Problem-Solving in Undergraduate Mathematics and Computer Aided Assessment

by

Matthew Badger

A thesis submitted to
The University of Birmingham
for the degree of
Doctor of Philosophy

School of Mathematics
The University of Birmingham
February 2013

Abstract

Problem solving is an important skill for students of the mathematical sciences, but traditional methods of directed learning often fail to teach students how to solve problems independently. To compound the issue, assessing problem-solving skills with computers is extremely difficult. In this thesis we investigate teaching by problem solving and introducing aspects of problem solving in computer aided assessment.

In the first part of this thesis we discuss problem solving and problem-based pedagogies. This leads us, in the second part, to a discussion of the Moore Method, a method of enquiry-based learning. We demonstrate that a Moore Method course in the School of Mathematics at the University of Birmingham has helped students' performance in certain other courses in the School, and record the experiences of teachers new to the Moore Method at another U.K. university.

The final part of this thesis considers word questions, in particular those involving systems of equations. The work discussed here has allowed the implementation of a range of questions in the computer-aided assessment software STACK. While the programmatic aspects of this work have been completed, the study of this implementation is ongoing.

Acknowledgements

First and foremost I would like to thank my supervisor, Dr. Christopher Sangwin, for the considerable amount of work that he has put in in assisting me with both the work which comprises this thesis, and the thesis itself. I have also been ably assisted by my co-supervisor, Dr. Dirk Hermans, whose input at points has been invaluable. Both Dr. Sangwin and Dr. Hermans join a group of teachers and lecturers without whose guidance I would not have found myself in this position.

I owe a considerable debt of gratitude to both my examiners for their work in assessing the original version of this thesis, and the valuable suggestions that came from that.

Finally, my friends and family have been a great support in helping me through the extended process of pursuing postgraduate education and writing a Ph.D. thesis. I am particularly grateful for the encouragement of those who have been here before: Tom, Angela, James, and Katherine.

Contents

1 Introduction 1

I Problem-based Learning 3

2 Problem-solving 4
2.1 Thinking Mathematically . . . 5
2.2 Problem-solving . . . 12
2.2.1 Teaching Problem-solving . . . 18

3 Problem-based Learning 22
3.1 Directed Learning . . . 22
3.1.1 Lectures . . . 23
3.1.2 Example Classes . . . 25
3.1.3 The structure of directed learning . . . 27
3.2 Problem-based Learning . . . 30
3.2.1 Arguments for and against PBL . . . 34

4 Problem-based Learning in England and Wales 36
4.1 The HE Mathematics Curriculum Summit and the Problem-solving Project . . . 36
4.2 The Questionnaire . . . 40
4.2.1 Summary of Questionnaire Responses . . . 41
4.3 Case Study Methodology . . . 44
4.4 Approaches to Teaching Problem-solving . . . 45
4.5 Case Studies . . . 48
4.6 Durham University . . . 48
4.6.1 Teaching . . . 49
4.6.2 Assessment . . . 52
4.6.3 Motivations and Development . . . 52
4.6.4 Problem Selection . . . 53
4.6.5 Outcomes and the Future . . . 56
4.7 University of Manchester . . . 56
4.7.1 Teaching . . . 57
4.7.2 Assessment . . . 59
4.7.3 Motivations and Development . . . 60

4.7.4 Problem Selection . . . 61
4.7.5 Outcomes and the Future . . . 63
4.8 Queen Mary, University of London . . . 63
4.8.1 Teaching . . . 64
4.8.2 Assessment . . . 66
4.8.3 Motivations and Development . . . 67
4.8.4 Problem Selection . . . 67
4.8.5 Outcomes and the Future . . . 68
4.9 University of Warwick . . . 69
4.9.1 Teaching . . . 70
4.9.2 Assessment . . . 72
4.9.3 Motivations and Development . . . 73
4.9.4 Problem Selection . . . 75
4.9.5 Outcomes and the Future . . . 76

II The Moore Method 78

5 The Moore Method 79

5.1 Robert Lee Moore and his Method . . . 79
5.1.1 Moore's Moore Method . . . 80
5.2 The Moore Method Today . . . 85
5.2.1 Course Material . . . 86
5.2.2 Presentations and Written Solutions . . . 89
5.2.3 Use in Teaching . . . 91
5.2.4 Criticisms and Limitations of the Method . . . 92
5.3 Relationship to other PBL, and Research on the Method . . . 94

6 Determining the Effectiveness of the Moore Method at Improving Mathematics Performance 98

6.1 Introduction . . . 98
6.2 The Module . . . 99
6.2.1 Teaching and Examination . . . 100
6.2.2 Self Selection . . . 102
6.3 Methodology . . . 103
6.3.1 The Data . . . 103
6.3.2 The Method . . . 105
6.4 Results . . . 108
6.4.1 First Year Results . . . 108
6.4.2 Second year analysis . . . 112
6.4.3 Third Year and Mean Results . . . 119
6.5 Conclusion . . . 120

7 Approaches to the Moore Method: Staff Experiences with a Moore Module 125

7.1 The Module . . . 126
7.2 Methodology . . . 129
7.3 Analysis of Results . . . 132
7.3.1 Module Planning and Motivation . . . 134
7.3.2 Choices of Problems . . . 140
7.3.3 Events During Class . . . 144
7.3.4 Benefits to Students . . . 149
7.3.5 Benefits to Staff . . . 152
7.3.6 Difficulties and Drawbacks . . . 154
7.3.7 Changes to the Module . . . 157
7.3.8 Reflections of Staff . . . 158
7.4 Conclusions . . . 160

III Computer Aided Assessment 163

8 The solution and comparison of equations 164

8.1 STACK and Computer-Aided Assessment . . . 164
8.2 Validity and Correctness . . . 166
8.2.1 Validity . . . 167
8.2.2 Correctness . . . 169
8.3 Equivalence of Equations . . . 170
8.4 Systems of Equations and their Varieties . . . 172
8.5 Gröbner Bases . . . 175
8.5.1 Basic Theory . . . 175
8.5.2 Hilbert Basis Theorem . . . 180
8.5.3 Systems of Linear Equations . . . 182
8.5.4 Systems of Univariate Polynomials . . . 184
8.6 Systems of Multivariate Polynomials . . . 192
8.6.1 Term Orders . . . 192
8.6.2 A Division Algorithm . . . 195
8.6.3 Gröbner Bases . . . 198
8.6.4 Comparing Systems of Equations . . . 202
8.6.5 Limitations of the System . . . 204
8.7 Conclusion . . . 205

9 Applications of Gröbner Bases 206

9.1 Potential Response Trees . . . 206
9.2 Using Gröbner Bases to assess questions in STACK . . . 209
9.2.1 Alternate Answers . . . 213
9.2.2 Incorrect Responses . . . 214
9.3 Testing and Distribution . . . 217

9.4 Answer Test Code 218
9.4.1 The `ATSysEquiv` function 219
9.4.2 The `ATSysEquivVars` Function 221
9.4.3 The `ATSysEquivGrob` Function 222
9.5 Example Questions 223
9.5.1 Simple Pythagorean 223
9.5.2 Equal Time 226
9.5.3 Frame Absolute 226
9.5.4 Three Relative Amounts 228

10 Conclusions and Future Opportunities 230
10.1 Conclusions 230
10.1.1 Part I 230
10.1.2 Part II 233
10.1.3 Part III 237
10.2 Omissions and Opportunities 239
10.2.1 Part I 239
10.2.2 Part II 241
10.2.3 Part III 243
10.3 Final Word 246

A Results 247

B Code 250

C Pólya's Advice 256

List of References 257

Major Corrections

The most major correction to this thesis is the addition of Chapter 10, starting on page 230. Section 10.1 summarises our conclusions from the three parts of the thesis, while Section 10.2 highlights the limitations of the work and the possible directions for future research.

Section 9.4 starting on page 218, has been added to explain the workings of the answer test code and its related functions.

Section 9.5 starting on page 223, presents examples of four further questions taken from Mayer (1981) and implemented using the SysEquiv answer test.

Minor Corrections

A summary of minor corrections by in each part of the thesis. Errors in language or slight rewordings are not included.

Part 1

Page 4: Rewrote introduction to include quote from the report of the HE Mathematics Curriculum Summit, clarified remarks on our definition of problem-solving at the start of Section 2.1

Page 7: Added an example to illustrate Tall's *concept image*.

Page 14: Added a footnote on Brousseau's use of *devolution*.

Example 2.2.1 (Page 15): Replaced $-\pi$ with 0 so there was no 'trick' solution to the question.

Page 22: Added a clarifying remark on our comparison of directed learning and discovery learning, in particular that these do not cover *all* mathematical pedagogies, and nor are they necessarily completely distinct.

Page 24: Added comments on Bligh's dataset, which mainly, but not exclusively, used STEM subjects.

Page 29: Corrected miscounting of universities and highlighted limitations of the questionnaire answers.

Page 31: Added footnote on the definition of *ontogeny*.

Page 44: Added remarks on case study universities' membership of the Russell Group. Corrected Macondo to membership of the 1994 Group, but no the Russell Group.

Page 48: Expanded introduction and explanation of the Case Studies.

Part 2

Page 81: Added a brief description of Dewey's empirical pedagogy.

Page 92: Added mention of the use of the Socratic Method at London South Bank University, as an example of an alternative pedagogy aimed at large groups.

Page 103: Added comment on the proportion of MSci students in 1Y before 2011/12.

Section 6.3.2 (Page 105 and onward): Added discussion of the results of Levine's Test and the inhomogeneity of the data. Clarified definitions of null and alternative hypotheses. Rewrote significances.

Page 113: Made clear that the 1st year mean mark was calculated without including the 1Y module result for 1Y students.

Page 121: Made clearer the language on students who performed least well on average on the Class Test scored better than they had on first year modules, while those who performed most well scored worse than they had.

Page 125: Macondo is a member of the 1994 Group of universities, but not the Russell Group. Rephrased the research questions in terms of staff *perceptions* of efficacy.

Page 130: Made clear that the author conducted and transcribed the staff interviews.

Page 132: Added comment on the selection of quotes.

Part 3

Page 164: Noted Sangwin's review of current computer-aided assessment software.

Page 166: Made clear that this work enables the assessment of some questions previously unassessable with CAA systems.

Page 169: Drew connection between formal undecidability and the halting problem.

Page 178: Removed duplicate definition!

Page 196: Corrected polynomials in Example 8.6.1.

Page 206: Made it clear that the answer test code was written almost entirely by the author.

Page 207: Made reference to buggy-/mal-rules.

Chapter 1

Introduction

This thesis comprises three parts. In Part I we discuss *problem-based learning*, a range of student-centred pedagogies whose aim is to develop mathematical thinking. We begin, in Chapter 2, with a brief review of work on mathematical thinking and problem solving. Here we discuss what it means to view the world with a mathematical eye and the proclivities that are particular to practitioners of the mathematical sciences, and how these ways of thinking apply to mathematical problem solving. Chapter 3 covers two broad categories of pedagogies. The first, directed learning, refers to the traditional lecture-based approach that the majority of courses in U.K. mathematics departments employ to teach students. The second, discovery learning, describes a range of student-centred pedagogies which rose to prominence in the 1960s. In particular, we discuss enquiry-based learning. In the final chapter in the first part of the thesis, Chapter 4, we survey the state of problem-solving teaching in mathematics departments in England and Wales.

The focus of Part II is the *Moore Method*, a pedagogy developed by the Texan mathematician R.L. Moore during the first half of the twentieth century. Chapter 5 discusses the history of

Moore and his method, and compares it to other problem-based learning pedagogies. Chapter 6, is a report on an implementation of the Moore Method in a first year course at the University of Birmingham's School of Mathematics. The report draws on over 6,000 exam results for quantitative analysis of the course; demonstrating that participation in this course serves to aid students' work in certain other courses in the School. Chapter 7 records the experiences of staff teaching the Moore Method at another U.K. university, many for the first time, to discover those aspects that are important to the Method.

The final part of this thesis covers computer-aided assessment (CAA). Working with the computer-aided assessment system STACK, developed at Birmingham, we investigate what it means for two systems of polynomial equations to be the same, and how we can use this information to assess questions in CAA. Chapter 8 looks at the mathematical basis for the comparison of equations and systems of equations – Gröbner bases. This is a topic in algebraic geometry which uses the geometric properties of systems of equations to provide information about their algebraic solutions. In Chapter 9 we make use of the theory of Gröbner bases to assess systems of equations in STACK. This extends to both finding equivalence between systems, and to using properties of inequivalent systems to provide feedback to students and teachers.

Part I

Problem-based Learning

Chapter 2

Problem-solving

In the report of the HE Mathematics Curriculum Summit, held in January 2011 at the University of Birmingham, it is argued that (Rowlett, 2011, p. 19):

> “Problem-solving is the most useful skill a student can take with them when they leave university. It is problematic to allow students to graduate with first class degrees who cannot handle unfamiliar problems.

The aim in the first part of this thesis is to determine the extent and variation of problem-solving teaching in England and Wales. Each chapter is dedicated to a particular aspect of problem-solving, and the part as a whole lays the groundwork for discussions in the second part of the thesis. We being, in this chapter, with a review of the literature on mathematical thinking and problem-solving (not explicitly defined by Rowlett, 2011). In Chapter 3 we give an overview of the history of teaching, and teaching *with*, problem-solving, and contrast it to traditional *directed learning*. In Chapter 4 survey the state of problem-solving teaching in mathematics departments in England and Wales, and give four case-studies highlighting the variations in such teaching.

2.1 Thinking Mathematically

We will discuss the efforts of others to define mathematical thinking shortly, though we begin with our own attempt, purely as a basis for comparison:

> 'Mathematical thinking' refers to any thought processes which involve the navigation, representation, alteration, or creation of some structure or concept, or conjecturing, justifying, verifying, or refuting such actions.

This definition is so broadly applicable it may be argued it encompasses all human thought, and if we try to use it to identify a clear dividing line between mathematical thinking and 'non-mathematical' thinking, we will not be successful. That being granted, it is certianly the case that mathematical thinking extends beyond mathematics and other STEM subjects. Relatively few people *do* mathematics on a daily basis, but the opportunities for engaging in mathematical thinking are all around us. Whenever we plan a route, follow a recipe, or work to a budget, we are involved in a process that has some structure to it, and whose structure we must navigate in order to achieve a goal. Making use of structure requires us to identify relationships between objects, and determining the pattern of relationships between several such objects allows us to derive something about a structure. Both these processes see us engaging in mathematical thinking. Counting, or enumeration, is a basic example of mathematical thinking that is demonstrated to some degree even by insects (Dacke and Srinivasan, 2008) and songbirds (Hunt et al., 2008).

Mason et al. (2010)[1] offered the following four processes as forming the bulk of mathematical thinking: specializing, generalizing, conjecturing, and convincing. These are described by Burton (1984, p. 38) as follows:

Specializing. When one is faced with a question or problem, a powerful way to explore its meaning is by examining particular examples. Such specializing is the key to an inductive approach to learning and is observed as natural to the learning of children. Each example provides the opportunity for manipulating elements that are concrete in the child's thinking, whether they are physical manifestations or ideas.

Conjecturing. When enough such examples have been examined, conjecturing about the relationship that connects them happens almost automatically. Through conjecturing, a sense of any underlying pattern is explored, expressed, and then substantiated.

Generalizing. The recognition of pattern or regularity provokes the statement of a generalization. Such statements appear to be the building blocks used by learners to create order and meaning out of an overwhelming quantity of sense data, and it is on such generalizations that much behavior depends.

Convincing. To become robust, a generalization must be tested until it is convincing. First the thinker convinces himself or herself and then the world outside. The convincing process is the means by which a generalization moves from being personal to being public. A picture of the deductive approach is obtained by inverting the order of the processes.

[1] The first edition is Mason et al. (1982).

> Beginning with a generalization, one explores the web of conjectures it provokes and tests them against particular specializations.

What separates mathematics, and pure mathematics in particular, from other subjects or tasks is that the structures in question are frequently abstract. Many structures, such as numbers or groups, are motivated by the real-world but themselves exist only as concepts. We understand such abstractions through *concept images*, which are neither constant nor necessarily consistent:

> “We shall call the portion of the concept image which is activated at a particular time the evoked concept image. At different times, seemingly conflicting images may be evoked. Only when conflicting aspects are evoked simultaneously need there be any actual sense of conflict or confusion. — Tall and Vinner, 1981

By way of example, a student who begins learning about groups from the symmetries of a triangle may initially hold in their mind the image of a group being related to that particular shape, or regular polygons in three or fewer dimensions. As their understanding develops, this image will have to change to conform to the new types of groups they encounter, or they risk either failing to identify a particular object as a group, or to make false generalisations based on their image of all groups being dihedral groups.

Specialising is a particular way of representing a mathematical structure that allows us to start building up our concept image, while the recognition of similarity between two concept images may lead to a more general concept image.

The QAA Benchmark (2007) for undergraduate degree programmes in mathematics highlights the importance of this ability to generalise:

> Graduates will also appreciate the power of generalisation and abstraction in developing mathematical theories or methods to use in problem solving. Theory-based programmes may tend to emphasise the role of logical mathematical argument and deductive reasoning, often including formal processes of mathematical proof; practice-based programmes may tend to emphasise understanding and use of structured mathematical or analytical approaches to problem solving. — Lawson et al., 2007

Part of the process of learning to think mathematically is tied to the idea of 'becoming a mathematician':

> ...becoming a good mathematical problem solver—becoming a good thinker in any domain—may be as much a matter of acquiring the habits and dispositions of interpretation and sense-making as of acquiring any particular set of skills, strategies, or knowledge. — Resnick, 1988, p. 58

As educators we want to ensure undergraduates are "acquiring the habits and dispositions" of mathematicians so that they learn to "see the world in ways like mathematicians do" (Schoenfeld, 1992, p. 16); we want them to think mathematically. Thinking mathematically becomes a natural process for someone experienced in the mathematical sciences, to the extent that they may not identify it as different to any other thinking, however "the predilection to quantify and model is certainly a part of the mathematical disposition, and is not typical of those outside mathematically oriented communities." (Schoenfeld, 1992, p. 21)

In most cases the aim of developing students' mathematical thinking remains part of the "hidden curriculum" (Snyder, 1973) about which they are not explicitly told. Instead, mathematical degree programmes concentrate on knowledge and technique which, while of their own importance, are "[often] presented as the **only** factor. Drawing attention to the process of enquiry and the emotional and psychological states they provoke, and focusing on these factors, seems to me to be a necessary part of helping people towards a more useful and more creative view of mathematical thinking" (Mason et al., 2010, p. 133).

We want students to view the world with a mathematician's eye, not with the aim that they all continue to postgraduate study, but because mathematical thinking has wider application than mathematical research:

> Thinking mathematically is not an end in itself. It is a process by which we increase our understanding of the world and extend our choices. Because it is a way of proceeding, it has widespread application, not only to attaching problems which are mathematical or scientific, but more generally. — Mason et al., 2010, p. 140

Mathematicians apply their thinking to making and proving conjectures. The process of proving something is often hidden by the final proof itself, a "communicable, decontextualized, depersonalized, detemporalized form" (Brousseau, 1997, p. 227) of the original 'proof', if one can be said to exist. In *Proofs and Refutations*, Lakatos et al. (1976) demonstrated, through a Socratic dialogue, how mathematicians in the 18^{th} and 19^{th} centuries developed proofs of Euler's formula for the Euler characteristic. While the proof that resulted from these discussions was eventually correct, the heuristic process involved many incorrect claims and false 'proofs'. As Pólya (1954a, p. 159) says:

> It may appear a little more surprising to the layman that the mathematician is also guessing. The result of the mathematician's creative work is demonstrative reasoning, a proof, but the proof is discovered by plausible reasoning, by guessing.

This 'plausible reasoning' does not mean that guesses should be unfounded, but be based on experience. Underlying the processes of specializing, generalizing, conjecturing, and convincing, is a process of fundamental importance: reflection.

> As long as our activity glides smoothly along from one thing to another...there is no call for reflection. Difficulty or obstruction in the way of reaching a belief brings us, however, to a pause. In the suspense of uncertainty, we metaphorically climb a tree; we try to find some standpoint from which we may survey additional facts and, getting a more commanding view of the situation, decide how the facts stand related to one another. — Dewey, 1910, p. 11

Reflection is just one type of thinking *about* thinking, where we make careful assessment of our own thought processes or difficulties when engaging in learning and problem-solving. The American developmental psychologist J. H. Flavell termed this insight *metacognition*:

> I am engaging in metacognition if I notice that I am having more trouble learning A than B; if it strikes me that I should double check C before accepting it as fact. — Flavell, 1976, p. 232

A key aspect of metacognition in mathematical problem-solving is the mastery of negative emotions:

> Mathematical thinking is not only improved by learning how to conduct an enquiry, but also by recognising and harnessing to your advantage the feelings and phycological states that accompany it. At the most basic level, there are negative emotions to control. — Mason et al., 2010, p. 135

Part of the necessity for controlling one's negative emotions is the fact that mathematical problems do not always yield in a short space of time, requiring as they do tenacity and patience to be solved. Discussing a range of famous unsolved problems, Schoenfeld (1992, p. 15) notes that "…they differ only in scale from the problems encountered in the day-to-day activity of mathematicians. Whether pure or applied, the challenges that ultimately advance our understanding take weeks, months, and often years to solve". He continues on to claim that mathematics educators need to improve students' understanding of the nature of mathematics and problem-solving, prime among which is an appreciation of the time and patience that must be applied to seemingly intractable problems.

The impression one gets of someone who is *thinking mathematically* is of a person that seeks out patterns in the world around them, who conjectures theories and reflects on their own thought processes. By dint of this, they are a member of the mathematical community, having adopted its processes and customs. None of these features is peculiar to mathematical thinking however, and none absolutely defines it. Sternberg and Ben-Zeev (1996, p. 303) write:

> To the extent that one's goal is to understand *mathematical thinking* in terms of a set of clearly defining features that are individually necessary and jointly sufficient for understanding the construct, one is going to be disappointed. Indeed, it is difficult to find any common features that pervade all of the various kinds of mathematical thinking...

And so, while we have a general sense of the predispositions of someone thinking mathematically, we cannot clearly define the concept itself. In the following section we consider what it is that a person with such predispositions is *doing* when engaging in mathematical thinking.

2.2 Problem-solving

In the report of the HE Mathematics Curriculum Summit, held in January 2011 at the University of Birmingham, it is argued that (Rowlett, 2011, p. 19):

> Problem-solving is the most useful skill a student can take with them when they leave university. It is problematic to allow students to graduate with first class degrees who cannot handle unfamiliar problems.

The report concludes with 14 recommendations for developing higher education teaching, the first three of which relate to problem-solving, specifically sharing good teaching practice and developing problems and resources for educators. The work resulting from these recommendations is discussed in Chapter 4.

The importance of problem-solving in mathematics education is summed up by the Hungarian mathematician Paul Halmos in his article *The Heart of Mathematics*:

> “What does mathematics really consist of? Axioms (such as the parallel postulate)? Theorems (such as the fundamental theorem of algebra)? Proofs (such as Gödel's proof of undecidability)? Definitions (such as the Menger definition of dimension)? Theories (such as category theory)? Formulas (such as Cauchy's integral formula)? Methods (such as the method of successive approximations)?
>
> Mathematics could surely not exist without these ingredients; they are all essential. It is nevertheless a tenable point of view that none of them is at the heart of the subject, that the mathematician's main reason for existence is to solve problems, and that, therefore, what mathematics really consists of is problems and solutions.
>
> — Halmos, 1980, p. 519

Brousseau, the first to expound the theory of the *didactic contract* between the student and teacher in learning mathematics, is similarly strident when making the case for problem-solving:

> We know that the only way to "do" mathematics is to investigate and solve certain specific problems and, on this occasion, to raise new questions. The teacher must therefore arrange not the communication of knowledge, but the *devolution* of a good problem[2]. If this devolution takes place, the students enter into the game and if they win learning occurs. — Brousseau, 1997, p. 31

It follows immediately that to develop the art of thinking mathematically that we discussed above, one must learn to approach unfamiliar problems and solve them. Indeed, there is a general consensus among researchers of mathematics education that problem-solving and modelling are important skills for students of the mathematical sciences. Blum and Niss offered an overview of arguments for and against such training, and concluded that:

> Mastering mathematics can no longer be considered equivalent to knowing a set of mathematical facts. It requires also the mastering of mathematical *processes*, of which problem-solving—in the broadest sense—occupies a predominant position. — Blum and Niss, 1991, p. 44

The above extracts argue the case *for* problem-solving, though none explicitly defines the phrase. To some, problem-solving may mean solving word-problems (Gerofsky, 1996), to others, the questions on undergraduate 'problem sheets' suffice. While consensus is divided, George Póla wrote much on the subject and, in his two-volume work *Mathematics and Plausible Reasoning*, defined problem-solving as meaning:

> To search consciously for some action appropriate to attain a clearly conceived but not immediately attainable aim. — Pólya, 1962, p. 1.117

Pólya's definition of problem-solving demonstrates that trying to define a question as a problem without consideration of the student is fruitless: a problem for one person is not necessarily a problem for another. A *problem* then is an ordered pair of a question and an answerer; the question may be anything, so long as the answerer finds themselves not immediately knowing how to solve it. Because we are interested in problem-solving at undergraduate level, it is helpful if we make the assumption that 'answerer' can be replaced with 'student', where our student is a first-year undergraduate with an A in A-Level mathematics (or equivalent).

We contrast problems with *exercises*, a mathematical exercise being a question whose solution involves only routine procedures. Many exercises may therefore be automated with computer algebra systems. Having learned the relevant techniques, our student will be able to follow an obvious strategy and systematically apply the techniques in a sequence of steps to reach a correct solution. Example 2.2.1 will be an exercise for them:

▼ Example 2.2.1.

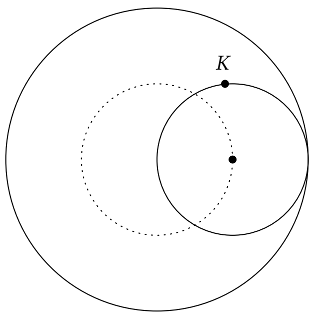
Calculate

$$\int_0^\pi \sin(2x)\, \mathrm{d}x.$$

A question being an exercise does not mean that there is not work to do—here, the student must correctly identify the symbols above and divine their meaning, apply the rule of integrating the function of a function, and know the various values of sine for integer multiples of π—but there is no mystery as to what that work is. Contrast this with Example 2.2.2, where it is unlikely that our student will immediately know how to proceed:

▼ Example 2.2.2 — *Circles*.

Let K be a point on a circle sitting inside and touching a stationary circle of twice its diameter. Describe the path of K as the smaller circle rolls round the larger one without sliding.



The defining characteristic of problems as we speak of them is *novelty*; there has to be something about them that a student has not seen before. Were our idealised student to have studied geometry widely, Example 2.2.2 may prove only as challenging as the integration exercise above. The amount of novelty required to present students with a challenge need only be limited; Clement et al. (1981) gave the example of the Students-and-Professors Problem:

▼ **Example 2.2.3 — *Students and Professors*.**

Write an equation for the following statement: "There are six times as many students as professors at this university". Use S for the number of students and P for the number of professors. — Clement et al., 1981

In their paper the problem was given to 150 calculus level students, 37% of whom answered incorrectly. The answer $6S = P$ accounted for two thirds of all errors. This type of word or story problem is not something we would consider a problem in the truest sense; once students have learned the technique of transliterating words into mathematics, repeating the process for new questions is straightforward. Mayer (1981) classified 1097 (U.S.) high-school algebra problems and found that they contained only three kinds of proposition:

1. Values are assigned to variables, e.g. time taken to walk up = 10.
2. Relationships between variables, e.g. distance up = distance back.
3. Identification of the unknown or goal, e.g. GOAL = distance up.

Word problems have a long history (Mason, 2001), having been used to teach mathematics in Mesopotamia (Robson, 2008), and Sangwin (2011) makes the case that they can form the first step on the road to mathematical research. However, "this is not what is meant by *problem-solving* at tertiary level, although it would be greatly advantageous if students on entry were better at it." (Mason, 2012)

2.2.1 Teaching Problem-solving

While problems have been posed to students since antiquity (Boyer and Merzbach, 1991; Robson, 2008), the modern problem-solving movement within mathematics education can be traced directly to the work of George Pólya. In *How to Solve It*, Pólya (1945) was the first to discuss problem-solving as a discipline in its own right and the first to offer a procedure for solving mathematics problems. Pólya suggests the following process for finding solutions:

1. Understanding the Problem.
2. Making a plan.
3. Carrying out the plan.
4. Looking Back.

See Appendix C for the expanded version of this advice. It is important to note that Pólya is not so much promoting a particular heuristic or algorithm, as one learns in mathematical topics, but a more general approach. Indeed, were we able to settle on a particular algorithm for solving a set of problems, they are no longer novel to us and so no longer problems. Instead, Pólya focuses more on powers that are used when we engage in mathematical thinking: specialising, generalising, conjecturing and convincing; he wanted not to teach 'problem-solving' but mathematical thinking.

Pólya expanded greatly on the ideas contained in *How to Solve It* in his two-volume work *Mathematics and Plausible Reasoning* (Pólya, 1954a,b)[3], which contained many concrete

[3] Throughout we cite the single volume Pólya (1962), with a 1. or 2. to denote which volume in the original books is being referenced.

mathematical examples to demonstrate the processes involved. It is important to note that the final stage, *Looking Back*, refers to the point at which we decide whether we have been successful or not; whether we need to attempt to solve the problem again, or if there is a more elegant way of solving it than that we have found. It is *not* the only stage at which reflection should occur, as reflecting on past experience is important at all stages of the problem-solving process. Understanding requires us to relate a problem to what we know already, and making a plan without considering what has worked or not worked in the past is surely a bad idea. Finally, when we carry out a plan we must continually reflect on whether the plan is working as we expect it to.

While Pólya's ideas have been continually used and developed, notably by Mason et al. (1982, 2010), and *How to Solve It* remains a set text for many mathematics education degrees, his strategies are not routinely taught to students. Begle (1979, p. 145) surveyed the empirical literature on these processes and came to the conclusion that:

> No clear-cut directions for mathematics education are provided by the findings of these studies. In fact, there are enough indications that problem-solving strategies are both problem- and student-specific often enough to suggest that hopes of finding one (or a few) strategies which should be taught to all (or most) students are far too simplistic.

In his review of the relevant research, Schoenfeld (1992) demonstrated three particular strategies useful in specific cases, such as when dealing with the roots of polynomials, and had the following to say:

> Needless to say, these three strategies hardly exhaust "special cases." At this level of analysis—the level of analysis necessary for implementing the strategies—one could find a dozen more. This is the case for almost all of Pólya's strategies. The indications are that students can learn to use these more carefully delineated strategies.

Developing strategies of more limited scope than those of Pólya demonstrate, Schoenfeld argues, a change from *descriptive* heuristics—that is to say, names for broad categories of processes—to *prescriptive* processes—what to do in specific cases—is a better approach to teaching students how to solve problems. Most recently, Lesh and Zawojewski (2007, p. 768) summed up the situation as follows:

> In mathematics education, Pólya-style problem-solving strategies—such as *draw a picture*, *work backwards*, *look for a similar problem*, or *identify the givens and goals*—have long histories of being advocated as important abilities for students to develop. Although experts often use these terms when giving after-the-fact explanations of their own problem-solving behaviors, and researches find these terms useful descriptors of the behavior of problem solvers they observe, research has not linked direct instruction in these strategies to improved problem-solving performance.

