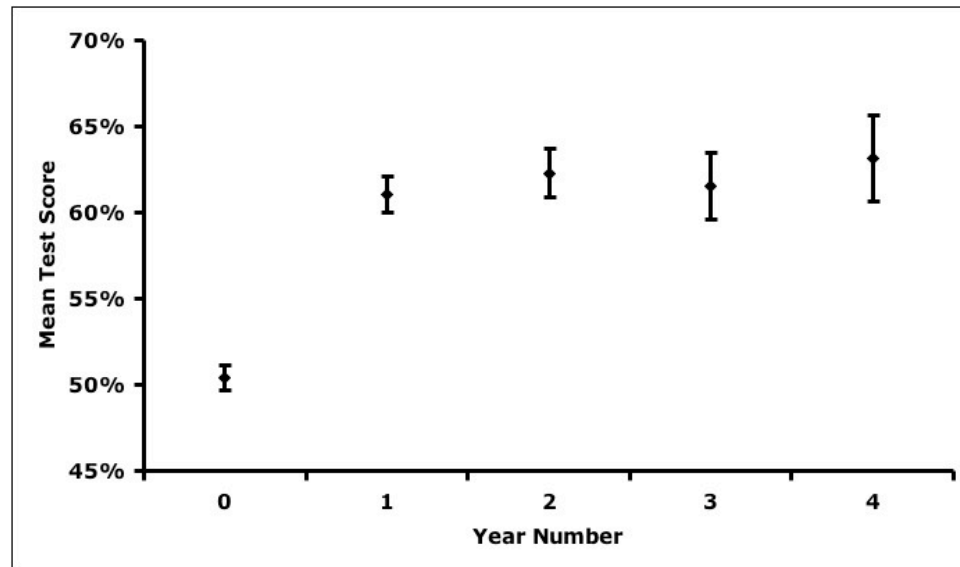


Figure 5.18: Mean test scores for each class in the physics program at institution A. The error bars show the standard error on the mean score. Figure and caption reproduced with permission. For full details see Reference [112]

Figure 5.18 shows the scores of the five year groups from Institution A, as a percentage score. There is a clear difference between those students in Year 0 (the Scottish first year students as have been considered through this thesis) and those in subsequent years, with the students in year 0 scoring 50% on the test and the following years all remarkably consistent between 61% and 63%. A single factor ANOVA test shows that there is a statistically significant difference in the data set ($p \leq 0.001$), however when an ANOVA test is run using only years 1-4 no difference is seen ($p = 0.81$) indicating that the difference in the overall scores is a result of the lower score for the first year. The students in year 0 were surveyed during their first laboratory session at the university and it is known that the subsequent course explicitly teaches data handling skills, perhaps accounting for the large increase in score for students tested at the beginning of their second year of study. The static nature of the later year scores is interesting as experimental work is a required part of the degree programs for all physics students at Institution A until the end of the third year of study (marked year

⁵The inclusion of five years of data indicates that Institution A must be a Scottish University with a five year MPhys Degree

2 in Figure 5.18) and even after this point the majority will go on to carry out experimental work until the end of their studies.

Of the remaining four institutions that collected pseudo-longitudinal data, only one shows any variation in scores between the year cohorts. An ANOVA test on the data collected at Institutions B, C and D show no significant difference between the year cohorts ($p=0.92$, $p=0.19$ and $p=0.28$ respectively) with remarkably consistent average scores, as shown in Table 5.1. Institution E did see an increase in scores between first and second year (from 44% to 53% $p=0.002$), although it should be noted that this was on a chemistry degree program that explicitly teaches data handling skills, rather than physics as is the focus of this thesis.

5.2.3 Section summary

A multiple choice diagnostic test has been design to diagnose gaps in student understanding of basic data handling skills. The instrument has undergone thorough tests with students and academic staff to establish face and construct validity before being trialed with 1164 students from ten academic institutions in the UK and Ireland. The results from the prototype trials have enabled the use of standard validity and robustness tests to be used, which have provided satisfactory results by all measures. Where questions were identified as unreliable or not fit for purpose they have been highlighted for potential future revisions as discussed in the original paper accompanying the publication of the test [112]. As half of the institutions taking part in the trial used the test with multiple year cohorts it has been possible to look at the changes in student performance over several years of a degree program. In three of the five institutions that have used the Data Handling Diagnostic with multiple year groups, no variation is seen in the results between years with little deviation from the average score on the test for all 1164 students. At Institution A, where the most complete pseudo-longitudinal data originates, a large step up in score is seen between the first year of study and the second, although after that the scores are static despite the inclusion of laboratory based courses throughout the degree.

5.3 Chapter summary

We have seen in Chapter 1 that experimental skills are highly prized in physics and laboratory work is considered a vital aspect of any physics degree. There are questions as to whether these skills are taught explicitly often enough as students are often assumed to learn such skills through working in the physics laboratory. In this chapter two different approaches have been used to consider the student experience of experimental physics, firstly student attitudes towards experimental work and secondly their ability to use the data handling skills needed to successfully navigate the undergraduate laboratory.

In terms of their attitudes towards and views of experimental work, the students at the University of Edinburgh are largely positive, although there is, in some areas, a sign that students are less positive in their views than they were of their school experimental experience. Comparing the experiences of students in schools and university is extremely difficult, due to the fact the survey was designed in such a way that the majority of the questions asked are very context specific, such as asking students if practical work is “well organised”. University students are seen to find practical work to be “less understandable” than those students in school, however it is difficult to know if this is a function of the teaching style or an increase in difficulty in the subject material. In all the questions asked, even when a difference was detected between the views of those in schools and those at university, the views were still very positive with the majority of students finding labs easy to understand, and offering a good linkage of theory and practical work.

In contrast, the use of a diagnostic test to examine the experimental skills of the students shows that students are not well equipped with the skills needed to navigate the experimental work. With an average score of 52% for all students over the whole test it is clear that there are large areas of knowledge that students have not managed to learn: bearing in mind that the test was devised with the input of academic staff all of whom teach such experimental classes, who contributed questions on subjects they felt that students should be able to do, the score is remarkably low.

The lack of increase in the scores over time is also of interest with three of the ten institutions that provided pseudo-longitudinal data showing no change in score at all between years. The data of Institution A shows the most complete picture

of the test scores of students, with a step up in score seen between the first and second years of study but with no change from second year through to the end of the degree program. This increase in scores between first and second year is also seen at Institution E, although this data is collected from a chemistry degree program. At both Institutions A and E data handling skills are explicitly taught in the first year of study and at both institutions gains are seen in overall test score. The other institutions in the survey are known to rely upon more implicit methods of teaching such skills, which may well explain the lack of overall gain. This may, therefore, have implications for potential curriculum reform if such skills are to be encouraged.

The two studies included in this chapter provide an interesting juxtaposition where students appear to be confident in their own ability but testing of the students suggests that this confidence is misplaced. While it is the case that data handling skills represent only one aspect of practical work they are, never the less, a key component. The results of this chapter suggest that, given that many undergraduate students will go on after graduation to either laboratory based careers or postgraduate study, there is a need to bridge the gap between student expectations and the reality of student ability in the undergraduate laboratory.

Chapter 6

Conclusions, Discussion and Future Work

The focus of this thesis has been on investigating the changes seen in student abilities and attitudes, during their transition from secondary school pupils through to the end of an undergraduate degree. This has been studied from three different perspectives, firstly the school to university transition, followed by student attitudes towards learning and, finally, student laboratory work (both from the perspective of perceptions towards practical work and the data handling skills required for such work). All of the work presented in this thesis has been carried out at the University of Edinburgh in Scotland. The field of Physics Education Research is still relatively new within the UK, which presented many opportunities for innovative research. The structure of the Scottish degree program, as extensively described in Chapter 2, presents a unique, within the UK, opportunity for data collection as the large first year physics class is made up of approximately half physics majors and half non-majors. In contrast to the non-majors seen in North American academic institutions these students must have school leaving qualifications in physics and mathematics and must be as well qualified in these areas as those who have chosen to study physics as their degree, the only thing that separates the two groups is their chosen subject of study.

In Chapter 3 it has been seen that due to the differing entrance qualifications that students entering the physics degree program at the University of Edinburgh may possess, as well as the variation in course content between different A-

Level examination syllabi there is no absolute definition of what students will have covered in school prior to university. For students studying the A-Level qualification we see that there is variation in content for both physics and mathematics, with differences seen in even the types of mathematics modules chosen for students to study. In this thesis it has been hypothesised that this variation in knowledge may lead to anxiety and loss of confidence for some students, however we have not investigated whether this is the case. Previous work conducted in the School of Physics and Astronomy¹ has shown that students that begin their degree with only SQA Higher qualifications tend to leave the degree program during the first year at a higher rate than those with Advanced Highers and A-Levels, however no difference is seen in the retention rate of the latter two groups. It would have been an interesting study to examine whether there is a link between which A-Level examination boards students have taken qualifications with and their performance on the degree program, unfortunately this was not possible as this level of detail was not stored by the university central registry. If this information could be found (either from existing records or through asking the students to provide such information) it could present an interesting potential area for future work. For academic staff to keep up with all the different syllabi incoming students may have studied is a large task to undertake alongside other teaching commitments. This task is further complicated by the fact that these syllabi are evolving with time and constantly changing (in fact a new SQA Higher syllabus was announced in 2010 with the first student examinations on the syllabus available for early adopters from Summer 2012 [113]). One of the ways that this task can be made easier for staff is through publications such as the review of pre higher education mathematics that was published by the UK Higher Education Academy [114] in 2010, although to the best of our knowledge there is not an equivalent resource detailing the physics qualifications.

The issue of instructor expectations of student ability was also discussed in Chapter 3. It was revealed that teachers have higher expectations of student mathematical ability than the realities of students' actual mathematical skills. This may be indicative of a wider issue, where instructors tend to form expectations of student ability based upon their own previous knowledge and

¹This work has not been published but has been carried out by members of academic staff within the School of Physics and Astronomy

through remembering back to their own education, without considering the fact that the level and content of the school leaving qualifications has changed, especially over the last twenty years. Whilst it is undoubtedly true that the work carried out in this thesis on the topic of staff expectations is a small scale study, both in terms of the data it aims to gather and the number of academic staff responding, conclusions from the data presented suggest that there is scope to carry out further work on the topic. If this area of study were to be investigated further a prudent first step would be to repeat the study with a diagnostic test that was known to be more reliable than the one that has been used in this thesis. As has been extensively discussed in Chapter 3 it is believed there were flaws with both the length and level of difficulty of the diagnostic test deployed, although this was outside of the control of the Physics Education Research group. If a more reliable test were used and a clear difference still remained between the staff expectations and student performance this would add credibility to the initial results included here. It could prove interesting to extend the study further by also looking at the physics skills and abilities of students compared to staff expectations. If a test such as the Force Concept Inventory [66], which examines student conceptual thinking rather than subject knowledge, could be used this would potentially remove any bias arising from instructors tending to judge student knowledge based on their own prior learning and would provide a clearer measure of how accurately instructors can estimate student ability. Results of using the Force Concept Inventory to gauge instructor expectations of student physics ability might also provide an opportunity for instructors to reflect on and perhaps reconsider the material they are teaching or the teaching methods they employ. The FCI has been extensively used at Harvard University as a measure of the effectiveness of the Peer Instruction method of teaching [115]. Mazur who pioneered the use of Peer Instruction in the undergraduate class room has stated that on viewing the test instrument for the first time he was confident that his students would not struggle with the questions and in fact was worried students would be “*offended by the simplicity*” of the instrument [116]. However he has stated that the results gathered from using the test with his students clearly showed that there was a large gap between his perceptions of the level of difficulty of the instrument and the performance of the students within his class [116].

Scientific Reasoning was examined on either side of the secondary school to university transition using selected questions from the Lawson Test of Scientific Reasoning[28]. It was found that, although both groups of students performed well on the test, there was a clear gap between the two cohorts with the secondary school students performing less well on all five questions used. In one of the five questions, where students were asked to consider the reaction of flies to light, not only was a gap observed in the answers of the two cohorts but, in addition, both groups were clearly seen to show an example of incomplete scientific reasoning. While both groups were able to answer the question correctly only a minority of the students were able to correctly articulate **why** the answer was the correct answer. It would appear that the test used was beginning to reach the limit of its usefulness for the university students involved in the study, as in some questions it appeared that all the students were answering the questions correctly, leaving little room for any interpretation of inter-cohort differences. In retrospect it may have been a mistake to use only selected questions from the test as this limited the usefulness of the data for comparing to previously conducted studies. This decision was made when this body of research was still intended as a one year Masters by Research degree and opportunities to meet with and survey students (school pupils in particular) were few and time restricted. If the body of work in this thesis were to be repeated from the beginning it would be more beneficial to collect, if possible, answers to the full test so that this could be compared to other published studies. Using the full test would also remove the possibility that the questions chosen either over or underestimated the gap in reasoning between the two cohorts.

The topic of attitudes towards study has been a major theme of this thesis with results of studies carried out into student attitudes first presented in Chapter 3. Here it is seen that students attitudes towards studying physics become less positive during the first year of university study, when measured using the Colorado Learning Attitudes about Science Survey [34], in fact (using the terminology associated with the assessment instrument) they are said to become less expert-like in their physics thinking. As has been discussed within this thesis this result is in line with previous use of the assessment tool where students have consistently been seen to become less expert-like in their thinking after a year (or even a semester) of physics instruction. The data and results presented in

this thesis are the first known use of the CLASS survey in the UK and the fact that the pattern of a decline in expert-like thinking is repeated in such a different educational system to that of North America suggests that this a symptom of an overall way of thinking rather than a function of the teaching style or content. In this thesis it is suggested that this decline in expert-like thinking may be linked to the works of Perry [51] which outlines stages of learning for undergraduate students.

The results found from initial usage of the CLASS instrument led to further investigation into student attitudes towards study and expert-like thinking, which is found in Chapter 4 of this thesis. This body of work contains two branches of interest; a more in depth study of the first year experience and a longitudinal study into the changing attitudes of students as they progress through the physics degree program. In the first of these areas, the first year experience, the first year class as a whole was split into sub cohorts in order to examine any differences that existed between different parts of the class. The difference between physics majors and non-majors was examined with those students who were not intending to study a physics degree seen, prior to university instruction, to be significantly less expert than those who were enrolled on a physics degree - a gap which only widened after instruction. This difference exists despite the fact that there is no significant difference in the school leaving qualifications of the two cohorts and all those taking the first year physics course had chosen to do so². Further examination of this data revealed that there is also no link between CLASS score and school leaving grades, removing any suggestion that the non majors are less expert-like thinkers as they are 'weaker' students.

The Scottish degree system presents a unique opportunity to examine the differences between physics majors and non majors as, unlike the North American education system, all those students taking physics as an outside subject in their first year of study must be as well qualified to study on the physics program as the physics majors. To further illustrate the uniqueness of this cohort further data was collected with biology students who were studying a course of physics and mathematics as part of their degree. The expert-like thinking of these students was found to be far lower than those in the physics class defined as non-majors and was in line with the scores of non-majoring students from North American

²The only degree program for non majors that requires their students to take the first year physics class is Geophysics who make up approximately 10% of the physics class.

institutions.

Another way that the first year physics cohort was subdivided was to examine the differences in expert-like thinking between female and male students. It was found that while no difference in expert-like thinking existed prior to university instruction there was a statistically significant difference between the two groups after instruction, with the female students seen to be significantly less expert than their male peers³. This gap in expert-like thinking is also seen in the literature with males tending to be more expert-like than female students, as the gap does not exist prior to instruction it suggests that this difference is due to something that changes during the semester, which again could suggest a drop in confidence for the female students.

The second area of research found within Chapter 4 is a longitudinal study of student attitudes over the course on an undergraduate degree, again using the CLASS instrument. Two approaches were used to investigate this topic; firstly a pseudo-longitudinal study, where all data was collected in the same year and intra-cohort comparisons were drawn and secondly a fully longitudinal study where the same students were surveyed on numerous occasions as they progressed through their undergraduate degree. Both the methods have advantages and disadvantages, as discussed in Chapter 4, however it would appear that both methods produce consistent and therefore reliable results. As well as confirming the validity of each of the methods, the two techniques used also present the same overriding picture; that there is remarkably little overall change in student attitudes over the course of an undergraduate degree. Students come into the degree program with relatively high levels of expert-like thinking that remain high throughout their degree (with the first year dip in expert-like thinking at the end of first year study having recovered to previous levels when students are surveyed at the end of their second year of study).

This surprising lack of change in expert-like thinking is made perhaps even more remarkable by the fact that the category scores for each of the year cohorts are showing some variation, as discussed in Chapter 4. In both the fully longitudinal and pseudo-longitudinal methods some variation in category

³An exception to this was in the 2010-11 data where the female students were found to remain constant in their expert-like thinking. It is not known whether this is due to changes in the recruitment and teaching of the first year cohort or an anonymously, this is discussed in further detail in Chapter 4.

scores is seen for each different year cohort, however the overall level of expert-like thinking remains the same. The reasons for this are not known although it may, in part, be due to the way the questions have been categorised, with some questions fitting into more than one category and some into none at all. However the number of questions that do not fit into any category are small and it is unlikely that these would mask changes happening in the categorised questions to keep the overall score the same. Relatively few questions feature in each of the CLASS categories, with the most questions in any category being eight out of the 42 available statements and the fewest being only four statements. For this reason a change in the answer of any one statement from positive to negative (or visa versa) will have a large effect on the category score but will have a minimal effect on the score over all 42 statements. For this reason it is the view of the thesis author that any comparisons made between cohorts on a per category basis should be treated with extreme caution.

The fully longitudinal method involved collecting data only from students enrolled in the degree program whereas the pseudo longitudinal method allowed the consideration of people outside this range, for example students in schools, and those graduates that have enrolled on PhD programs as well as members of academic staff. By extending the sampled range to include those outside of the undergraduate degree two clear transition regions have been observed. A significant step up in expert-like thinking was seen when students decided to study physics at university and a further increase was observed when those who have completed their degree go on to study for post-graduate qualifications in the subject. This has led to the hypothesis that what is being observed is a selection effect where the most expert-like thinkers are staying on in the field of physics whilst those with less expert-like views leave at the potential exit points.

Attempts to prove the selection effect hypothesis are most certainly a priority area for future work emerging from this thesis, although any attempt to do so could meet with several potential difficulties. Firstly the number of students in the final years of study of the undergraduate degree is much smaller than those beginning the degree, this means that any attempt to collect CLASS survey data with a large sample size is difficult. In addition students are often undecided on their career path at the point of graduation, which is generally the last possible point where survey data could be collected for these students.

Presented in Chapter 4 is an attempt to test the selection effect hypothesis by trying to correlate intended career destination data with CLASS scores. The data is suggestive of a trend but with the small sample of students that responded to surveys about their post graduation plans no statistical significance could be established. Another issue that complicates any attempt to test the hypothesis is the possibility that the survey may not be valid for individuals. In the original paper that accompanied the release of the CLASS survey it was noted that the instrument had not been validated for individuals and this was outlined as an area for future work [34]. To the best of our knowledge, at the time of writing this thesis, no such study had been published. Furthermore the analysis of the data collected in the fully longitudinal section of this thesis has revealed a large amount of change in the rankings of individuals in the class in terms of their CLASS score on a year to year basis, despite the fact that the overall score has remained the same. This leads us to believe that in fact the CLASS instrument is not valid for individuals, a view that is shared by one of the original survey authors [80]. Whilst the survey not being valid for individuals would not have an effect on any of the work presented in this thesis, which looks at cohorts of students rather than individuals, it would effectively make it impossible to carry out any studies where the scores of individuals were compared to their career destinations, thus necessitating a very large sample size, so that the scores of whole cohorts with the same intended destinations could be considered.

If it does prove possible to test the selection effect hypothesis and does indeed prove to be the case that those students with the most expert-like thinking on entering a degree program are the same students that go onto become physics postgraduate students and beyond, then this could potentially be expected to have implications on student recruitment. Certainly it would provide another potential source of information for those responsible for selecting new students, although questions would need to be asked as to whether students attitudes towards study should have an effect on whether they are allowed to embark on a physics degree. At the very core of the CLASS survey lies one key question, who are the expert physicists that are being held up as an example of perfect expert-like thinkers? It is known from the original paper that accompanied the release of the CLASS survey[34] that the expert opinion was collected from twelve physics academics and instructors at the University of Colorado (see Chapter 2

for further details) who came up with a general consensus to the survey questions through group discussion. It is however interesting to note that when the CLASS survey was given to members of staff as part of the pseudo-longitudinal survey in Chapter 4 of this thesis the staff scored 83% which even taking into account the few questions in the survey that do not have an overall expert response is still quite different to the expert response included in the survey. This has been further investigated in the work carried out as part of another PhD project in the Physics Education Research group[117] where the survey was given to over 150 physics academics from institutions around the UK and again the score was well below 100% expert agreement. This could perhaps suggest that a researcher at a British academic institution has a different expert opinion to those in North America. If this is true it surely therefore raises the question of which levels of expertise should be expected of students in a UK academic institution? An alternate explanation could be that rather than aiming for a score of 100% expert-like thinking, it would perhaps be more appropriate to consider an expert region where over a certain score the survey taker is classed as an expert thinker. Questions as to where this line would be drawn though are an area that would need careful consideration and could potentially be an area for future work ⁴.

As well as potential cultural differences in expert-like thinking there is, in addition, a question to be asked as to whether there would be any benefit in using CLASS scores as any form of recruitment test. The purpose of a physics degree is not just to create the next generation of physics academics, many graduates leave to work in industry or take the skills they have learned from their degree to work in an entirely different field. Those people may have a different way of thinking to that of physics academics but they are not necessarily any less suited to study a physics degree.

It is the personal view of the author of this thesis that CLASS is at its most useful in allowing the changes in student attitudes to be monitored, either of one cohort over time or looking at the differences between two different groups. Regardless of whether the expert opinions provided by the survey designers are truly the correct answers to the questions (if such a thing can even be said to exist for a survey that examines people's personal attitudes) they do provide a baseline against which attitudes can be measured. By comparing all cohorts to

⁴This work has already begun, in part, through work carried out within the Physics Education Research group at the University of Edinburgh see [117]

the same baseline this allows the comparison of the changes in attitudes of groups over time that would not otherwise be possible

The final results chapter of this thesis considered student laboratory skills and student expectations of the undergraduate teaching laboratory. This work again consists, in part, of work that initially began when this project was a one year Masters by Research project and so has links back to the first year experience and the school university transition as discussed in Chapter 3.

When considering student expectations of experimental work the students were surveyed on two occasions, once at the beginning of their degree, when they were asked to retrospectively consider their experiences of school practical work, and once at the end of their first year of university teaching when they were asked to consider the teaching they had just experienced in their undergraduate laboratory classes. This work was carried out as part of a collaboration lead by the University of Glasgow and the data was collected with the research aims of another project in mind. As such the surveys were not ideal for the research questions addressed in this thesis, as they were not designed to serve as pre and post instruction surveys, but rather as two separate surveys addressing two different lab experiences. However three of the five attitude questions on the surveys were used to compare the expectations of students in both the school and university laboratory environment. It was seen that both cohorts were generally positive about their laboratory experiences, although in all cases where a statistically significant difference existed between the two groups, students had a more favourable view of their school practical work than that at university. It is interesting to note that both groups are confident of their ability stating that they found laboratory work understandable. Students also appear to have an understanding of the bigger picture behind the necessity of laboratory work with students suggesting labs are important because they illustrate theory as well as because they are where new discoveries are made. These are perhaps both quite sophisticated ways of thinking showing a higher level of comprehension of some of the reasons why students are compelled to do at least some practical work during their degree.