Briefly then, Pólya's heuristic is too general to be prescriptive whilst Schoenfeld's processes are too numerous to be decided between, effectively doing little to solve the problem at hand.

Perhaps as a result of the apparent difficulty in *teaching* problem-solving, there has been over the last several decades a "pendulum of curriculum change" (Lesh and Zawojewski, 2007, p. 764) swinging between problem-solving and basic technical skills. The pendulum currently appears to be swinging in the direction of problem-solving once again, though its teaching remains in something of a difficult position. The importance of problem-solving in developing mathematical thinking is appreciated by many, and the process that mathematicians go through when solving problems can be retrospectively described in common terms. One possible solution to the seeming impasse is to refer back to Pólya's *Problems and theorems in analysis* (Pólya and Szegö, 1925), co-authored with Szegö, in which is found the following:

> "The independent solving of challenging problems will aid the reader far more than the aphorisms which follow, although as a start these can do him no harm. — Pólya and Szegö, 1925, p. vii

This was written two decades before the publication of *How to Solve It*, and has an appealing clarity. If we want students to solve problems on their own, perhaps the best solution is to give them problems to solve with minimal concentration on strategy and process. In the following chapter we look at the rise of *problem-based learning*, a range of pedagogies aimed at teaching students through their own problem-solving and other mathematical enquiries.

Chapter 3

Problem-based Learning

In this chapter we draw a distinction between two broad pedagogical categories. The first, *directed learning*, describes the lecture based approaches traditionally used in university mathematics courses. The second, *discovery learning*, refers to a set of student-centred pedagogies that became popular in the 1960s. A subset of the discovery pedagogies, called *problem-based learning* (PBL), is the focus of our discussion.

We stress that directed learning and discovery learning do not cover all pedagogies employed in mathematics teaching, nor that each is necessarily distinct from the other.

3.1 Directed Learning

Direct instruction is a term coined by Siegfried Engelmann to describe teaching by lectures or the demonstration of material (Bereiter and Engelmann, 1966). We define *directed learning* to be learning where the majority of teaching time is handed over to direct instruction. We use the School of Mathematics at the University of Birmingham as our specific example; other U.K. mathematics departments are considered in Section 3.1.3.

The first year, first semester, module MSM1Aa – *Calculus and Algebra I* is typical of Birmingham's first year modules. In eleven weeks of study there are 44 lectures, 20 hours of computer lab work and 10 hours of example classes. Each lecture is 50 minutes and it is anticipated "that each 'one hour' lecture requires another two hours of private study" (University of Birmingham, 2010, p. 21). Few other modules have this amount of computer work but the ratio of four lectures to one example class is usual.

3.1.1 Lectures

Lectures are where the course leader expounds the theory that the students are required to learn for the course. This theory will cover definitions, techniques, theorems and proofs. Each lecturer may take a slightly different approach to this teaching, with some posing questions to the class, but lecturers will spend almost all the lecture time at the board or presenting to students in some other way, and there will be little or no assessment of students in the lectures themselves.

Each lecture includes everyone taking the module in the same room; it is not unusual that a first year lecture have more than 150 students sitting in on it, while fundamental modules include lectures to nearly 250 people. Students are expected to take notes during lectures, and use these notes for reviewing learning and answering questions in example classes. Some lecturers provide typeset lecture notes to students, either at the end of a lecture or before the examination period.

Lectures are as effective as other methods for transmitting information (Bligh, 1998). Table 3.1 gives a comparison of different teaching methods where acquisition of information is the main criterion, and is read as follows. Each of the five types teaching methods is compared to lectures by counting the number of studies that show them to be either more effective, less effective, or not significantly different in effectiveness at transmitting information. It shows that besides the *Personalised System of Instruction* (P.S.I.), a method centred around students that stresses written work and self-pacing, lectures are not significantly better or worse at transmitting information to students. Bligh compiled this information from an extensive meta-analysis of 298 studies in the educational literature spanning 71 years of research (1925–1996), the majority of which were from the 1970s and 1980s. The majority, but not all, of these studies focussed on STEM subjects; the comparative studies used for the meta-analysis are at Bligh (1998, pp. 269-276).

Teaching Method	Lectures Less Effective	No Significant Difference	Lectures More Effective
Programmed learning and P.S.I.-related	20	17	8
Discussion (various)	18	54	22
Reading and independent study	10	21	9
Inquiry (e.g., projects)	6	6	3
Other (mostly audio, TV, computer-assisted learning	27	57	20

Table 3.1: Comparison of lectures with other teaching methods in promoting information acquisition. (Bligh, 1998, p. 11)

Given their effectiveness at transmitting information, and the relatively low-cost, it is unsurprising that lectures are the main conduit for learning concepts at undergraduate level.

3.1.2 Example Classes

In example classes students work on tasks given to them by the lecturer. Example classes divide the students taking a module into smaller groups, typically around 20 people in size, and include a number of postgraduate demonstrators to assist students. At Birmingham example classes are taken every other week in almost all modules.

These tasks are handed out in lectures and students are expected to work on them, alone or in groups, before example classes. The classes are then used to concentrate on those questions that the students have failed to answer on their own. Answers are submitted either at the end of the examples class or in the days following it.

The following example is from the fourth assessment of the first year module MSM1B – *Sequences and Series*.

▼ Example 3.1.1.

Q1.

1. *Define what it means for a series*

$$\sum_{n=1}^{\infty} a_n$$

of real numbers to converge.

2. *Using the definition, prove that the following series converge, and find their limits.*

   (a)

$$\sum_{n=1}^{\infty} \frac{1}{4n^2 - 1}.$$

   (b)

$$\sum_{n=1}^{\infty} \frac{1}{(n+1)(n+3)}.$$

   (c)

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)}.$$

[In this question you may wish to use partial fractions.]

Example 3.1.1 is typical of those found on assessment sheets; the first part requires students to give a definition that is fundamental to their learning. The second part uses the definition and applies a theory; in particular identifying the partial sum and applying the algebra of limits to satisfy the definition of a convergent series.

Another example of a question from the same assessment sheet is:

▼ **Example 3.1.2.**

R1. *Study carefully the proof (from lectures) that*

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$$

diverges. For which values of $\alpha \in \mathbb{R}$ does this divergence proof adapt to the series

$$\sum_{n=1}^{\infty} \frac{1}{n^{\alpha}}?$$

Write down a complete proof for such values of α.

These questions require students to look back at the proofs that appear in their lecture notes. In this way the definitions, techniques, theorems and proofs that their lectures cover are reiterated to students through their example classes.

3.1.3 The structure of directed learning

Lectures are an effective way of transmitting information to students, but they are less effective than discussion, enquiry and various other methods at promoting *thought*. Table 3.2, also taken from Bligh (1998), is read in the same manner as Table 3.1.

Teaching Method	Lectures Less Effective	No Significant Difference	Lectures More Effective
Discussion	29	1	2
Reading and independent study	18	3	1
Inquiry	10	1	1
Other methods	6	17	0

Table 3.2: Comparison of lectures with other teaching methods in promoting thought. (Bligh, 1998, p. 5)

Again, this table is the result of an extensive meta-analysis of 89 studies published from 1942–1996. Bligh admits that measuring thought is less straightforward than measuring transmitted information, though he also states that "in spite of the variety, with few exceptions discussion is consistently more effective than lectures in getting students to think." (Bligh, 1998, p. 14)

Directed learning is therefore structured to promote the learning of concepts and developing student's abilities to apply these concepts in practice. Examples classes complement lectures by providing students with the necessary time to think independently and use the knowledge they have gained. Both these methods of direct interaction with students are in turn complemented by students working in their own time to cover material and practice answering questions.

To assess the fraction of time given to lectures in U.K. mathematics departments, BSc Mathematics undergraduate programmes (UCAS code G100) at the 24 Russell Group universities were considered. Information was gained from course brochures and departmental websites. Five universities, Edinburgh, Leeds, Queen's University Belfast, Southampton, and York gave no information on their typical course structures. Further, five universities did not mention the ratio of lectures to tutorials or example classes but did give more detail. They were as follows:

King's College London

> Teaching is predominantly by lectures, supplemented with tutorials and problem classes. (London, 2011)

Newcastle

> The material will be delivered through the medium of lectures. (University of Newcastle,

> 2011)

Nottingham

> Teaching is primarily by lectures, supported by workshops (sessions in which you practise solving problems with support from academic staff), small group tutorials and private study. (University of Nottingham, 2011, p. 6)

Oxford

> Mathematicians from across all the colleges come together for lectures which are arranged by the University. This is usually how students first meet each new topic of mathematics. (University of Oxford, 2011, p. 2)

Warwick

> The most common way of imparting knowledge is the 50-minute lecture, with audiences ranging from upwards of 250 down to 5 students. (University of Warwick, 2011)

Of the 14 remaining universities, the ratio of lectures to example classes in first year courses was as follows:

The 14 universities shown in Table 3.3 average just over 3 lectures to one tutorial in their first-year mathematics undergraduate teaching. Of the departments considered, it is clear that directed learning is the dominant pedagogy for teaching first year students. Expanding this analysis beyond first year modules and Russell Group universities may reveal different trends, however, and presents an opportunity for future research.

University	Lectures:Examples classes	Source
Birmingham	4:1	University of Birmingham (2010, p. 21)
Bristol	2 or 3:1	University of Bristol (2011, p. 2)
Cambridge	6:1	University of Cambridge (2011, p. 5)
Cardiff	4:1	Cardiff University (2011, p. 26)
Durham	2:1	Durham University (2012)
Exeter	2:1	University of Exeter (2012)
Glasgow	2.7:1	University of Glasgow (2011, p. 4)
Imperial	2.6:1	Imperial College London (2011, p. 4)
UCL	3:1	University College London (2011)
Liverpool	3:1	University of Liverpool (2011, p. 4)
LSE	2:1	London School of Economics (2011)
Manchester	3 or 4:1	University of Manchester (2011, p. 2)
Queen Mary	3:1	Queen Mary, University of London (2012)
Sheffield	2:1	University of Sheffield (2011)

Table 3.3: Lecture:Example class ratios in Russell Group universities' mathematics departments

3.2 Problem-based Learning

Discovery learning refers to a number of alternative pedagogies whose central feature is learning by doing. They are named for psychologist Jerome Bruner's 1961 paper *The Act of Discovery* (Bruner, 1961). Bruner argued that "practice in discovering for oneself teaches one to acquire information in a way that makes that information more readily viable in problem solving" (Bruner, 1961, p. 26). In *pure* discovery learning, students operate without assistance of any kind, either being set a number of questions or simply an area to explore. By the near-complete removal of the teacher in any capacity, pure discovery learning is *minimally guided*; students decide which avenues to travel down entirely on their own. Bruner was particularly influenced by the work of the John Dewey, who in *My Pedagogical Creed* stated that:

> The teacher is not in the school to impose certain ideas or to form certain habits in the child, but is there as a member of the community to select the influences which shall affect the child and to assist him in properly responding to these influences.
>
> — Dewey and Small, 1897, p. 9

Note that Dewey's stance was less rigid than Bruner's, offering more guidance to students than did Bruner.

At a similar time as Bruner, the Swiss psychologist Jean Piaget, working with South African mathematician Seymour Papert, developed *constructionism.* This pedagogy was related to the constructivist epistemology, developed by Piaget, who "arrived at his view of cognition as a biologist who looked at intelligence and knowledge as biological functions whose development had to be explained and mapped in the ontogeny[1] of organisms" (von Glasersfeld, 1990, p. 22). The central aim of constructionism was to give students an experience in which they could construct truths for themselves, and lead Papert to the creation in 1967 of the Logo programming language, designed to help students explore mathematical ideas for themselves.

In addition to the work of Piaget and Papert, Lev Vygotsky worked in the early twentieth century on child psychology and learning. Vygotsky did most of his work in the 1920s and early 1930s, this was subsequently drawn upon during the development of the constructivist pedagogies in the 1960s. In particular, the concept of the *zone of proximal development* (ZPD) in student learning came to the fore:

[1]emphn. The origin and development of an individual organism from embryo to adult.

> It is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers.
> — Vygotsky, 1978, p. 86

Vygotsky's ideas hugely influenced Bruner et al, and lead Wood et al. (1976, p. 90) to define *scaffolding*:

> Discussions of problem solving or skill acquisition are usually premised on the assumption that the learner is alone and unassisted. If the social context is taken into account, it is usually treated as an instance of modelling and imitation. But the intervention of a tutor may involve much more than this. More often than not, it involves a kind of "scaffolding" process that enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts. This scaffolding consists essentially of the adult "controlling" those elements of the task that are initially beyond the learner's capacity, thus permitting him to concentrate on and complete only those elements that are within his range of competence.

Constructivism appears in many forms throughout the educational literature, principally as *enquiry-based learning* (EBL) and *problem-based learning* (PBL). In both, students are presented with tasks by the teacher and through working on these tasks acquire the knowledge that the teacher is trying to convey. It is the scaffolding that the teacher puts in place that differentiates

EBL and PBL from minimally-guided pedagogies such as discovery learning. EBL is sometimes termed simply *enquiry learning*, while in the United States it is called *inquiry learning* (IL); the form that students' 'enquiries' take can vary greatly, and so there is no clear distinction between the two:

> As we have examined the broad variety of instantiations of PBL and IL, we have not uncovered any dimensions that consistently distinguish between PBL and IL. Indeed, we think there are no clear-cut distinguishing features. PBL frequently engages students in explorations and analyses of data, such as one would expect IL environments to do, and IL frequently poses problems and asks students to consult various resources to solve them as PBL environments do. — Hmelo-Silver et al., 2007, p. 100

We prefer the term *problem-based* learning, as this gives a clearer indication of what is involved. In PBL, students are presented with problems through which they learn both concepts and techniques. It is not prescriptive however; PBL may include lectures or other presentations by a teacher; what is key is the general constructivist approach, where students have the opportunity to learn and create for themselves. In the following chapter we will look more closely at the use of PBL in four university mathematics departments, while Part II describes a particular pedagogy—the Moore Method—used at two other universities.

3.2.1 Arguments for and against PBL

Since the introduction of discovery learning and the constructivist pedagogies, debate regarding their effectiveness as teaching tools has been ongoing. Kirschner et al. (2006) make the case against discovery learning, viewing constructivism, discovery, enquiry-based and problem-based learning as a single pedagogy, and argue that there is ample evidence that these minimally guided teaching methods are ineffective. They agree with Mayer's assertion that we should "move educational reform efforts from the fuzzy and unproductive world of ideology—which sometimes hides under the various banners of constructivism—to the sharp and productive world of theory-based research on how people learn" (Mayer, 2004, p. 18). Kirschner et al. (2006) argue that minimally guided techniques are only effective when teachers construct suitable scaffolding to assist students who are not making progress, in particular citing Aulls (2002) which describes qualitative studies of teachers using discovery learning.

Klahr and Nigam (2004) report a study on 9 and 10 year-old students learning about the effect of slope with rubber balls on ramps. Half of the students were taught with direct instruction whilst the other half were taught by discovery learning, with students receiving no direction or feedback. Students taught via discovery learning performed roughly one third as well as their peers taught by directed learning.

Given so withering an attack as Kirschner et al one may expect that discovery learning and its siblings would fall from favour within educational establishments. However, Hmelo-Silver et al. (2007) issue a rebuttal in the case of problem- and enquiry-based learning. Firstly, they argue, viewing distinct pedagogies as one and the same is misguided, and that judging problem-and enquiry-based learning as minimally guided is particularly flawed. The scaffolding which Kirschner et al decry as evidence that EBL and PBL are ineffective is fundamental to the process, and not demonstrative of any shortcoming:

> “We agree with Kirschner et al. (2006) that there is little evidence to suggest that unguided and experientially-based approaches foster learning. However, IL and PBL are not discovery approaches and are not instances of minimally guided instruction, contrary to the claims of Kirschner et al. Rather, PBL and IL provide considerable guidance to students.

Secondly, Hmelo-Silver et al cite a large evidentiary base demonstrating the effectiveness of PBL/EBL. Studies have shown that students who experience PBL or EBL score higher than their peers in both university- (Palmer, 2002) and lower-level (Marx et al., 2004) courses.

Looking to investigate the effects of combining discovery learning and direct instruction, Dean Jr. and Kuhn (2007) studied three groups of 10 year old students learning by three different methods. The first two groups were taught as those in Klahr and Nigam's paper, while the third was engaged in discovery learning that was preceded by a direct instruction session on relevant material. It was this third group which outperformed the two taught by a single pedagogy, supporting the case for a combination of direct instruction and EBL where appropriate.

Chapter 4

Problem-based Learning in England and Wales

In this chapter we discuss the place of problem-solving teaching in the curricula of undergraduate degree programmes in England and Wales. This work was motivated by the outcomes of the HE Mathematics Curriculum Summit, held at the University of Birmingham in January 2011, and funded by the National HE STEM Programme as part of the Mathematical Sciences HE Curriculum Innovation Project. The funding was administered by the Mathematics, Statistics and Operational Research (MSOR) subject centre of the Higher Education Academy.

4.1 The HE Mathematics Curriculum Summit and the Problem-solving Project

The HE Mathematics Curriculum Summit took place in January 2011 at the University of Birmingham, organised by the MSOR Network. The heads of mathematics of 26 universities were represented, along with educational representatives from several organisations (IMA, RSS, ORS, the Council for Mathematical Sciences, **sigma**, MSOR and the National HE STEM Programme), and a number of individuals. The principal aim of the summit was to determine the important aspects of a mathematics degree, and to identify opportunities for curriculum development that may then be funded by the Curriculum Innovation Project.

Part of the summit included discussion groups on three aspects of the curriculum, the third of which considered the question "If most maths graduates "aren't confident" in handling unfamiliar problems – should we care?". Chaired by Professor Tony Croft, the report of this discussion includes the following:

> " A different approach to teaching may be required in order to develop skills and confidence in unfamiliar problems. Students can be focused on learning and applying methods and a change is needed so that they begin to think creatively for themselves. — Rowlett, 2011, p. 11

After further presentations, the groups reconvened with the task of identifying priorities for funding projects in curriculum development. This group made the following three recommendations for work in problem-solving:

1. Sharing good practice: Collect case studies for how to embed problem solving into curricula. Develop a good practice guide for problem solving and assessment of problem solving. Consult existing sources, including George Pólya's 'How to Solve it'. Consider the questions: what is a problem and what makes a problem a useful teaching tool? Consider the teaching and assessment of unfamiliar problems and problems that are not

easily solved, including lecturer and student confidence in approaching such problems. Consider the tension between rigorous proof and ways of approaching problems to get a ‘useful’ answer; does insistence on rigorous proof interfere with students’ confidence in approaching unfamiliar problems?

2. Development of a bank of problems with solutions and extensions. Including unfamiliar problems, problems that are not easily solved, problems that have a correct answer but not a single best approach.

3. Development of a collection of teaching resources on the development of mathematics – stories from history and more recent development of the discipline. These should aim to counter a view of mathematics as a static, completed body of knowledge and instead encourage awareness of the process of doing mathematics. They should develop students’ awareness of the culture of mathematics.

After the summit, calls for funding were made for projects to undertake the recommended work (other work related to industry, assessment, skills, and sustainability). The Problem-solving Project, a joint project of **sigma** and the MSOR Network was awarded the majority of the funding to complete the work in problem-solving. The project’s principal investigator was Trevor Hawkes of **sigma**, Dr. Christopher Sangwin of MSOR was a primary contributor, and Matthew Badger, also working at **sigma**, the research assistant.

The first two of the project’s aims were to:

- Using online search, emails, and MSOR and personal contacts, the RA will carry out a survey of the extent to which problem-solving activities inform the curricula in English and Welsh[1] HE mathematical sciences departments. (Note: Problem-solving is cited 5 times in the QAA MSOR subject benchmark.)

- In the light of this survey, we will select a cross-section of different approaches to problem-solving as candidates for case studies (for instance, Birmingham, Warwick, and Manchester are likely targets). Using interviews with practitioners, their learning resources, assessment methods and student feedback, the RA will produce at least 6 (and up to 10) case studies which will be carefully written up and cross-checked to form an integral part of the project output.

The result of this work was the questionnaire and case-studies that we discuss in this chapter. Further case-studies were created for the Moore Method modules at two other universities—Birmingham, and one which will remain anonymous—however each of these is looked at in greater depth in the following part of this thesis.

The project ran from November 2011 to June 2012, its main outcome being the guide *Teaching Problem-solving in Undergraduate Mathematics* (Badger et al., 2012), which included contributions by John Mason, Bob Burn, and Sue Pope. Sue Pope's joint project between Liverpool Hope University and nrich ran parallel to the Problem-solving Project. The project website, including problems and a PDF version of the guide, is at `http://mathcentre.ac.uk/problemsolving`.

[1]The National HE STEM Programme does not cover Scotland or Northern Ireland.

4.2 The Questionnaire

The first project task was to determine the extent to which explicit problem-solving teaching is found in university mathematics departments in England and Wales. Given the tight timescale, an online questionnaire distributed to heads of departments via email was decided on as the fastest and most effective approach.

A list of mathematics departments was collated from the information available on the HoDoMS website (`http://www.coventry.ac.uk/ec/HODOMS/`); email addresses for the heads of each department were then found via their websites. A questionnaire was authored, kept short to encourage a large number of responses, and distributed with Google Docs. Because interpretations vary, it began with the following description of problem-solving as:

> “ ...any substantial task or activity that calls for original, lateral or creative thinking by students; brings several ideas or techniques together in a surprising way; introduces something new; illuminates some topic, e.g. with a helpful counterexample.

This was followed by five questions on problem-solving in the respondent's department:

1. Does your institution offer a module in any of its Mathematics Degree Programmes which requires students to engage in problem-solving?
    - No
    - Yes, an optional module
    - Yes, a compulsory module
2. What is the module code and title (choose the best example in your programme)?
3. Is problem-solving the central aim of the module?
4. What year is the module taken in?

5. Please give a brief description of the module you offer.

Respondents' department and contact details were asked for to determine who had answered the quesitonnaire, this was also used for follow-up contact with those who agreed to it.

Individual emails were sent to the heads of department of the 59 English and Welsh HEIs offering a Mathematics BSc. These included a brief description of the project and a link to the questionnaire. After a month, attempts were made to contact by telephone those who had not responded to either the email or questionnaire.

4.2.1 Summary of Questionnaire Responses

The questionnaire was completed by 35 heads of department or their representatives. Responses indicated an absence of explicit problem-solving teaching in the mathematics degrees of English and Welsh HEIs. Twelve of the respondents acknowledged that they had no explicit problem-solving in either compulsory or optional modules. In half of these cases, problem-solving was mentioned in the published aims of the programme.

Of the remaining 22 departments, 5 offered an optional module to students while 17 stated that problem-solving was included in one or more compulsory modules. From the module descriptions in the responses, and further information from departments' module handbooks, these were divided as follows:

- Seven modules on modelling, which were seen as distinct from problem-solving.
- Five project modules (four of them final-year), involving students researching and reporting on various topics, but not explicitly solving problems.

- Three modules on operational research, numerical methods, and numerical analysis that again did not include explicit problem-solving.
- One response naming no particular module but stating that problem-solving was inherent to a mathematics degree.
- Six modules (four optional, two compulsory) where explicit problem-solving was taught as the central aim of the module.

The responses to the questionnaire demostrate the difficulty in discussing ‘problem-solving’ in the context of a mathematics degree. The description we used was necessarily narrow, given the project remit, and while we used this in the questionnaire ‘problem-solving’ was interpreted more broadly by many respondents. The six modules identified as including a significant amount of explicit problem-solving in the sense that we intended subsequently became the subjects of our case-studies. They were as follows:

University of Birmingham Problem-solving is taught to first-year students in their first-term using the Moore Method. There are no lectures, and the only materials students are given are the problems they are required to solve and definitions they might need to solve them. No examples or model answers are available to students. There are currently two groups using questions on different topics – geometry and set theory. Students do not get to choose which group they are in. They meet for two hours per week to present and discuss their solutions. The module is optional for BSc students but now expected for students enrolled on the 4-year Mathematics MSci programme. This is discussed in Chapter 6.

Durham University Delivered in the first-term of the first-year, Durham's compulsory module, entitled *Problem-Solving*, is based on *Thinking Mathematically* (Mason et al., 2010). It is taught using both lectures and problems classes, in which students work through problems in groups, recording the progress they are making by using *rubrics*[2].

University of Macondo (Pseud.) Macondo's compulsory second-year module, *Investigations in Mathematics*, sees students working in groups of around a dozen on different mathematical topics at an appropriate level. It is taught using a modified Moore Method by several members of staff working in parallel. This is discussed in Chapter 7.

University of Manchester The *Mathematics Workshop* is a first-year, first-term module again using *Thinking Mathematically*, though to a lesser degree than at Durham. Students begin the year working in computer labs, before concentrating on modelling and problem-solving after the mid-term break. A compulsory module, the *Workshop* has both lectures and classes.

Queen Mary The only third-year module in the study, Queen Mary's *Mathematical Problem-Solving* is an optional module taken by a dozen students each year. Each student has a different set of problems from a range of topics in pure mathematics. Although there are no lectures and only a single class each week, students may seek help from members of staff at other times.

Warwick Warwick's first-year *Analysis 1* module has been taught using problem-solving for the past 15 years. This unique example among our case-studies is a core module taken by

[2]Discussed in Section 4.6

all its 320 first-year mathematics students, taught by problem-based learning. Students have one lecture and four hours of group work in problems classes each week.

Five of the universities deemed to have an eligible problem-solving module were members of the Russell Group, or (in Queen Mary's case), *about* to become members. Macondo is a member of the 1994 Group of universities. Were the questionnaire to have had more responses, we may have found other suitable modules and been able to choose a better cross-section of UK mathematics departments. As it is, we must note that this is the case and use those data available.

4.3 Case Study Methodology

The six modules above were identified as case-study candidates from the questionnaire reseponses. In each instance, the person who completed the questionnaire was happy for further contact to be made, and each subsequently agreed to an interview to document the problem-solving module in their department. Interviews were semi-structured, to ensure that all aspects of the modules were covered, and allowing the member of staff to make comments outside the questions we had. These interviews were conducted and transcribed by Matthew Badger in March and April 2012.

Interviews and available departmental literature (module descriptions, problem sheets, etc.) were used to record the teaching practices in each of the six modules. As each module was taught using a distinct approach, comparison between modules was limited. Further, the efficacy of each module was only inferred—in all cases but Warwick's—from the opinions of the member of staff and not from other analyses.

4.4 Approaches to Teaching Problem-solving

Subsequent sections in this chapter deal individually with the four non-Moore Method modules at Durham, Manchester, Queen Mary and Warwick. Here we give an overview of some of the similarities and differences between all six modules.

We first note that teaching problem-solving is not the primary aim of the module at Warwick, rather it uses problem-solving to teach first-year analysis. Instead of a traditional lecture course (three hours of contact a week in a large lecture theatre), the students have a 1-hour lecture and two 2-hour classes each week. In these classes, students work at tables in small groups through a series of ten workbooks, that contain the definitions and problem sequences that form the bulk of the material they need to learn. Students from other departments who take *Analysis 1* in their first-year are taught by the standard lecture method, and do not perform as well on the final examination as their peers in mathematics (Alcock and Simpson, 2001).

The inclusion of lectures is not unique to Warwick; modules at Durham and Manchester also use lectures for some of their teaching. Here we come upon the issue of resources: Macondo has the fewest students with around 100 a year, while the largest, Warwick, has as many as 320 to cater for. They need far more teachers, rooms, time-table slots, and administration than a standard lecture module, and to help reduce the teaching burden supplementary lectures are used at these three institutions. Table 4.1 outlines the use of resources in our six case-study modules, and clearly demonstrates that a compulsory module taught using problem-based learning in a large department imposes a considerable burden on the organiser.

Module	Classes	Class Size	Staff	Assistants	Class hours	Lecture hours
Birmingham	2	14	2	0	2.5[1]	0
Durham	7	18	7	0	2	1
Macondo	9	10	9	0	1	0
Manchester	9	30	9	9	2	0.5[2]
Queen Mary	1	12	1	0	1	0
Warwick	11	30	11	11	4	1

Table 4.1: Division of labour in six problem-solving modules

All six modules focused on students solving problems for themselves, though each approach was distinct. Students at Birmingham and Queen Mary work individually; the other four modules include a group-work component. At Birmingham, collaboration occurs in the classroom, while at Queen Mary each student has their own set of problems to work through, preventing plagiarism. At the four universities that include group work in their problem-solving activities, those at Durham and Warwick have no group component to their mark, while Macondo and Manchester have only 30% (from a poster and presentation) and 10% (from completed projects' group average) respectively.

Besides solving problems, students at Warwick learn the necessary material from the notes that they are presented with. At Macondo students must research their topic, which contributes to their individual workbooks and the group poster and presentation. Both Durham and Manchester reference *Thinking Mathematically* (Mason et al., 2010); at Durham students

[1] This represents two and three hours of classes, alternating weekly.

[2] This represents a single, one-hour lecture a fortnight.

work through some chapters closely. This is done to encourage students to document carefully the process by which they solve problems; at Durham students are also marked on their approach to solving problems. At Birmingham students present their solutions to problems to their peers, though this is seen as an integral part of the problem-solving process. At Queen Mary problem-solving is the sole activity.

With the exception of Warwick, all modules use submitted solutions to problems as a large part of their assessment. At Birmingham and Queen Mary, straightforward solutions to problems make up 50% and 100%, respectively, of the module credit. At Durham 25% is awarded for a problem solution, and a further 50% for a project including an extended problem. At Macondo students' workbooks, including solutions to problems, are worth 50% of the module mark, while at Manchester project reports are worth 60% total, of which 40% comes directly from problem solutions. At Warwick, students submit solutions to problems on a weekly basis, and these are marked but not used for assessment. With the exception of Warwick again, the ways in which the modules were assessed differs greatly from the majority of mathematics university modules in England and Wales (Iannone and Simpson, 2012a, p. 5).

Presentations are used at Birmingham and Macondo as a key component of the Moore Method. The module with the largest presentation component is *Developing Mathematical Reasoning* at Birmingham, where students receive 50% of their module marks for their best two presentations of solutions to problems. At Macondo, students in groups give presentations on the topics that they have been working on for 20% of their final mark.

4.5 Case Studies

We dedicate the remaining sections of this chapter to documenting the problem-solving modules at each of the four universities of Durham, Manchester, Queen Mary, and Warwick. Chapters 6 and 7 cover the modules at Birmingham and Macondo, respectively, in greater detail.

It is important to note that each of these case studies resulted from reviewing the module literature, and interviews with lead teaching staff. As such, these case studies do not constitute an objective analysis of the modules discussed, and so where opinions of their effectiveness are expressed by staff, we are not in a position to attest to their veracity. The realities they describe are the realities for the staff teaching them.

4.6 Durham University

Problem Solving is a first-year, first-term module in Durham University's Department of Mathematical Sciences[3]. The module has been run since 2006/07 and aims to teach students to solve problems and to write proofs, and draws much inspiration from Mason et al's *Thinking Mathematically* (2010). Students spend time discussing the ideas in the book, besides solving problems themselves. Its entry in the first-year module handbook is as follows:

[3]*Problem Solving* is the first half of the two-term module *Problem Solving and Dynamics*. This is purely for administrative purposes; mathematically, the modules are entirely separate.

> This module gives you the opportunity to engage in mathematical problem solving and to develop problem solving skills through reflecting on a set of heuristics. You will work both individually and in groups on mathematical problems, drawing out the strategies you use and comparing them with other approaches.

4.6.1 Teaching

Problem Solving is compulsory for all of the 130 first-year mathematics undergraduates at Durham. The cohort is divided arbitrarily into seven groups of around 18 students, each taught by a separate member of staff. Teaching alternates weekly between a one-hour lecture followed by a one-hour seminar, and a single, two-hour seminar. The basic materials for the module are a set of problems that students work through, both individually and in groups.

Lectures are used to draw out ideas that have been presented by students during the seminars, or in their submitted answers to the problems in the module. Lectures refer to relevant sections of *Thinking Mathematically*, and students are encouraged to read these along with working on the module's problems.

Two key ideas that students take from *Thinking Mathematically*—and on which the book has a lot to say—are the *schema* and a *rubric*. The *schema* is used to describe the process of problem-solving itself, which it divides into three phases: 'entry', 'attack', and 'review'. These three phases bear close resemblance to Pólya's four-step process of:

1. Understand the problem.

2. Make a plan.

3. Carry out the plan.

4. Looking back.

Students are taught to identify which phase of the problem-solving process they are in, in the hope that this will give them a clearer idea of how to proceed. The process is not linear, however, and students may return from the attack phase to the entry phase if an approach appears not to work.

A *rubric* is the process by which students document their *own* problem-solving process, with reference to Mason et al's schema. At each phase students write down their thoughts while working on a problem, reflecting on previous problems' rubrics to prompt thinking on the current problem. For example, during the *entry* phase, students are encouraged to write "I know X", where X is something that they are given by the question, and "I want Y", where Y is something that they are looking to find. The intention in getting students to write these rubrics is so that they develop their own 'internal monitor', allowing them to make the most of their experience in problem-solving, and to give a more structured approach to later problem-solving endeavours.

In the first two lectures students are introduced to the schema and rubrics, while later lectures cover each phase of the problem-solving process in more detail. Lectures also cover related topics, such as how to write mathematics formally.

Seminars constitute three-quarters of the module's contact time. In these, students begin by discussing a problem they will have been given earlier in the week, before moving on to new problems. New problems are discussed and students present potential solutions to them;

they are also encouraged to suggest their own extensions of the problems. Most problems will take twenty minutes or more to be discussed, solved, and extended, though a few each year take the whole of a two-hour seminar. The point at which the group moves on to a new problem is at the discretion of the seminar leader, as appropriate to the level of interest or progress made by the class. Problems are selected for each individual seminar, enough to ensure that students will not run out, and those problems are not reached in that seminar are not used in following seminars.

The ability to write a correct mathematical proof is one of the key aims of the module. Students are required to apply what they have been taught in the lectures, and discussed in the seminars, to create proofs that are largely correct, in an appropriate format and with suitable mathematical language (including correct grammar).

Solutions to some problems are uploaded to the university's intranet for students to review once they have submitted their own solutions during seminars. For other problems, staff upload their own rubrics, for students to see that the problems are not necessarily immediately solvable, even for experienced mathematicians. New problems are included each year that the teaching staff do not attempt to answer before students do.

Group Work

Students are encouraged to work together in seminars, though this is not compulsory; if a student wishes to work alone they are free to do so. The majority choose to work with their peers.

4.6.2 Assessment

The module mark is based on three pieces of submitted work. The first, worth 25% of the total, is the submission of a rubric that explains how they approached solving a particular problem, the choices they made and their reasons for doing so. The second piece of work, also worth 25%, is a solution to a particular problem that has to meet the standards of a substantially complete mathematical proof. The final piece of work is a project done over the Christmas holiday on an extended problem from the module; this is submitted at the beginning of the mock exam period after the Christmas break.

For its first three years, the second 50% of the module mark was awarded for an examination on a subset of the problems in the module. Because this relied on examined problems being reached by all groups, students were able to determine those problems that would be on the exam, and to learn their solutions by rote. The project work was put in place to prevent such problems occurring.

4.6.3 Motivations and Development

Adrian Simpson joined Durham from the University of Warwick in 2006/2007, becoming the Principal of Josephine Butler College and a Reader of Mathematics Education. He joined the university with the expressed purpose of setting up the new module. He had taught a similar module in the third-year of Warwick's mathematics degree programme for nearly a decade, and taken the same module as an undergraduate. Dr. Simpson was previously involved with the first-year analysis module that used problem-solving, described (with Lara Alcock) in

(Alcock and Simpson, 2001) and in Section 4.9.

The aims of the module are to develop students' problem-solving skills, to improve their mathematical writing, and their rigour in creating proofs. *Thinking Mathematically* is used to help students understand the process by which problems may be solved, and to improve their confidence when approaching unfamiliar problems. Besides the change from an examination to a project for its assessment, the module is now taught to more groups of fewer students. There have not been any other changes, though Dr. Simpson is no-longer the module leader.

4.6.4 Problem Selection

Problems are drawn from a number of sources, notably *Thinking Mathematically*, though these have been further developed over the time that the module has been running. Other problems have come from correspondence with colleagues in the department and beyond. Problems are not related to one-another, though most are from pure mathematics and, in particular, number theory. The module has no prerequisites and problems are not dependent on learning from other modules in the first term.

▼ **Example 4.6.1 — *Square Differences*.**

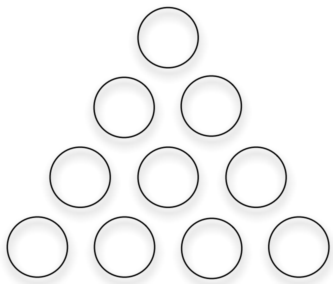
Which numbers can be written as the difference of two perfect squares? E.g.

$$6^2 - 2^2 = 32.$$

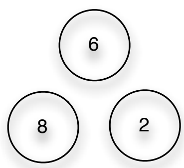
Example 4.6.1 is the type of question that would take students around 20 minutes to solve in their groups, and a little more time outside of a seminar to write a correct proof. Example 4.6.2 below shows a problem that students will be given for homework, in addition to completing the proofs of problems given in seminars. Here they would not be expected to find a general solution for the numbers 1 to n where n is a triangle number.

▼ **Example 4.6.2 — *Triangle of Differences*.**

Consider the array below:



Can you enter the numbers 1 *to* 10 *into the array so that each number is the difference of the two below it, e.g.*



A problem that usually takes students two-hours to complete an investigation of is given in Example 4.6.3. It also demonstrates that not all the problems are 'questions' as such, but aim to prompt open-ended investigation by students.

▼ **Example 4.6.3 — *Fibonacci Factors*.**

The Fibonacci sequence F_n is defined by $F_1 = F_2 = 1$ and $F_{n+2} = F_{n+1} + F_n$ for all positive integers n.

Look at what factors the sequence has. By choosing different starting numbers, or the recurrence relation we can generalise Fibonacci sequences, e.g. Lucas numbers have $L_1 = 2, L_2 = 1$ and $L_{n+2} = L_{n+1} + L_n$; Pell numbers have $P_1 = 1, P_2 = 2$ and $P_{n+2} = 2P_{n+1} + P_n$.

Regarding this particular problem, Adrian Simpson recalled the following story:

> “In both directions; occasionally you see students will struggle with basic stuff. Every lecturer says this. But I do remember very vividly this Fibonacci factors problem where you have, eventually you show that F_n is always divisible by F_k for all n, and that's a big double induction. And this girl was working on it and working on it, she was really struggling, but eventually she got it, and I was sitting and chatting to her saying how impressed I was, it was really good, saying "Have you ever done a double induction before?" "A double what?" "A double induction. You've done this induction there and then you've done this induction that fits inside it." "What's induction?" "You must have done induction!" It turns out she had never done further maths so she had never done induction. She had effectively invented induction for herself and in doing so did it as a nested induction, one

inside the other. And I was just going "My God that's just brilliant!".

4.6.5 Outcomes and the Future

Dr. Simpson reports that students, in general, enjoy the module. Having no previous experience of university level mathematics, students in general take very well to the module—university mathematics is "what we tell them it is". There are always a few who fail to grasp the ideas contained in *Problem Solving*, for whom the module is less enjoyable, however these students make up a small minority of the cohort.

Having led the module for the last five years, Adrian Simpson is moving to the teaching of a different module next year, and so its leadership will be passed onto another member of staff, albeit one who has taught *Problem Solving* before.

4.7 University of Manchester

The first-year, first semester, module *Mathematics Workshop* has been running at the University of Manchester since the merger of Victoria University of Manchester[4] and University of Manchester Institute of Science and Technology (UMIST) in 2004. It aims to develop students' modelling and problem-solving skills, to improve their ability to write and present mathematics, and to teach them to use the software MATLAB. The module is compulsory for all 270 single and some joint honours mathematics students, who take the module in ten separate groups,

[4]Victoria University of Manchester was colloquially known as the University of Manchester.

each led by a postgraduate facilitator. Run by Dr. Louise Walker, the module is described in the first-year handbook as follows:

> These weekly classes are intended to help students experience a wide range of mathematical topics. The course unit includes a number of projects to be worked on individually and in groups. The projects are assessed via a written report. Marks will be awarded for presentation as well as mathematical content to encourage the development of good writing habits.

4.7.1 Teaching

There are a one hour lecture and a two hour workshop every week; the style in which the workshops are taught varies over the course of the module. The content is as follows:

- Week 1: Introduction to the course
- Weeks 2–5: Introduction to MATLAB and numerical methods
- Week 6: Mid semester break
- Week 7: Introduction to modelling and problem-solving
- Weeks 8–9: Project 1
- Weeks 10–11: Project 2
- Week 12: Test

During the first five weeks of the module, workshops are held in computer labs where students work on individual problems. Here, the focus is not so much on problem-solving but rather becoming familiar with using MATLAB, which students then use later in the module to help solve the problems they are given. During these weeks, assessment is by worksheets in workshops and homeworks, each of which are marked by hand by the postgraduate facilitators.

Following the mid-semester break, the focus of the module shifts to modelling and problem-solving. After an introductory week, students complete two group projects lasting two weeks each, covering topics from applied mathematics. In 2011/12, the first project covered difference equations; the second, graph fitting. *How to Solve It* (Pólya, 1945) and *Thinking Mathematically* (Mason et al., 2010) are set texts for the module, and students are referred to relevant chapters in the latter book at appropriate times in lectures.

Lectures cover mathematical material and include an example of the type of problems included in workshops. Workshops include two problems, the first of which will be a relatively straightforward example close to those seen in lectures; the second of which relies on modelling and an exploration of the model's limitations.

Students work in groups in the two-hour workships to answer problems. Most complete draft solutions in that time. Postgraduate facilitators spend time with each of the five or six groups in their room to ensure that students understand the problems and are getting on with work.

Students complete solutions to problems for home work; these then form part of their project report. They also complete a separate homework task, often using MATLAB to model scenarios from the second worksheet problem. Completing the project report is also included as homework.

Group Work

Group work is central to *Mathematics Workshop*. While students submit individual reports for each of the projects that they complete during the module, 5 of the 30 marks for each project are from the average for the group. Students are warned that:

> Even though you have worked as a group on this project, the report should be **all your own work**. There are marks for the clarity as well as the correctness of your mathematical arguments.

4.7.2 Assessment

Project reports comprise 60% of the module mark, each is worth 30%. These are marked by the postgraduate facilitators, and reviewed by Dr. Walker. The remainder of the marks are awarded for MATLAB assignments and the class test in the final week of term. The breakdown of marks is as follows:

- MATLAB Assignments – 30% total
- Two project reports – 30% each, made up from:
  - 20% for solutions to problems
  - 3% for homework
  - 2% for the presentation of the report
  - 5% from the group mean of the first three items

- Class test – 10%

Solutions to worksheet problems form most of the content of the project reports.

4.7.3 Motivations and Development

Mathematics Workshop is derived from a previous module that was taught to first-year engineering students at Victoria University of Manchester. In 2000, the School of Engineering began teaching its first-year undergraduates primarily using problem-based learning, and the mathematics service module provided by the School of Mathematics was required to conform to the new standard. There the format of one one-hour lecture, one two-hour workshop was begun, and proved popular with students who had previously learned mathematics from a standard lecture-based module.

Dr. Walker had been involved with the service module from its inception, and when Victoria University of Manchester merged with UMIST in 2004, she was keen to introduce a module along similar lines for first-year mathematics students. As a result of the two universities' merger, the mathematics degree programme was redesigned and finding space within an existing programme was not an issue. The new department's teaching committee saw an opportunity to introduce a module that covered a wider range of skills than were in a typical mathematics course. The aim of the new module was not to teach a particular area of mathematics, but to give students an apprenticeship in doing mathematics for themselves.

Both the group-based teaching and project-based assessment of the workshop satisfy a number of internal assessment criteria for the undergraduate mathematics degree programme as a whole, specifically the departmental employability audit.

4.7.4 Problem Selection

All problems are written specifically for the module; the topics for which are chosen each year by Dr. Walker. Problems in early iterations of the module frequently proved too challenging for students. Since then the difficulty of problems has been reduced, in order to encourage students to complete the work as much as possible without intervention from teaching staff.

In 2011/12 the module had no pure mathematics problems, partly as a result of its involvement with an HE STEM Programme project on modelling, run by Prof. Mike Savage at the University of Leeds; in 2012/13 a number of pure topics were to be reintroduced. Example 4.7.1 gives the main problem from the first week of a 2011/12 project on difference equations.

▼ **Example 4.7.1 — *Difference Equations*.**

We want to study the populations of rabbits and foxes that live in a certain region, from a given starting point. Let the population of rabbits after n months be denoted by r_n *and the population of foxes be* f_n*. The initial populations are* r_0 *and* f_0*.*

We model the populations using the following coupled first order difference equations:

$$r_{n+1} = ar_n - bf_n$$

$$f_{n+1} = cf_n + dr_n$$

where $a; b; c; d > 0$ *are parameters.*

1. *Explain why these equations seem a reasonable way to model the problem. What real life properties of the problem could the parameters* $a; b; c; d$ *represent?*

2. *Eliminate* f *from the set of equations to find a single linear second-order difference equation in* r*. Solve this equation to get a formula for* r_n *in the case when* $a = 2;$ $b = 1 : 3;$ $c = 0 : 5$ *and* $d = 0 : 4$ *with initial populations* $r_0 = 500; f_0 = 100.$

3. *What are the limitations of this model?*

4.7.5 Outcomes and the Future

Dr. Walker reports that most students enjoy the module, though some are uncomfortable with the teaching style, and in particular the open-endedness and unfamiliarity of problems. Each year, every module in the department is assessed by a focus group to address concerns and improve outcomes for future students. While a self-selecting, the students who attend the focus group are generally positive about the module.

Managing the module takes considerable time investment, but Dr. Walker is clear that she thinks this pays dividends for students' learning.

4.8 Queen Mary, University of London

Mathematical Problem-Solving is a third-year undergraduate module at Queen Mary, University of London. It aims to develop students' problem-solving skills, their creativity and logical thought processes, and proof writing. The module does not have explicit prerequisites but some problems require knowledge from core second-year modules to be solved. *Mathematical Problem-Solving* is an optional module, taken by around a dozen students a year and currently taught by Dr. David Ellis. The module is advertised to students in the third-year handbook as follows:

> "The module is concerned with solving problems rather than building up the theory of a particular area of mathematics. The problems are wide ranging with some emphasis on problems in pure mathematics and on problems that do not require

> knowledge of other undergraduate modules for their solution. You will be given a selection of problems to work on and will be expected to use your own initiative and the library. However, hints are provided by staff in the timetabled sessions. Assessment is based on the solutions handed in, together with an oral examination.

4.8.1 Teaching

Each student who takes the module is given a set of a dozen problems taken from a range of topics in pure mathematics. To prevent plagiarism, each pupil is given a different set of problems, and problem sets have roughly equal numbers of problems on geometry, combinatorics, analysis and number theory. Complexity theory and combinatorial geometry problems are also present in some problem sets. The only check that students' work is there own is an oral examination, discussed in the following section. Students enrolled on the module are given practice problems to try during the Christmas holiday, to give them an idea of the structure and difficulty of the module.

Example 4.8.1 is a number theory question from the module that is worth 8 marks, the sixth in its particular progression.

▼ **Example 4.8.1.**

Do there exist three natural numbers $a, b, c > 1$, such that

$a^2 - 1$ *is divisible by* b *and* c*,*

$b^2 - 1$ *is divisible by* a *and* c*, and*

$c^2 - 1$ *is divisible by* a *and* b*?*

A one-hour class takes place each week, where students can get direction from the module leader on particular problems, or work on problems while others ask questions. Students are encouraged to contact the module leader in person and by email if they require assistance outside of class. Other members of staff may offer students assistance with a problem, in which case they report to the course leader exactly what assistance was given.

Students receive problems individually, with five or six days between problems; if they submit a solution to a problem early, they will be given the next problem to work on. Solutions can be submitted to the module leader at any time, and these will be returned with appropriate comments. If there is an issue with a student's solution they can work on it more and resubmit their solution with no mark penalty. Solutions which students submit are expected to be mathematically complete, to a reasonable degree as defined by the module leader. There is no group work in the module.

4.8.2 Assessment

Each problem in *Mathematical Problem-Solving* is worth between 6 and 17 marks, a student is awarded those marks if they produce a correct solution. Marks are given for partial solutions, and the marks for a correct solution may be reduced if a student is given assistance, at the teacher's discretion. The final mark for the module is the sum of the marks gained from a student's ten highest scoring solutions, thus if a student's ten highest scoring answers are worth seven marks each, they get a score of seventy. The distribution of available marks means that to get 100% for the module a student will need to answer several questions of considerable difficulty worth 13 or more.

Final solutions to all twelve problems are submitted at the end of the first week after Easter. Thus, students will have twelve weeks working on the problems during term time, and four further weeks of the Easter holiday to refine their solutions (though no assistance is given by staff once the Easter break begins). There follow individual oral examinations to ensure that students understand their solutions and that their work is their own. No marks are awarded for the exam, though it gives the opportunity for the module leader to clarify ambiguities in a students' solutions and adjust marks accordingly, if needed. The oral examination normally takes a quarter of an hour; there is no written examination.

Model solutions to problems are never handed out, the only solution a student sees is the one they produce.

4.8.3 Motivations and Development

Mathematical Problem-Solving has been running in its current form for over 20 years, though some of the problems have changed. Professor Charles Leedham-Green created the module in the late 1980s to give students an experience of learning mathematics by a means other than lecture-based modules; the purpose and format of the new module was agreed by the then Head of Department. Problems were intially shared between students, but have been individual to each student for more than 15 years. During its early years, the module was the only one in the department that consistently required no scaling [private correspondence]; furthermore students frequently produced solutions to problems that were unique.

4.8.4 Problem Selection

The problems currently used in *Mathematical Problem-Solving* are a superset the originals, added to over the intervening years by the members of staff who have lead the module. Individual problems generally consist of one or two parts; in problems with two or more parts, these individual parts are related in some way. The following is an eight mark question with two parts; the first of which is relatively straightforward but which prompts one to consider possible solutions to the second part:

▼ **Example 4.8.2.**

Find all positive integers n such that $(n-1)!$ is not divisible by n. Repeat this for n such that $(n-1)!$ is not divisible by n^2.

The following example is worth 15 marks, and presents a considerably greater challenge to students. Here it may not be immediately clear what the question is asking:

▼ **Example 4.8.3.**

If $2n$ points are evenly placed on the circumference of a circle, and then a (non-convex) $2n$-gon is drawn joining up these points, prove that it must have two parallel sides.

4.8.5 Outcomes and the Future

The module is aimed at the strongest students. Dr. Ellis reports that most who take the module perform relatively well and a few excel, requesting new problems frequently and scoring close to full marks. Some struggle with the module, however, while others are put off by the practice problems over Christmas. Students who need the most assistance accept that they will not get full marks for a correct solution to a problem, and so their potential mark for the module as a whole will be lower.

The module is popular among teaching staff, not least because, with a good bank of problems from which to work, the time required to teach the module is much less than that of a standard lecture course. Furthermore, staff recognise the benefits that the module brings to students who take it and work hard.

4.9 University of Warwick

Analysis 1 is a first-year, first-term, analysis module at the University of Warwick, taken by all 320 single-honours mathematics undergraduate students and taught using a problem-solving approach. The module is followed in the second-term by *Analysis 2*, a traditional lecture-based module.

Analysis 1 was the subject of Alcock and Simpson's chapter *The Warwick Analysis Project: Practice and Theory* (Alcock and Simpson, 2001), that explored the origins of the module, the experiences of students and teachers, and its efficacy.

The first-year handbook advertises the module to students as follows:

> " With the support of your fellow students, lecturers and other helpers, you will be encouraged to move on from the situation where the teacher shows you how to solve each kind of problem, to the point where you can develop your own methods for solving problems. You will also be expected to question the concepts underlying your solutions, and understand why a particular method is meaningful and another not so. In other words, your mathematical focus should shift from problem solving methods to concepts and clarity of thought.

4.9.1 Teaching

The module is taught with a single, one-hour lecture and two, two-hour classes every week. Lectures proceed in the traditional manner, though given that there is only one lecture a week, time during them is spent on the most important concepts and difficult proofs that students will have to tackle during classes. Each week students are given a workbook of roughly ten pages, and a set of assignments to complete for handing in the following week.

A single weekly lecture caters for all 320 students who take the module; in classes students are divided arbitrarily at the beginning of the module into groups of around 30. Each group is assisted by a postgraduate student or member of teaching staff, and a second- or third-year undergraduate student, whose jobs are to prompt students that are stuck, to answer questions (without revealing answers), and to encourage students as necessary. The first class of a week is before the lecture, and so students first experience the module material in classes.

Workbooks are distinguished from traditional lecture notes by much of the theory being contained in the assignments that students complete. Example 4.9.1 is taken from the second page of the fifth workbook, covering completeness:

▼ **Example 4.9.1 — *Completeness.***

Theorem

Between any two distinct real numbers there is a rational number.

[i.e. if $a < b$, there is a rational $\frac{p}{q}$ with $a < \frac{p}{q} < b$.]

Assignment 1

Prove the theorem, structuring your proof as follows:

1. *Fix numbers $a < b$. If you mark down the set of rational points $\frac{j}{2^n}$ for all integers j, show that the one lying immediately to the left of (or equal to) a is $\frac{[a2^n]}{2^n}$ and the one lying immediately to the right is $\frac{[a2^n]+1}{2^n}$.*
2. *Now take n large enough (how large?) and conclude that $a < \frac{[a2^n]+1}{2^n} < b$.*

Corollary

Let $a < b$. There is an infinite number of rational numbers in the open interval (a, b).

Assignment 2

Prove the corollary.

We have shown that the rational numbers are spread densely over the real line. What about the irrational numbers?

Group Work

While the entire cohort is initially divided into groups of 30, further subdivision happens between students organically. Students are encouraged to work in smaller groups however, and over time this generally results in three or four students working together in classes. Each group is visited either by the postgraduate or undergraduate assistant to make sure that they are working on the problems.

4.9.2 Assessment

For the purposes of assessment, *Analysis 1* and *Analysis 2* are considered a single, two-term module. The majority of the marks for this combined module, 60%, are awarded for the three-hour end-of-year examination, which includes some questions relating to *Analysis 1*, though mostly focused on the second-term. The total credit for the module is made up as follows:

- *Analysis 1* – Mid-term exam: 7.5%
- *Analysis 1* – Final Examination: 25%
- *Analysis 2* – Weekly assignments: 7.5%
- Final examination: 60%

The mid-term examination, covering the first five weeks' content, is taken by students during a two-hour class. The final exam is taken in the first week of the second-term and lasts 90 minutes.

Besides the mid-term and final examinations, students have weekly assignments in *Analysis 1* that they hand in at the start of class on the following Monday. These assignments are those from the workbooks, numbering around sixteen.

4.9.3 Motivations and Development

The origins of *Analysis 1*'s existence as a problem-based learning module are documented by Alcock and Simpson (2001):

> "Recognising that students continued to perform poorly despite his efforts at improved presentation, one member of the Warwick staff with over 30 years of traditional teaching experience chose to adopt a new lecturing style. In an initial modification of a second-year metric spaces course, he encouraged students to develop much of the mathematics for themselves by posing central, illuminating questions. While this course met with limited success and even more limited appreciation from the students, he persevered to initiate the Warwick Analysis Project, approaching two members of the mathematics education group to help teach it in the first-year.
>
> The project is based on students' completion of a carefully structured sequence of questions, through which the syllabus of the Analysis 1 course is developed. An excellent textbook bringing together mathematical, pedagogical and mathematics education research knowledge had been developed over a number of years at another university and this text (Burn, 1992) is central to the course.

In 1996/97 the new-style module was piloted with 35 volunteer students and, after positive results, was taken by all 235 single-honours mathematics students from 1997/98. Roughly 150 joint-honours students continued to be taught in the traditional manner, and as a result comparisons could be drawn between the examination performances of the two cohorts.

	Foundations	Analysis 1	Analysis 2
New Module Students	60.0	76.5	63.5
Traditional Module Students	57.7	64.6	51.1

Table 4.2: Examination results in Alcock and Simpson (2001, p. 106)

Examination results for single- and joint-honours students taking the new- and traditional-style modules are given in Table 4.2. The *Foundations* and *Analysis 1* examinations were those taken in the first week of the second-term (the same exam structure remains today); the *Analysis 2* examination is from the third term. Comparisons between assessments within these modules were not made because 1) assignments were different in the new and traditional modules, and 2) both analysis modules are counted as one for the purposes of a final mark. While one may expect the examination marks of single-honours students to be higher than those of joint-honours students, the difference is dramatic in the analysis modules, and furthermore *Foundations* marks—a module where both cohorts are taught in the same manner—are much closer to one-another. Covering only one year's students, who studied different material to subsequent cohorts, with data now over 15 years old, this study holds little relevance to the current module, however.

4.9.4 Problem Selection

As mentioned in the previous section; the first few iterations of the new-style *Analysis 1* used Burn's *Numbers and Functions: Steps into Analysis* (1992) as the source of the workbooks and assignments. Original notes were then written for the module by Alyson Stibbard, who was at the time a member of the teaching staff at Warwick; these workbooks were tailored to the curriculum at the Mathematics Department. They have been used, largely unchanged, since 1998/99. Problems in these workbooks come directly from the material being taught; they are not selected so much as they are necessary elements of the process.

Besides the sections of exposition and problems that build up students' understanding of analysis, all but the first workbook finish with one or two sections on applications. For instance, the fourth workbook covers more advanced sequences (continuing on from Workbooks 2 and 3), and ends with a section on the application of the theory of sequences of numbers to sequences of more general objects, in the case of Example 4.9.2, polygons.

▼ **Example 4.9.2.**

Example *Let P_n be the regular n-sided polygon centred at the origin which fits exactly inside the unit circle. This sequence starts with P_3, which is the equilateral triangle.*

Let $a_n = a(P_n)$ be the sequence of areas of the polygons. In Workbook 1 you showed that $a_n = \frac{n}{2}\sin(\frac{2\pi}{n})$ and in Workbook 2 you showed that this converges to π as $n \to \infty$ which is the area of the unit circle.

Exercise 6 *Let $p_n = p(P_n)$ be the perimeter of P_n. Show that $p_n = 2n\sin(\frac{\pi}{n}) = 2n\sin(\frac{2\pi}{2n}) = 2a_{2n}$.*

Assignment 14 *Show that $\lim p_n = 2\pi$.*

4.9.5 Outcomes and the Future

The sole lecturer for *Analysis 1* is Dr. John Andersson. Classes are assisted by both postgraduates, and second- and third-year undergraduates who themselves performed in the module. Dr. Andersson reports that, in general, students engage well with the module and participate in class discussions, though in the first weeks they may be reticent to do so. The teaching method appeals to Dr. Andersson as, he says:

> “...it’s impossible to teach people mathematics; I think the only way you learn mathematics is by doing it. However, I think enthusiasm is important, because they don’t have to turn up at lectures. If your lectures aren’t interesting then they

> won't turn up, so hopefully you have something to say in the lectures that can help them in the right direction, in developing.

The module's source material has changed little since original problems were written in 1998/99. The teaching method is unmodified and, according to Dr. Andersson, single-honours mathematics students still perform appreciably better (an estimated 7-8%) than joint-honours students taking a traditionally-taught analysis module.

Part II

The Moore Method

Chapter 5

The Moore Method

In this part of this thesis we are interested in the *Moore Method*, a problem-based learning pedagogy for university mathematics teaching, developed in the United States in the first half of the 20$^{\text{th}}$ Century. We want to know whether the Method, which is popular in the United States but has not been extensively analysed, can help to develop students' mathematical problem-solving skills. In this chapter we give an overview of the Method and its background; subsequent chapters cover two studies of modules taught using the Moore Method at the University of Birmingham and another university.

5.1 Robert Lee Moore and his Method

The Moore Method is named after the influential topologist Robert Lee Moore (1882–1974) (Parker, 2005), who developed it for teaching his graduate-level mathematics modules. Moore was born in Dallas in 1882 and, having taught himself calculus at home, entered the University of Texas at Austin at the age of 16. Moore graduated with a BSc in three years, taught at a Texan high school for a year, and completed a doctorate at the University of Chicago in 1905.

While at Chicago, and just 19 years old, Moore proved that Hilbert's 21st axiom was redundant. After brief periods at Tennessee, Princeton and Northwestern universities, Moore moved, in 1911, to the University of Pennsylvania.

At Pennsylvania Moore developed his method of teaching. He used an axiomatic approach; students were presented with definitions and 'axioms'—not necessarily axioms in the truest sense—and they used these to solve problems and prove theorems. After nine years at Pennsylvania, Moore returned to an associate professorship at the University of Texas in 1920, becoming a full professor three years later. He was at Texas for the remainder of his career, eventually retiring in 1969 aged 86. He was elected to the National Academy of Sciences in 1931, was the editor-in-chief of the American Mathematical Society's (AMS) Colloquium Publications in 1930–33, and in 1936–38 president of the AMS. He died in 1974.

Moore guided fifty doctoral students to Ph.D.s, and now has nearly 3,000 mathematical descendants. Three of his graduate students served as presidents of the AMS and five as presidents of the MAA Parker (2005).

5.1.1 Moore's Moore Method

Moore first used his Method while teaching point set topology to graduate students at the University of Pennsylvania, as an introduction to the material they would need for their research work. He subsequently used it for undergraduate teaching, including with students whose major was not mathematics. It was Moore's time as a graduate student at the University of Chicago, however, where he first came into contact with the development of matheamtical pedagogies.