In the second element of Chapter 5 the design and use of a diagnostic instrument to measure student data handling skills was discussed. The Data Handling Diagnostic was designed by the Physics Education Research group at

the University of Edinburgh and was deployed to over 1000 students in the UK and Ireland. The instrument was subject to a battery of tests designed to assess the validity and robustness of the instrument, as has been outlined in Chapters 2 and 5. The instrument was seen to pass all tests and thus is considered satisfactory for use with students.

The data collected during the initial trials of the Data Handling Diagnostic present a stark picture, with students performing generally fairly poorly on the test. In addition little improvement is seen as students progress through their degree at institutions where several cohorts have been surveyed. A step up in skills is generally only seen if such skills are specifically taught, despite the common view amongst instructors that students will pick up such skills through working in a laboratory environment.

There is clear difference between student expectations of their own ability in the undergraduate laboratory and the actual level of their skills as measured by the Data Handling Diagnostic. A situation is seen where students believe they find the lab understandable and straight forward but tests to measure such skills show that this confidence is misplaced. There are several avenues for future projects that could be explored from this initial work. The data diagnostic is to be deployed with more students to look at the year on year consistency of the test. It has also already been given to students at Harvard University, further testing of other international institutions would provide a clearer picture as to whether the low scores on such a test are an international phenomenon or whether it is specific to the UK and Ireland. If more time were available and it was possible to gather the data it would be advantageous to link this work back to that discussed in Chapter 3 where the content of school leaving qualifications was discussed. In Chapter 3 it was hypothesised that students could be disadvantaged by the examination board chosen or qualification studied at school. If the information regarding the examination board studied could be collected (perhaps even by simply asking the students when they start their degrees) then it would be possible to examine whether any correlation existed between those who have done less practical work at school and their performance on the Data Handling Diagnostic test.

Within this thesis are three results chapters, all relating to some aspect of the transition from secondary school through to graduation. The results

chapters can be read as three distinct bodies of work, however the theme, which runs through them all is always examining how students change during their degree. Despite such a general overarching theme of research and such separate results chapters within the thesis there are common themes that run through the findings of this project. Firstly, it can be seen that perceptions of study play an important role, whether this be the instructors' perceptions of their students, student consideration of their own skills or students attitudes towards science study in general, it can be seen that the way people think about doing science is inherently linked to how they actually do science. In all chapters we see some form of student testing designed to measure either how such attitudes or student skills and abilities are evolving over time. The results of these tests, regardless of whether any measurable scores have changed over time provide a clear set of benchmarks against which student performance can be compared. In this sense what has been created is a map of some of the student experience, focusing on the three main areas of attitudes towards study, the first year experience and undergraduate laboratory work. With so many factors contributing towards the undergraduate experience it would perhaps never be possible to map the entire experience and the works included in this thesis should be considered a starting point for many other potential research projects.

When this project was originally conceived it considered only the school to university transition, with the aim that if changes experienced by students on either side of the school to university boundary could be identified it might be possible to improve the student experience and therefore student retention by better understanding the challenges students are facing. While this project has grown and now considers the whole undergraduate experience thus considering changes in students over a much longer period of time, the aims of the project have not altered. Through better understanding the changes that occur in student thinking and abilities it should be possible for instructors to more effectively support their students. There are many avenues for future research presented by the results of this thesis, as well as many areas which have not yet even been considered but, hopefully, with each piece of research that is carried out on the subject, the undergraduate experience will be better understood by instructors which can only help them to aid students in their success.

Appendix A

Colorado Learning Attitudes about Science Survey

On the following four pages are a copy of the Colorado Learning Attitudes About Science Survey. For further details of the design of the survey see [34].



Name: _____

Last 6 digits of your Student ID #: _____

Introduction

Here are a number of statements that may or may not describe your beliefs about learning physics. You are asked to rate each statement by circling a number between 1 and 5 where the numbers mean the following:

- 1. Strongly Disagree
- 2. Disagree
- 3. Neutral
- 4. Agree
- 5. Strongly Agree

Choose one of the above five choices that best expresses your feeling about the statement. If you don't understand a statement, leave it blank. If you understand, but have no strong opinion, choose 3.

Survey

- 1. A significant problem in learning physics is being able to memorize all the information I need to know.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

- 2. When I am solving a physics problem, I try to decide what would be a reasonable value for the answer.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

- 3. I think about the physics I experience in everyday life.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

- 4. It is useful for me to do lots and lots of problems when learning physics.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

- 5. After I study a topic in physics and feel that I understand it, I have difficulty solving problems on the same topic.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

6. Knowledge in physics consists of many disconnected topics.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

7. As physicists learn more, most physics ideas we use today are likely to be proven wrong.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

8. When I solve a physics problem, I locate an equation that uses the variables given in the problem and plug in the values.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

9. I find that reading the text in detail is a good way for me to learn physics.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

10. There is usually only one correct approach to solving a physics problem.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

11. I am not satisfied until I understand why something works the way it does.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

12. I cannot learn physics if the teacher does not explain things well in class.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

13. I do not expect physics equations to help my understanding of the ideas; they are just for doing calculations.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

14. I study physics to learn knowledge that will be useful in my life outside of school.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

15. If I get stuck on a physics problem my first try, I usually try to figure out a different way that works.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

16. Nearly everyone is capable of understanding physics if they work at it.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

17. Understanding physics basically means being able to recall something you've read or been shown.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

18. There could be two different correct values to a physics problem if I use two different approaches.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

19. To understand physics I discuss it with friends and other students.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

20. I do not spend more than five minutes stuck on a physics problem before giving up or seeking help from someone else.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

21. If I don't remember a particular equation needed to solve a problem on an exam, there's nothing much I can do (legally!) to come up with it.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

22. If I want to apply a method used for solving one physics problem to another problem, the problems must involve very similar situations.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

23. In doing a physics problem, if my calculation gives a result very different from what I'd expect, I'd trust the calculation rather than going back through the problem.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

24. In physics, it is important for me to make sense out of formulas before I can use them correctly.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

25. I enjoy solving physics problems.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

26. In physics, mathematical formulas express meaningful relationships among measurable quantities.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

27. It is important for the government to approve new scientific ideas before they can be widely accepted.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

28. Learning physics changes my ideas about how the world works.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

29. To learn physics, I only need to memorize solutions to sample problems.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

30. Reasoning skills used to understand physics can be helpful to me in my everyday life.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

31. We use this statement to discard the survey of people who are not reading the questions. Please select agree-option 4 (not strongly agree) for this question to preserve your answers.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

32. Spending a lot of time understanding where formulas come from is a waste of time.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

33. I find carefully analyzing only a few problems in detail is a good way for me to learn physics.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

34. I can usually figure out a way to solve physics problems.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

35. The subject of physics has little relation to what I experience in the real world.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

36. There are times I solve a physics problem more than one way to help my understanding.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

37. To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

38. It is possible to explain physics ideas without mathematical formulas.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

39. When I solve a physics problem, I explicitly think about which physics ideas apply to the problem.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

40. If I get stuck on a physics problem, there is no chance I'll figure it out on my own.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

41. It is possible for physicists to carefully perform the same experiment and get two very different results that are both correct.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

42. When studying physics, I relate the important information to what I already know rather than just memorizing it the way it is presented.

Strongly Disagree	1	2	3	4	5	Strongly Agree
-------------------	---	---	---	---	---	----------------

Appendix B

Glossary of Terms

Within this thesis there are many acronyms and abbreviations that have been used, as well as terms relating to the UK education system that may be unfamiliar to those from different educational backgrounds. Every attempt has been made to explain each term when first used within the text but the following list is available for quick reference for the reader. All terms are included in alphabetical order.

Term	Definition
A-Level	The highest school leaving qualification offered by schools in England, Wales and Northern Ireland. A full A-Level consists of six modules and they are usually studied over two years. Pupils will generally study three or four A-Levels and they are generally considered a prerequisite for university study. These are not fully nationalised examinations and a variety of different syllabi are available in each subject dependent on the examination board setting the exam.
Advanced Highers	Qualification offered by the Scottish Qualifications Authority (SQA). Pupils in Scotland have the option of leaving school after study for Higher examinations or staying on for an optional final year and working towards Advanced Highers. Both qualifications can be used to gain entry to university study.
AQA	Assessment and Qualifications Alliance, one of the three English (or five UK excluding Scotland) examination boards.
AS-Level	Advanced subsidiary level. The first three modules of the A-Level examination, which are usually examined at the end of the first year of study. If only the first three modules are completed the qualification gained is known as the AS-Level.
CCEA	Council for the Curriculum, Examinations and Assessment, one of the five examination boards available to students in the UK (outside of Scotland). Covers qualifications, curriculum and assessment in Northern Ireland.
CLASS	The Colorado Learning Attitudes About Science Survey.
Edexcel	One of the three English (or five UK excluding Scotland) examination boards.
FCI	The Force Concept Inventory, a diagnostic multiple choice text which examines students understanding of Newtonian physics. Continued over page

Term	Definition
GCSE	General Certificate of Secondary Education - School Leaving Qualification in England, Wales and Northern Ireland usually taken at the age of 16. 16 years of age marks the end of compulsory education although it is not compulsory for the examinations to be passed before leaving school.
Highers	School leaving qualifications offered by the Scottish Qualifications Authority (SQA). Examinations are taken at the end of the fifth year of secondary education in Scotland and five or six subjects are generally studied. They are considered sufficient to go on to university study although many pupils chose to stay on for a final sixth year of study to work towards Advanced Highers.
OCR	Oxford Cambridge and RSA Examinations, one of the three English (or five UK excluding Scotland) examination boards.
STEM	Science, Technology, Engineering and Mathematics.
WJEC	Welsh Joint Education Committee one of the five examinations boards available to students in the UK (outside of Scotland). Covers qualifications, curriculum and assessment in Wales.

Appendix C

School Leaving Qualifications

In this appendix are details of the structure of each of the seven physics A-Levels and both the Scottish Higher and Advanced Higher qualifications. A summary of the modules of each of the qualifications is seen in Figure I. There is virtually no traditional course work in any of the qualifications. There are two exceptions to this; in the Edexcel A-Level 10% of the total grade can come from either a case study or a written report of an experiment seen on a visit to a working laboratory. The other exception is OCR B, which comprises approximately 10% of the total grade from a written summary of an area of physics, looking at student abilities to reason and defend arguments.

In a wider sense of the term coursework, several of the A-Levels assess practical skills through lab experiments with a written report, carried out in the students own time and marked by teachers in schools. As is often the case with A-levels there is a huge amount of variation between the various exam boards, the details of practical/research work for all seven A-levels and SQA Highers and Advanced Highers is found below.

In all cases the total amount of experimental work is never more than a third of the final A-level grade.

AQA A 20% in total (over the whole A-level) of practical work. This can either be centre marked or an externally marked practical exam. The decision is left to each individual school.

AQA B As for AQA A.

CCEA All students take a practical exam.

Edexcel 10% of total A-Level mark (taken at AS level) comprises either an

experiment based on a visit or a case study involving an application of physics. Students write a report, which is then assessed.

10% of total A-Level (taken at A2 level) is the experimental physics unit. Students plan and carry out an investigation and write a report, which is then assessed.

OCR A One third of the total A-Level is lab work, which is carried out in school and marked internally with moderation from OCR.

OCR B At AS level 16.5% of the marks come from a lab practical with a submitted student lab report.

At A2 level 16.5% of the total mark comes from a unit entitled Researching physics. Two thirds of this unit is a practical investigation but the final third is a short written report summarising an area of physics.

WJEC Practical skills assessed under exam conditions

Higher None, practical work is included alongside other modules.

Advanced Higher Long practical investigation.