The University of Chicago had been founded in 1894, and among its first faculty members was the psychologist and education reformer John Dewey, who, in 1896, founded the University of Chicago Laboratory Schools, which aimed to teach subjects using Dewey's empirical pedagogy. Dewey believed that experience was the *only* source of knowledge, and his pedagogy focussed on letting students experiment with whatever topic they were working on, though the limits of this were not clearly defined. The Head of the Department of Mathematics at Chicago, E. H. Moore (no relation), who was also the president of the AMS at the time, developed his own ideas based on Dewey's theories. In his retirement address to the AMS in December 1902, E. H. Moore proposed the 'Laboratory Method' of teaching mathematics and physics to undergraduates, stressing the following:

> "The teacher should lead up to an important theorem gradually in such a way that the precise meaning of the statement in question, and further, the practical — i.e. computational or graphical or experimental — truth of the theorem is fully appreciated; and furthermore, the importance of the theorem is understood, and indeed the desire for the formal proof of the proposition is awakened, before the formal proof itself is developed. Indeed, in most cases, much of the proof should be secured by the research work of the students themselves. — Moore, 1967, p. 419

Dewey and E. H. Moore's laboratory system of instruction did not meet expectations, and in 1904 Dewey left Chicago. By that point, however, the laboratory system had come to the attention of R. L. Moore, who subsequently took an interest in it (Parker, 2005, p. 55). While Moore quickly dismissed the idea of the laboratory system, it represented the first contact that

Moore had with those working on the forefront of pedagogical theory.

E. H. Moore further influenced R. L. Moore by being the principle member in a triumvirate of mathematicians at Chicago, the other two being Oskar Bolza and Heinrich Maschke. Between them, these three fostered a competitive environment among the students in the department, one which Moore's Method sought to emulate. So, while Moore did not teach at Chicago, it was undoubtedly an important time for the young postgraduate student and one that influenced his own teaching.

Moore never explicitly codified his Method, and the 'Moore Method' as it exists today is the result of work by his students that have subsequently used it in their own teaching. We rely on the accounts of his graduate students to describe it, in particular those of Jones (1977, pp. 273–278) and Coppin et al. (2009). The basic rules of the course were as follows (adapted from Parker, 2005, p. 100):

1. There were no textbooks for the course and none was to be consulted.

2. Students would prove theorems from given axioms and present their proofs to the class. Students were not to discuss their proofs between themselves *until* such a presentation has taken place.

3. Once a student had presented a proof the rest of the class would offer comment and criticism.

4. Competitiveness was key, and students were encouraged to come up with alternative proofs to the same theorems.

5. Much emphasis was placed on logic.

To foster fair competition between students, those who took his courses were required to have no background in the topic being studied. Moore would expel students whom he discovered to have read material outside that which they were given in class. On receiving a letter from his future graduate student Mary-Elizabeth Hamstrom, asking for advice on reading to do the summer before she joined the University of Texas in 1948, Moore's reply included the sentence "I wish you had never taken a course in Real Variable Theory and that you had read *even less* about point set theory than I imagine you have." (Parker, 2005, p. 245)

Once a class was selected, it would begin with the Method being outlined to students:

> There were certain undefined terms (e.g., "point" and "region") which had meaning restricted (or controlled) by the axioms (e.g., a region is a point set). He would then state the axioms that the class was to start with[...] An example or two of situations where the axioms could be said to apply (e.g., the plane or Hilbert space) would be given. — Jones, 1977, p. 274

After students had a written down the axioms and motivating examples, Moore would read definitions and theorems from his books. He then let students go away and find proofs of theorems, and to construct examples demonstrating the important aspects of a theorem's statement.

At subsequent meetings, students would be called upon to prove theorems in an order chosen by Moore. These proofs were completed at the board, with weaker students being given the opportunity to prove a theorem before a stronger student. Moore was strict in preventing students from interrupting the presenter; only once they had completed a proof or got stuck was the class invited to comment. If a student failed to complete a proof Moore would invite another student to offer a proof, or allow the original student to return in the following class to present a solution:

> “Quite frequently when a flaw would appear in a proof everyone would spend some time (possibly in class) trying to get an example to show that it couldn't be "patched up," i.e., a counterexample to the argument (even though the theorem might be correct). — Jones, 1977, p. 275

Classes proceeded in this way for the rest of the course. Students submitted written solutions to all problems, but took no final examination. The course mark awarded by Moore took into account both their written solutions and presentations. While Moore did not impose solutions on students, and made sure each put in a lot of effort to solve the problems with which they were presented, he was not a silent partner in their teaching:

> “Moore helped his students a lot but did it in such a way that they did not feel that the help detracted from the satisfaction they received from having solved a problem. He was a master at saying the right thing to the right student at the right time. — Mahavier, 1999, p. 339)

The motivation for the Method was to develop in students an understanding of the way in which mathematical research proceeds, where the development of one's own ideas is central. At times in his graduate topology courses students would be presented with Moore's own theorems to prove, and where a student found a more elegant proof than Moore's, would be credited in his subsequent writing.

5.2 The Moore Method Today

The Moore Method, as it became to be known, is essentially that laid-out in Section 5.1.1: mathematical problems are posed by the lecturer to the whole class, and students try to solve these problems independently of one-another. They then present their solutions to the rest of the class, the class discusses each solution and decides whether it is correct and complete, and all students submit written solutions for assessment. While these are the essential details of the Method, Moore finely controlled each aspect to the extent that both Parker (2005) and Coppin et al. (2009) say that the only person ever to teach using *the* Moore Method was Moore himself.

Some variations on the Method are described as *Modified* Moore Methods (Cohen, 1982; Chalice, 1995), though all versions have the following five common principles, as given by Coppin et al. (2009, p. 13):

1. The goal of elevating student from recipients to creators of knowledge.

2. The commitment to teaching by letting students discover the power of their minds.

3. The attitude that every student can and will do mathematics.

4. The careful matching of problems and material to students.

5. The material, varying widely in difficulty, to cover a significant body of knowledge.

5.2.1 Course Material

The material from which students learn is of prime importance to the Method, and its structure is significantly different to that found in traditional undergraduate texts in mathematics. Exposition and examples are kept to a minimum, with problems constituting the bulk of the material. To ensure that students are not overwhelmed at the start of a course, problems begin very simply; the complexity of problems increases to keep pace with students' developing understanding and knowledge of a topic.

The *Journal of Inquiry-Based Learning in Mathematics* (JIBLM) offers Moore Method course material for a range of topics in pure, applied, and statistical mathematics. The following extract, taken from one of the analysis courses available (Neuberger, 2009), demonstrates the level at which students begin solving problems:

▼ **Example 5.2.1.**

Definition 1 *The statement that S is a segment means that there are points a and b such that S is the set of all points between a and b.*

Definition 2 *The statement that I is an interval means that there are points a and b such that I is the set consisting of a, b, and all points between a and b.*

Definition 3 *Suppose M is a point collection. The statement that the point p is a limit point of M means that every segment containing p contains a point of M different from p.*

Definition 4 *Suppose M is a point collection. The statement that the point p is a boundary point of M means that every segment containing p contains a point of M and a point not in M.*

Theorem 5 *If a and b are two points, then a is a limit point of the interval $[a, b]$.*

Students must use the four definitions to solve Theorem 5. Later in the course, students are required to prove the following theorem:

▼ **Example 5.2.2.**

Theorem 98 *Suppose f is a function whose domain includes the interval $[a, b]$ and c is in (a, b). If the domain of each of $f', f'', \ldots$ includes $[a, b]$, and there is an number M such that $|f^{(n)}(x)| \leq M$ for all x in $[a, b]$, $n = 1, 2, \ldots$, then*

$$f(x) = \sum_{i=0}^{\infty} f^{(i)}(c) \frac{(x-c)^i}{i!}$$

for all x in $[a, b]$.

Coppin et al. (2009, pp. 57–58) recommend that those teaching using the Moore Method either take material from existing resources, such as the JIBLM, written specifically for the Method, or write their own. They cover four axioms for writing material for Moore Method courses. These are:

1. Study the masters.
2. Progress from the simple to the complex.
3. Calibrate the Zone of Proximal Development.
4. Write well.

Coppin's third axiom, that course leaders should calibrate the Zone of Proximal Development (Vygotsky, 1978), relates to the progression of problems from the simple to the complex, and illustrates the tension inherent to the Method:

> You, as author, have a measure of control over the level of sophistication students will need, how fast you will want them to move through the material and how deeply you will want them to learn. In particular, you may ask yourself the following questions:
>
> - How much time do we need to spend here on this material? Is it worth it?
> - If we move quickly through this material, will my students know enough to handle such-and-such later on?
> - Are we proceeding at the right depth for later material?
>
> — Coppin et al., 2009, p. 58

This issue is not unique to the Moore Method—any pedagogy whose pace is not determined by the teacher, but the speed at which students solve problems, has the potential for serious slow-downs—but it is important. When a student is struggling to solve a problem with the given information, it is then that the teacher must decide how much assistance they should give; to be sure they are "saying the right thing to the right student at the right time".

5.2.2 Presentations and Written Solutions

Classes in Moore Method courses centre on students presenting their solutions to problems to the rest of the group. These presentations are frequently not polished, though the proofs that they give are expected to be complete and correct Parker (2005). When a student has completed their presentation, or reached a point in their solution beyond which they cannot continue, the class engages in discussion of its validity, or methods of completing or improving the solution. This process may repeat, with a student presenting an updated solution to a problem at a subsequent class, but the final result must be a solution that is mathematically complete.

In some instances students may agree on an incorrect solution, or disagree that a correct solution is as such:

> “There are also occasions when the class (or a large subset of it) refuses to accept a perfectly good proof. A case in point is the proof that there is no set of all sets. The majority of the class were deeply perturbed by the definition of the set $A = \{x \in X \parallel x \notin x\}$ and the meaning of the question ‘A in A?’ Such situations can be harder to deal with. — Good, 2006, p. 37

Moore put heavy emphasis on the quality of written solutions that students were required to submit, and the Moore Method and its modifications give similar weight to written solutions. Full marks are only awarded to a solution if it is not only mathematically but also grammatically correct, as one would expect from a teacher’s solutions. In most instances of the Method the submission of written solutions iterates; students have solutions returned with comments and marks, and are able to resubmit these again for assessment. Moore would allow any number of resubmissions, though these are sometimes limited to one or two times (Good, 2006, p. 37).

5.2.3 Use in Teaching

The Educational Advancement Foundation (EAF) is a philanthropic organisation whose aim is "the development and implementation of inquiry-based learning at all educational levels in the United States, particularly in the fields of mathematics and science, and the preservation and dissemination of the inquiry-based learning methodology inspired by Dr. R. L. Moore (1882–1974), famed professor of mathematics at The University of Texas at Austin." (The Educational Advancement Foundation, 2006) The EAF is based at the R. L. Moore house in Austin, Texas, and is the principle proponent of the Method. It has strong links to the University of Texas at Austin, where the Moore Method is used to teach number theory, real analysis, and point-set topology.

The Moore Method has been spread through the United States, and abroad, by Moore's mathematical descendants, and is principally used outside Texas for teaching the subjects in which Moore worked: analysis and topology. Besides these topics, at the University of Chicago the Method is used to teach calculus, algebra, geometry, and number theory.

The EAF currently funds projects for teaching both mathematicians and mathematics teachers at the universities of Michigan, Chicago, Texas, and California, Santa Barbara, though these come under the umbrella of 'Inquiry-Based Learning' and are not all taught using the Moore Method or its modifications.

The modules discussed in the following two chapters are the only Moore Method modules identified in our survey, however other courses have presaged them with similar teaching methods; notably the Warwick Analysis Project, previously discussed and analysed by Alcock and Simpson (2001). The written material for that course, and Burn's *Steps into Analysis* (1992) that it originally used, bear a striking resemblance to the Moore courses available from the JIBLM.

5.2.4 Criticisms and Limitations of the Method

As each student needs the opportunity to present their work to the class, preferably several times a term, and the class needs to engage in discussion of these presentations, it is not feasible to teach large groups with the Method. Typically, class sizes of ten to twenty students are recommended (Coppin et al., 2009), though Moore preferred smaller classes of four to eight students (Parker, 2005, p. 259). Fewer and discussion is difficult to get going; more and not everyone has the chance to participate, or can get away without doing so. Thus the Method is unsuitable for teaching a typical U.K. cohort of 100–250 undergraduate students as a single class. In such cases alternative pedagogies to directed learning remain, such as the Socratic Method practised at London South Bank University (Crisan et al., 2010), in which students in classes of up to 100 engage with the teacher on a individual level.

The amount of material that can be covered in a Moore Method course is limited, even if the balance mentioned by Coppin et al in Section 5.2.1 is successfully struck. Students need to take time—sometimes lots of time—to get used to working through problems on their own, and progress is usually slower than in courses taught using traditional directed learning. Furthermore, the students who make up each class will influence the rate of progress and so judging how much material can be covered in a term can be difficult:

> “Progress can be slow (it took possibly five hours of class time to get through the first sheet to my satisfaction). It also depends to an extent on the individuals in the class, so it might well be difficult to judge exactly what material the group will have worked through by the end of a semester. — Good, 2006, p. 37

The modules that we discuss in Chapters 6 and 7 overcome these two issues by dividing the cohort that they teach into groups of appropriate size, and do not aim to cover a large amount of material. Further, and unlike most of the Moore Method courses at the universities of Texas and Chicago, their aim is to develop students' problem-solving and communication skills, and not principally the learning of specific mathematical material.

Competition is inherent in the Method; healthy competition between students to see who can prove a theorem first encourages them to work hard, however this too has its downsides: “Friendly competition spurs many students to work harder than they would otherwise, and occasionally to work above and beyond reasonable limits (i.e., to the detriment of other courses).” (Dancis and Davidson, 1970) Besides being potentially detrimental to their other courses, the competitive aspect may go too far, putting off some students from engaging in the process: “The Moore Method sometimes creates an unhealthy atmosphere of competition and isolation among students.” (Cohen, 1982, p. 474) Weaker students, or those that do not flourish in a competitive atmosphere, tended to struggle in Moore's own classes:

> “But most of them dreaded and hated class for the simple reason that is was a painful experience for them. Because they had difficulty presenting things and because the things that they conjectured were often wrong, they were used as

> an example of how you could go wrong. — Moore's student Mary Ellen Rudin in Parker, 2005, p. 271

While undoubtedly an important aspect of the Moore Method, teachers have a measure of control over the competition in class, by deciding who presents solutions to particular problems, and choosing how much assistance they give to different students. Moore's behaviour towards students was at times claimed to be deliberately antagonistic (Parker, 2005, p. 271), something unlikely to be encouraged in a modern setting.

5.3 Relationship to other PBL, and Research on the Method

While the ideas of Dewey and E. H. Moore undoubtedly spurred R. L. Moore into thinking about pedagogical strategies, he developed his method of teaching in isolation, and as such the 'Moore Method' is seen as distinct from a range of pedagogies with which it has a lot in common. Taxonomically, the Moore Method is a type of problem-based learning, though its proponents such as the Educational Advancement Foundation refer to it as *Inquiry-based Learning*; we point the reader back to Section 3.2 and, in particular, Hmelo-Silver et al. (2007)'s comments on the lack of clear distinction between EBL and PBL.

The roles of the written material, the class, and the teacher combine to ensure that the Moore Method is clearly not 'minimally guided', though before they present to the class each student has the opportunity to direct their own investigations into a problem. It is this lack of discussion before the presentation that best sets apart the Method from other PBL pedagogies, and something that requires the complicity of students taking a Moore Method course.

The emphasis on writing mathematics further defines the method as different from other PBL pedagogies. Neither aspects of independent study nor a concentration on proof writing are necessarily absent in other methods of PBL, and at its heart the Moore Method has much in common with other student-centred PBL pedagogies. The strong culture surrounding the teaching of the Method, at the University of Texas at Austin in particular, has helped it to remain a distinct pedagogy over a century of educational reform.

Given this relationship to other student-centred PBL pedagogies, much of the research discussed in the previous part of the thesis is relevant to the Moore Method. The central principle of student discovery relates directly to the work of Bruner (1961, p. 26) and so research on discovery learning, combined with that on constructivism, may be used in support of the efficacy of the Method.

Brousseau (1997, p. 229) notes the importance of working alone in learning mathematics:

> One of the fundamental contributions of the modern *didactique* consists of showing the importance of the role played in the teaching process by the learning phases in which the student works almost alone on a problem or in a situation for which she assumes maximum responsibility.

While there is an amount of indirect support of the Moore Method in the literature, there exists a dearth of studies aimed directly at it. Smith (2006) compared the Moore Method with traditional direct instruction methods of teaching and their effects on approaches to proofs. The author asserts that there were marked differences in the approaches of the students in the two groups, though they were only two students and three students in size, respectively, and:

> the [Moore] students' approach to proof is reminiscent of Weber and Alcock (2004, p. 210)'s notion of a semantic proof production: "a proof in which the prover uses instantiation(s) of the mathematical object(s) to which the statement applies to suggest and guide the formal inferences that he or she draws". The students' use of initial strategies, notation, prior experiences, and examples could be considered as such instantiations of mathematical concepts, meaningful ways of thinking about mathematical objects.

Perhaps one reason for the lack of research on the Method is the way it has spread. The Moore Method is still centred around the University of Texas in Austin, and its dissemination has largely taken place by passing the Method from teacher to student. Often these teachers are mathematics researchers who may not have the time or inclination for scrutiny as part of an educational study. By means of an example, of the four authors of Coppin et al. (2009), Coppin was a student of Hubert Stanley Wall, a colleague of Moore and a leading proponent of the method, while the other three authors were students of John Neuberger, another student of Wall's. Mahavier's father, furthermore, was a student of Moore. Smith (2006) is also based at Austin. At the Annual Moore Legacy Conference held each June, presenters describe their position in the Moore family tree; grandson, or grand-niece, for example. Students who subsequently become teachers themselves are generally predisposed to teach in the manner in which they were taught, and so the way in Moore Method has propagated through the mathematical community is natural:

> “The student’s learning about teaching, gained from a limited vantage point and relying heavily on imagination, is not like that of an apprentice and does not represent acquisition of the occupation’s technical knowledge. It is more a matter of imitation, which, being generalized across individuals, becomes tradition. Lortie, 1975, p. 63

Chapter 6

Determining the Effectiveness of the Moore Method at Improving Mathematics Performance

6.1 Introduction

This chapter is a longitudinal study of "1Y", a first year Moore Method module in the School of Mathematics at the University of Birmingham. In the chapter we demonstrate a correlation between participation in 1Y and performance in a number of other first- and second-year modules in the School.

Entitled "Developing Mathematical Reasoning", the optional module started in the academic year 2004/05 and is taken by roughly a dozen students each year. Its aim is to improve students' problem solving in a process driven by the students themselves. The data for the study are the module marks of students over the period 2005/06–2011/12; 99 students who took 1Y in that time are compared to their peers in the rest of the cohort.

6.2 The Module

1Y is a *Module Outside the Main Discipline* (MOMD), optional modules that students choose to meet the credit requirements to progress to the following year. In the first year in the School of Mathematics undergraduate students take 100 credits of prescribed modules and are free to take 20 credits of a MOMD, which need not be mathematics.

The 1st year module descriptions booklet describes 1Y as follows:

> “By engaging with interesting elementary problems and by reflecting on the experience, students will learn certain specific techniques and will gain insight into the nature of mathematics in the context of elementary material. Mathematical reasoning skill will be developed through an extended period of directed student centred learning in which students are expected to work through a list of problems, present solutions on the board to class mates, and to comment on their own work and the work of others. To succeed in this module students will need to be able to the work independently outside class hours, without the use of reference materials.

The module began in 2004/05, covering naïve set theory and being taught by Dr. Good. From 2006/07–2010/11 it was a geometry module taught by Dr. Sangwin. In 2011/12 it was taught to two distinct groups, one covering set theory (again taught by Dr. Good), one covering geometry (Dr. Sangwin). Each of the module leaders uses their own problems, that are designed to form a complete (though limited) part of their topic. There is very limited exposition in the module material.

Example 6.2.1 is the first problem that the geometry class has to answer. It is originally taken from *Lines and Curves* (Gutenmakher et al., 2004), a book that was previously used as the original source of problems in the geometry module, until 2009/10:

▼ **Example 6.2.1 — *Cat on a Ladder*.**

A ladder standing on a smooth floor against a wall slides down to the floor. Along what curve does a cat sitting in the middle of the ladder move? Along what curve does the cat move if it does not sit in the middle of the ladder?

Most modules in the School have the same structure: two lectures weekly and a fortnightly tutorial each lasting 50 minutes, continuing for the whole of the 11 week term. While 1Y nominally has the same format, in practice there is no distinction made between the lectures and tutorials, and student participation is expected at all 27 class meetings.

6.2.1 Teaching and Examination

The module is taught using a close approximation to the original *Moore Method*. Students begin with a set of 'axioms', a list of assumed results that can be taken for granted and used to solve the problems in the module. In the geometry group, this is presented as a sheet titled *Some Facts from School Geometry*, of which Example 6.2.2 is one:

▼ **Example 6.2.2 — *Fact B.1.4.***

A theorem on proportional segments in a circle. *If two chords AB and CD of a circle intersect at the point E, then*

$$|AE|.|BE| = |DE|.|CE|.$$

The first week of lessons are then used to discuss these results, with students demonstrating their meaning on the board without having the pressure of presenting their own work to the class. This also gives students enough time to begin working on problems outside of class, so time is not wasted.

Once students begin working on the problem sheets, the class takes on a format that it maintains for the rest of the term: a student will present a solution at the board and the rest of the class ask questions or make comments about the solution, and decide whether it is correct and complete. The module leader also participates in discussions, though their role is to act as chairperson, and clarify the objections of others. They are responsible for keeping discussion focused and at a professional level. It is always the aim that the teacher is the last to speak, so that all students have the opportunity to contribute to the discussion first. If the class has agreed on an incorrect solution, it is the job of the teacher subtly to direct them in recognising the issue.

Problems are solved in order and students are chosen by the module leader in an ostensibly random order to present solutions to the class. If a student fails to offer a solution to a number of problems in a row the module leader will likely ask them, sometimes descreetly, to work on a few particular problems for the following class.

When a correct solution is reached it may be tided up and the whole of the class makes notes to enable them to submit written solutions later. Students are required to submit solutions to all problems, not just those that they solved themselves, and they may resubmit solutions as many times as they like on receiving feedback for them.

For purely administrative reasons, 1Y is bound to another module, 1X, titled "The Impact of Mathematics", which shares its final mark equally with 1Y; departmental regulations previously limited the amount of coursework that can be used in the examination of a module. Until the academic year 2011/2012 the module mark consisted of the following:

- Quality of two best presentations to the class (26%)
- Individual written solutions to all problems (24%)
- Written examination (50%)

From 2011/12 the written examination was dropped and the marks distributed equally between students' two best presentations and their written solutions to problems.

6.2.2 Self Selection

There is an element of self selection in 1Y participation. As a set theory course, 1Y was limited to students who scored the equivalent of an A in A-Level mathematics, however from 2006/07 this restriction was lifted. The number of students in the department who did not meet this requirement has been small and decreasing, to the extent that the 2009/10 and later cohorts consist entirely of grade A students. As 1Y was optional, it was usual that a number of students drop out in the first few weeks in favour of a different MOMD.

From 2011/12, the module became expected of students enrolled on the Mathematics MSci programme, G103. This requirement was put in place to reflect the difference in offers to BSc and MSci students (AAA/AAB for BSc and A*AA for MSci). MSci students who wish to study a foreign language are encouraged to do so, and BSc students wanting to take 1Y are still welcome on the module, space permitting. While this means that the group is less self-selecting, it is more clearly distinct from the rest of the cohort. Before 2011/12, MSci students made up a similar proportion of 1Y students as in the cohort as a whole.

6.3 Methodology

6.3.1 The Data

Raw data for the study are the final marks awarded by the examinations board in the courses considered. These marks, from 19 modules, include those of the 99 1Y students and the 1638 remaining students from the same years. Each of these marks includes an exam result and any coursework submitted; if students have retaken a module their highest mark is used. These data include nearly fifteen thousand exam results and give a firm statistical basis on which to draw conclusions. Modules in the second and third years of the degree programme are included to determine whether any follow-on effects are present.

Table 6.1 shows the number of data points used in the analysis of each year and level, i.e.

the number of marks recorded by students at each level. These results, 6867 for the first year, 5085 for the second year and 1383 for the third year were distributed over six, six and seven modules, respectively, in each of the three years.

Academic Year	Level		
	1st	2nd	3rd
2005/06	896		
2006/07	907	807	
2007/08	864	793	219
2008/09	1003	723	282
2009/10	1087	901	236
2010/11	1004	959	342
2011/12	1106	902	304

Table 6.1: Data points for analysis by level and year

Each of the 18 years' data were contained in separate spreadsheets that listed student ID, module code and final mark. Thus a student who gained a mark in more than one module was listed more than once in the same list. To arrange the data into a usable form, all 18 spreadsheets were concatenated into a single sheet and a macro written (see Listing B.1) that put the marks for each student on a single row arranged so that the marks for each module were in a single column. Results were initially ordered by student ID, then module code and final mark, all in ascending order. Thus, for modules which students have re-sat examinations or retaken the module, the higher mark was used for analysis. One module initially included for analysis, MSM3P04, was taken by no 1Y students and so was left out.

To these results were added the results of the 1Aa class test, a 50 minute examination taken halfway through the first semester. These data, from 1498 students who had a mark recorded, were also given their own column. Finally the gender of the student and whether or not they were a 1Y student were given columns of their own. The resulting table recorded the results of 1737 students across all years, and was copied into an SPSS data set for analysis.

6.3.2 The Method

Standard statistical methods were used in the study; in particular Lomax (2001) was used as a guide for the methods themselves (Chapter 12 covers multivariate regression analysis) while Field (2005) informed the use of SPSS. Whilst we are nominally interested in *attainment*, our primary concern is *progress*; Plewis (1997, pg. 24) states that:

> " The distinction between progress and attainment is an important one in educational research. Progress is a dynamic concept needing longitudinal data for its measurement, attainment is a static concept needing only cross-sectional data.

As our dataset comprises results from individual modules, it is cross-sectional, though we argue, given that for the majority of students we have more than an single year of results, our study is still dynamic to a degree.

Initial work used independent t-tests to determine whether 1Y students performed better than their peers to a statistically significant degree (5% or under). Tables A.1 and A.2 in the Appendix show the results of the t-tests on eleven of the modules considered, using data up to 2009/2010. In most modules, however, the data failed Levene's Test for homogeneity of variances, and so we cannot base solid conclusions on these results. Furthermore, the mean of 1Y students' scores on the 1Aa class test was 12.85% higher than their peers. (None of the 1Y students had completed the module before taking the class test, though roughly a fifth would have taken the module in their first semester and so had *some* 1Y experience.)

As t-test proved inappropriate for our dataset, it was decided to perform multivariate regression analysis to predict students' scores on other modules. Multivariate analysis using SPSS allows us to determine, with reasonable margin of error, the contribution that 1Y participation makes to other modules. The decision to use the 1Aa class test data for regression analysis of first year modules was made because complete transcripts of A-Level results for students were not available. This being the case, the only other variable that could be used would be a student's A-Level grade. When the overwhelming majority of students in the dataset recorded an A in mathematics this does not give sufficient distinction between students for reasonable coefficients of determination. Overall correlation between the class test mark and students' mean marks in first and second year modules was reasonably high with a coefficient of determination of 0.36. The 1Aa class test data contributes 10% to each student's 1Aa mark, with the remaining 90% coming from the end of year examination (70%) and continuous assessment (20%). Given that the contribution of the 1Aa class test mark to the 1Aa module mark is relatively small, it was decided that in the particular case of the 1Aa module regression, we would still be able to draw conclusions on the result of the regression. Thus we have the equation:

$$\text{Module mark } = \beta_0 + \beta_1 CT + \beta_2 1Y. \tag{6.1}$$

In Equation 6.1, β_1 is the unstandardised coefficient of the class test mark (the Class Test mark being a percentage) and β_2 is the unstandardised coefficient of the 1Y variable. In this way the contribution 1Y makes to a student's performance in a module is isolated from other factors that influence a student's accomplishment. The 1Y variable is 0 for students who did not take 1Y and 1 for students who did take 1Y; this gives us the added advantage that the value of β_2 is the increase in mark gained by being a 1Y student.

With such models we have two hypotheses that must be rejected in order for positive conclusions to be drawn. The first, the null hypothesis, states that the model itself does not predict the outcome variable, i.e. $H_0 : \beta_1 = \ldots = \beta_i = 0$. To reject this hypothesis we calculate the F statistic, given by the equation

$$F = \frac{R^2/m}{(1 - R^2)/(n - m - 1)}, \tag{6.2}$$

where R^2 is the coefficient of multiple determination, m is the number of predictors and n is the sample size. The F statistic is then usually compared to a critical value to determine whether it is significant at 5%, though SPSS gives the exact significance when performing regression analysis. If this hypothesis is dismissed, the model is a good predictor of the outcome variable, and one of the β_i may be statistically significantly different to zero.

The second, alternative, hypothesis applies to each individual coefficient and states that it is not statistically significantly different to zero. This test statistic t is given by the parameter estimate, i.e. the unstandardised coefficient, divided by its standard error

$$t = \frac{\beta_i}{s(\beta_i)}. \tag{6.3}$$

The calculation of the standard error of β_i can be found in Lomax (2001, pg. 244). The t-statistic is also usually compared to a critical value but once again SPSS gives its exact significance.

For each module on which regression analysis was performed, conclusions could only be drawn if both the F- and t-statistics were suitably significant, and so the two questions that essentially had to be answered were *Is the model a good predictor of the module mark?* and *Does the* $1Y$ *coefficient contribute a significant amount?*

6.4 Results

6.4.1 First Year Results

Table 6.2 shows the details of the regression analysis for the first year courses MSM 1Aa, 1Ab, 1B, 1C and 1D. For each module the first model uses only the 1Aa class test mark while the second uses the class test mark and 1Y participation. For each coefficient their value and the standard error are given, then the standardised b value, the t-statistic and its significance. There are then listed the model's R^2 value, indicating how well the data fits the curve defined by the model (which for the first model, having only one variable, will be its Pearson's r statistic) and its standard error, the change in R^2, F and the significance of the F-test change. It is important to note that the changes given are from the previous model, so for the second model listed these changes represent the improvement *over* the first model. For the first model these are the improvement over the line of gradient zero going through the mean of the output variable, and so R^2 and $R^2\Delta$ will be the same.

Model		Unst. Coeff.					Data		Std.	Change Statistics		
		β_i	Std. Error	b	t	Sig.	Points	R^2	Error	$R^2\Delta$	F Δ	Sig. F Δ
MSM 1Aa – Calculus and Algebra I												
1	(Const.)	30.37	1.28	0.00	23.70	0.00	1494	0.29	14.57	0.29	605.25	0.000
	Test Mark	0.47	0.02	0.54	24.60	0.00						
2	(Const.)	30.66	1.26	0.00	24.34	0.00	1494	0.31	14.31	0.03	54.63	0.000
	Test Mark	0.46	0.02	0.52	24.07	0.00						
	1Y Student	11.52	1.56	0.16	7.39	0.00						
MSM 1Ab – Calculus and Algebra II												
1	(Const.)	22.84	1.55	0.00	14.71	0.00	1484	0.23	17.41	0.23	436.96	0.000
	Test Mark	0.49	0.02	0.48	20.9	0.00						
2	(Const.)	23.12	1.54	0.00	15.04	0.00	1484	0.24	17.23	0.02	33.65	0.000
	Test Mark	0.47	0.02	0.46	20.37	0.00						
	1Y Student	10.88	1.88	0.13	5.8	0.00						
MSM 1B – Foundation and Abstraction												
1	(Const.)	27.78	1.86	0.00	14.96	0.00	951	0.22	16.1	0.22	266.82	0.000
	Test Mark	0.44	0.03	0.47	16.33	0.00						
2	(Const.)	27.91	1.82	0.00	15.37	0.00	951	0.25	15.75	0.03	43.96	0.000
	Test Mark	0.42	0.03	0.45	15.9	0.00						
	1Y Student	11.57	1.75	0.19	6.63	0.00						
MSM 1C – Vector Algebra, Elementary Mechanics and Computational Mathematics												
1	(Const.)	18.68	1.69	0.00	11.08	0.00	1032	0.33	15.33	0.33	498.48	0.000
	Test Mark	0.57	0.03	0.57	22.33	0.00						
2	(Const.)	19.01	1.68	0.00	11.30	0.00	1032	0.33	15.27	0.01	9.45	0.002
	Test Mark	0.56	0.03	0.56	21.67	0.00						
	1Y Student	5.32	1.73	0.08	3.07	0.00						
MSM 1D – Discrete Mathematics and Statistics												
1	(Const.)	21.63	1.57	0.00	13.74	0.00	1214	0.27	15.95	0.27	444.3	0.000
	Test Mark	0.49	0.02	0.52	21.08	0.00						
2	(Const.)	21.84	1.55	0.00	14.05	0.00	1214	0.29	15.75	0.02	32.69	0.000
	Test Mark	0.48	0.02	0.50	20.58	0.00						
	1Y Student	9.88	1.73	0.14	5.72	0.00						

Table 6.2: Regression for first year modules over all students 2005/06–2011/12

For each of the models the F-statistic is, as we may reasonably expect, significant at 5%. While the R^2 values of the ten regressions are not particularly large, ranging from 0.22 to 0.33, they give large improvement in predicting a student's score over taking the mean for the class. What interests us the most is the change in the F-test value and its associated significance when we add in the 1Y variable to each of the models. Related to this we are also interested in the value of the unstandardised coefficient of the 1Y variable (β_2 from Equation 6.1) to see the direct impact that 1Y participation has on the mark for the course.

Table 6.2 shows the F-test value for the 1Y variable is significant in all cases to greater than 0.1%, except for 1C where it is significant to 0.2%. For 1C the 1Y coefficient is 5.32, roughly half what it is for the other modules, but its standard error is similar at 1.73. This will account for the slightly larger t-test statistic, and lower F-test change over the first model.

The problem with these results is that they include data from students who took 1Y in 2011/12, after it had become compulsory for students on the MSci programme. As the set of 1Y students is almost exactly the set of first-year MSci students (others may opt in, or a student doing an MSc may opt to do a language module instead), we cannot say whether their improved performance is a result of being a 1Y student or being an MSci student. In spite of performing better on the 1Aa class test to a similar degree to previous years, these students skew the model outputs and cause the 1Y coefficient to become significantly larger, and the class test mark to become smaller, across all modules. A further confounding factor is that 2011/12 1Y students, who number 42, make up a far larger proportion of the cohort in this than previous years. While module marks are moderated and have similar descriptive statistics across all years, we cannot rule out the possibility of this skewing results considerably. Instead we rely—for first year analysis only—on Table 6.3, which shows the same tests as Table 6.2 using data only from 2005/06 to 2010/11. With these data, including the 1Y variable in the models again gives significant improvement to the F-test value for all modules, though the coefficient of the 1Y variable is smaller than before. With smaller data sets, the standard errors are a little larger, but still considerably smaller than the coefficients themselves.

It is of course rather more difficult to assert the existence of a causal relationship than merely a correlative one, however these data strongly indicate that participating in 1Y improves students' performances in other modules in the first year.

		Unst. Coeff.					Data		Std.	Change Statistics		
Model		β_i	Std. Error	b	t	Sig.	Points	R^2	Error	$R^2\Delta$	F Δ	Sig. F Δ
MSM 1Aa – Calculus and Algebra I												
1	(Const.)	22.20	1.26	0.00	17.69	0.00	1291	0.42	13.04	0.42	930.08	0.000
	First Year	0.56	0.02	0.65	30.50	0.00						
2	(Const.)	22.62	1.25	0.00	18.12	0.00	1291	0.43	12.94	0.01	21.92	0.000
	First Year	0.55	0.02	0.63	29.75	0.00						
	1Y Student	7.79	1.66	0.10	4.68	0.00						
MSM 1Ab – Calculus and Algebra II												
1	(Const.)	18.47	1.70	0.00	10.89	0.00	1284	0.27	17.40	0.27	463.52	0.000
	First Year	0.53	0.02	0.52	21.53	0.00						
2	(Const.)	19.02	1.69	0.00	11.27	0.00	1284	0.28	17.27	0.01	19.95	0.000
	First Year	0.52	0.02	0.50	20.82	0.00						
	1Y Student	9.92	2.22	0.11	4.47	0.00						
MSM 1B – Foundation and Abstraction												
1	(Const.)	19.59	1.95	0.00	10.05	0.00	821	0.31	15.29	0.31	370.01	0.000
	First Year	0.53	0.03	0.56	19.24	0.00						
2	(Const.)	20.22	1.93	0.00	10.50	0.00	821	0.33	15.08	0.02	24.51	0.000
	First Year	0.51	0.03	0.54	18.54	0.00						
	1Y Student	9.76	1.97	0.14	4.95	0.00						
MSM 1C – Vector Algebra, Elementary Mechanics and Computational Mathematics												
1	(Const.)	17.79	1.92	0.00	9.28	0.00	886	0.33	15.68	0.33	421.32	0.000
	First Year	0.58	0.03	0.57	20.53	0.00						
2	(Const.)	18.27	1.92	0.00	9.49	0.00	886	0.33	15.65	0.00	4.66	0.031
	First Year	0.57	0.03	0.56	19.73	0.00						
	1Y Student	4.58	2.12	0.06	2.16	0.03						
MSM 1D – Discrete Mathematics and Statistics												
1	(Const.)	16.44	1.70	0.00	9.68	0.00	1041	0.32	15.53	0.32	499.12	0.000
	First Year	0.54	0.02	0.57	22.34	0.00						
2	(Const.)	16.85	1.69	0.00	9.96	0.00	1041	0.33	15.45	0.01	12.22	0.000
	First Year	0.53	0.02	0.56	21.69	0.00						
	1Y Student	6.99	2.00	0.09	3.50	0.00						

Table 6.3: Regression for first year modules excluding 2011–12

6.4.2 Second year analysis

Regression analysis for second year courses required more consideration than that for the first year courses. Each of the first year courses' results may be used as a variable for predicting second year course marks, and it is tempting to take this approach. This methodology presents two problems, however, which when combined decrease the potential validity of results. The first problem is that the correlation between results in the five first year modules is high, as shown in Table 6.4:

Module	Statistic	MSM1Aa	MSM1Ab	MSM1B	MSM1C	MSM1D
MSM1Aa	Correlation	1.000	0.828	0.838	0.775	0.812
	Significance (2-tailed)	.	0.000	0.000	0.000	0.000
	df	1328	1318	843	892	1068
MSM1Ab	Correlation	0.828	1.000	0.840	0.801	0.821
	Significance (2-tailed)	0.000	.	0.000	0.000	0.000
	df	1318	1320	842	890	1065
MSM1B	Correlation	0.838	0.840	1.000	0.763	0.856
	Significance (2-tailed)	0.000	0.000	.	0.000	0.000
	df	843	842	850	558	823
MSM1C	Correlation	0.775	0.801	0.763	1.000	0.775
	Significance (2-tailed)	0.000	0.000	0.000	.	0.000
	df	892	890	558	916	706
MSM1D	Correlation	0.812	0.821	0.856	0.775	1.000
	Significance (2-tailed)	0.000	0.000	0.000	0.000	.
	df	1068	1065	823	706	1081

Table 6.4: Correlation between marks in first year modules

When the correlation between two variables in a regression is high, it is not possible to confidently determine the effect of the two variables on the outcome variable: does the first have more effect, or the second? To solve this problem we could apply backwards stepwise analysis to remove variables that are not contributing sufficiently to the outcome, i.e. those whose t-test statistic is not significant. However, when the correlation is high the variables that are removed are, to a certain extent, arbitrary.

The second problem this approach presents is that not all of the students for whom we have second year marks have recorded results for each of the five first year modules that we are using for our analysis. Only 651 students have a full complement of first year marks, while 1605 have at least one (the vast majority of these having three or more; some joint-honours students do not take two of MSM 1B, 1C or 1D).

Given these two problems it was decided that using the mean of the first year marks (excluding the 1Y mark) would be the fairest and most defensible way to proceed with analysis. This keeps the number of variables used to a minimum and, by using more data points, decreases the standard error of our results. We therefore use a modified version of Equation 6.1,

$$\text{Module mark } = \beta_0 + \beta_1 M + \beta_2 Y, \tag{6.4}$$

in which the class test mark CT is replaced by the mean of the first year modules, M. Results for the six second year modules' regression analysis are in tables 6.6 and 6.7 and are read in the same way as Table 6.2.

Outliers

For the second year regression six outliers with particularly high (≥ 0.025) Cook's distance on the MSM 203 regression analysis were removed from the dataset. These were all students who had performed well in first year modules but then very poorly (averaging below 25%) in all second year modules, possibly having dropped out. Being unrepresentative cases with large influence on the results of the regression analysis they were removed; three other outliers with large Cook's distance were not removed as, whilst the models did not predict these scores particularly well, they did give relatively sensible estimates. Two of these had scored very well whilst one had scored relatively poorly. Table 6.5 shows these outliers with their first and second year module marks; the three rows shaded grey are those outliers that were not removed for the analysis.

	Module Mark										
Student	1Aa	1Ab	1B	1C	1D	201	203	2A	2B	2C	2D
1	58.0	53.9	52.2	15.0	45.9	16.2	4.4	40.1	22.2	41.7	12.6
2	51.8	55.8	59.7	53.4	59.7	55.6	79.1	56.7	78.3	60.9	66.9
3	63.0	7.5	22.4	40.6	9.5	3.2	0	0	1.5	0	2.0
4	66.8	65.6	72.3	59.5	55.1	49.2	90.0	77.0	88.6	93.2	74.3
5	76.1	77.6	77.2		72.0	11.8	9.5	8.1	11.7	22.3	7.8
6	93.1	80.1	81.2	80.8	84.7	12.3	11.8	10.6	10.2	7.2	10.9
7	88.5	85.2	75.3	74.6	69.2	14.3	12.7	9.6	10.2	12.9	9.2
8	62.2	65.6	73.5	73.6	68.8	51.8	40.5	55.5	56.1	55.9	59.2
9	83.0	82.0	78.5	77.7	82.8	10.0	7.3	8.4	8.7	9.6	5.6

Table 6.5: Outliers in the MSM 203 regression

Second Year Results

The second year results are rather more intesting than those of the first year. Instead of across-the-board improvement, we see a very clear distinction between modules where the 1Y variable is significant and those where it is not. MSMs 201, 2A and 2D have very small 1Y coefficients that are smaller (for 201 and 2D, considerably so) than the standard error. Meanwhile, MSMs 203, 2B and 2C have modest but significant coefficients of between three

Model		Unst. Coeff. β_i	Std. Error	b	t	Sig.	Data Points	R^2	Std. Error	Change Statistics $R^2\Delta$	F Δ	Sig. F Δ
MSM 201 – Applied Mathematics II												
1	(Const.)	10.02	2.63	0.00	3.81	0.00	673	0.38	15.24	0.38	409.72	0.000
	1st Yr. Mean	0.87	0.04	0.62	20.24	0.00						
2	(Const.)	10.07	2.65	0.00	3.81	0.00	673	0.38	15.25	0.00	0.04	0.839
	1st Yr. Mean	0.87	0.04	0.61	19.88	0.00						
	1Y Student	0.43	2.10	0.01	0.20	0.84						
MSM 203 – Polynomials and Rings & Metric Spaces												
1	(Const.)	-14.38	3.57	0.00	-4.03	0.00	365	0.52	15.81	0.52	393.86	0.000
	1st Yr. Mean	1.15	0.06	0.72	19.85	0.00						
2	(Const.)	-13.26	3.60	0.00	-3.69	0.00	365	0.53	15.74	0.01	4.46	0.035
	1st Yr. Mean	1.12	0.06	0.70	18.82	0.00						
	1Y Student	5.22	2.47	0.08	2.11	0.04						
MSM 2A – Analytical Techniques												
1	(Const.)	6.52	1.98	0.00	3.30	0.00	1150	0.39	15.25	0.39	723.54	0.000
	1st Yr. Mean	0.88	0.03	0.62	26.90	0.00						
2	(Const.)	6.67	1.99	0.00	3.35	0.00	1150	0.39	15.25	0.00	0.50	0.481
	1st Yr. Mean	0.88	0.03	0.62	26.39	0.00						
	1Y Student	1.37	1.95	0.02	0.70	0.48						

Table 6.6: Regression for second year modules MSM 201, 203 and 2A

and five percent. For these modules, even taking into account students' first-year means, that include improvements partially attributable to participation in 1Y, taking 1Y in the previous year appears to have further influenced their results.

Two highlights of these results are as follows: firstly, a set of students from across the cohorts considered that was selected for better performance in the 1Aa class test, or over the first year as a whole, would not produce such a clear set of improvements in some second year modules but not others. It remains to be seen whether the 1Y students of 2011/12 continue this trend, given the different selection criteria. These data show clear evidence that there is

Model		Unst. Coeff. β_i	Std. Error	b	t	Sig.	Data Points	R^2	Std. Error	Change Statistics $R^2\Delta$	F Δ	Sig. F Δ
MSM 2B – Real and Complex Variable Theory												
1	(Const.)	9.26	2.25	0.00	4.12	0.00	776	0.40	13.95	0.40	524.67	0.000
	1st Yr. Mean	0.84	0.04	0.64	22.91	0.00						
2	(Const.)	9.70	2.26	0.00	4.30	0.00	776	0.41	13.94	0.000	2.90	0.089
	1st Yr. Mean	0.83	0.04	0.63	22.21	0.00						
	1Y Student	3.09	1.81	0.05	1.70	0.09						
MSM 2C – Linear Algebra and Programming												
1	(Const.)	17.23	2.07	0.00	8.32	0.00	1057	0.35	15.36	0.35	567.18	0.000
	1st Yr. Mean	0.82	0.03	0.59	23.82	0.00						
2	(Const.)	17.73	2.08	0.00	8.51	0.00	1057	0.35	15.34	0.000	4.00	0.046
	1st Yr. Mean	0.81	0.04	0.58	23.08	0.00						
	1Y Student	3.94	1.97	0.05	2.00	0.05						
MSM 2D – Linear Programming and Symmetry & Groups												
1	(Const.)	6.45	2.09	0.00	3.09	0.00	730	0.44	12.52	0.44	567.31	0.000
	1st Yr. Mean	0.81	0.03	0.66	23.82	0.00						
2	(Const.)	6.48	2.10	0.00	3.08	0.00	730	0.44	12.53	0.000	0.01	0.916
	1st Yr. Mean	0.81	0.03	0.66	23.36	0.00						
	1Y Student	0.17	1.64	0.00	0.11	0.92						

Table 6.7: Regression for second year modules MSM 2B, 2C and 2D

something different about the 1Y students from the years 2005/06–2010/11; though we cannot say whether it is students who perform better in these modules are predisposed to choosing a module like 1Y, or whether 1Y has impacted their thinking and learning in such a way that they *become* able to perform better in these modules.

The second aspect of these results that is worthy of attention is the modules in which the 1Y coefficient is and is not significant. 1Y aims to teach mathematical thinking and problem-solving skills to students who have not previously been presented with an explicit approach to either. We previously hypothesised that pure mathematics modules would benefit from such teaching, while applied mathematics would be largely unaffected. However, the results do not bear this hypothesis out; they are more nuanced. The modules without significant 1Y variables cover the following:

MSM201 – Applied Mathematics II Second-order non-linear ODEs; vector fields; vector calculus.

MSM2A – Analytical Techniques Calculus of functions of several real variables; numerical techniques; boundary value problems for ODEs.

MSM2D – Linear Programming & Symmetry and Groups First semester covers basic linear programming; linear programming models; graphical and simplex methods. Second semster covers the basics of symmetry and group theory.

The topics covered in the first two modules cover the learning of basic techniques in applied mathematics, and were not expected to be impacted by 1Y participation. The last, however, spends an entire semester covering group theory, yet the 1Y variable in its model is the smallest recorded in our analysis and has the lowest significance among second-year modules. Because students sit only one exam, we cannot break-down the module marks into their two halves, though to mask any improvement in Symmetry & Groups so completely 1Y students would have also to perform significantly worse in Linear Programming. This is unlikely, given that the other first-year MOMD, 1X, does not cover the topic. Perhaps there is something to be said for the way in which group theory is taught to second-year students: by geometric example. Students are introduced to symmetries of familiar objects (triangles, circles, cubes, icosahedron, etc.), before representing their rotations and reflections through matrices or the permutation of vertices. This process of discerning a structure through examples is essentially *reductive*; the way in which 1Y is taught, with both set theory and geometry, is cleary *constructive*. This is conjecture; there could be many other confounding factors at play, but none that seems so immediately obvious an answer.

Modules whose models *did* have significant 1Y coefficients were as follows:

MSM 201 – Polynomials and Rings & Metric Spaces Axiomatic approach to ring theory; quotents of polynomials; finite fields; metric spaces; analytical concepts such as convergence and continuity.

MSM 2B – Real & Complex Variable Theory Limits, continuity and differentiabilty of real-valued functions; intermediate value theorem, mean value theorem and Taylor's Theorem; Analytic complex-valued functions; Cauchy's integral theorem; the residue theorem.

MSM 2C – Linear Algebra & Programming First semester covers the basics of linear algebra and matrices; linear transformations, eigenvectors and the characteristic polynomial; inner products. Second semester covers the basics of programming with C++.

Each of these modules' models has a 1Y coefficient significant at 5%. Besides the programming component of MSM 2C, are all strongly pure mathematical. The second-semester of 2C includes 22 hours of computer labs, in which students spend time solving programming problems for themselves; while this is not mathematical problem-solving, the parallels are clear. The module with the largest 1Y coefficient, and with the largest significance, is MSM 201. Its model also has the largest R^2 value seen; 0.53, remarkably high when considering the range of first year marks for each individual student have a mean standard deviation of 7.63%.

Of further interest are the constant coefficient and that of the first year mean; being -13.3% and 1.12 respectively; the only instance of a negative constant coefficient and mean coefficient greater than 1 over the first two years' models.

6.4.3 Third Year and Mean Results

A similar approach to second year data was taken with third year data. First and second year means were initially used for regression, but when backwards stepwise regression was performed it removed the first-year mean, whose coefficient's F-test significance was only 87.9%. Again, this is due to the high significance of the second-year mean, and its high correlation to the first-year mean.

Statistics for the five third-year modules' models are shown in Table 6.8. There is little here to discuss; none of the 1Y coefficients presented are statistically significant to any reasonable degree. Of most interest is the MSM 3P08 model, whose 1Y coefficient is 4.41%, though because of the relatively small number of data points (86), its significance is only 29.2%. This is still the largest significance presented, however, and of the 21 1Y students who went on to take 3P08 in their third year, 12 scored 70% or more. We speculate that this may become significant with more data, though it is currently not enough to base conclusions on.

Further to the results for individual third-year modules, the means for all three years were set as the dependent variables for linear regression analysis, using the 1Y variable and the previous year's mean (1Aa class test mark for the first year). The results are given in Table 6.9. The first and second year models tally with expectations from individual modules: significant contributions from the 1Y variable in both cases; 8.11% in the first-year model and 3.25% in the second-year model.

For the third year model, the 1Y coefficient was small but negative; -1.48%. This was only significant at 47.8%, however. There is a weak negative correlation between second-year mean and the improvement between second- and third-year means: students who perform worse in the second year improve in the third year more on average than those who perform better. The correlation of the second year mean with third year improvement has a Pearson's r value of only 0.153, but the coefficient of the second year mean in the model is -0.127 with a significance of 1.7%. As 1Y students on the whole perform better than average, they improve less between their second and third years, which may account for the negative coefficient in the model. We reiterate, however, that the coefficient is not significant over 538 data points and as such unlikely to be of importance.

6.5 Conclusion

The results presented above clearly demonstrate a correlation between participation in MSM 1Y and higher performance in some modules in the School of Mathematics. The regression analyses take into account—and demonstrate this to be independent of—any original advantage that 1Y students may have over the rest of the cohort.

First-year results show improvement over all modules, when compared to the 1Aa class test mark. This mark has a coefficient of correlation of between 0.27 and 0.42 with first year module results, and taking 1Y participation into account gives signficant improvements in the correlation across all modules. Its unstandardised coefficient, which represents the percentage increase in mark from being a 1Y student, lies between 4.58% (MSM 1C) and 9.92% (MSM 1Ab); a large increase. The unstandardised coefficient of the 1Aa class test mark, meanwhile, is 0.51–0.57 for all modules, and the constant coefficient between 16.85–22.62%. This means that students who perform weakly on the relatively low-stakes class test improved more on average across all modules; conversely students who score 80–100% on the class test are unlikely to do as well over all first-year modules: there is a bunching-up of results in the 40–70% region. This in turn means that a randomly selected group of high-achievers from the class test would be expected to perform worse on average over all modules; such sizable unstandardised coefficients in the models would not be seen. 1Y students, who perform well on average in the class test, are less likely to be affected by the more difficult summer examinations than their peers. For the first year at least, participation in 1Y offers clear material benefits.

Second-year results are more modest; by using the first-year mean mark, the models for second-year modules already account for most of the benefits of 1Y participation. Even here, however, the effects of 1Y are felt. In three modules covering topics in pure mathematics and programming, the unstandardised coefficient of the 1Y variable added 3.09–5.22% to a student's mark. In three other modules there is no effect evident; a combination of results highly unlikely to happen randomly, or with a selected group of high-performing first-year students. Third-year and year-mean results do not further support the case for 1Y, but neither do they diminish it.

We believe that there is a causal relationship being presented here, but we cannot be entirely sure of it. Overall, the results show that 1Y participation contributes significantly to marks in several first- and second-year modules, however being entirely optional until 2011–12 the effect may be due to the groups' enthusiasm or greater interest in mathematics. Neither can we say for certain that it is the Moore Method, and not the small group size, the intense concentration on a smaller area of mathematics, or the quality of the teachers, that we have to thank for the better results. That said, it is the case that students who take the module perform better in modules that are not directly related to the subject they have studied in 1Y, but which do include an axiomatic approach to pure mathematics. We can furthermore say that the use of the Moore Method has certainly not hindered students' progress, and in that way support the case for Hmelo-Silver et al (2007)'s view that problem-based learning courses which offer suitable scaffolding to students are a worthwhile investment for mathematics teachers.

		Unst. Coeff.					Data		Std.	Change Statistics		
Model		β_i	Std. Error	b	t	Sig.	Points	R^2	Error	$R^2\Delta$	F Δ	Sig. F Δ
MSM 3A02 – Applied Mathematics II												
2	(Const.)	27.97	7.02	0	3.99	0	102	0.23	16.88	0.01	0.72	0.397
	First Year	0.64	0.12	0.48	5.39	0						
	@1YStudent	-14.66	17.22	-0.08	-0.85	0.4						
MSM 3A03 – Continuum Mechanics												
2	(Const.)	4.94	3.41	0	1.45	0.15	353	0.45	14.35	0	0.31	0.579
	2nd Yr. Mean	0.94	0.06	0.68	16.93	0						
	1Y Student	-1.74	3.13	-0.02	-0.55	0.58						
MSM 3A04 – Partial Differential Equations												
2	(Const.)	13.69	3.94	0	3.47	0	248	0.4	13.24	0	0.49	0.486
	2nd Yr. Mean	0.77	0.06	0.62	12.44	0						
	1Y Student	2.23	3.19	0.03	0.7	0.49						
MSM 3A05 – Perturbation Theory & Chaos												
2	(Const.)	1.81	5.49	0	0.33	0.74	157	0.42	14.38	0	0.2	0.652
	2nd Yr. Mean	0.88	0.08	0.65	10.48	0						
	1Y Student	-1.89	4.18	-0.03	-0.45	0.65						
MSM 3P05 – Number Theory												
2	(Const.)	15.55	2.2	0	7.06	0	315	0.59	10.07	0	0.09	0.761
	2nd Yr. Mean	0.73	0.04	0.77	20.57	0						
	1Y Student	-0.57	1.87	-0.01	-0.3	0.76						
MSM 3P08 – Group Theory & Galois Theory												
2	(Const.)	-11.35	7.93	0	-1.43	0.16	86	0.52	16.03	0.01	1.13	0.292
	2nd Yr. Mean	1.02	0.11	0.7	8.91	0						
	1Y Student	4.41	4.15	0.08	1.06	0.29						

Table 6.8: Regressions for third year modules (Model 1 figures omitted)

Model		Unst. Coeff.		b	t	Sig.	Data Points	R^2	Std. Error	Change Statistics		
		β_i	Std. Error							$R^2\Delta$	F Δ	Sig. F Δ
First Year Mean												
1	(Const.)	18.49	1.33	0.00	13.93	0.00	1294	0.38	13.83	0.38	807.51	0.000
	Test Mark	0.55	0.02	0.62	28.42	0.00						
2	(Const.)	18.93	1.32	0.00	14.34	0.00	1294	0.39	13.73	0.01	21.10	0.000
	Test Mark	0.54	0.02	0.61	27.67	0.00						
	1Y Student	8.11	1.77	0.10	4.59	0.00						
Second Year Mean												
1	(Const.)	4.52	1.50	0.00	3.02	0.00	937	0.59	10.32	0.59	1374.73	0.000
	1st Yr. Mean	0.92	0.02	0.77	37.08	0.00						
2	(Const.)	4.95	1.51	0.00	3.29	0.00	937	0.60	10.30	0.00	5.14	0.024
	1st Yr. Mean	0.91	0.03	0.76	36.09	0.00						
	1Y Student	3.25	1.43	0.05	2.27	0.02						
Third Year Mean												
1	(Const.)	12.62	2.20	0.00	5.73	0.00	538	0.47	12.62	0.47	479.81	0.000
	2nd Yr. Mean	0.79	0.04	0.69	21.90	0.00						
2	(Const.)	12.44	2.22	0.00	5.61	0.00	538	0.47	12.62	0.00	0.50	0.478
	2nd Yr. Mean	0.79	0.04	0.69	21.64	0.00						
	1Y Student	-1.48	2.09	-0.02	-0.71	0.48						

Table 6.9: Regression for year means

Chapter 7

Approaches to the Moore Method: Staff Experiences with a Moore Module

In the previous chapter we investigated the progress of Moore Method students within a cohort at the University of Birmingham, through quantitative analysis of students' module results; in this chapter we look at Moore Method module taught by a team of staff at another U.K. university.

This study has three main aims. We want to record the experiences of staff members who have taught the module, and through doing so evaluate the how effective they saw the Moore Method as being. Furthermore, we want to determine the aspects of the teaching staff felt were beneficial, and those aspects that were less so. In particular, the following questions are important to us:

1. What are teachers' preconceptions of Moore Method classes, and to what extent were these met?

2. What were perceived as the objectives of the class, and to what extent were these met?

3. How were the problems selected for the class?

4. What are teachers' attitudes to student progress during the class?

5. What is the efficacy of the Method/module?

6. How has its teaching changed since it was introduced?

7. Was anything surprising or unexpected encountered?

7.1 The Module

Investigations in Mathematics is a compulsory 10-credit module, taken by all single-honours mathematics undergraduates in the Autumn term of their second year. There are one hour of classes every week. The cohort is divided into nine groups of roughly ten students each, working on problems in one of five topics. Students chose a group on a first-come, first-served basis, so while there is a strong element of self-selection, not all students attended their first choice course.

Throughout this chapter, to protect our sources, Macondo[1] will serve as an alias for the university whose teaching we have studied. The University of Macondo is a member of the 1994 Group and home to world-class teaching and research in mathematics. The typical offer for a place on the G100 undergraduate mathematics programme is AAB, with an A in mathematics.

Investigations has been a part of the mathematics degree programme at Macondo for nearly a decade, having been taught using a range of approaches, before becoming a Moore

[1]In honour of *One Hundred Years of Solitude* (Márquez, 2003), which is not about writing a Ph.D. thesis.

Method module in 2009/10. Given the many changes that the module has undergone in terms of students, staff, and teaching methods, it is not possible to establish a causal relationship between the move to the Moore Method and the success of the module; we therefore deal with this 'new' *Investigations* as an isolated module in its own right.

All topics but one are studied by two groups. The material for each topic is set by one staff member; four topics have an additional member of staff who teaches the same material to a second group. The lead staff members choose to author new material or use existing sources as they see fit. The following topics were used in the years 2009/10 and 2010/11 (not all topics were taught both years); the letters in brackets identify the staff members who taught them:

- Dynamical systems (A)
- Special Relativity (B) and (D)
- Algebraic coding theory (C),
- Chaos Theory (D),
- Graph Theory (D) and (E),
- Mathematical Theory of Communication (F),
- Numerical methods and integration (G) [Not interviewed].

Lecturers A and C are responsible for the module in its current form. Eight other members of staff were involved in the teaching in 2009/10 and 2010/11, whose teaching experience ranged from none at all to over 40 years.

The module is taught using the Moore Method, though it differs slightly from the 'traditional' Method as taught at Birmingham and described in Chapter 5. In particular, students are encouraged to look up information in books and online outside of classes, though in a way that better ties together the problems that they work on, rather than to look for answers to a particular problem. Lecturer C explains:

> "We took a policy decision early on that it would be hopeless to follow the classic Moore approach that said that no student would be allowed to talk to another student, that just was not going to work and we were far from convinced that we would wish it to work were it even possible, we wanted the students to talk inside and outside classes. (C25)

Groups working together to solve problems, and research on topics outside class, set the module apart from the traditional Moore Method, though there is still a heavy emphasis on written mathematics and presentations, in common with other modified Moore Methods. Group work was not compulsory, however, and students were free to work alone if they wished to. Unlike the majority of courses in this department, there is no final exam, rather assessment is entirely dependent on work done in classes and at home. Each student's mark for the module was made up as follows:

Assessment	Percentage
Mid-term notebook	10%
Poster	10%
Presentation	20%
Final notebook	60%

The opportunity to analyse a module such as *Investigations* is unique. In a survey of 1843 modules from mathematics degree programmes in England and Wales, closed-book examinations accounted for 75% or more of the final module mark in nearly 70% (1267) of modules (Iannone and Simpson, 2012b, p. 5). Modules taught with a significant amount of problem-solving are rarer still (see Section 4.2.1), and none discovered to employ as many members of staff as that at Macondo.

7.2 Methodology

Because we are interested in the development of the module and the experience of teaching it, and given the rarity of such an opportunity in interviewing several staff teaching the same module, it was decided that individual interviews would be the best way to answer our research questions. In particular, semi-structured interviews were chosen to allow participants to express their views on the course and introduce topics that we as researchers may not have considered, yet still making sure all the questions we wanted to answer were included in discussion.