Board	Module 1	Module 2	Module 3	Module 4	Module 5	Module 6	Last Revised
Advanced Higher	Mechanics	Electric Phenomena	Wave Phenomena	Investigation	N/A	N/A	2004
Higher	Mechanics and Properties of Matter	Electricity and Electronics	Radiation and Matter	N/A	N/A	N/A	2004
AQA-A	Particles, Quantum Phenomena and Electricity	Mechanics, Materials and Waves	Investigative and Practical Skills	Fields and Further Mechanics	Nuclear, Thermal Physics and an Optional Topics	Investigative and Practical Skills	2008 (first AS 09 A 10)
AQA-B	Harmony and structure in the universe	Physics keeps us going	Investigative and Practical Skills	Physics inside and out	Energy under the microscope	Investigative and Practical Skills	2008 (first AS 09 A 10)
CCEA	Forces, Energy and Electricity	Waves, Photos and Medical Physics	Practical Techniques	Momentum, thermal physics, circular motion, oscillations and atomic and nuclear physics	Fields and their application	Practical Techniques	2008 (first AS 09 A 10)
Edexcel	Physics on the go	Physics at work	Exploring Physics (practical)	Physics on the move	Physics from creation to collapse	Experimental physics	2008 (first AS 09 A 10)
OCR-A	Mechanics	Electrons, Waves and Protons	Practical skills in physics	The Newtonian World	Fields, Particles and Frontiers in Physics	Practical Skills	2008 (first AS 09 A 10)
OCR-B	Physics in Action	Understanding Processes and Experimentation and Data Handling	Physics in Practice	Rise and fall of the clockwork universe	Field and Particle Pictures	Researching Physics	2009
WJEC	Motion, Energy and Charge	Waves and Particles	Practical Physics	Oscillations and Fields	Electromagnetism and nuclei and options	Experimental Physics	2008 (first AS 09 A 10)

Figure C.1: Grid showing the structure of each of the seven physics A-Level syllabi and the two SQA physics qualifications.

Appendix D

Mathematics Diagnostic

On the following two pages are a copy of the mathematics diagnostic test given to first year students at the University of Edinburgh in September 2009. The test was not designed by the Physics Education Research Group and as such only a paper copy was made available to the group. If the scanned copy is unclear please contact the Physics Education Research Group for a photocopied version.

DIAGNOSTIC TEST PAPER

15th, September SEPTEMBER 2009

11:30am

- Answer as many questions as you can.
- Do the questions quickly and accurately.
- No electronic calculators are allowed (or necessary) for solving these problems.
- You do not have to do the questions in any particular order.
- You do not have to show all your working; write down as much or as little as you find necessary for solving the problem.
- For each of the 20 questions, 1 mark will be given for a correct answer, no marks for a wrong answer. No marks are awarded for incomplete solutions.

*Let me see: four times five is twelve, and four times six is thirteen,
and four times seven is – oh dear! I shall never get to twenty at that rate!*

Alice

- 1.1 If $F = G \cdot \frac{Mm}{(R+h)^2}$, solve this equation for h . You may assume that all quantities are positive.
- 1.2 Solve the equation $E^3 + E^2 - 6E = 0$ for the real variable E .
- 1.3 Find all real ϕ for which $\phi^3 + \phi^2 - 6\phi \leq 0$.
- 1.4 For real variable σ , find the smallest value of $\sigma^2 - 6\sigma + 14$.
- 1.5 Simplify as far as possible $\frac{\lambda(1-\lambda^2)}{(\lambda-1)^2} + \frac{\lambda^2+1}{\lambda-1}$.
- 1.6 If $\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$, solve this equation for R .
- 1.7 Solve the quadratic equation $2\kappa^2 - 6\kappa + 3 = 0$ for the real variable κ .
- 1.8 Solve the quadratic equation $\tau^2 \ln 2 - \tau \ln 6 + \ln 3 = 0$ for the real variable τ .
- 1.9 Simplify as far as possible $5 \ln e + \ln \sqrt{e} + \frac{5}{2} \ln \frac{1}{e}$.
- 1.10 Solve the following equation for the real variable $\frac{u}{v}$
- $$4u^2 = 9v(u + v).$$
- 1.11 Express the expression $\frac{\eta-2}{\eta^2-\eta}$ as partial fractions.
- 1.12 Simplify as far as possible $\left(\frac{1}{\cos x} + \tan x\right)\left(\frac{1}{\cos x} - \tan x\right)$.
- 1.13 Order the following three vectors by magnitude
- $$\underline{k} = (1, 3, 4), \quad \underline{l} = (2, 1, 3), \quad \underline{m} = (3, 3, 2).$$
- 1.14 Sketch the function $y(x)$ defined by the equation
- $$(x-1)^2 + y^2 = 1.$$
- 1.15 Differentiate with respect to m the function $T(m) = m^4 + 2 + \frac{3}{m^2}$.
- 1.16 Differentiate with respect to m^3 the function $T(m) = m^4 + 2 + \frac{3}{m^2}$.
- 1.17 Differentiate twice with respect to the variable Φ the function $f(\Phi) = \frac{1}{e^{-\Phi}}$.
- 1.18 Find the constant C such that $n(t) = Ct^2 + 3t^{-3}$ solves the equation
- $$\frac{dn(t)}{dt} + \frac{3}{t}n(t) = t, \quad t > 0.$$
- 1.19 Find the function $\phi(\theta)$ such that the derivative of $\phi(\theta)$ with respect to θ is $\phi'(\theta) = \theta^{-1} + \sqrt{\theta}$ and $\phi(1) = -1$.
- 1.20 Find a value for α such that $\sin\left(\alpha - \frac{\pi}{4}\right) = 1$.

Appendix E

Staff Expectations Survey

Question Number	Question
1	To the nearest 10%, what do you think the mean test score will be when all our first year students take this test?
2	To the nearest 5%, what do you think that the standard deviation of scores will be when all our first year students take this test?
3	Do you think there will be a correlation between diagnostic test mark and maths grade at A-level / Higher / Adv Higher, and if so, how strong?
4	To the nearest 10%, what percentage of the first year class do you think will get the first question correct?
5	Please identify what you think are the three questions that they will find the easiest.
6	Please identify the three they will find the most difficult.

Appendix F

Lawson Test of Scientific Reasoning

On the following ten pages are a copy of the Multiple Choice version of the Lawson Test of Scientific Reasoning (see [28]). Selected questions were given to students in the final year of secondary school and the first year of study at the University of Edinburgh.

**CLASSROOM TEST OF
SCIENTIFIC REASONING**

Multiple Choice Version

Directions to Students:

This is a test of your ability to apply aspects of scientific and mathematical reasoning to analyze a situation to make a prediction or solve a problem. Make a dark mark on the answer sheet for the best answer for each item. If you do not fully understand what is being asked in an item, please ask the test administrator for clarification.

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO

Revised Edition: August 2000 by Anton E. Lawson, Arizona State University. Based on: Lawson, A.E. 1978. Development and validation of the classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15(1): 11-24.

1. Suppose you are given two clay balls of equal size and shape. The two clay balls also weigh the same. One ball is flattened into a pancake-shaped piece. Which of these statements is correct?
- The pancake-shaped piece weighs more than the ball
 - The two pieces still weigh the same
 - The ball weighs more than the pancake-shaped piece

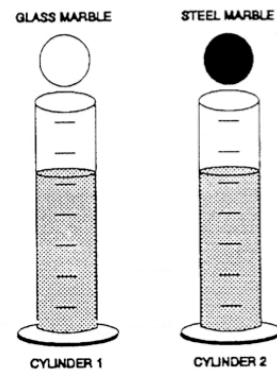
2. *because*

- the flattened piece covers a larger area.
- the ball pushes down more on one spot.
- when something is flattened it loses weight.
- clay has not been added or taken away.
- when something is flattened it gains weight.

3. To the right are drawings of two cylinders filled to the same level with water. The cylinders are identical in size and shape.

Also shown at the right are two marbles, one glass and one steel. The marbles are the same size but the steel one is much heavier than the glass one.

When the glass marble is put into Cylinder 1 it sinks to the bottom and the water level rises to the 6th mark. *If we put the steel marble into Cylinder 2, the water will rise*

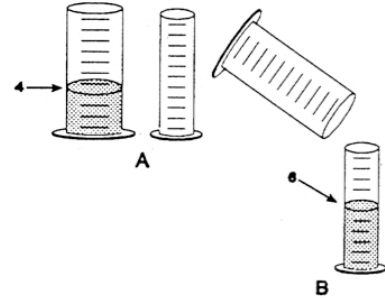


- to the same level as it did in Cylinder 1
- to a higher level than it did in Cylinder 1
- to a lower level than it did in Cylinder 1

4. *because*

- the steel marble will sink faster.
- the marbles are made of different materials.
- the steel marble is heavier than the glass marble.
- the glass marble creates less pressure.
- the marbles are the same size.

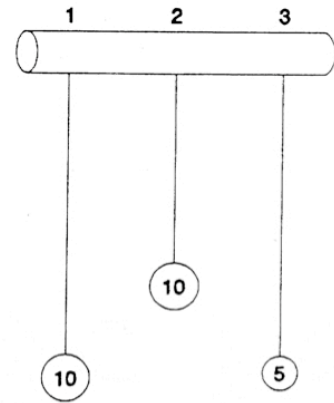
5. To the right are drawings of a wide and a narrow cylinder. The cylinders have equally spaced marks on them. Water is poured into the wide cylinder up to the 4th mark (see A). This water rises to the 6th mark when poured into the narrow cylinder (see B).



Both cylinders are emptied (not shown) and water is poured into the wide cylinder up to the 6th mark. *How high would this water rise if it were poured into the empty narrow cylinder?*

- a. to 8
 b. to 9
 c. to 10
 d. to 12
 e. none of these answers is correct
6. *because*
- a. the answer cannot be determined with the information given.
 b. it went up 2 more before, so it will go up 2 more again.
 c. it goes up 3 in the narrow for every 2 in the wide.
 d. the second cylinder is narrower.
 e. one must actually pour the water and observe to find out.
7. Water is now poured into the narrow cylinder (described in Item 5 above) up to the 11th mark. *How high would this water rise if it were poured into the empty wide cylinder?*
- a. to 9
 b. to 8
 c. to $7\frac{1}{2}$
 d. to $7\frac{1}{3}$
 e. none of these answers is correct
8. *because*
- a. the ratios must stay the same.
 b. one must actually pour the water and observe to find out.
 c. the answer cannot be determined with the information given.
 d. it was 2 less before so it will be 2 less again.
 e. you subtract 2 from the wide for every 3 from the narrow.

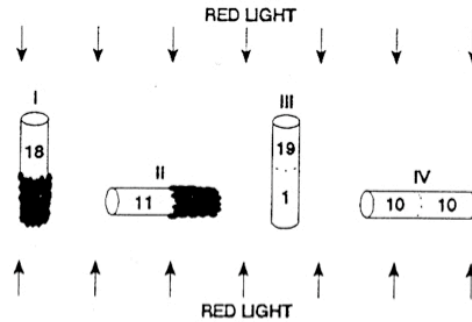
9. At the right are drawings of three strings hanging from a bar. The three strings have metal weights attached to their ends. String 1 and String 3 are the same length. String 2 is shorter. A 10-unit weight is attached to the end of String 1. A 10-unit weight is also attached to the end of String 2. A 5-unit weight is attached to the end of String 3. The strings (and attached weights) can be swung back and forth and the time it takes to make a swing can be timed.



Suppose you want to find out whether the length of the string has an effect on the time it takes to swing back and forth. *Which strings would you use to find out?*

- only one string
 - all three strings
 - 2 and 3
 - 1 and 3
 - 1 and 2
10. *because*
- you must use the longest strings.
 - you must compare strings with both light and heavy weights.
 - only the lengths differ.
 - to make all possible comparisons.
 - the weights differ.

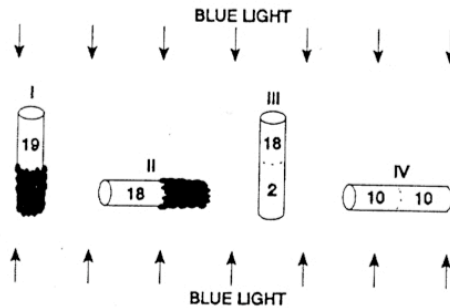
11. Twenty fruit flies are placed in each of four glass tubes. The tubes are sealed. Tubes I and II are partially covered with black paper; Tubes III and IV are not covered. The tubes are placed as shown. Then they are exposed to red light for five minutes. The number of flies in the uncovered part of each tube is shown in the drawing.



This experiment shows that flies respond to (respond means move to or away from):

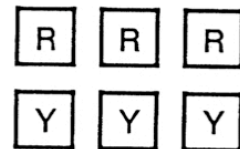
- red light but not gravity
 - gravity but not red light
 - both red light and gravity
 - neither red light nor gravity
12. *because*
- most flies are in the upper end of Tube III but spread about evenly in Tube II.
 - most flies did not go to the bottom of Tubes I and III.
 - the flies need light to see and must fly against gravity.
 - the majority of flies are in the upper ends and in the lighted ends of the tubes.
 - some flies are in both ends of each tube.

13. In a second experiment, a different kind of fly and blue light was used. The results are shown in the drawing.



These data show that these flies respond to (respond means move to or away from):

- blue light but not gravity
 - gravity but not blue light
 - both blue light and gravity
 - neither blue light nor gravity
14. *because*
- some flies are in both ends of each tube.
 - the flies need light to see and must fly against gravity.
 - the flies are spread about evenly in Tube IV and in the upper end of Tube III.
 - most flies are in the lighted end of Tube II but do not go down in Tubes I and III.
 - most flies are in the upper end of Tube I and the lighted end of Tube II.
15. Six square pieces of wood are put into a cloth bag and mixed about. The six pieces are identical in size and shape, however, three pieces are red and three are yellow. Suppose someone reaches into the bag (without looking) and pulls out one piece. *What are the chances that the piece is red?*

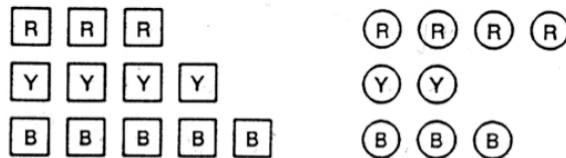


- 1 chance out of 6
- 1 chance out of 3
- 1 chance out of 2
- 1 chance out of 1
- cannot be determined

16. *because*

- a. 3 out of 6 pieces are red.
- b. there is no way to tell which piece will be picked.
- c. only 1 piece of the 6 in the bag is picked.
- d. all 6 pieces are identical in size and shape.
- e. only 1 red piece can be picked out of the 3 red pieces.

17. Three red square pieces of wood, four yellow square pieces, and five blue square pieces are put into a cloth bag. Four red round pieces, two yellow round pieces, and three blue round pieces are also put into the bag. All the pieces are then mixed about. Suppose someone reaches into the bag (without looking and without feeling for a particular shape piece) and pulls out one piece.



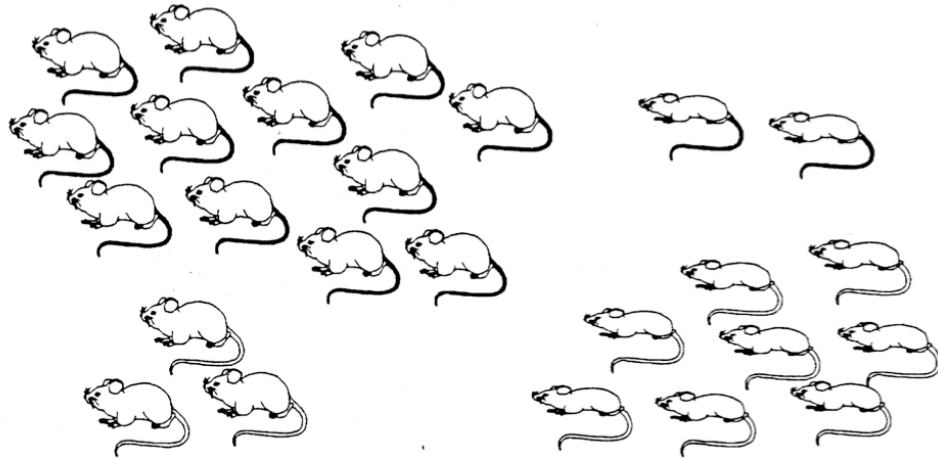
What are the chances that the piece is a red round or blue round piece?

- a. cannot be determined
- b. 1 chance out of 3
- c. 1 chance out of 21
- d. 15 chances out of 21
- e. 1 chance out of 2

18. *because*

- a. 1 of the 2 shapes is round.
- b. 15 of the 21 pieces are red or blue.
- c. there is no way to tell which piece will be picked.
- d. only 1 of the 21 pieces is picked out of the bag.
- e. 1 of every 3 pieces is a red or blue round piece.

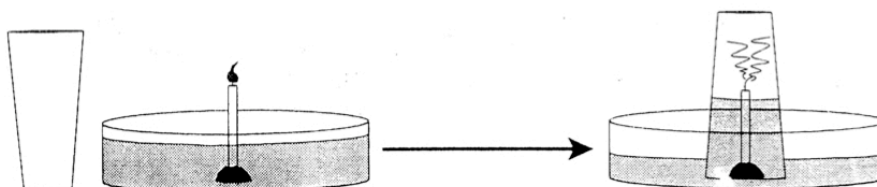
-
19. Farmer Brown was observing the mice that live in his field. He discovered that all of the mice were either fat or thin. Also, all of them had either black tails or white tails. This made him wonder if there might be a link between the size of the mice and the color of their tails. So he captured all of the mice in one part of his field and observed them. Below are the mice that he captured.



Do you think there is a link between the size of the mice and the color of their tails?

- a. appears to be a link
 - b. appears not to be a link
 - c. cannot make a reasonable guess
20. *because*
- a. there are some of each kind of mouse.
 - b. there may be a genetic link between mouse size and tail color.
 - c. there were not enough mice captured.
 - d. most of the fat mice have black tails while most of the thin mice have white tails.
 - e. as the mice grew fatter, their tails became darker.

21. The figure below at the left shows a drinking glass and a burning birthday candle stuck in a small piece of clay standing in a pan of water. When the glass is turned upside down, put over the candle, and placed in the water, the candle quickly goes out and water rushes up into the glass (as shown at the right).



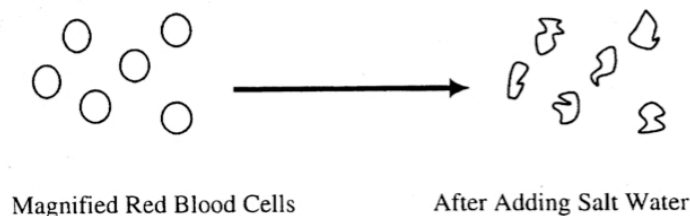
This observation raises an interesting question: Why does the water rush up into the glass?

Here is a possible explanation. The flame converts oxygen into carbon dioxide. Because oxygen does not dissolve rapidly into water but carbon dioxide does, the newly formed carbon dioxide dissolves rapidly into the water, lowering the air pressure inside the glass.

Suppose you have the materials mentioned above plus some matches and some dry ice (dry ice is frozen carbon dioxide). *Using some or all of the materials, how could you test this possible explanation?*

- Saturate the water with carbon dioxide and redo the experiment noting the amount of water rise.
 - The water rises because oxygen is consumed, so redo the experiment in exactly the same way to show water rise due to oxygen loss.
 - Conduct a controlled experiment varying only the number of candles to see if that makes a difference.
 - Suction is responsible for the water rise, so put a balloon over the top of an open-ended cylinder and place the cylinder over the burning candle.
 - Redo the experiment, but make sure it is controlled by holding all independent variables constant; then measure the amount of water rise.
22. What result of your test (mentioned in #21 above) would show that your explanation is probably wrong?
- The water rises to the same level as it did before.
 - The water rises less than it did before.
 - The balloon expands out.
 - The balloon is sucked in.

23. A student put a drop of blood on a microscope slide and then looked at the blood under a microscope. As you can see in the diagram below, the magnified red blood cells look like little round balls. After adding a few drops of salt water to the drop of blood, the student noticed that the cells appeared to become smaller.



This observation raises an interesting question: Why do the red blood cells appear smaller?

Here are two possible explanations: I. Salt ions (Na^+ and Cl^-) push on the cell membranes and make the cells appear smaller. II. Water molecules are attracted to the salt ions so the water molecules move out of the cells and leave the cells smaller.

To test these explanations, the student used some salt water, a very accurate weighing device, and some water-filled plastic bags, and assumed the plastic behaves just like red-blood-cell membranes. The experiment involved carefully weighing a water-filled bag, placing it in a salt solution for ten minutes, and then reweighing the bag.

What result of the experiment would best show that explanation I is probably wrong?

- a. the bag loses weight
 - b. the bag weighs the same
 - c. the bag appears smaller
24. *What result of the experiment would best show that explanation II is probably wrong?*
- a. the bag loses weight
 - b. the bag weighs the same
 - c. the bag appears smaller

Appendix G

Focus Group on expert-like thinking

In the following appendix are notes taken during a focus group on expert-like thinking held with final year students at the University of Edinburgh. Two focus groups were held in total, the first on 26/04/2010 and the second on 28/04/2010.

G.0.1 Focus Group 26th April 2010

Approximate timeline from CLASS focus group, Friday 26th March 2010, 2.00pm, room 4309 JCMB.

2:08 Start

Orientation and motivation.

2:11 What makes an expert in physics?

Independence, subject knowledge, critical thinking, intuition, self-reliance, applying existing knowledge to unfamiliar problems.

2:13 Are you experts?

No. An expert knows one thing extremely well. Expertise is a way of thinking. Some people have an aptitude naturally. Think some other people in the class are experts, but probably wouldnt classify themselves that way. Relative views of expertise. Ways of thinking inextricably bound up with subject knowledge.

2:16 Some debate about the importance of original thought.

2:20 Self-identification as physicists.

Senior honours projects help with this. Equate skills with being a physicist.

2:22

Without realising it we have developed a different approach to problems. You only realise this when talking with other people [non-physicists].

2:24

A physics degree is not enough to make someone an expert. You have to be a certain sort of person. It cant be taught / yes it can. Debate. Genetic pre-dispositions.

2:25 Things in degree that have affected expertise. [No prompting at all from interviewers.]

1st and 4th year enhanced this. 2nd and 3rd year detracted from this. Massive dip in 2nd and 3rd year. Physics Skills helped offset this. Every year became increasingly less confident. 3rd year was the best year of the degree. Expertise linked to enjoyment feeds in to effort, self-worth, ways of thinking. Hated lab in 3rd year. Liked choosing own courses in 4th year. Many agree. 4th year projects you know one particular thing better than anyone. Can now read papers and journals. Top of class at school, at university less so. Confidence wavers.

2:29 [Now looking at longitudinal results graphs.]

Not surprising. Dip in 3rd year not surprising. 1st year you think youre an expert because you dont know enough about it. Started enjoying it a lot less in 3rd year. By 4th year starting to learn a lot more again.

2:33 Selection effect going into PhD.

The kind of people who want to do PhDs. Self-selection.

2:35

Can only talk about how good you are with reference to other people.

2:37 Dip in 3rd year.

3rd year courses compulsory. 4th year courses are things you're interested in its psychological. You realise even after studying it for hours and hours, you're not going to do well in the exam. I lost confidence and enthusiasm in 3rd year. Liked doing projects in 4th year.

2:39

Being good in other areas besides your core focus an indicator of expertise.

2:43 Expertise in industrial research.