Another advantage of this approach was the ability for participants to express themselves in their own words, allowing for subtleties of meaning that a questionnaire may fail to pick up. Given that three of the participants spoke English as a second language (though all perfectly fluently), interviews allowed them to correct themselves immediately if they were unhappy with the wording of a particular phrase.

Six members of the *Investigations* teaching staff were recruited for interviews, of a total of ten (four were on sabbatical or otherwise unavailable). Interviews, conducted in May 2011, were recorded and transcribed, totalling a little over three hours of audio and 20,000 words. Interviews were conducted and transcribed by Matthew Badger (R1), Christopher Sangwin (R2) collaborated on analysis. Interviewees were assigned the names Lecturer A,…, Lecturer F in an arbitrary order, and a seventh lecturer referred to during interviews but not interviewed was titled Lecturer G. Students were unavailable for interviews, and securing approval from two universities' ethics committees to interview students would have taken a considerable amount of time.

In March 2012 a follow-up interview was conducted with Lecturers A and C who were responsible for *Investigations*, however Lecturer C had recently returned from a sabbatical and had not taught it that year. This interview was used to record any changes that had been made in the year after the original interviews were conducted, or any that were planned in the subsequent year. This was also transcribed and coded for analysis by R1.

A thematic analysis of the seven interviews was carried out and its results constitute our findings. The strengths of thematic analysis, as given by Guest et al. (2012, p. 17), demonstrate that it is suitable for the work:

- Well suited to large data sets.
- Good for team research.
- Interpretation supported by data.
- Can be used to study topics other than individual experience.

In particular, the iterative nature of thematic analysis fitted well with the approach of using semi-structured interviews: we were able to discover themes that we had not considered before the interviews had taken place, and to tune finely the coding in light of the data.

While there is "no accepted, standardised approach to carrying out a thematic analysis" (Howitt and Cramer, 2011, p. 329), Braun and Clarke (2006) describe an approach that is recommended in a number of texts (Howitt and Cramer, 2011; Matthews and Ross, 2010), and shown in Figure 7.1.

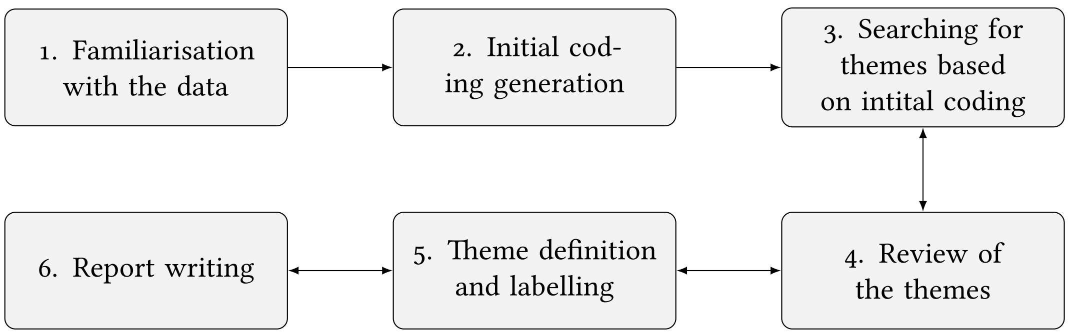


Figure 7.1: The structure of Braun and Clarke (2006)'s thematic analysis, adapted from Howitt and Cramer (2011, p. 336)

Work on the interview transcripts began with R1 and R2 independently coding four of the interviews (A–D), and subsequently grouping these codes into themes. At this stage, R1 had 18 codes, while R2 had 15. Of these, 14 were common to both, although some were subsets of the other's. For example, R1's "Differences to 'The' Moore Method" was R2's "Relationship to Moore Method". Some codes of one researcher were subdivided by the other, for example R2's "View of students' performance" was subdivided by R1 into "Efficacy of the method", "range of student response" and "staff understanding of student knowledge". From these independent codings a new coding scheme was developed to reflect their similarities and differences. Two transcripts (C and F) were then coded with this new scheme and the two codings of Lecturer C's transcript were compared. Both R1 and R2 used 48 codes on the transcript, half of which agreed exactly, half being changed by one or other researcher after discussion. These resulted from differing interpretations of the language of the coding scheme, which was then revised as a result. Two extracts on other colleagues' teaching and students' comments on the modules were unable to be coded, so codes were added for these. The particular problem of nomenclature was resolved by referring to the course since its move to the Moore approach as *New Investigations*. From here, *Investigations* refers to the module in its original form as a problem-solving course.

The final coding scheme was drawn up from these discussions and is given in Table 7.1.

7.3 Analysis of Results

Sections of interviews were grouped into themes according to the coding scheme developed and discussed above. We present these results as a narrative, making extensive use of quotations from our interviewees, chosen as representative examples of the relevant codes in our coding scheme. The letter and number in brackets after a quotation show the lecturer and line-number from the interview transcript: (B29). The follow-up interview is denoted *A&C* with the letter

Before *New Investigations*	
	Motivation for New Investigations
B1	Dissatisfaction with current learning/shortcomings within current programme.
B2	Purpose of *New Investigations*.
	Planning of the module by the department
B3	Description of the course setup, technical details.
B4	Differences to "The" Moore Method.
	Planning of each section
B5	Staff background. Their learning as a Moore Method student. Research area, previous teaching experience with and without Moore Method.
B6	Importance of "ownership" by staff of their course, e.g. variation on "The" Method.
B7	Designing course problem sets.
B8	Staff expectations of the course.
B9	Support, help, advice or problem sets from other colleagues.
During *New Investigations*	
D1	Descriptions of events in class.
D2	Staff view of students' performance.
.1	Range of individual students' performance.
.2	Efficacy/group performance.
D3	Problems encountered during the course.
D4	Perceived changes in student behaviour, or student experiences.
D5	Surprises or unexpected events.
After *New Investigations*	
A1	Staff understanding of student performance in general, e.g. appreciation of their knowledge.
A2	Reflections and recommendation to others.
A3	Changes made (after year 1), or proposed.
A4	Changes to other courses as a result.
A5	Student comments about *New Investigations*.
A6	Discussion of others' teaching *New Investigations*.
Outside *New Investigations*	
O1	Discussion of other courses and teaching unrelated to *New Investigations*.

Table 7.1: Final coding scheme agreed by R1 and R2.

of the lecturer who was speaking written before the quote: (A&C73).

7.3.1 Module Planning and Motivation

To address our primary research question, *to determine the efficacy of the Moore Method as implemented at the University of Macondo*, we have first to understand what the desired effect was. The module, presided over by Lecturers A and C, wanted to deal with the continuing problem of making *Investigations in Mathematics* meet its goals, namely to introduce students to problem-solving and proof-writing:

> [...] for me one of the primary purposes really was for them to try to develop the ability to do rigorous mathematics and to understand what proof really is. (A51)

Lecture D echoed this sentiment:

> I think that the main thing is that it forces students to think about things that they don't normally think about, for example how to construct a proof, what really counts as a proof. But more important how to construct one, how to build it yourself rather than just work through something that has been given to you. (D19)
>
> [...] this has got the additional factor that I've just mentioned of forcing thinking on difficult topics. (D27)

Other colleagues stressed the importance of students driving learning for themselves:

> I was really trying to plant that seed of questioning in them, like, okay, you can use some basic formula, but do you know how one gets to this formula? I think that to a certain extent I succeeded in that goal, at least for some students. I'm not sure if everybody would really share that opinion that you have to understand the things you are using, but that was one of the outcomes. (F86)

> In terms of the goals we tried to explain to them as early as possible that the only outcomes would be the ones they created so if they wanted to take the topic to the end it was for them to get there. (B39)

At its most fundamental level, the module aims to teach students both to think and communicate mathematically:

> C: What do we say? To my mind one of the most important objectives is it is training people to think. How can I? It's, do you have a phrase?
>
> A: Training people to think logically through an argument and to develop the ability to present their arguments.
>
> C: It's thinking skills and presenting skills, but the thinking skills are...
>
> A: To get people to engage on a more fundamental level and to see how ideas develop from the ground up, as opposed to starting at the level where you're just manipulating...
>
> C: It's the distinction between mathematics being developed in order to develop a

> tool to solve something as opposed to learning about a finished product. So it's seeing mathematics in a more live way.
>
> C: Maybe rather than presenting we're talking about communicating. That's the skill we want them to leave with, their ability to be able to explain something whether it be verbally or in writing or... so it's a communication skill. (A&C33)

Previous incarnations of the module had not been entirely successful at meeting these goals:

> “ ...they've all been aiming at trying to get students to digest work in different ways and think about it in a deeper way than the one digested lecture-note course tends to lead to. And one by one our various attempts were unsuccessful, I would say, they were not as successful as one would have hoped. (C19)

Part of the the dissatisfaction was the way in which the module presented uncontextualised problems to students, instead of including any narrative:

> “ ...before it was organised very loosely as just a problem solving course. [...] It was just a list of individual problems that had no overarching theme and the students were asked to approach those problems as best they could... (A19)

When it fell to Lecturers A and C to develop a new version of the module, the aim was to use the Moore Method as a basis for developing ideas on how the module was to be run. This came in part from Lecturer A's own experiences of the Moore Method as an undergraduate in the United States:

> ...it wasn't so much an intention that we would do a Moore Method course but we would attempt to wrestle with this ongoing problem perhaps using some of the ideas that he [A] and I've subsequently picked up. (C21)
>
> As much as anything our driving desire was not to follow received wisdom but to follow in my own case my own instincts from 12 years of seeing things that worked and didn't work. (C23)

Group work was encouraged, but students were free to work alone if they wished. Furthermore, students were only assessed in their groups for the presentations to other groups:

> C: It's sort of implicitly there that it would be to their benefit if they have [worked in groups], but it's not one of the...
>
> A: In fact one of the comments that we get from the panel that was reviewing our curriculum change was that we should institute some way of having groups assessing each other. Or assess within the group, but we don't see how to do this in a reasonable way. (A&C51)

Besides group work, the other significant difference to a traditional Moore Method module was the encouragement students were given to research around their problems either from books or online. Steps were taken to prevent them being able to find solutions to problems, however, and so they still had to solve problems for themselves:

> C: Yes, I thought about that because you [R1] said something very similar on your report and we have various different schemes to prevent that here, but, a variety of different things. What did you do?
>
> A: I did something similar,
>
> C: You didn't tell them what the big theorem was…
>
> A: Yeah … and I think another lecturer did catalogue numbers and not telling them that they were…
>
> C: Yes. In others you prevent that by doing such very specific problems that, so going back to the special relativity, it was step-by-step with very specific situations and names of particular people, you would not be able to go to the internet and find "Fred set out at this time…"
>
> A: What I did this year was kind of interesting. There are lots of different fractal dimensions and lots of different equivalent definitions of the same one, so I picked just one specific one and it's one that virtually you'll never find people using in that formulation. It makes it easy also to spot because people will frequently use one of the more common ones.
>
> A: I guess Moore had something similar, I don't know if he intended, 'cause he…
>
> C: He didn't have the Internet! (A&C223)

The individual staff teaching the module had a degree of autonomy but the progression of problems seen in the Moore Method was highly important:

> I think some people have just thought of it as again a problem-solving course, but with just a theme developed throughout, whereas for me one of the advantages of the Moore Method is that it can teach a student how to really develop something from ground zero and understand it in a more fundamental way and to be able to actually see the rigorous development of the topic. (A33)

Lecturer E was particular forthright in questioning the use of isolated puzzles, but found the structured approach of *New Investigations* more appealing:

> if you spend time teaching [...] puzzle solving, you spend time on puzzle solving and you don't teach students proper mathematics. What does it mean "proper mathematics", it means the students need to know mathematics theories, and so what is mathematics? It is some sort of building, right, and there are very deep theories, and very nice constructions and so on, and the students don't see the beauty of mathematics, don't see this construction, [...] Okay, the students can do this in their free time but what is the point in teaching them to solve these puzzles at university? [...] This criticism is reasonable because we should not overestimate these things, problem solving. This is fine to teach students a bit but we need also to pay attention to this classical university education of course. (E58)
>
> Yes, I think there should be something like that. It shouldn't be too much, right, [...] but it is reasonable to have a few. The students should have maybe one or two times through their studies here a couple of modules like that. (E132)

Interestingly, he cited this course structure in other guises, drawing a comparison with similar methods not referred to as the 'Moore Method':

> I wouldn't call it Moore Method, it's an old idea. [...] I came from the Soviet Union and I know some lecturers did use this method, it's quite standard, and I know some books are written in this way [...] (E42)

Lecturer B, who had been teaching for six years, including time as a graduate student, was particularly motivated to teach in a different way:

> For me it was curiosity to teach it precisely for this investigations aspect, to see how you could teach it in a classroom without a strict text book or me lecturing at them. (B19)

As demonstrated by the above extracts, all the lecturers interviewed showed an interest in the teaching of the module, and were generally enthusiastic in embracing the idea of teaching in this non-standard manner, at least for a single module.

7.3.2 Choices of Problems

As discussed above, problems were chosen for each topic by the topic leader, four topics had a second group taught by a different member of staff using the same material. All those interviewed except Lecturer D had been the topic leader, and so chosen the material.

An immediate issue that presents itself in choosing a topic and problems to go with it is the level of abstraction, and the speed with which students can begin to work on problems themselves:

> [...] I think the students always respond better with something they perceive more directly so if it's something where they can draw pictures or do something directly without too much difficulty on the computer to supplement their understanding then it tends to go much more quickly. I don't think most of our students are comfortable with abstraction and so the topics that are more abstract are just more difficult for them and they go much more slowly. (A75)

> So this is why, for example, I've chosen Graph Theory. It may be a bit more suitable for this method. [...] there is not so much background they need to know. (E44)

Lecturer E used pre-existing Moore Method course notes from the *Journal of Enquiry Based Learning in Mathematics* (Clark, 2007):

> Yes, everyone got a copy because it's free online and this was specifically designed for students, I printed it for everyone and then we just went through problems. (E76)
>
> I didn't know how to do most of the problems, although they are elementary, and [...] so it was okay for me. I managed to solve all of them, or most of them, right. So I didn't need to go through Graph Theory books or whatever. (E62)

Lecturer D was happy to teach a topic with which he was not familiar:

> “ But I wasn’t very familiar with the more pure mathematical side of it that the material we had steered us along. In a way I didn’t think that was a bad thing because it put me in the same position as the students and I could see a bit of the difficulties they had and also I was less tempted to give them my already pre-thought-out answers. (D49)

Instead of building a topic up, Lecturer F opted to give students a paper (Shannon et al., 1949), in the hope that they would be able to bridge the gaps in it:

> “ [...] every student was given the paper. It didn’t contradict to the whole purpose of the course because the paper’s written in such a way that every formulation is not rigorous enough, and the proofs are not rigorous enough either. So what we’ve done, okay the idea is there, and the outline of this theory is there, so my plan was “Okay, why don’t we build a Mathematical Theory of Communication?” using Shannon’s work as a background. And of course depending on the level of detail and the level of rigour, you can arrive at different goals. (F52)

This was less successful than planned, mainly because of the difference between the expectation and reality of students’ abilities:

> “I wanted the idea of a hook, and so I gave them Shannon’s paper, but in fact it was a little bit too difficult for them to get through. (F107)

Reflecting on this experience, he went on to comment:

> “But problem classes must not be done like “Here is a set of problems, I’ll see what you do.” It must be really carefully planned and supervised. [...] if the students are stuck for more than five minutes, then my experience suggests that you need to interfere, you need to step in and give a hint. [...] you need to make sure they’re not sleeping so you need to keep their competitiveness up, their desire for mark up. Okay for deep learners you don’t need to stimulate them, they’ll do anything you ask, but the rest of the group need to be also stimulated somehow. (F132)

Students were given the choice of which topic to sign up for; this was cited as being significant:

> “I deliberately allowed them to sign up for what they wanted [...] so where possible folks were with their friends so they were comfortable with the people they were talking to. I think that was important. (C37)

> “I think it played a role for this group because they all signed up for the topic they wanted. They also signed up with peers whom they shared that interest with, so pairs and groups that signed up at the same time. [...] I do think that was [...] why the group work and the discussion groups went so well, because I had students

> who had chosen the topic that they wanted to and they were all sort of curious about it in some way. (B73)

Due to the limited space available, not all students managed to get on their first-choice course, however this did not necessarily devalue the experience they had.

> “So I definitely had a student in my group who told me that he had tried to sign up for a different one, but he was gracious enough to say that he was quite pleased that he'd come to this one and that he'd learned something from it. (D113)

The fundamental difference between *New Investigations* and its previous versions, the interconnectedness of the problems given to students, was seen as an important part of its success:

> “I think this was the first time the course had worked satisfactorily, or more than satisfactorily, and I think one of the reasons why was that, we'd tried PBL before but they tended to be disconnected problems, problems for their own sake, whereas what we were doing here was we were attempting to teach a particular topic through a sequence of problems, and I think that gave a narrative to the module that I think engaged people more. (C81)

7.3.3 Events During Class

Besides the first session, in which the way in which the module was taught was explained to students, and the relevant material distributed, the aim was that students would do the majority of the talking in class:

> C: You want to be taking as little part as possible, but the criteria I used is to, if at all possible, say nothing other than asking questions. So I think I would aim never to explain anything during the term if I can get by without doing it, I would only... now that might mean that I have to ask blindingly obvious questions! Making smaller, and smaller, and smaller steps, but I think it is that you want to encourage thinking. (A&C117)

Before the module began, staff expected progress would be slow; all were surprised at how slow, however:

> I had expectations beyond what was realistic, I think, at the very beginning as to how much they could get through [...] And in a way I'm happy that they follow at whatever pace happens because I think that we are going into things more deeply, [...] if you skip to the next thing then it's almost been a waste of time. So I'd rather go very slowly and try to achieve something; It's not a way to teach a large quantity of material, one gets through a tiny proportion of what you would do in a lecture course but you are teaching something else and hopefully skills that are just as worthwhile. (C51)

> I would say on average we had about two problems per week [...] (B51)

Because the topics being studied were not essential for third-year modules, however, this proved to be an opportunity for students to develop their understanding to a deeper level than recalling definitions:

> Quite slowly I think. They take time to grasp that they are going to have to do them [the problems] themselves. Then with a new topic it's often quite hard for them to get their heads round the definitions, which is actually what I spent a lot of time on with them. We got the definition that was given to them and tried to unpack it. [...] they just couldn't grasp how they would prove anything about isomorphism and it really was a question of getting them to unpack the definition and getting them to see that what you're saying is that there is a one-to-one map from one set to another set with certain properties being preserved. And although they would probably be able to produce that definition on paper they had no idea how you might use it in a particular example. It's almost as if you were saying that the simplest things were hardest for them. (D57)

Lecturers A–D reported positive experiences during classes:

> I got the impression that everyone was contributing, [...] they had some [...] people who were more involved with the debates, would more easily be the ones to have two opposing opinions, and therefore debate them out. Nevertheless everyone else still actively participated, which even in a normal classroom if I just had to teach at the blackboard for 10 people, just a small class, it's difficult to get that sort of attention. So that worked well. (B39)

The activity of the classes themselves centred around the cut and thrust of mathematical discussion:

> That seemed to be working quite well because there was a lot of discussion going round about those solutions and how you would approach them and what you would need and the initial solution was proposed and then ten minutes later taken down because they realised there seemed to be an aspect in their discussion that they didn't seem to overcome so they actually went back. (D33)

However, this was not always easy to initiate. Peer-pressure, rather than expectations of the lecturer concerned, seems to have been an important motivating factor for at least some of the students.

> the students were actually pretty reluctant to get themselves involved in the early days. It took a bit of work to get the idea of the game. As a result; towards the end of the term the group had to make a presentation on what they'd learnt to other groups who had been studying other topics, and when that began to come over the horizon it fired them into life and they realised that there was something here that they had to do or they were going to look like fools in front of their colleagues. (C39)

Lecturer E was less satisfied:

> In general, I think my expectation was that the students would participate a bit more. And would go forwards rather than backwards. (E116)
>
> Some students did [...] solve problems, but in general what happens eventually you come to the class and they start to think about the problems in the class [...] the students are not prepared very well and the time is limited. (E118)

While Lecturer E was disappointed, B had a much more positive experience:

> I think I was quite fortunate with my group, they were very motivated, worked well as an overall group. (B37)
>
> The students were all interested in the topic itself and took it up [...]. As I said it varied depending on the student, in the discussion obviously you might have the more vocal ones to cover more, from the notebooks you could see more how the level of understanding went. I would still say it was a very good level, there didn't seem to be students who just said "okay it's hopeless." I'm not sure how they managed to get through the ten weeks of the discussions without having understood a bit. I didn't have any ones who you'd say the setup was not suitable for them. (B43)

This lack of uniformity in the groups was discussed by others:

> I had two significantly different groups, first year they went very slowly. Extremely slowly. There was another group using the same topic and they went through it more quickly, I think it had to do with the composition of the group more than anything else. (A63)

> …maybe it's pure luck, but it just seems to me that the groups we are given, they are random. In some groups you might have two students who will be driving the whole group, but in some group there are no obvious drivers, no obvious leaders, no obvious outstanding students who would be able to really break through. At the end of the day such a person would emerge, but not at the beginning. (F109)

7.3.4 Benefits to Students

Most lecturers found that once their group had got started, students engaged well with the material and each other:

> However participation-wise and enthusiasm-wise I was surprised. Maybe because I had a specifically good group, having spoken to some others there were groups where it worked well and groups where it didn't work well, so maybe it was due to my group, but certainly they really took the topic for themselves I didn't have to give any pushes and say "come on you need to get on with this", it was their initiative they were quite self-motivated as a group and worked together very well. So that was beyond what I expected I thought it would be more difficult to get to

> that point. (B83)
>
> There were two notebooks that really stood out because without having given them a set of lecture notes the progression in the notebook was still nevertheless one which you could have used as a sort of mini course. (B69)

> “ I wouldn’t have wanted to say that the proofs were always as good as they could be, or even necessarily correct, in some cases they proved things that were false, but in a way I didn’t think that mattered - they’d got the idea of what you were trying to do. I think they had a feel for what a proof was about, what a counterexample was about. And some of them of course were better than others. (D73)

Presentations, a key component in a traditional Moore Method module, was important in *New Investigations*:

> “ ...it was developing presentation skills and that proved to be a more significant part than I’d perhaps appreciated because [...] it’s one thing for a student to prepare a presentation and then go up there with his or her piece of paper and write it on the board and say “I’ve done my presentation”, but when there’s active discussion and they’re trying to answer questions and think at the board that’s a rather different skill and it’s one that takes a lot of practice and so I think that took off as something that the students were able to get something more out of it than I’d perceived at the beginning. (C29)

> Once we got going, once they realised what was happening and once they realised that it didn't matter if they said something stupid, and they were, I think they'd done quite a bit of work some of them, in between the sessions. They'd come along with questions and ideas. (D67)
>
> The other group was more difficult. Some of them did do some work, between sessions, but it was hard to capitalise on that, to feed it in and to get everybody up to speed on it. (D69)

However, some students and at least one member of staff would have preferred more formal lectures:

> [In 2009/10] A couple of the students I think were a bit frustrated and wished to be given a bit more direct instruction, though generally I think they were happy with it. This year the response was much more positive, I didn't really get any negative feedback. All the students seemed to enjoy the module, that was my impression. (A67)

> ...if we were allowed to introduce changes then I would propose to use at least a part of this formal lecturing in the module. It's quite difficult because the whole idea is to pick different topics and then excel in them; how can you find so many different lecturers for those topics? (F133)

7.3.5 Benefits to Staff

One of the strongest themes that appeared was that the module gave staff a much clearer idea of what students did and did not understand:

> Well I think they improved. They didn't gain the ability to do as well as I thought they would with making a rigorous argument, but they did make significant progress so I was pleased in that sense. But I think what Lecturer C and I both noticed from doing these modules is that you actually confront much more directly what the students do know and understand, and it's always a bit shocking as you stand in front of a large group lecturing you assume as they're sitting there that they're following what you're saying but when you start interacting directly with the students you begin to appreciate that their understanding is very limited, [...] (A71)

> [...] if I were to teach the same course as lectures now, I would use precisely those problems to identify what I need to teach differently, because I saw better what the students were capable of and what they found quite hard. (B79)

> I think it's pointed to how limited other forms of teaching are in terms of how much they take away as transferable thinking skill and analytic skills. (C41)
>
> I'm thinking about what they take away as cognitive skills for analysing mathematics and writing mathematics well. As well as a certain amount of pure factual

> knowledge although I would say that that was the least of the problems [...]. But you begin to realise it's fragile knowledge and as soon as they are moved to use the same skills but in a slightly different context it's as if the skills disappear. (C45)

> They can mechanically do things but they very rarely understand what's going on behind. (F28)

The lack of understanding that the was demonstrated at times led Lecturer F to question the position of the module in the second year:

> Independent thinking. I doubt that I was successful in this, because in my experience I'm not sure if [they] are capable of doing that. [...] It's because at this point they still lack information, fundamental information, and mathematics is a cumulative science. [...] So in this respect I think it is a little bit too early for them to have such a course. (F88)

Lecturer F was more positive about the discussion that took place; Lecturer D meanwhile was drawn into thinking about the ideas more deeply too:

> Well, see, this an interesting version of scientific seminars. Scientific seminars are where you really do discover, do research, and then you're being judged by your peers, and some other people, and there is room for discussion, for argument. It's an idea like that. But what I found was that it doesn't appear out of no-where if you leave the students as they are. You must create a framework for that; say

> "Okay, this is how we're going to work." Then they would accept it in the beginning and their expectation would be in this framework. So I would do that again, yes. (F117)

> I realised that in all the many [42] years that I'd been looking at all these essays on the history of maths I'd never fully got my head around some of these proofs on the map colouring theorems. In a way you don't have to when you're marking an essay on the history of it, but it forced me to look hard at some things that you just take for granted. (D85)

7.3.6 Difficulties and Drawbacks

Lecturer F initially had problems with students engaging in the process:

> [...] what I found out was that if I just leave things as I was recommended initially, don't interfere, just observe and monitor what's going on; nobody goes anywhere. (F60)
>
> So what was happening in the first two classes, is that when I set up these tasks, I found out after the first week that not everyone, well basically nobody, was successful in solving them. And then I asked students one-by-one to go to the board and present what they'd done. I found that really that was dissatisfactory. Next week, same story. (F62)

> How do you let students solve a problem without interfering? I think that's a completely different skill, I don't know how to do that, even with my own children I don't know how. (F28)

To address the issue of students being unwilling to engage in discussion, they introduced some structure, though they had reservations about doing so:

> "I assigned a presenter for every task, in every week, and in addition to assigning a presenter, I assigned a question master. For each topic, for each question there was a presenter and two question masters. The question master's job was to ask questions, which makes sense, and also, should the presenter have difficulty, to take over the presentation and present their own solutions. (F64)
>
> I partially destroyed the whole point of the exercise. But, at the end of the day, we moved and we actually progressed, at a certain point, because everybody in the class was listening to the presentations. (F66)
>
> Once you engage, it starts on its own. [...] At the end of the day everybody was engaged and [...] they did make a reasonable effort. (F78)

The most significant theme to emerge was the difficulty of dealing with differences within a group, and engaging students from the beginning:

> So one of the things I do find problematic with the method in general is that with such a small group of students, it's the luck of the draw. If you have a group of people with no motivation [...] the whole process can grind almost to a halt and go very slowly. And other times you may have someone that's way ahead of everybody else and that student may become very frustrated with the rest of the group. (A63)

> And probably the other thing to say, and I guess this is true of any problem-based learning, those that get enthusiastic and are able throw themselves into it and have a wonderful time, and those that don't, don't. There is quite a polarity, of the folks that don't, there's only a relatively small number but there are those that don't, never get it, and probably for them it's a dreadful experience. (C47)

> The main problem is to force students to do something, to motivate them. This is what happens in teaching anywhere, the students before an exam only, they don't do anything unless there is some assessment. (E120)

> I think that you need a lot more, you need to think a lot more about how to find ways of getting students to participate and getting them to participate early on. It's not too difficult to get the quieter student to speak and chip in sixth week into the course but you want to try and get something happening earlier than that I think. (D125)

> C: The rest [three students in a class of six] were good, solid, 2:2/3rd students who were trying to do specially relativity, and they may or may not have been working outside of class, but if they were they weren't getting very far. They may genuinely may not have been getting very far. So yeah, there it was a question of increasing the level, working with the three that did turn up, step-by-step, usually with someone at the board, trying to prompt the others to basically help them, but that would be such a relatively different group to the others that I did. (A&C151)

Engagement proved particularly difficult for some students whose English was not sufficiently developed:

> It had four Chinese students and they were not very articulate students, very difficult to get them to say anything in the class, possibly because they weren't confident in the use of English, not in that sort of conversational setting. (D61)

This problem was previously encountered at Birmingham:

> Students who are not fluent in English clearly find the course hard, at least the way I teach it, with a good deal of often fast moving class discussion. — Good, 2006, p. 38

7.3.7 Changes to the Module

In its second year, changes were made to New Investigations to mitigate some of the problems experienced in its first year, mainly to increase student engagement during the first half of the term:

> They had to design a poster as a team, about what the topic was about in the first few weeks. There was a notebook component that they had to keep and again in the second year we had an interim hand-in halfway through [...]
>
> It was a very small assessment of it and was a means of giving them feedback as to how they were doing and what the quality of writing was and how they should align themselves for the second half of the module and their final main write-up. I think that probably worked quite well as well. (C73)

In the follow-up interview, it was revealed that the number of teaching staff was to reduce for the following year—2012/13—to make organisation among other modules easier:

> A: I don't know the specifics of that but there may just be two or possibly three people doing the... and that's not to do with the experiences that we've had but it's more a question of efficiency for staff. Some people don't like the idea of having to do many little different tasks, and so I think it's more efficient that if just a couple of people take take care of Investigations. (A&C177)

7.3.8 Reflections of Staff

In general staff saw the module as a success, though with caveats mentioned in the previous section. Some aspects of the problem-solving involved in the module have been used in other modules in the mathematics department. When asked if the course was worthwhile, Lecturer C responded:

> Oh, absolutely, yes, I'm enthusiastic. We have attempted to embed in small ways the module as parts of some later year modules. I think Lecturer A tried in his level 4 module to have some component, he was teaching a dynamical systems module and was trying to get some component run in this fashion. And we had a 3rd year module which also included a stream. I think they were met with moderate success but that is whole new step again to try to embed it as part of another module. (C85)

When others were asked if the course was worthwhile, Lecturer B responded with an enthusiastic "Yeah!" (B87); Lecturer A said "I do, I think we're still tweaking it but so far it has been worthwhile." (A95) and Lecturer D offered:

> I think it was, yes. I don't know how the department view it, but they've got to accept the fact that you're assessing very different things from the traditional [modules]. (D121)

> Yes, I think there should be something like that. It shouldn't be too much, right, [...] but it is reasonable to have a few. The students should have maybe one or two times through their studies here a couple of modules like that. (E132)

7.4 Conclusions

One of the strongest themes to emerge from all six interviews was an enthusiasm for teaching, and a willingness on the part of all colleagues to try a non-standard pedagogy that all but one had no experience with. None advocated using the Moore Method as the principle means of teaching undergraduate mathematicians—Lecturers E and F both explicitly warned against this—but even those who had been assigned the teaching (B, D, and E) were happy to 'give it a go'. Lecturer D found themself working on proofs that had previously featured in their history of mathematics teaching, but that they had not considered as carefully before. Naturally our sample is self-selecting; someone showing little enthusiasm would be unlikely to be assigned the module and even less likely to agree to an interview; we must bear this in mind.