Different attitudes in different settings. If asked about graphs before being shown them, would have expected roughly same shape but lower absolute values. A few agree. Conversely, other would have expected them to start low and rise up. A few agree. Maps perfectly how I felt about physics. Many agree. Maps perfectly how I thought the rest of the year felt about physics.

2:45 Does degree cause strategic, non-expert-like working?

Yes. But all stages of education do this. Learning stuff to pass exams, not for the sake of it. Are aware of what expertise is, and know that they are not working like an expert.

2:47

Will have to learn a new skill set in PhD.

2:48 Where does expertise come from?

Autonomy and personal engagement. Selection effects. Many agree.

2:50 Did you answer CLASS with what you thought you should have said?

Moderate agreement. Frustration leads to answering in a particular way. You need knowledge from an undergrad degree to become an expert. But it wont make you an expert.

2:52 Disagreement from a few about whether selection effects apply.

2nd 3rd year drop seen as a counterexample. If people dropping out led to an increase in cohort expertise, would expect a low-to-high progression over degree. Physics 1A promotes expert-like thinking. Clicker questions with unexpected answers make you critically reflect on own knowledge.

2:54

Dont think I will ever become an expert.

2:55

Graduating this year and wishes felt more like an expert. Feel like Im getting there. Really slowly.

2:56

Expertise in physics seems to be a bar that is always going up and up [in later years] can anyone really become an expert in physics?

2:57

Theres a tendency to define people better than yourself as an expert and yourself not.

2:59

Standing on the shoulders of giants. Some emphasis on being research active / at the forefronts being equivalent to being an expert.

3:02

You don't need a PhD to be an expert, but if you're an expert you're likely to do a PhD.

3:0

4 Last section of exam questions (unseen material) picks out experts.

3:05

Skills do not define an expert. Experts know where to look [in research context].

3:08

End

G.0.2 Focus Group 28th April 2010

Approximate timeline from CLASS focus group, Monday 29th March 2010, 2.00pm, room 3216 JCMB.

2:09 Start

2:10 Orientation

2:12 What makes an expert?

Ability to solve a wide range of problems. Apply known skills to new problems. Wide range of subject knowledge. Apply basics to give insight into problems.

2:13 Are you experts?

Not yet. Closer than when I started. Never feel confident tackling any problem without spending time studying beforehand. Are they experts because they've seen all the problems before, or are they really applying knowledge to new problems?

2:14 What things at university have affected your expertness?

My degree has definitely influenced the way I think. Physics Skills focuses on problems. With some exams there's a tendency to drill for them. Progressive difficulty of exams. Feel more confident before you begin your revision.

2:16

When you go to courses and try really hard and still struggle. Lose expertness when you cram for exams. You're not learning it for the fun of it, you're learning to do the exams. Really like the subject when not having to do exams.

2:18 Leaving aside factual recall / knowledge, how do your problem solving skills compare to experts?

Skills have improved, but not to level of expert. Experts have more exposure to problems. Understand it so well they can teach it.

2:20 Levels of natural ability.

Maxwell's equations and deriving the speed of light: Feels like the same level of revelation as for all physicists.

2:21

Experts have broader view of physics as a whole.

2:22 Self-identification as physicists.

Feels on the outside of the physics community. On the cusp of it. Veering more towards feeling like a physicist. Research projects feels more like doing physics work. Mathematical Physics hated group project, did not feel like a physicist.

2:23 Research methods.

Hated it at the time but on reflection was really useful. Can now read research papers critically. Don't have to be taught now, can go and find out for ourselves. Important trait of experts: don't have to ask someone for help, can be self-reliant.

2:25 What do you expect expertness graphs to look like?

Have seen before, but naively would have expected an increase.

2:26 [Now looking at graphs.]

2:28

Surprised that 3rd year is so low. Increase from 3rd year to PhD not surprising. Early drop-outs less likely to think like an expert. 3rd year is a hard year: Getting jaded with exam after exam. Around 2nd to 3rd year you figure out how to pass exams strategic approach.

2:30

Does shape of graph reflect your own experience? No. Expected to be trained and learn a scientific mindset doesnt seem to have really happened. Felt more confident in 3rd year but tailed off in 4th year.

2:33 Selection effects?

Dont know. Popular perspective is that physics is a hard degree. You could get through a physics degree without feeling like an expert just by doing problems.

2:34 Are there people in your year who are experts?

Yes. They have the mindset / are more interested. Expertness linked to personal interest.

2:36

Some people have a natural ability, which is not the same as being an expert. Perceive a couple of lecturers as not necessarily being experts. Depends on whether they approach teaching in the same way as their research. Depends if they actually care about teaching. Would class majority of academics as experts / some disagreement.

2:38 What about physicists in industry?

Not really. Maybe. More applied expertise? Not really comparable.

2:40

Physics as a discipline is about finding something new. Original thinking? Yes / no. Understanding would be enough to be classed as an expert. When you understand something you get a buzz out of conveying it. If you teach something well you must understand it well, so must be an expert.

2:43

Maybe the best teachers are the ones who struggled the most themselves so have empathy with their students difficulties. Sometimes the loftiest professors are hardest to understand.

2:45 Long term plans?

Undecided. Maybe PhD. Dont know what area. Maybe academia. Still enjoying it, but potential competition is off-putting.

2:46 Will you ever become experts?

Probably, maybe after 20 years. Only been taking physics seriously for about 2 years.

2:48

PhD will likely make or break you as an expert.

2:49 What makes an expert / where do they come from?

Personal qualities / natural ability. Hard work. Self-driven alteration / improvement of personal approach.

2:51

Know how experts think but dont necessarily think like that. A bit hazy on why this is. Non-committal. Suggest need more things like Physics Skills, things where you need a process rather than memorised information. Has had preconceptions challenged: sometimes lecturers struggle too, maybe only perceive them as experts because theyve put the work in?

2:55

Degree process is filtering out the non-physics-enthusiast. People who like physics stay in it largely they are the ones who are good at it. Maybe a physics degree is not conducive to bringing out someones interest. Doing exams: Thinking this is not why I wanted to do physics.

2:57

Some people dont do physics degree for the physics. Intending to go into banking etc.

2:58

Experts not necessarily experts in all areas. But general approach should apply to all areas: Transcends specific disciplines.

2:59 End.

Appendix H

Perceptions of School Practical Work Survey

On the following two pages are a copy of the Perceptions of School Practical Work Survey given to first year students at the University of Edinburgh in November 2008.

Attitudes to Practical Physics Work At School

Practical work - in the form of laboratory experiments - forms a major component of your studies of physics at university. In this survey, we are interested in learning something of your experiences and views of the practical side of the subject from your time at school. Completing this will be invaluable in insuring that future laboratory classes are as useful and educational as possible.

Your responses will be treated confidentially.

Your experiences of practical physics.

Q1 Here is a way to describe a racing car.

quick <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> slow important <input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> unimportant safe <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> danger	The position of the cross in the box between the word pairs show that you consider it as very quick, slightly more important than unimportant and quite dangerous.
---	--

Use the same method to answer this question.

What are your opinions about your experiences of your practical work in physics at school?
(Cross ONE box on each line.)

Useful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Useless
Not helpful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Helpful
Undstandable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Not understandable
Satisfying	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Not satisfying
Boring	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Interesting
Well organised	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Not well organised
The best part of physics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The worst part of physics
Not enjoyable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Enjoyable

Q2. Think about your experiences in practical physics work.

(Cross the box which best reflects your opinion).

	<i>Strongly Agree</i>	<i>Agree</i>	<i>Neutral</i>	<i>Disagree</i>	<i>Strongly Disagree</i>
(a) I prefer to have written instructions for experiments.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(b) Practical work helps my understanding of physics topics.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(c) Discussions in the practical work enhance my understanding of the subject.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(d) I had few opportunities to plan my experiments during the laboratory work.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(e) I felt confident in carrying out the experiments in physics.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(f) The experimental procedure was clearly explained in the instructions given.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(g) I was so confused in the laboratory that I ended up following the instructions without understanding what I was doing.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(h) There was good linkage between experiments and the relevant theory.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Q3. Imagine you have been asked to run the experiments for the final year of your school's physics class.

What changes might you introduce in order to make the practical experience more helpful so that future students can be helped to understand and apply the ideas of physics.

PTO

Q4. Here are several reasons why experimental work is an integral part of a physics course.

- A. Physics is a practical subject. E. New discoveries are made by means of experiments.
 B. Experiments illustrate theory for me. F. Experimental skills can be gained in the laboratory.
 C. Experimental work allows me to test out ideas. G. Experimental work allows me to think about physics.
 D. Experiments assist me to plan and organise. H. Experimental work makes physics more enjoyable for me.

Pick the THREE which you consider to be integral and rank them in descending order of importance in the boxes below.

First	Second	Third
<input type="text"/>	<input type="text"/>	<input type="text"/>

If you can think of another reason why experimental work is integral please specify in the box below.

Q5. In what ways do you think university practical work will differ this year from the practical work you experienced at school? (Choose up to three answers)

- Use of more complicated equipment I will do the experiments myself instead of watching them being done
 Use of modern equipment I will have more time for each experiment
 I will get less guidance than at school I will have more choice in the experiments I do

Tell us about yourself.

Thank you for completing the first five questions in this survey. I'm glad you've made it this far. Not long to go now! To help us put your information into context, we'd appreciate it if you could tell us a bit about yourself. This will let us see if there are any commonalities between men and women, which degrees you are studying and so on.

Q6. Are you looking forward to the practical work this year?

Very much Not at all

Q7. What is your gender?

Female Male

Q8 What degree do you intend to study at university?

- Physics Mathematical Physics
 Astrophysics Computational Physics
 Other (please specify)

Q9. What is the highest level you have studied physics at school?

- Higher Advanced Higher A-Level Other (Please specify)

Q10. Which of the following would best describe your practical physics experience at school?

- I did most of my experiments myself, either alone or in groups
 Most of my experiments were done as computer-based simulations
 My teacher carried out most of my experiments as demonstrations.

Thank you for your help.

Appendix I

Perceptions of University Practical Work Survey

On the following two pages are a copy of the Attitudes to University Practical Work Survey given to first year students at the University of Edinburgh in March 2009. The test was sent to the University of Edinburgh in paper format and no copies of the test remain. The version included here was designed for use at the University of Glasgow. It differs only in Questions 2, where the final two options were not included, and 7 where the degree disciplines were corrected to be those relevant to the University of Edinburgh.

Attitudes to Laboratory Physics In First Year at Glasgow

When you began your physics course we asked you about your experiences of the practical side of physics at school. Now, as you complete your first year of studying physics at university, we'd like to know what you've thought of the laboratory component of the Physics 1 course. Some of these questions are the same as those we asked in the first survey - this allows us to see how your attitudes to practical physics have changed. As before, your responses will be treated confidentially and will be of great value to us.

Here is a way to describe a racing car.

quick	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	slow
important	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unimportant
safe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	danger

The position of the cross in the box between the word pairs show that you consider it as very quick, slightly more important than unimportant and quite dangerous.

Use the same method to answer question 1.

Q1. What are your opinions about your present university laboratory experiences in physics?
(Cross ONE box on each line)

Useful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Useless
Not helpful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Helpful
Understandable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Not understandable
Satisfying	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Not satisfying
Boring	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Interesting
Well organised	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Not well organised
The best part of physics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	The worst part of physics
Not enjoyable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Enjoyable

Q2. Think about your experiences in laboratory work in physics.
(Cross the box which best reflects your opinion).

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
(a) I prefer to have written instructions for experiments.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(b) I was unsure about what was expected of me in writing up my experiments.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(c) Laboratory work helps my understanding of physics topics.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(d) Discussions in the laboratory enhance my understanding of the subject.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(e) I only understood the experiment when I started to write about it afterwards.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(f) I had few opportunities to plan my experiments during the laboratory work.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(g) I felt confident in carrying out the experiments in physics.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(h) The experimental procedure was clearly explained in the instructions given.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(i) I was so confused in the laboratory that I ended up following the instructions without understanding what I was doing.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(j) There was good linkage between experiments and the relevant theory.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(k) The amount of time given for each experiment was sufficient	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(l) The demonstrators provided valuable assistance with my work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(m) I found the inclusion of a break beneficial.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(n) Attempting the tutorial questions before the lab was very helpful to perform the experiment.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Q3. Here are several reasons why laboratory work is an integral part of a physics course.

- | | |
|---|--|
| A. Physics is a practical subject | E. New discoveries are made by means of |
| B. Experiments illustrate theory for me | F. Experimental skills can be gained in the laboratory |
| C. Laboratory work allows me to test out ideas | G. Experimental work allows me to think about Physics |
| D. Experiments assist me to planning and organise | H. Experimental work makes Physics more enjoyable for me |

Pick the THREE which you consider to be integral and rank them in descending order of importance in the boxes below.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------

PTO

2330133443

Q4. Think back over the experiments which you have completed during this year.

(a) Which experiment did you find most useful or enjoyable ?

.....

(b) What was it about that experiment that made it particularly useful or enjoyable ?

.....

(c) Did you find the experiment easy or challenging ?

.....

(d) What did it teach you ?

.....

(e) List any skills which improved as a result of doing the experiment.

.....

Q5. Imagine you have been asked to run the laboratories for your year group next session. What changes might you introduce in order to make the laboratory experience more helpful so that future students can be helped to understand and apply the ideas of physics.

Tell us about yourself.

Thank you for completing the first five questions in this survey. I'm glad you've made it this far. Not long to go now! To help us put your information into context, please tell us a bit about yourself. This will let us see if there are any commonalities between male and female students, or which degrees you are studying.

Q6. What is your gender?

Female Male

Q7 What degree do you intend to study at university?

Physics Physics and Astronomy Physics and Astrophysics
Physics and (Applied) Maths Astronomy and (Applied) Maths Chemical Physics
Other (Please specify)

Q8. Has your intended degree changed during the course of 1st year? And if it has, did your experiences in the lab influence your decision?

- Degree choice not changed
- Degree choice has changed – my experience in the labs had nothing to do with the decision.
- Degree choice has changed – my experience in the labs was so good I want to keep taking physics.
- Degree choice has changed – my experience in the labs has put me off physics for good.

Thank you for your help.

Bibliography

- [1] S.W. Hawking. *A brief history of time: A reader's companion*. Bantam Dell Pub Group, 1992.
- [2] J. Gleick. *Genius: The life and Science of Richard Feynman*. Vintage, 1993.
- [3] C. Beckett. *Thatcher - 20 Prime Minsisters of the 20th Century*. Haus Publishing, 2006.
- [4] V. Peterson. *J.K. Rowling: Extraordinary Author*. Russell Sage Foundation, 2001.
- [5] S.C. Barrie. A research based approach to generic graduate attributes policy. *Higher Education Research and Development*, 23(3):261–275, 2004.
- [6] R. Bridgstock. The graduate attributes we've overlooked: enhancing graduate employability through career management skills. *Higher Education Research and Development*, 28(1):31–44, 2009.
- [7] G. Brunello and L. Cappellari. The labour market effects of alma mater: Evidence from Italy. *Economics of Education Review*, 27:564574, 2008.
- [8] J.E.Rosenbaum. *Beyond college for all: Career paths for the forgotten half*. Russell Sage Foundation, 2001.
- [9] L.W.Perna. The benefits of higher education: Sex, racial/ethnic and socioeconomic group differences. *Review of Higher Education-The Johns Hopkins University Press*, 29(1):23–52, 2005.
- [10] J. Lechat. University Influence on Regional Development: Example of a new creation-the Louvain Catholic University at Louvain-La-Neuve. *European Journal of Education*, 14(3):239–250, 1979.
- [11] H. Crabtree, C. Roberts, and C. Tyler. Understanding the problems of transition into higher education. *Education in a Changing Environment - Conference Proceedings*, pages 339–351, 2007.
- [12] A. Cook, B.S. Rushton, and K.A. Macintosh. *Student Transition and Retention (STAR)*. University of Ulster, Coleraine, Northern Ireland, 2006.
- [13] J.C. Blickenstaff. Women and science careers: leaky pipeline or gender filter? *Gender and Education*, 17(4):369–386, 2005.

-
- [14] M. Yorke. *Leaving early: undergraduate non-completion in Higher Education*. Falmer, 1999.
- [15] V. Johnstone. Improving student retention by accident or design? *Exchange Magazine*, 1:9–11, 2002.
- [16] L. Harvey and S. Drew with M. Smith. The first-year experience: a review of literature for the Higher Education Academy. *The Higher Education Academy - Report*, pages 1–181, 2006.
- [17] M. Birch and N.R. Walet. How do we identify ‘at risk’ students upon entry to university? *GIREP-EPEC and PHEC Internal Conference - Selected Contributions*, pages 185–195, 2009.
- [18] M.M. Casey and S. McVitie. The impact of low attendance on the success and progression of level 1 undergraduates: Analysis and intervention. *GIREP-EPEC and PHEC Internal Conference - Selected Contributions*, pages 196–201, 2009.
- [19] A. Harrison. A-level passes rise amid university places pressure. *BBC News*, 2011. <http://www.bbc.co.uk/news/education-14558490> Page Retrieved 19/05/2012.
- [20] B. Goldacre. Are exams getting easier? Nobody knows. *The Guardian*, 2011. <http://www.guardian.co.uk/commentisfree/2010/aug/21/ben-goldacre-bad-science-exams> Page Retrieved 19/05/2012.
- [21] K Birbalsingh and C. Keates. A-level results: are exams getting easier? *The Telegraph*, 2011. <http://www.telegraph.co.uk/education/educationnews/8708052/A-level-results-are-exams-getting-easier.html> Page Retrieved 19/05/2012.
- [22] D.P. Ausubel. *Educational psychology: A cognitive view*. San Fransicso: Holt, Rinehart and Winston, 1968.
- [23] R.C. Rist. Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review (Education, General Education, Learning and Instruction and Education and Society)*, 40(3):411–451, 1970.
- [24] J.E. Brophy. Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, 75(5):631–66, 1983.
- [25] D. Lawson. What can we expect from A-level mathematics students? *Teaching Mathematics Application*, 16(4):151–156, 1997.
- [26] Engineering Council. Measuring the mathematics problem. *Engineering Council Report*, 1996. <http://www.engc.org.uk/ecukdocuments/internet/document/%20library/Measuring%20the%20Mathematic%20Problems.pdf> Page Retrieved 19/05/2012.

- [27] M.R. Cohen and E. Nagel. *An Introduction to Logic and Scientific Method*. Simon Publications, 2002.
- [28] A. E. Lawson. The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15:11–24, 1978.
- [29] V. P. Coletta and J. A. Phillips. Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics*, 73(12):1172–1182, 2005.
- [30] M. Niaz. The role of cognitive style and its influence on proportional reasoning. *Journal of Research in Science Teaching*, 26(3):221–235, 1989.
- [31] M. Wallace. *Private Communication in relation to PhD research*. PhD thesis, 2012.
- [32] M.C. Linn and N.B. Songer. Cognitive and conceptual change in adolescence. *American Journal of Education*, 99:379–417, 1991.
- [33] J S. Eccles and J E. Jacobs. Social forces shape math attitudes and performance. *Journal of Women in Culture and Society*, 11:367–380, 1986.
- [34] W.K. Adams, K.K. Perkins, and N.S. Podolefsky. A new instrument for measuring student beliefs about physics and learning physics the Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics-Physics Education Research*, 2:010101, 2006.
- [35] M. Besterfiel-Sacre, C.J. Atman, and L.J. Shuman. Engineering student attitudes assessment. *Journal of Engineering Education*, April:133–141, 1998.
- [36] S.R. Sankaran, D. Sankaran, and T.X. Bui. Effect of student attitude to course format on learning performance: An empirical study in web vs. lecture instruction. *Journal of Instructional Psychology*, 27(1):62980730, 2000.
- [37] R.M. Felder, G.N. Felder, M. Mauney, C.E. Hamrin, and E.J. Dietz. A longitudinal study of engineering student performance and retention. iii. gender differences in student performance and attitudes. *Journal of Engineering Education*, 84(2):151–163, 1995.
- [38] J. Osborne, S. Simon, and S. Collins. Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9):1049–1079, 2003.
- [39] P.L. Gardner. Attitudes to Science. *Studies in Science Education*, 2:1–41, 1975.
- [40] N. G. Lederman, F. Abd-El-Khalick, R. L. Bell, and R. S. Schwartz. Views of Nature of Science questionnaire: Toward valid and meaningful assessment of learners conceptions of nature of science. *J. Res. Sci. Teach*, 39:497, 2002.
- [41] I. Halloun and D. Hestenes. Interpreting VASS dimensions and profiles for physics students. *Science Education*, 7:553–577, 1998.

- [42] E. F. Redish, R. N. Steinberg, and J. M. Saul. Student expectations in introductory physics. *American Journal of Physics*, 66:212–224, 1998.
- [43] E.F Redish. The distribution and change of student expectations in introductory physics. *E.F. Redish homepage University of Maryland*, 1996. <http://www.physics.umd.edu/perg/papers/redish/icupeex.htm> Page retrieved 19/05/2012.
- [44] H. Alhadlaq, F. Alshaya, S. Alabdulkareem, K. K. Perkins, W. K. Adams, and C.E. Wieman. Measuring students beliefs about physics in Saudi Arabia. *2009 Physics Education Research Conference Proceedings*, 1179:69–72, 2009.
- [45] E. Brewster, L. Kramer, and G. O’Brien. Modeling instruction: Positive attitudinal shifts in introductory physics measured with CLASS. *Physical Review Special Topics-Physics Education Research*, 5:013102, 2009.
- [46] E.F. Redish and D. Hammer. Reinventing college physics for biologists: Explicating an epistemological curriculum. *American Journal of Physics*, 77:629–642, 2009.
- [47] V.K. Otero and K.E. Gray. Attitudinal gains across multiple universities using the physics and everyday thinking curriculum. *Physical Review Special Topics-Physics Education Research*, 4, 2008.
- [48] V.K. Otero and K.E. Gray. Learning to think like scientists with the PET curriculum. *Physics Education Research Conference 2007*, 951:160–163, 2007.
- [49] C. Bally and N.D. Finkelstein. Student perspectives in quantum physics. *Physics Education Research Conference, Edmonton*, 1064:67–70, 2008.
- [50] C. Anders and R. Berg. Factors related to observed attitude change toward learning chemistry among university students. *Chemistry Education: Research and Practice*, 6:1–18, 2005.
- [51] W.G. Perry. *Forms of intellectual and ethical development in the college years: A scheme*. Holt, Rinehart and Winston, 1970.
- [52] L.E. Kost, S.J. Pollock, and N.D.Finklestein. Characterizing the gender gap in introductory physics. *Physical Review Special Topics-Physics Education Research*, 5:010101, 2009.
- [53] E. Gire, B. Jones, and E. Price. Characterizing the epistemological development of physics majors. *Physical Review Special Topics-Physics Education Research*, 5:010103, 2009.
- [54] K.K Perkins and M. Gratny. Who becomes a physics major? A long-term longitudinal study examining the roles of pre-college beliefs about physics and learning physics, interest, and academic achievement. *2010 PERC Conference Proceedings*, 1289:253–256, 2010.