Another strong theme was the gulf between the expectation and reality of the mathematical knowledge that students had at their disposal. All members of staff had been warned that progress would be slow, especially at the beginning of a Moore Method course, yet all were surprised (and some somewhat frustrated) with exactly how slowly progress was made. The 'shock' expressed by Lecturer A was representative of all. Progress was made in all groups, though not uniformly so. Both Lecturers A and D each taught two groups dramatically different in the speed of their progress. Lecturer F eventually resorted to assigning questioners and answerers, and including an amount of exposition at the beginning of a class, but none of the others altered the intended approach to teaching to such an extent. It is telling that Lecturer F began with material that was considerably different to that seen in a traditional Moore Method course. Taking an existing paper and attempting to 'plug the gaps' is at odds with Coppin et al.

(2009, p. 57)'s advice to "progress from the simple to the complex", and it is perhaps the case that the process suffered more as a result of this. Their method of teaching by the end of the module would be difficult to describe as a modified Moore Method, but something entirely different.

Besides progression through the problems themselves, student engagement in the whole process was difficult to begin with, especially in the first year that *New Investigations* was taught. The lack of examinations is unusual, and introducing the poster and mid-term notebook played an important part in encouraging students to engage with the module earlier. Moore was able to choose his students, and fostered a highly competitive environment that not all were comfortable with (see Section 5.2.4), but slow progress is inherent to the Method and covering a large amount of material was not an aim of the module.

There were two principle differences between 'the' Moore Method and the approach taken in *New Investigations*. Firstly, there was an emphasis on group work that is not present in the Moore Method *until* students have tried to solve problems for themselves. Students were not required to work in groups, however, though most chose to do so. Secondly, finding information outside course material was not proscribed but encouraged. Both of these changes reflected the aims of the module itself; students would otherwise not have the opportunity to work in groups during the undergraduate degree, and information gathering is an important skill for a graduating mathematician. In the second instance it was not the intention that students would be able to find solutions to their specific problems online, however; lecturers obfuscated problems and used non-standard definitions and terminology to prevent plagiarism.

In 2011/12 all the staff involved with the module had experience with teaching *New In-*

vestigations and Lecturer A was once again positive about the way it had run. The reduction in the numbers of staff teaching the module in 2012/13 aimed to make organising staff more straightforward, and allowed Lecturers A and C greater control. While changes to the module were made over the four years that it had been taught using the Moore Method, overall teaching remained unchanged. In that regard it appears as though the module has been a success for most, if not all, students.

Part III

Computer Aided Assessment

Chapter 8

The solution and comparison of equations

In Part I of this thesis we investigated the state of problem-solving teaching in universities in England and Wales. In Part II, we looked at the *Moore Method*, and its use in two universities in England. In this part we look at a separate area of mathematics education, namely *computer-aided assessment* (CAA), though with the same purpose—improving students' problem-solving skills—in mind.

8.1 STACK and Computer-Aided Assessment

The work discussed in this part of this thesis focusses on STACK, a *System for Teaching and Assessment using a Computer-algebra Kernel*, is a computer-aided assessment (CAA) system for mathematics. A review of other computer-aided assessment software is beyond the scope of this thesis, we recommend Chapter 8 of Sangwin (2013, pp. 127–161). Figure 8.1 shows an example question with a student's response and feedback. A student's answer is entered as a mathematical expression, whose mathematical properties are then determined by STACK in order to mark it and give appropriate feedback.

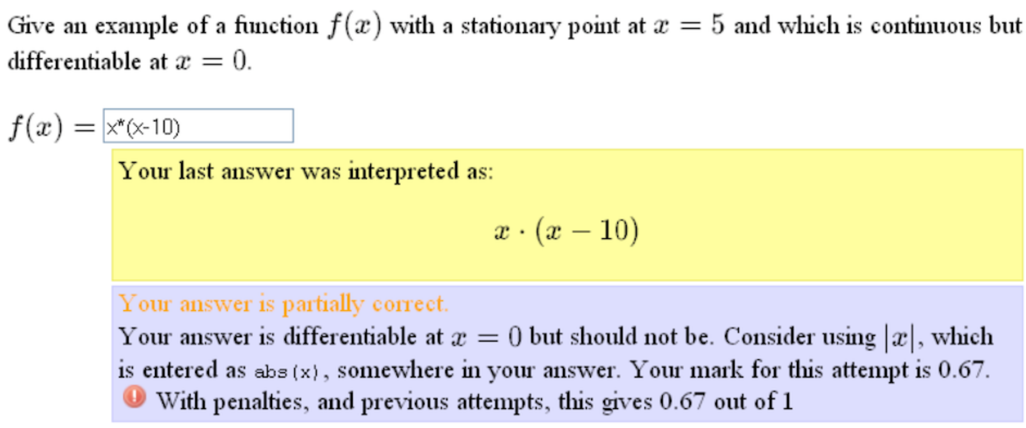


Figure 8.1: An example STACK question

For many questions a student's answer will have to be both algebraically *equivalent* to the correct answer, and in the appropriate *form*. A correct answer need not be unique, however, and STACK establishes the relevant properties of expressions to test objectively a student's answer. A distinguishing feature of STACK is that the feedback it gives may include calculations based directly upon the answer that the student has entered. Figure 8.1 gives an example of this type of feedback.

STACK establishes the properties of a student's answer and generates necessary feedback using the computer-algebra system (CAS) Maxima. It may also use Maxima to randomly generate specific instances of a more general question, so in the example above it could produce several versions of the question with a stationary point at $x \in \{1, \ldots, 10\}$.

The prototype test of correctness seeks to establish algebraic equivalence between a student's and teacher's answers, however this is limited to instances where there is only one correct answer (up to algebraic equivalence). The example in Figure 8.1 cannot be assessed using only algebraic equivalence. STACK provides a library of *answer tests* for assessing different types of equivalence, and more than one test can be applied to a student's answer (we cover this in Section 9.1).

The motivation for the work documented here was the desire to introduce a new answer test that would allow teachers to assess systems of equations, primarily because of their relevance to modelling and word questions (Sangwin, 2011). Assessing problem-solving is currently well beyond the capabilities of computer systems, but by requiring students to interpret information correctly and form it into coherent mathematics, modelling uses aspects of mathematical thinking also seen in problem-solving. In particular, we want to be able to assess systems of *multivariate polynomial equations*, i.e. polynomial equations in more than one variable, not something currently assessible in any other CAA systems. Before we introduce the mathematics involved in doing this, we begin by discussing the way in which STACK determines the correctness of an answer in general.

8.2 Validity and Correctness

When a student submits a response in CAA, it is their mathematics that we want to judge, not their ability to write correct syntax. For this reason, and others discussed below, submitting answers in STACK is a two step process. In the first step, the system checks that the student's answer is *valid*, and does not penalise a student whose answer fails to be adjudged as such. In the second step, the system checks the answer's *correctness* by establishing suitable equivalence with the teacher's answer.

8.2.1 Validity

The first job in deciding an answer's correctness is to determine whether it can be parsed into a valid mathematical expression. A mathematical expression is defined to be either an atom, e.g. a number or variable, or an operator with the correct number of operands. This definition is recursive: an operand may be another atom, or an operator itself. For example, $3x^3 + \ln(x) = 0$ is an expression, consisting of the operands $3x^3 + \ln(x)$ and 0, and the operator $=$. Its first operand is another operator, $+$, with the operands $3x^3$ and $\ln(x)$; its second operand is the number 0. $3x^3$ and $\ln(x)$ are also operators with operands.

In Maxima, an expression is either a function (operator) or an atom (number or variable), where the arguments of any function are themselves expressions. Thus, the expression $3x^3 + \ln(x) = 0$ can be represented with the following tree structure:

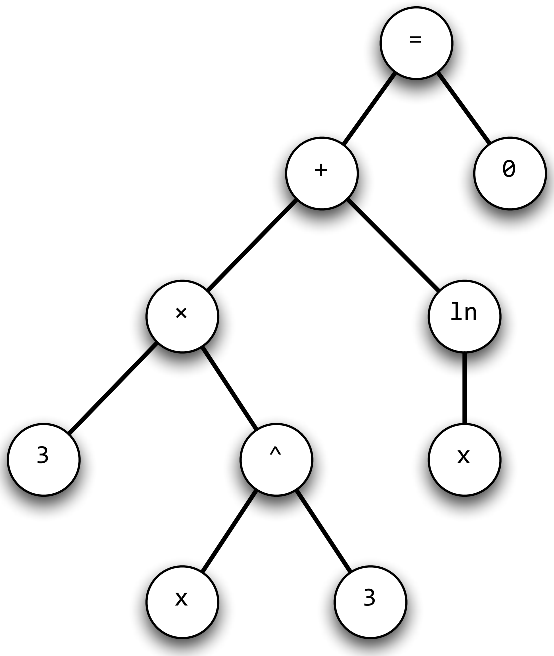


Figure 8.2: A Maxima expression tree

In this example, we have the unary function ln, the binary functions = and ^, and the n-ary functions + and ×. Were any function to have too many or too few arguments, for instance $3x =$, we would not have a valid expression. The first task in determining whether a student's answer is valid is to determine its *syntactic* validity: is it a valid expression in Maxima? This is a complicated process, but includes check that each function has the correct number of arguments, and brackets match. Once this is complete, the system checks that an answer will not compromise the system when input to Maxima, and that it does not circumvent the question by employing Maxima fuctions that it should not, e.g. `int` in an integration question.

Checks for syntactic validity and malicious code are performed on all answers input to STACK. Restrictions on functions are determined per question. Besides disallowing students to use certain functions, a teacher may choose to impose further restrictions on students' answers. Primary among these is checking that a student's answer is of the correct type: that is to say, is it the right kind of mathematical object? Two other validity checks relate to numbers in answers. A teacher may choose to forbid floating-point numbers in an answer, requiring students to use only fractions; further they may require that all fractions be in lowest terms.

When a student submits their answer to STACK, the required validity tests are run on their input and if their answer passes all the tests, it is presented to them as typeset mathematics with the message "Your answer was interpreted as:". This allows the student to check that they have entered what they intended to—as answers are submitted in Maxima syntax but presented back to students as typeset mathematics—and that any ambiguities in their answer have been correctly interpreted. A student may alter their answer and resubmit it for validation as many times as they like, without losing marks for an invalid answer.

8.2.2 Correctness

Once a student's answer has passed the relevant validity tests and they are happy with it, they submit it again for marking. In Section 9.1 we discuss the path that their answer takes through the system, but for now we assume that the student and teacher each have a single answer. This comparison is done using an *answer test*.

An answer test is a predicate function that returns `true` if the student's and teacher's answers are 'the same' to some suitable degree, and `false` if they are not. The prototype test is one of mathematical equivalence, taking the two answers `SA` (from the student) and `TA` (from the teacher) and determining whether

$$\texttt{simplify(SA - TA)} = 0.$$

This test only works in the case where the answers submitted are mathematical expressions without equals signs, and despite its simplicity this test has its pitfalls. Even in cases where the expressions used are elementary functions of a single real variable, establishing equivalence with zero is formally undecidable (i.e., equivalent to the halting problem) (Richardson, 1968). Formal undecidability notwithstanding, for any answer that a student is ever likely to submit, the equivalence or otherwise of $\texttt{SA} - \texttt{TA}$ to zero is established.

Besides either `true` or `false`, answer tests return two further pieces of information. The first is *feedback*, which can be displayed to the student at the teacher's discretion. The second is a *note*, which records the properties of the student's answer identified by the system. The note is stored in the system should the teacher wish to identify trends in the answering of questions. So, besides deciding whether a student's answer is correct or not, it is also preferable if an answer test is able to return worthwhile information on an answer that is incorrect, much as a teacher may when marking work.

8.3 Equivalence of Equations

An equation is an expression whose operator is equality; a statement, saying *these two things are equal to one another.* Some equations, such as $3 + 4 = 7$ and $11 + 12 = 25$, can be said to be true or false. We cannot say whether or not the equation $x + 4 = 11$ is true, however, because we do not know the value of x. Solving an equation involves the symbolic manipulation of its elements, subject to the algebraic rules within which one is working, to determine the values of the variables in the equation for which the equation holds true. The set of all such values is called the equation's *variety.*

As we said in the previous section, our prototype test is insufficient for determining the equivalence of two equations, and so we cannot use it to assess an answer that is an equation. (For the sake of simplicity, we specify that contained operands are not equations themselves, so $3 \times (9 = x)$ is not a valid expression.) If we have student's and teacher's equations $s_l = s_r$ and $t_l = t_r$, we could apply the prototype test to s_l, t_l and s_r, t_r independently, but this would fail correctly to identify the equvialence of $x = 3y$ to $x/3 = y$.

To compare two equations, STACK first gathers the terms on one side: $s_l - s_r = 0$ and $t_l - t_r = 0$. This addresses cases where $s_l = t_l + c$ and $s_r = t_r + c$.

Now we want to know when two expressions, or functions, are equivalent. If $s_l - s_r = s(X)$ and $t_l - t_r = t(X)$, where $X = \{x_1, \dots, x_k\}$, we certainly require them to have the same varieties, but testing for this property is not enough: were we only to look at the varieties of functions then x^n would be counted as the same function for all $n \in \mathbb{N}\backslash 0$. Instead we need to check whether $ks(X) = t(X)$ for some $k \in \mathbb{F}\backslash 0$, where $\mathbb{F}$ is the field from which the coefficients of $s(X)$ and $t(X)$ are taken. Clearly, $s(X)$ and $ks(X)$ have the same solutions, and their behaviour is algebraically equivalent; $ks(X)$ defines an equivalence class of functions over $\mathbb{F}[X]$ for all $k \in \mathbb{F}\backslash 0$.

To check whether $ks(X) = t(X)$ for some $k \in \mathbb{F}\backslash 0$, STACK divides $s(X)$ by $t(X)$ and checks whether the answer is a non-zero constant. Taking all this together, suppose that $s_l = kt_l + c$ and $s_r = kt_r + c$. Then

$$\frac{s_l - s_r}{t_l - t_r} = \frac{kt_l + c - kt_r + c}{t_l - t_r} = \frac{k(t_l - t_r)}{t_l - t_r} = k. \tag{8.1}$$

If the result of this division is not an element of the field $\mathbb{F}$ in which we are working, then the student's and teacher's equations are defined not to be 'the same'.

For example, were a teacher's answer to be the equation $x + 5 = \frac{y}{3} + 5$ and the student's $y = 3x$, it is clear that these are the 'same' equation, however neither side of the teacher's answer is equal to the student's. The process shown in Equation 8.1 is as follows:

▼ **Example 8.3.1.**

$$\frac{y-3x}{x-\frac{y}{3}} = \frac{3y-9x}{3x-y} = -3.$$

Both simple expressions and single equations are dealt with using STACK's *algebraic equivalence* answer test. Further tests exist for other types of questions or answers. If a student's answer is incorrect, STACK uses the various properties of the student's and teacher's to generate feedback. For instance given an indefinite integral, the integration answer test first determines if the student's answer is algebraically equivalent to the teacher's answer. If it is not, it then checks to see whether the student has missed the constant of integration in their answer.

8.4 Systems of Equations and their Varieties

To implement an answer test in STACK that compares a student's system of equations with a teacher's, we need some method of comparing two systems of equations algorithmically. Formally, a *system of equations* is a set of equations, each of which is asserted to be true. For example,

$$\left.\begin{array}{rcl} 4+x & = & 11, \\ x & = & 7. \end{array}\right\}$$

is a system of equations. For the purposes of the work we are doing, we define the canonical form of a system of equations to be a list of equations whose right hand sides are all zero. Thus our example reduces to the somewhat straightforward

$$\left.\begin{array}{rcl} x-7 & = & 0, \\ x-7 & = & 0. \end{array}\right\}$$

Note that we have also ordered the expressions from largest power of x to smallest; this is important later when we manipulate these expressions with algorithms. Once we begin discussing systems of multivariate polynomials, *term order* becomes highly important, and is discussed in Section 8.6.1. If we wrote our system as a set we would have the *reduction*:

$$\{x+4=11, x=7\} \to \{x-7=0\}.$$

Before this project was begun, STACK did not have the ability to assess systems of equations in any capacity. Two types of equations that would immediately prove useful for assessment are linear equations and polynomials. Linear equations, frequently referred to as simultaneous equations, are first taught to students at Key Stage 3 in the U.K., where they are taught to solve these equations via elimination (Qualifications and Curriculum Authority, 2007a). Key Stage 4 sees the introduction of the polynomials in the form of quadratic equations (Qualifications and Curriculum Authority, 2007b).

Gaussian elimination and the Euclidean Algorithm allow us to solve linear equations and systems of polynomials in one variable with relative ease. As we shall see in Theorem 8.5.14, a system of univariate polynomials is equivalent to a single polynomial. Each of these techniques is useful, but the applications are very limited. For example, this system:

$$\begin{cases} 90 &= vt \\ 90 &= (v+5)(t-\frac{1}{4}) \end{cases} \tag{8.2}$$

is easily solved through basic substitution, though neither algorithmic approach that we have mentioned can find its solutions.

The theory of Gröbner bases was developed by Bruno Buchberger in the 1960s for his PhD Thesis (Buchberger, 1965a,b) to give algorithmic solutions to problems in commutative geometry. The algorithms developed from the theory allow for efficient computation to be performed on polynomials and their ideals.

Using this theory we will expand the range of questions that can be dealt with by the Computer Aided Assessment (CAA) system STACK. Given a model answer and a student's answer, STACK implements a test (chosen by the teacher) to mark the student's answer and give feedback. By implementing Gröbner bases in Maxima, the Computer Algebra System (CAS) used by STACK, we can create a test to tell us when two systems of polynomial equations have the same set of solutions.

The mathematics found in the following sections has become relatively standardised over the last 40 years; the proofs of the theorems here are generally adapted versions from Adams and Loustaunau (1994), however any basic text on Gröbner bases should contain them.

8.5 Gröbner Bases

8.5.1 Basic Theory

We begin the discussion of Gröbner bases by formally defining varieties. We assume for this section a grounding in fields and the associated theory of commutative rings. Our exposition is relatively terse during its initial stages.

When we deal with an equation with just one variable, we write the variety as a set of numbers without explicitly mentioning the variable. With n variables, we write the variety as a set of n-tuples, given an ordering of the variables. For example, the variety of the equation $x^2 + x - 6 = 0$ is $\{-3, 2\}$ while the variety of the equation $x^4 + y^4 - 2(x^2 + y^2) + 2 = 0$ is $\{(1,1), (1,-1), (-1,1), (-1,-1)\}$, where each tuple reads (x-value, y-value).

To define a variety formally, we first define the *affine-n space* of a field:

Definition 8.5.1. *For positive $n \in \mathbb{N}$, the* ***affine*** ***n-space*** *of a field $\mathbb{F}$ is*

$$\mathbb{F}^n = \{(x_1, \dots, x_n) \mid x_i \in \mathbb{F}\}.$$

An expression $f \in \mathbb{F}[x_1, \dots, x_n]$ is a function from the affine n-space of $\mathbb{F}$ to $\mathbb{F}$ itself. When values are assigned to the variables of the expression it is called an ***evaluation****.*

There are two things to note here. Firstly, the functions we are dealing with are *expressions*, not equations. All manipulation that our software will perform on an equation, or system of equations, is preceded by converting an equation to an expression by subtracting one side from the other, for each individual equation. Secondly, all the expressions that we use are indeed functions of their variables, so this language is not ambiguous.

We may now define the *variety of a function.*

Definition 8.5.2. *Given some mathematical expression f with variables $x_1, \dots, x_n$, its* ***variety****, $V(f)$, is the set of solutions to the equation $f = 0$, i.e.*

$$V(f) = \{(a_1, \dots, a_n) \in \mathbb{F}^n \mid f(a_1, \dots, a_n) = 0\}.$$

The choice of field $\mathbb{F}$ is an important one. In our example above, the equation $x^4 + y^4 - 2(x^2 + y^2) + 2 = 0$ has other solutions, such as $(i, -i)$, if we allow values for our variables to have non-zero imaginary parts. While one may solve equations over many fields, both finite and infinite, we shall only consider the field of complex numbers and its subfields.

Given a system of equations $F = \{f_1, \dots, f_t\}$, we define its variety as follows, before relating it to the solutions of its constituent equations:

Definition 8.5.3. *Given the system of equations $F = \{f_1, \dots, f_t\}$ with variables $x_1, \dots, x_n$, the variety $V(F)$ is the set of all solutions of the system of equations*

$$f_1 = 0, f_2 = 0, \dots, f_t = 0.$$

Equivalently,

$$V(F) = \{(a_1, \dots, a_n) \in \mathbb{F}^n \mid f(a_1, \dots, a_n) = 0 \ \forall f \in F\}.$$

Lemma 8.5.4.
$$V(F) = \bigcap_{i=1}^{t} V(f_i).$$

Proof. Take an element $\bar{a} = (a_1, \dots, a_n) \in \mathbb{F}^n$ which is in $V(F)$. As $\bar{a} \in V(F)$ it is a solution of the system of equations

$$f_1 = 0, f_2 = 0, \dots, f_t = 0.$$

Then for all $1 \leq i \leq t$, $f_i(\bar{a}) = 0$, so $\bar{a} \in V(f_i)$. Hence $\bar{a}$ is in

$$\bigcap_{i=1}^{t} V(f_i).$$

Conversely, suppose $\bar{a} = (a_1, \dots, a_n) \in \mathbb{F}^n$ and $f_i(\bar{a}) = 0$ for all $1 \leq i \leq t$. Thus $\bar{a}$ is a solution to the original system of equations, and so $\bar{a} \in V(F)$. □

This is an elementary result, but one which will be of importance later when we start comparing systems of equations. This result allows us to determine whether or not an equation 'fits' in a system simply by looking at its variety and comparing it to that of the system concerned.

Definition 8.5.5. *Let $\mathbb{F}$ be a field. A **polynomial** is a finite sum of **terms** of the form $ax_1^{\beta_1} \dots x_n^{\beta_n}$ where $a \in \mathbb{F}$ and $\beta_i \in \mathbb{N}$[1]. The x_is are **variables**; a polynomial in only one variable is **univariate**, otherwise it is **multivariate**.*

We denote by $\mathbb{F}[x_1, \dots, x_n]$ the set of all polynomials in n variables whose coefficients are in $\mathbb{F}$. This set forms a commutative ring and is an $\mathbb{F}$-vector space with basis $\mathbb{T}^n$, the set of all

[1] We always take $\mathbb{N}$ to be the non-negative integers.

power products:

$$\mathbb{T}^n = \{x_1^{\beta_1} \dots x_n^{\beta_n} \mid \beta_i \in \mathbb{N}\}.$$

Definition 8.5.6. *The **ideal** generated by polynomials $f_1, \dots, f_t$ is written $\langle f_1, \dots, f_t \rangle$ and given by*

$$I = \langle f_1, \dots, f_t \rangle = \left\{ \sum_{i=1}^{t} u_i f_i \mid u_i \in \mathbb{F}[x_1, \dots, x_n] \right\}.$$

It is clear that the ideal as we have defined it is indeed an ideal in the ring theoretic sense and the set $\{f_1, \dots, f_t\}$ is a generating set of the ideal I. Finding a "better" generating set for the ideal I is the aim of the theory of Gröbner bases.

Lemma 8.5.7. *Given a finite system of equations*

$$f_1 = 0, f_2 = 0, \dots, f_t = 0, \tag{8.3}$$

the variety of the ideal generated by the f_i, $V(\langle f_1, \dots, f_t \rangle)$, is equal to the variety of the original system $V(f_1, \dots, f_t)$.

Proof. Let I be the ideal generated by the polynomials f_i. The variety of I, $V(I)$, is given by

$$V(I) = \{(a_1, \dots, a_n) \in \mathbb{F}^n \mid f(a_1, \dots, a_n) = 0, \forall f \in I\}. \tag{8.4}$$

Clearly any $(a_1, \dots, a_n) \in V(I)$ will be a solution to System 8.3, as each of the f_i are in the ideal I. On the other hand, if $(a_1, \dots, a_n) \in \mathbb{F}^n$ is a solution to System 8.4 and we choose a polynomial

$f \in I$,

$$f = \sum_{i=1}^{t} u_i f_i,$$

for some suitable polynomials $u_i \in \mathbb{F}^n$. Each f_i evaluated at $(a_1, \dots, a_n)$ is zero, so the total sum is zero, hence $f(a_1, \dots, a_n) = 0$, as required. $\square$

Let V be a set of points of the affine space $\mathbb{F}^n$ and define the set $I(V)$ to be the polynomials in $\mathbb{F}[x_1, \dots, x_n]$ for which each point in V is a root. Formally, we write

$$I(V) = \{f \in \mathbb{F}[x_1, \dots, x_n] \mid f(a_1, \dots, a_n) = 0 \ \forall \ (a_1, \dots, a_n) \in V\}.$$

The choice of I as notation here was not coincidence; naturally the multiplication of any of its elements by any other polynomial in $\mathbb{F}[x_1, \dots, x_n]$ will yield another polynomial whose variety includes V. Thus $I(V)$ is a ring ideal.

We have two maps,

$$\begin{array}{ccc} \text{Subsets of } \mathbb{F}[x_1, \dots, x_n] & \longrightarrow & \text{Varieties of } \mathbb{F}^n \\ S & \longmapsto & V(S) \end{array}$$

and

$$\begin{array}{ccc} \text{Subsets of } \mathbb{F}^n & \longrightarrow & \text{Ideals of } \mathbb{F}[x_1, \dots, x_n] \\ V & \longmapsto & I(V). \end{array}$$

It is the relationship between these two maps that Gröbner bases allow us to understand and exploit.

8.5.2 Hilbert Basis Theorem

The Hilbert Basis Theorem is crucial to the theory of Gröbner bases, as it guarantees that the algorithms that we will be creating terminate and that every variety is the solution set of a finite set of polynomials. First we must introduce a few more definitions and fix notation.

Definition 8.5.8. *Suppose that we have an ideal $\mathcal{J}$ of $\mathbb{F}[x_1, \ldots, x_n]$ and two polynomials $f, g \in \mathbb{F}[x_1, \ldots, x_n]$. We say that f and g are **congruent modulo** $\mathcal{J}$ if $f - g \in \mathcal{J}$ and write $f \equiv g \pmod{\mathcal{J}}$.*

By virtue of the fact that $\mathcal{J}$ is an ideal, and not simply a set of elements in $\mathbb{F}[x_1, \ldots, x_n]$, this gives us an equivalence relationship on $\mathbb{F}[x_1, \ldots, x_n]$ and we denote the set of equivalence classes by $\mathbb{F}[x_1, \ldots, x_n]/\mathcal{J}$. Elements of this equivalence class take the form $f + \mathcal{J}$ and are called the *cosests* of $\mathcal{J}$. $\mathbb{F}[x_1, \ldots, x_n]/\mathcal{J}$ has some interesting properties: it is a vector space over $\mathbb{F}$ and a commutative ring, called the *quotient ring* of $\mathbb{F}[x_1, \ldots, x_n]$ by $\mathcal{J}$.

Finding a basis for $\mathbb{F}[x_1, \ldots, x_n]/\mathcal{J}$ as a vector space over $\mathbb{F}$ and a set of coset representatives of the quotient ring are two problems that must be solved to find our generating sets. Further we will need to know, given a polynomial f and an ideal I, whether f is in I. This is called the *ideal membership problem.*

Theorem 8.5.9 (Hilbert Basis Theorem)**.** *In the ring $\mathbb{F}[x_1, \ldots, x_n]$ we have the following:*

1. *If I is any ideal of $\mathbb{F}[x_1, \ldots, x_n]$ then there exist polynomials $f_1, \ldots, f_t \in \mathbb{F}[x_1, \ldots, x_n]$ such that $I = \langle f_1, \ldots, f_t \rangle$. (That is to say, I is* finitely generated*).*

2. *For an ascending chain of ideals of* $\mathbb{F}[x_1, \dots, x_n]$, $I_1 \subseteq I_2 \subseteq \dots$, *there exists some* $n \in N$ *such that, for all* $m > n$, $I_n = I_m$. *(That is to say,* $\mathbb{F}[x_1, \dots, x_n]$ *is a* Noetherian Ring*).*

We shall prove Theorem 8.5.9 in the following way— firstly we will see that for any ring, each part of Theorem 8.5.9 implies the other. Secondly, we prove that if R is a Noetherian ring, so too is $R[x]$. Since every field is trivially a Noetherian ring, the result follows. The following two proofs are adapted from those in Adams and Loustaunau (1994, pg. 6).

Theorem 8.5.10. *For a commutative ring* R*, the following are equivalent:*

1. *If* I *is any ideal of* R *then there exist polynomials* $f_1, \dots, f_t \in \mathbb{F}[x_1, \dots, x_n]$ *such that* $I = \langle f_1, \dots, f_t \rangle$.

2. *For an ascending chain of ideals of* R, $I_1 \subseteq I_2 \subseteq \dots$, *there exists some* $n \in N$ *such that, for all* $m > n$, $I_n = I_m$

Proof. (1 ⇒ 2). Let

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots \subseteq I_n \subseteq \dots$$

be an ascending chain of ideals in R. Let I be the ideal $I = \bigcup_{n=1}^{\infty} I_n$, which by condition (1) is finitely generated by some $f_1, \dots, f_t \in R$. As each f_i is in I, there exists some N_i such that $f_i \in I_{N_i}$. Let $N = \max_{1 \le i \le t} N_i$, then each f_i is in I_N, so $I \subseteq I_N$, $I = I_N$ and the result follows.

(2 ⇒ 1). Let us assume that there exists some ideal I of R that is not finitely generated. Let $f_1 \in I$; there exists some $f_2 \in I$ such that $f_2 \notin \langle f_1 \rangle$, and $\langle f_1 \rangle \subsetneq \langle f_1, f_2 \rangle$. Continuing in this way gives us a strictly ascending chain of ideals, contradicting (2). □

Theorem 8.5.11. *If* R *is a Noetherian ring, so too is* $R[x]$.

Partial Proof. Let R be a Noetherian ring and let $\mathcal{J}$ be an ideal of $R[x]$. By Theroem 8.5.10, if we can show that $\mathcal{J}$ is finitely generated then $R[x]$ is Noetherian. We define

$$I_n = \{r \in R \mid r \text{ is the leading coefficient of an n-degree polynomial in } \mathcal{J}\} \cup \{0\},$$

for each $n \geq 0$. Each I_n is an ideal in R and $I_n \subseteq I_{n+1}$. I_n is an ideal of R and $I_n \subseteq I_{n+1}$, as $xf \in \mathcal{J}\ \forall f \in \mathcal{J}$ because $\mathcal{J}$ is an ideal. R is Noetherian, so there exists N such that $I_n = I_N$ for all $n \geq N$. By Theorem 8.5.10 each I_i is finitely generated, $I_i = \langle r_{i1}, \dots, r_{it_i} \rangle$. For $i = 1, \dots, N$ and $j = 1, \dots, t_i$, let f_{ij} be a polynomial in $\mathcal{J}$ of degree i with leading coefficient r_{ij}. Let

$$\mathcal{J}^* = \langle f_{ij} \mid 1 \leq i \leq N, 1 \leq j \leq t_i \rangle.$$

It suffices to show that $\mathcal{J} = \mathcal{J}^*$. Clearly $\mathcal{J}^* \subseteq \mathcal{J}$; the proof of the reverse inclusion involves induction on n and can be found in Adams and Loustaunau (1994, pg. 7). □

Combining Theorem 8.5.11 with simple induction on the variables $\{x_1, \dots, x_n\}$ it is easy to see that if $\mathbb{F}$ is Noetherian, $\mathbb{F}[x_1, \dots, x_n]$ is too. By combining this with Theorem 8.5.10 we see that the Hilbert Basis Theorem holds.