- [55] Institute of Physics. The physics degree - graduate skills base and the core of physics. *Institute of Physics*, 2010. http://www.iop.org/education/higher_education/accreditation/file_43311.pdf Page Retrieved 19/05/2012.
- [56] Institute of Physics. First destination statistics. *Institute of Physics website*, 2008. http://www.iop.org/policy/statistics/education/page_43191.html.
- [57] A. Richardson, M. Sharma, and J. Khachan. What are students learning in practicals? a cross study in university physics laboratories. *CAL-laborate International*, 16:20–27, 2008.
- [58] American Association of Physics Teachers. Goals of the introductory physics laboratory. *Physics Teacher*, 35:5468, 1997.
- [59] OFQUAL. Review of standards in physics, GCSE 2002-2007, GCE 2001-2007. *OFQUAL Report*, 2009. <http://webarchive.nationalarchives.gov.uk/+http://www.ofqual.gov.uk/2231.aspx> Page Retrieved 19/05/2012.
- [60] D. Herrington and M.B. Nakleh. What defines effective chemistry laboratory instruction? teaching assistant and student perspectives. *Journal of Chemical Education*, 80:1197, 2003.
- [61] K. Fox. Authentic alternatives to practical work. *School Science Review*, 322:45–52, 2006.
- [62] A. H. Johnstone R. J. Sleet and J. F. Vianna. An information processing model of learning: its application to an undergraduate laboratory course in chemistry. *Studies in Higher Education*, 19:77–88, 1989.
- [63] National Research Council (USA). National Science Education Standards. *National Academy Press*, 1996.
- [64] A.B. Arons. Guiding insight and inquiry in the introductory physics laboratory. *Physics Teacher*, 31:278–282, 1993.
- [65] R. Trumper. The physics laboratory: a historical overview and future perspectives. *Science Education*, 12:645–670, 2003.
- [66] D. Hestenes, M. Wells, and G. Swakhamer. Force Concept Inventory. *Physics Teacher*, 30(3):141–158, 1992.
- [67] L. Ding, R. Chabay, B. Sherwood, and R. Beichner. Evaluating an electricity and magnetism assessment tool: Brief Electricity and Magnetism Assessment. *Physical Review Specialst Topics - Physics Education Research*, 2:010105, 2006.
- [68] C. Murphy and J.Beggs. Children’s perceptions of school science. *School Science Review*, 84(308):109–116, 2003.
- [69] W. Speering and L. Rennie. Students’ perceptions about science: The impact of transition from primary to secondary school. *Research in Science Education*, 26(3):283–298, 1996.

- [70] L.D. Shulman and P. Tamir. Research of teaching the natural sciences (in R.M.W. Tavers Second Handbook of research on teaching). *Chicago Rand McNally*, pages 1098–1140, 1973.
- [71] M Hanif, P.H. Sneddon, and N. Reid. The perceptions, view and opinions of university students about physics learning during undergraduate laboratory work. *European Journal of Physics*, 30:85–96, 2008.
- [72] The Russell Group. *The The Russell Group website*, 2011. <http://www.russellgroup.ac.uk/PageRetrieved19/05/2012>.
- [73] The Times Higher Education - The World University Rankings 2011-2012. *The Times Higher Education*, 2011. <http://www.timeshighereducation.co.uk/world-university-rankings/2011-2012/top-400.html> Page Retrieved 19/05/2012.
- [74] QS World University Rankings 2011/12. *QS World University Rankings website*, 2011. <http://www.topuniversities.com/university-rankings/world-university-rankings/2011>Page Retrieved 19/05/2012.
- [75] University of Edinburgh. University of Edinburgh factsheet. *The University of Edinburgh website*, 2011. <http://www.ed.ac.uk/schools-departments/governance-strategic-planning/facts-and-figures/university-factsheet/> Page Retrieved 19/05/2012.
- [76] S.P. Bates. Reshaping large-class undergraduate science courses: the weekly workshop. *CAL-laborate, UniServe Science International Newsletter*, 14, 2005. http://science.uniserve.edu.au/pubs/callab/vol14/cal14_bates.pdf Page Retrieved 19/05/2012.
- [77] P. Denny, J. Hamer, A. Luxton-Reilly, and H. Purchase. PeerWise: Students sharing their multiple choice questions. *ICER '08 Proceeding of the Fourth International Workshop on Computing Education Research*, pages 51–58, 2008.
- [78] E. Mazur. *Peer Instruction: A User's Manual*. Series in Educational Innovation (Prentice Hall, Upper Saddle River, NJ), 1997.
- [79] W.K. Adams, K.K. Perkins, M.Dubson, and C.E Wieman. The Colorado Learning Attitudes about Science Survey. *Proceedings of the 2004 Physics Education Research Conference*, AIP Conf. Proc. No. 790, 2005.
- [80] W.K. Adams. Personal communication. 2011.
- [81] K. Bartel Sheehan. Online research methodology: Reflections and speculations. *Journal of Interactive Advertising*, 3(1):56–61, 2002.
- [82] K.A. Slaughter, S.P. Bates, and R.K. Galloway. The changes in attitudes and beliefs of first year physics undergraduates: A study using the CLASS survey. *International Journal of Innovation in Science and Mathematics Education*, 2011.

- [83] W.K. Adams, C.E. Weiman, and K.K. Perkins. Modifying and validating the Colorado Learning Attitudes about Science Survey for use in chemistry. *Journal of Chemical Education*, 85(10):1435, 2008.
- [84] K. Semsar, J.K. Knight, G. Birol, and M.K. Smith. The Colorado Learning Attitudes about Science Survey (CLASS) for use in Biology. *CBE-Life Sciences Education*, 10:268–278, 2011.
- [85] R. Cudeck. *Exploratory Factor Analysis Chapter from Handbook of applied multivariate statistics and mathematical modeling*. San Diego, California USA: Academic Press, 2000.
- [86] K.E Gray, W.K. Adams, C.E. Wieman, and K.K. Perkins. Students know what physicists believe, but they dont agree: A study using the CLASS survey. *Physical Review Special Topics-Physics Education Research*, 4:020106, 2008.
- [87] A. Field. *Discovering Statistics using SPSS*. Sage, 2009.
- [88] G.D. Ruxton. The unequal variance t-test is an underused alternative to student’s t-test and the Mann-Whitney U test. *Behavioral Ecology*, 17(4), 2006.
- [89] S. Wuttiprom, M.D. Sharma, I.D. Johnston, R. Chitaree, and C. Soankwan. Development and use of a conceptual survey in introductory quantum physics. *International Journal of Science Education*, 13(5):631–654, 2009.
- [90] R. Chabay and B. Sherwood. Qualitative understanding and retention. *AAPT Announcer*, 27:96, 1997.
- [91] D. Doran. *Basic Measurement and Evaluation of Science Instruction*. NSTA, Washington DC, 1980.
- [92] P. Kline. *A Handbook of Test Construction: Introduction to psychometric testing*. Methuen, London, 1986.
- [93] 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.
- [94] A. F. Heckler and E.C. Sayre. What happens between pre- and post-tests: Multiple measurements of student understanding during an introductory physics course. *American Journal of Physics*, 78(7):768–777, 2010.
- [95] Joint Council for Qualifications. Joint council for qualifications, national results by subject 01-09. *Joint Council for Qualifications Website*.
- [96] London Mathematical Society. Measuring the mathematics problem. *London Mathematical Society Report*, page 2, 1999.
- [97] N. Frederiksen. The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39(3):193–202, 1984.

- [98] M.L. Smith. Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5):8–11, 1991.
- [99] Warwick Mansell. Are teachers choosing ‘easy’ exam boards? *Newspaper Article, The Guardian*, 2009. <http://www.guardian.co.uk/education/2009/aug/25/teachers-choosing-exam-boards-gcse> Page Retrieved 19/05/2012.
- [100] The University of York. Welcome to Salters Horners Advanced Physics (SHAP). *Departmental Website*, 2011. <http://www.york.ac.uk/org/seg/salters/physics/index.html> Page Retrieved 19/05/2012.
- [101] L. Bao, T. Cai, and K. Koenig. Learning and scientific reasoning. *Science*, 323:586–587, 2009.
- [102] D. Kuhn and D. Dean. Scientific reasoning - Where are we now? *Journal of Cognitive Development*, 23:431–434, 2008.
- [103] E.F. Redish. *Teaching Physics with the Physics Suite*. Wiley, 2003.
- [104] A. Miyake, L.E. Kost-Smith, N.D. Finkelstein, S. J. Pollock, G. L. Cohen, and T. A. Ito. Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330:1234–1237, 2010.
- [105] A. Barrantes, A. Pawl, and D. Pritchard. What do seniors remember from freshman physics? *2009 PERC Conference Proceedings*, 1179:47–50, 2009.
- [106] Quantification in the life sciences 1. *The University of Edinburgh Website*, 2011. http://student.biology.ed.ac.uk/undergraduate/courses/course_details.php?euccid=BILG08008 Page Retrieved 19/05/2012.
- [107] UCAS. Ucas tariff tables. *UCAS Website*, page Retrieved 11/05/2012, 2011. http://www.ucas.com/students/ucas_tariff/tarifftables/ Page Retrieved 19/05/2012.
- [108] P. Oliver, G. Marwell, and R. Teixeira. A theory of the critical mass. i. interdependence, group heterogeneity, and the production of collective action. *American Journal of Sociology*, 91(3):522–556, 1985.
- [109] A. Powell, B.M. Bagilhole, and A.R.J. Dainty. The problem of women’s assimilation into uk engineering cultures: can critical mass work? *Equal Opportunities International*, 25(8):688–699, 2006.
- [110] P. H. Sneddon, K.A. Slaughter, and N. Reid. Perceptions, views and opinions of university students about learning physics during practical work at school. *European Journal of Physics*, 30(5):1119–1129, 2009.
- [111] J. Day and D. Bonn. Development of the concise data processing assessment. *Physical Review Specialst Topics - Physics Education Research*, 7:010114, 2011.

- [112] R.K. Galloway, S.P. Bates, H.E. Maynard-Casely, K.A. Slaughter, and H. Singer. Data handling diagnostic: Design, validation and preliminary findings. *Physical Review Special Topics-Physics Education Research*, Submitted, 2012.
- [113] SQA. Scottish higher updates. *SQA Website*, 2012. <http://www.sqa.org.uk/sqa/42175.html> Page Retrieved 19/05/2012.
- [114] Higher Education Academy. Understanding the uk mathematics curriculum pre-higher education. *Higher Education Academy Report*, page Retrieved 11/05/2012, 2010. <http://www.heacademy.ac.uk/physsci/publications/miscellaneous#maths>.
- [115] C. H. Crouch and E. Mazur. Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69(9):970–977, 2001.
- [116] E. Mazur. Confessions of a converted lecturer. *YouTube video of lecture*, 13-16minutes, 2009. <http://www.youtube.com/watch?v=Wws1BPj8GgI> Page Retrieved 19/05/2012.
- [117] R. Donnelly. Private communication in relation to PhD research. 2012.

Publications

Galloway, R.K., Bates, S.P., Maynard-Casely, H.E., **Slaughter, K.A.** and Singer, H. **2012** Data Handling Diagnostic: Design, validation and preliminary findings. Submitted to Physical Reviews-Specialist Topics (PER)

Bates, S.P., Galloway, R.K., Lopston, C and **Slaughter, K.A.** **2011** How attitudes and beliefs about physics change from high school to faculty: an investigation with CLASS. Physical Reviews-Specialist Topics (PER) 7 020114

Slaughter, K.A., Bates, S.P. and Galloway, R.K. **2011** The changes in attitudes and beliefs of first year physics undergraduates: a study using the CLASS survey. International Journal of Innovation in Science and Mathematics 19(1) 29-42

Sneddon, P.H., **Slaughter, K.A.** and Reid, N **2009** Perceptions, views and opinions of university students about physics learning during practical work at school EUROPEAN JOURNAL OF PHYSICS 30(5), 1119-1129

Conference Publications

Slaughter, K.A., Bates, S.P. and Galloway, R.K. **2011** A longitudinal study of the development of attitudes and beliefs towards Physics. Physics Education Research Conference (PERC) Omaha Nebraska.

Bates, S.P. and **Slaughter, K.A.** **2010** Attitudes Towards Science Study: CLASS at the University of Edinburgh. GIREP 2009 Conference Proceedings 176-184.