8.5.3 Systems of Linear Equations

The two special cases of systems of polynomials that we have previously mentioned are linear equations, whose equations' variables have powers that are either 0 or 1, and univariate polynomials, whose equations have only one variable. Each of these cases is dealt with by existing theory, namely Gaussian elimination for the linear case, and the Euclidean algorithm for the univariate case. In both cases the theory of Gröbner bases simplifies to the existing theory.

In this section and in Section 8.5.4, we put each of these familiar approaches in terms of the language of Gröbner bases, namely of functions and the ideals that they define. In this way we introduce new notation to familiar topics before we cover the generalised theory.

▼ Example 8.5.1.

Let $f_1 = x + 2y - z$ and $f_2 = x + y + z$ be two linear polynomials in $\mathbb{R}[x, y, z]$. The ideal that they form is $I = \langle f_1, f_2 \rangle$ and their variety $V(f_1, f_2)$ is the set of solutions to the system

$$\begin{array}{rcrcrcl} x & - & y & - & z & = & 0, \\ 2x & - & y & + & z & = & 0. \end{array}$$

Putting this in the familiar matrix form and performing Gaussian elimination is done as follows

$$\begin{bmatrix} 1 & -1 & -1 \\ 2 & -1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 3 \end{bmatrix}$$

From the final matrix we solve the system to give $x = -2z$ and $y = -3z$. This process of Gaussian elimination is essentially a method for finding a simpler generating set for the ideal $I = \langle f_1, f_2 \rangle$. We have effectively created a new polynomial $f_3 = f_2 - 2f_1 = y + 3z$ and replaced f_2 with it. Clearly $f_3 \in I$, and as $f_2 = f_3 + 2f_1$, $f_2 \in \langle f_1, f_3 \rangle$ and $\langle f_1, f_3 \rangle = \langle f_1, f_2 \rangle = I$. By simplifying the generating set in this way we can find $V(I)$ more easily.

The process of replacing f_2 by f_3 in Example 8.5.1 is called the *reduction* of f_2 by f_1, and is written

$$f_2 \xrightarrow{f_1} f_3.$$

Repeated reduction is written

$$f_2 \xrightarrow{f_1, f_3}_{+} f_4.$$

The + symbol indicates that a polynomial may be reduced by each f_i more than once. In the above example we can view f_3 to be the remainder of the polynomial division of f_2 by f_1, as follows

$$\begin{array}{r|l} & \phantom{2x - {}} \hfill 2 \\ \cline{2-2} x - y - z & 2x - y - z \\ & 2x - 2y - 2z \\ \cline{2-2} & - y + 3z \end{array}$$

giving $f_2 = 2f_1 + f_3$.

Example 8.5.1 demonstrates a number of aspects that will be important in solving general multivariate polynomials. Firstly, variables were removed in the order x, y, z. Had we removed y before z we would have ended up with a different system of equations, though one that had the same variety as our final system. Secondly, it was the leading terms of the equations f_1 and f_2 that were used to create f_3. Finally, we have a straightforward description of the solution space of Example 8.5.1:

$$V(I) = V(f_1, f_2) = V(f_1, f_3) = \{\lambda(-2, -3, 1) \mid \lambda \in \mathbb{R}\}.$$

8.5.4 Systems of Univariate Polynomials

When faced with a system of univariate polynomials, we can find its solution set by using the Euclidean Algorithm. In this section we describe the Euclidean Algorithm in the terms of Gröbner bases. We begin by fixing some notation to aid our discussion of the process.

Definition 8.5.12. *If $f \in \mathbb{F}[x]$ is a non-zero polynomial, its **degree**,* $\deg(f)$, *is the largest* *exponent of x that appears in f. The* ***leading term*** *of f, $\mathrm{lt}(f)$, is the term of f with highest degree and the* ***leading coefficient****, $\mathrm{lc}(f)$, is the coefficient of the leading term. The* ***leading power product*** *of f, $\mathrm{lp}(f)$, is $\frac{\mathrm{lt}(f)}{\mathrm{lc}(f)}$.*

For example, given the polynomial $f = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$, $\deg(f) = n$, $\mathrm{lt}(f) = a_n x^n$, $\mathrm{lc}(f) = a_n$ and $\mathrm{lp}(f) = x^n$.

In the following example we demonstrate the central tool of the Euclidean Algorithm, namely polynomial division, and put it in terms of the language and notation of Gröbner bases.

▼ Example 8.5.2.

Given two univariate polynomials, $f = 6x^3 + 4x^2 + 2x + 1$ and $g = 2x^2 + 2x$, we divide f by g as follows:

$$\begin{array}{r|l}
 & \phantom{6x^3 + 4x^2 +{}} 3x - 1 \\ \cline{2-2}
2x^2 + 2x & 6x^3 + 4x^2 + 2x + 1 \\
 & 6x^3 + 6x^2 \\ \cline{2-2}
 & - 2x^2 + 2x + 1 \\
 & - 2x^2 - 2x \\ \cline{2-2}
 & \phantom{6x^3 - 2x^2 +{}} 4x + 1
\end{array}$$

We see that the quotient of f by g is $3x - 1$ and the remainder is $4x + 1$. Putting this another way, $f = (3x - 1)g + (4x + 1)$.

To begin the calculation we multiplied g by $3x$ and subtracted the result from f to give the polynomial $h = -2x^2 + 2x + 1$. $3x$ was chosen because multiplying the leading term of g by it gives the leading term of f, allowing us to reduce the degree of the resulting polynomial h.

In general, using the notation that we have introduced, for two polynomials $f = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$ and $g = b_m x^m + b_{m-1}x^{m-1} + \ldots + b_1 x + b_0$ with $n \geq m$, we get our new polynomial by subtracting $\frac{\mathrm{lt}(f)}{\mathrm{lt}(g)} g$ from f. In a similar way to the linear case, we call h the *reduction* of f by g and denote it as

$$f \xrightarrow{g} h.$$

Repeated reduction of f leads to a polynomial whose degree is less than that of g. This polynomial is r, the remainder of f divided by g. We have Euclid's Theorem:

Theorem 8.5.13. *Let g be a non-zero polynomial in $\mathbb{F}[x]$. For any $f \in \mathbb{F}[x]$ there exist unique polynomials q and r in $\mathbb{F}[x]$ such that*

$$f = qg + r, \quad \deg(r) < \deg(g).$$

The Theory of Gröbner bases is dedicated to finding *algorithmic* solutions to finding varieties, and so we present our first algorithm, for computing the two polynomials q and r from Theorem 8.5.13. The notation in our algorithms is relatively standard; **Loop** shows where a **While** or **For** loop repeats.

Algorithm 8.5.1 — *Division Algorithm*

Input: $f, g \in \mathbb{F}[x]$ **with** $g \neq 0$**.**

Output: $q, r \in \mathbb{F}[x]$ **such that** $f = qg + r$ **and** $\deg(r) < \deg(g)$**.**

Initialisation: $q := 0$**,** $r := f$**.**

While $r \neq 0$ **And** $\deg(g) \leq \deg(r)$ **Do**

$$q := q + \frac{\mathrm{lt}(r)}{\mathrm{lt}(g)}$$
$$r := r - \frac{\mathrm{lt}(r)}{\mathrm{lt}(g)}g$$

Loop

▼ **Example 8.5.3.**

Applying Algorithm 8.5.1 to Example 8.5.2 works as follows:

Initialisation: $q = 0$, $r = 6x^3 + 4x^2 + 2x + 1$

First Pass

$$q := 0 + \frac{6x^3}{2x^2} = 3x$$
$$r := (6x^3 + 4x^2 + 2x + 1) - \frac{6x^3}{2x^2}(2x^2 + 2x) = -2x^2 + 2x + 1$$

Second Pass

$$q := 3x + \frac{-2x^2}{2x^2} = 3x - 1$$
$$r := (-2x^2 + 2x + 1) - \frac{-2x^2}{2x^2}(2x^2 + 2x) = 4x + 1$$

Here the ***while*** *loop ends as* $\deg r < \deg g$, *leaving the same* q *and* r *as in Example 8.5.2.*

As with linear systems, given a polynomial reduction $f \xrightarrow{g} h$, the ideals $I = \langle f, g\rangle$ and $\langle h, g\rangle$ are the same. It follows that we can replace f by h in the generating set of I. From this fact we prove Theorem 8.5.14:

Theorem 8.5.14. *Every ideal of* $\mathbb{F}[x]$ *is generated by one element.*

Proof. Let I be a non-zero ideal of $\mathbb{F}[x]$ and g be a non-zero polynomial of least degree in I. By Theorem 8.5.13 we know that, for any $f \in \mathbb{F}[x]$, there exist polynomials q and r such that $f = qg + r$, where $\deg(r) < \deg(g)$ or $r = 0$. If r is not zero, then the fact that $\deg(r) < \deg(g)$ contradicts the minimality of the degree of g (because I is an ideal so $r = f - qg \in I$). Hence $r = 0$ and $f = qg$, so $I \subseteq \langle g \rangle$. g is in I so we have equality. □

One result of Theorem 8.5.14 is that any system of univariate polynomials is equivalent to a system defined by a single polynomial. In the theorem, the polynomial g is unique up to multiplication by a constant, as if $I = \langle g_1 \rangle = \langle g_2 \rangle$, g_1 divides g_2 and vice-versa. It follows that specifying that g is monic makes it unique. The rest of this section is dedicated to finding the polynomial g that generates the same ideal as the general system of polynomials $f_1, \dots, f_t \in \mathbb{F}[x]$.

Definition 8.5.15. *Given two polynomials f_1 and f_2 in $\mathbb{F}[x]$, their* ***greatest common divisor****, g, is the polynomial such that*

1. *g divides both f_1 and f_2.*

2. *If $h \in F[x]$ divides f_1 and f_2, then h divides g.*

3. *g is monic (i.e.,* $\mathrm{lc}(g) = 1$*).*

We denote the greatest common divisor of f_1 and f_2 by $\gcd(f_1, f_2)$.

Proposition 8.5.16. *Let $f_1, f_2 \in \mathbb{F}[x]$, with at least one of f_1 and f_2 not zero. Then* $\gcd(f_1, f_2)$ *exists and* $\langle f_1, f_2 \rangle = \langle \gcd(f_1, f_2) \rangle$.

Proof. By Theorem 8.5.14 there exists monic g such that $\langle f_1, f_2\rangle = \langle g\rangle$. As $f_1, f_2 \in \langle g\rangle$, g divides both f_1 and f_2. Let h be a polynomial which divides both f_1 and f_2. Since g is in the ideal $\langle f_1, f_2\rangle$, there exist l_1 and $l_2 \in \mathbb{F}[x]$ such that $g = l_1 f_1 + l_2 f_2$. Thus h divides g, and so $g = \gcd(f_1, f_2)$. □

Lemma 8.5.17. *Let* $0 \neq f_1, f_2 \in \mathbb{F}[x]$. *Then* $\gcd(f_1, f_2) = \gcd(f_1 - qf_2, f_2)$ *for all* $q \in \mathbb{F}[x]$.

Proof. Clearly, $\langle f_1, f_2\rangle = \langle f_1 - qf_2, f_2\rangle$. By Proposition 8.5.16,

$$\langle \gcd(f_1, f_2)\rangle = \langle f_1, f_2\rangle = \langle f_1 - qf_2, f_2\rangle = \langle \gcd(f_1 - qf_2, f_2)\rangle.$$

The generator of a principal ideal domain is unique up to constant multiples, and as the greatest common divisor of two polynomials is monic, it follows that

$$\gcd(f_1, f_2) = \gcd(f_1 - qf_2, f_2).$$

□

These facts are employed by the Euclidean Algorithm, Algorithm 8.5.2, which we explain below.

Algorithm 8.5.2 — *Euclidean Algorithm*

Input: $f_1, f_2 \in \mathbb{F}[x]$ **with** $f_1 \neq 0$.

Output: $f = \gcd(f_1, f_2)$.

Initialisation: $f := f_1$, $g := f_2$.

While $g \neq 0$ **Do**

$f \xrightarrow{g}_{+} r$**, where** r **is the remainder of the division of** f **by** g

$f := g$

$g := r$

Loop

$f := \frac{1}{\mathrm{lc}(f)} f$**.**

Algorithm 8.5.2 terminates because the degree of r in the **While** loop is strictly decreasing. By Lemma 8.5.17, when $g \neq 0$, the **While** loop gives $\gcd(f_1, f_2) = \gcd(f, g) = \gcd(r, g)$. At the final step, $g = 0$, so $\gcd(f_1, f_2) = \gcd(f, 0) = \frac{1}{\mathrm{lc}(f)} f$. Hence Algorithm 8.5.2 does indeed find the greatest common divisor of the polynomials f_1 and f_2.

In Example 8.5.4 we apply Algorithm 8.5.2 to $f_1 = x^3 - 3x^2 - x + 3 = (x-3)(x-1)(x+1)$ and $f_2 = x^2 - x - 2 = (x-2)(x+1)$.

▼ **Example 8.5.4.**

Initialisation: $f := x^3 - 3x^2 - x + 3$, $g := x^2 - x - 2$

First Pass

$$\begin{aligned}
&x^3 - 3x^2 - x + 3 \xrightarrow{x^2-x-2} -x - 1\\
&f := x^2 - x - 2\\
&g := -x - 1
\end{aligned}$$

Second Pass

$$\begin{aligned}
&x^2 - x - 2 \xrightarrow{-x-1} -2x - 2 \xrightarrow{-x-1} 0\\
&f := -x - 1\\
&g := 0
\end{aligned}$$

$f := \frac{1}{\mathrm{lc}(f)} f = \frac{1}{-1} f = x + 1$, *hence* $\gcd(f_1, f_2) = x + 1$.

Expanding the two polynomial case up to a system of equations $f_1, \dots, f_t$ is straightforward. It is clear that $\gcd(f_1, \dots, f_t) = \gcd(f_1, (f_2 \dots, f_t))$ and so finding the greatest common divisor of a system of univariate polynomials is a simple matter of induction. We now modify Proposition 8.5.16 to show that this greatest common divisor generates the same ideal as the system $f_1, \dots, f_t$.

Proposition 8.5.18. *Let* $0 \neq f_1, \dots, f_t \in \mathbb{F}[x]$. *The* $\gcd(f_1, \dots, f_t)$ *exists and* $\langle f_1, \dots, f_t \rangle = \langle \gcd(f_1, \dots, f_t) \rangle$.

Proof. By Theorem 8.5.14 we know that there exists some monic $g \in \mathbb{F}[x]$ such that $\langle f_1, \dots, f_t \rangle = \langle g \rangle$. Each $f_i \in \langle g \rangle$, $1 \leq i \leq t$ and so g divides each f_i. Suppose that there exists some $h \in \mathbb{F}[x]$ which divides each f_i. As $g \in \langle f_1, \dots, f_t \rangle$, there exist $l_1, \dots, l_t \in \mathbb{F}[x]$ such that

$$g = l_1 f_1 + \dots + l_t f_t.$$

Hence $h|g$, as required. $\square$

8.6 Systems of Multivariate Polynomials

In the previous two sections we introduced the bulk of the notation and principles that we will employ in finding generating sets for the ideals of systems of multivariate polynomials. Two important aspects of the multivariate case are the order of terms and the division algorithm. In Section 8.5.3 we used Gaussian elimination as our division algorithm, while in Section 8.5.4 polynomial division was used. We highlighted the fact that term orders were important in linear equations, as they are too when dealing with univariate polynomials. When dealing with *multivariate* polynomials term orders are equally important, but require a firmer grounding than we have previously given them. One may naturally choose that x is greater than y in a linear equation, or that x^5 is greater than x^3 in a univariate polynomial, but it is not clear how we should deal with terms such as x^2y and xy^2. The division algorithms that we use always need some term order to guarantee that they work consistently with the same polynomials.

8.6.1 Term Orders

Both Algorithm 8.5.1 and 8.5.2 relied on polynomials being written in order from largest power of x to smallest, to guarantee that the degree of the polynomials got smaller on every loop, and the algorithms eventually terminated. Before we look at a division algorithm for multivariate polynomials, we begin by considering term orders, the main aim of which is to define a well order $<$ on the set of power products

$$\mathbb{T}^n = \{x_1^{\beta_1} \dots x_n^{\beta_n} \mid \beta_i \in \mathbb{N}\}.$$

We write an element $x_1^{\beta_1} \dots x_n^{\beta_n}$ as $\boldsymbol{x}^{\boldsymbol{\beta}}$. A total order is required so that we know in what order to eliminate the terms of our polynomials, and such orders must be well orderings so that there exists no infinite descending chains $\boldsymbol{x}^{\boldsymbol{\beta}_1} > \boldsymbol{x}^{\boldsymbol{\beta}_2} > \dots$ which would prevent our algorithms from terminating.

Definition 8.6.1. *A* ***term order*** *on* $\mathbb{T}^n$ *is a total order* $<$ *on* $\mathbb{T}_n$ *which satisfies the following:*

1. $1 < \boldsymbol{x}^{\boldsymbol{\beta}}$ *for all* $\boldsymbol{x}^{\boldsymbol{\beta}} \in \mathbb{T}^n \backslash 1$*;*

2. *If* $\boldsymbol{x}^{\boldsymbol{\alpha}} < \boldsymbol{x}^{\boldsymbol{\beta}}$*, then* $\boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{x}^{\boldsymbol{\gamma}} < \boldsymbol{x}^{\boldsymbol{\beta}}\boldsymbol{x}^{\boldsymbol{\gamma}}$ *for all* $\boldsymbol{x}^{\boldsymbol{\gamma}} \in \mathbb{T}^n$*.*

Clearly, to define a term order we need first to order the variables of $\mathbb{F}[x_1, \dots, x_n]$. The canonical form of this ordering is that the first variable is defined to be the 'largest', so we have $x_1 > x_2 > \dots > x_n$. With three or fewer variables the order is $x < y < z$. A proof that a term order as we have defined it is a well-order can be found in (Adams and Loustaunau, 1994, p. 21).

Different term orders serve different purposes and are useful for different calculations in Gröbner bases. Our next definition gives the particular type of order that we are after; we shall justify its suitability when we have given a few examples.

Definition 8.6.2. *The* ***degree reverse lexicographic order*** *on* $\mathbb{T}^n$ *with* $x_1 > x_2 > \ldots > x_n$ *is*

$$\boldsymbol{x}^{\boldsymbol{\alpha}} < \boldsymbol{x}^{\boldsymbol{\beta}} \Leftrightarrow \begin{cases} \displaystyle\sum_{i=1}^{n} \alpha_i < \sum_{i=1}^{n} \beta_i, \\ \\ \textit{If the sums are equal, working from the right, the first exponents} \\ \alpha_i \textit{ and } \beta_i \textit{ which differ, satisfy } \alpha_i > \beta_i. \end{cases}$$

Degree reverse lexicographic order is so titled because it first considers the total degrees of the terms, and then looks 'backwards' at the terms. The *lexicographic order* considers only the exponents of the terms' variables, in ascending order. With two variables degree reverse lexicographic order looks like:

$$1 < x_2 < x_1 < x_2^2 < x_2 x_1 < x_1^2 < x_2^3 < \ldots .$$

This term order may appear to be somewhat contrived, but allows for the efficient calculation of Gröbner bases. One important aspect of all term orders is that if one term divides another, it appears in a term order before the term it divides:

Proposition 8.6.3. *Let* $\boldsymbol{x}^{\boldsymbol{\alpha}}, \boldsymbol{x}^{\boldsymbol{\beta}} \in \mathbb{T}^n$. *If* $\boldsymbol{x}^{\boldsymbol{\alpha}}$ *divides* $\boldsymbol{x}^{\boldsymbol{\beta}}$ *then* $\boldsymbol{x}^{\boldsymbol{\alpha}} \leq \boldsymbol{x}^{\boldsymbol{\beta}}$.

Proof. As $\boldsymbol{x}^{\boldsymbol{\alpha}}$ divides $\boldsymbol{x}^{\boldsymbol{\beta}}$ there exists some $\boldsymbol{x}^{\boldsymbol{\gamma}}$ such that $\boldsymbol{x}^{\boldsymbol{\beta}} = \boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{x}^{\boldsymbol{\gamma}}$ with $\boldsymbol{x}^{\boldsymbol{\gamma}} \geq 1$. Now since $1 \leq \boldsymbol{x}^{\boldsymbol{\gamma}}$, $\boldsymbol{x}^{\boldsymbol{\alpha}} \leq \boldsymbol{x}^{\boldsymbol{\alpha}}\boldsymbol{x}^{\boldsymbol{\gamma}} = \boldsymbol{x}^{\boldsymbol{\beta}}$, as required. □

With a chosen term order on $\mathbb{F}[x_1, \ldots, x_n]$ we can now correctly order the terms of $f \neq 0 \in$

$\mathbb{F}[x_1, \dots, x_n]$,

$$f = a_1\boldsymbol{x}^{\alpha_1} + a_2\boldsymbol{x}^{\alpha_2} + \dots + a_r\boldsymbol{x}^{\alpha_r},$$

where each a_i is non-zero, $\boldsymbol{x}^{\alpha_i} \in \mathbb{T}^n$ and $\boldsymbol{x}^{\alpha_1} > \boldsymbol{x}^{\alpha_2} > \dots > \boldsymbol{x}^{\alpha_r}$. The leading power product, leading coefficient and leading term of f are then suitably redefined as:

- $\mathrm{lp}(f) = \boldsymbol{x}^{\alpha_1}$,
- $\mathrm{lc}(f) = a_1$,
- $\mathrm{lt}(f) = a_1\boldsymbol{x}^{\alpha_1}$.

When $f = 0$ we define $\mathrm{lp}(f) = \mathrm{lc}(f) = \mathrm{lt}(f) := 0$. Further, each of $\mathrm{lp}(f)$, $\mathrm{lc}(g)$ and $\mathrm{lt}(f)$ are homomorphisms from $\mathbb{F}[x_1, \dots, x_n]$ to $\mathbb{F}[x_1, \dots, x_n]$ (i.e., $\mathrm{lp}(fg) = \mathrm{lp}(f)\mathrm{lp}(g)$).

8.6.2 A Division Algorithm

Once a term order has been settled on, it is possible to create a division algorithm for multivariate polynomials by working down the order, as we did with univariate polynomials. In this section we work towards the creation of our division algorithm, beginning with the division of one multivariate polynomial by another.

Definition 8.6.4. *Given polynomials $f, g, h \in \mathbb{F}[x_1, \dots, x_n]$, with $g \neq 0$, we say that f* ***reduces*** *to h modulo g in one step if and only if $\mathrm{lp}(g)$ divides some non-zero term $\boldsymbol{x}^{\alpha}$ of f and*

$$h = f - \frac{\boldsymbol{x}^{\alpha}}{\mathrm{lt}(g)}g.$$

As with the univariate case, we write $f \xrightarrow{g} h$.

▼ Example 8.6.1.

Let $f = 2y^2x - yx^2 + 3yx$ and $g = 2y + x + 1 \in \mathbb{Q}[x, y]$. We use degree lexicographic order with $y > x$. Dividing f by g gives us:

$$
\begin{array}{r|l}
 & \quad\quad\quad\quad\quad yx \quad\quad - x^2 + x \\
\cline{2-2}
2y + x + 1 & 2y^2x - yx^2 + 3yx \\
 & \underline{2y^2x + yx^2 + yx \quad\quad\quad\quad} \\
 & \quad - 2yx^2 + 2yx \\
 & \quad \underline{- 2yx^2 \quad\quad - x^3 - x^2} \\
 & \quad\quad\quad 2yx + x^3 + x^2 \\
 & \quad\quad\quad \underline{2yx \quad\quad + x^2 + x} \\
 & \quad\quad\quad\quad\quad x^3 \quad\quad + x
\end{array}
$$

In reduction notation this division looks as follows:

$$f \xrightarrow{g} -2yx^2 + 2yx \xrightarrow{g} 2yx + x^3 + x^2 \xrightarrow{g} x^3 - x.$$

Definition 8.6.5. *Let $f, f_1, \dots, f_s$ and h be polynomials in $\mathbb{F}[x_1, \dots, x_n]$ with each f_i non-zero. Let $F = \{f_1, \dots, f_s\}$. We say that f **reduces to** h **modulo** F if and only if there exist $h_1, \dots, h_t = h \in \mathbb{F}[x_1, \dots, x_n]$ and indices $i_1, \dots, i_t \in \{1, \dots, s\}$ such that*

$$f \xrightarrow{f_{i_1}} h_1 \xrightarrow{f_{i_2}} h_2 \xrightarrow{f_{i_3}} \dots \xrightarrow{f_{i_{t-1}}} h_{t-1} \xrightarrow{f_{i_t}} h_t = h.$$

When this is the case we write

$$f \xrightarrow{F}_{+} h.$$

*The polynomial h is said to be **reduced** with respect to F.*

Clearly, if $f \xrightarrow{g}_+ h$ and $g \in G$, then $f \xrightarrow{G}_+ h$.

We can now define our division algorithm for multivariate polynomials which, given $f, f_1, \dots f_s \in \mathbb{F}[x_1, \dots, x_n]$ returns quotients $u_1, \dots, u_s$ and remainder $r \in \mathbb{F}[x_1, \dots, x_n]$ such that

$$f = u_1 f_1 + \dots + u_s f_s + r.$$

Algorithm 8.6.1 — *Multivariate Division Algorithm*

Input: $f, f_1, \dots, f_t \in \mathbb{F}[x_1, \dots, x_n]$ **with** $f_i \neq 0$ $(1 \leq i \leq t)$.

Output: $u_1, \dots, u_t, r \in \mathbb{F}[x_1, \dots, x_n]$ **such that** $f = u_1 f_1 + \dots + u_t f_t + r$, r **is reduced with respect to** $\{f_1, \dots, f_t\}$ **and** $\mathrm{lp}(f) = \max(\mathrm{lp}(f_i), \mathrm{lp}(u_i), \mathrm{lp}(r))$.

Initialisation: $u_i := 0$, $r := 0$, $h = f$.

While $h \neq 0$ **Do**

If there exists i **such that** $\mathrm{lp}(f_i)$ **divides** $\mathrm{lp}(h)$ **Then choose the least such** i **and**

$$u_i := u_i + \frac{\mathrm{lt}(h)}{\mathrm{lt}(f_i)}$$
$$h := h - \frac{\mathrm{lt}(h)}{\mathrm{lt}(f_i)} f_i$$

Else

$$r := r + \mathrm{lt}(h)$$
$$h := h - \mathrm{lt}(h)$$

Loop

We now justify that Algorithm 8.6.1 does indeed terminate and finds the u_i and r that we are after.

Theorem 8.6.6. *Given a set of non-zero polynomials $F = \{f_1, \dots, f_s\}$ and f in $\mathbb{F}[x_1, \dots, x_n]$, Algo-*

rithm 8.6.1 produces polynomials $u_1, \ldots, u_t$ and r in $\mathbb{F}[x_1, \ldots, x_n]$ such that

$$f = u_1 f_1 + \ldots + u_t f_t + r,$$

with r reduced with respect to F and

$$\mathrm{lp}(f) = \max(\max_{1 \le i \le t}(\mathrm{lp}(u_i)\mathrm{lp}(f_i)), \mathrm{lp}(r)).$$

Proof. As the algorithm is subtracts leading terms at each iteration, and since the term order is a well-order, it is clear that the algorithm terminates.

We compute h_{i+1} from h_i by subtracting $\frac{\mathrm{lt}(h_i)}{\mathrm{lt}(f_j)} f_j$ from it, and as the algorithm begins with $h = f$, at any stage $\mathrm{lp}(h) \le \mathrm{lp}(f)$. For each i we obtain u_i by adding terms $\frac{\mathrm{lt}(h)}{\mathrm{lt}(f_i)}$, so $\mathrm{lp}(u_i)\mathrm{lp}(f_i) \le \mathrm{lp}(f)$. Since r is obtained by adding the terms $\mathrm{lt}(h)$, $\mathrm{lp}(r) \le \mathrm{lp}(f)$ too. □

8.6.3 Gröbner Bases

Definition 8.6.7. *A set of non-zero polynomials $F = \{f_1, \ldots, f_t\}$ in an ideal I is called a* ***Gröbner basis*** *for I if and only if for all $0 \ne f \in I$ there exists $1 \le i \le t$ such that $\mathrm{lp}(f_i)$ divides $\mathrm{lp}(f)$.*

Clearly every ideal I has a Gröbner basis, as the set of all non-zero polynomials in I satisfies our definition. Further, the definition says that F is a Gröbner basis for I if each $f \in I$ reduces to 0 modulo F. This is the key fact that we use when we wish to compare a student's system which we know to be wrong. We reduce each of the student's equations with respect to the Gröbner basis of the teacher's system; those that have non-zero remainder cannot be members of the ideal and so do not fit the teacher's variety.

Definition 8.6.8. *If F is a subset of $\mathbb{F}[x_1, \dots, x_n]$ we define the* ***leading term ideal*** *of F to be the ideal generated by the leading terms of the elements of F, i.e.*

$$\mathrm{Lt}(F) = \langle \mathrm{lt}(f) \mid f \in F \rangle.$$

With this definition in place we present Theorem 8.6.9 which gives us a set of equivalent statements regarding Gröbner bases.

Theorem 8.6.9. *Let I be a non-zero ideal of $\mathbb{F}[x_1, \dots, x_n]$. For a set of non-zero polynomials $F = \{f_1, \dots, f_t\} \subseteq I$ the following are equivalent:*

1. *G is a Gröbner basis for I.*
2. *$f \in I$ if and only if $f \xrightarrow{G}_+ 0$.*
3. *$f \in I$ if and only if*

$$f = \sum_{i=1}^{t} h_i g_i \text{ with } \mathrm{lp}(f) = \max_{1 \le i \le t}(\mathrm{lp}(h_i)\mathrm{lp}(g_i)), \qquad g_i \in G, h_i \in \mathbb{F}[x_1, \dots, x_n].$$

4. $\mathrm{Lt}(G) = \mathrm{Lt}(I)$.

Proof. $(1 \Rightarrow 2)$. Directly from our definition we see that if $f \in I$ then $f \xrightarrow{F}_+ 0$. On the other had, suppose that $f \xrightarrow{F}_+ 0$. This means that there exist $u_1, \dots, u_s$ such that

$$f = u_1 f_1 + \dots + u_t f_t.$$

As each $f_i \in I$ and I is an ideal of $\mathbb{F}[x_1, \ldots, x_n]$, $f \in I$ as required.

(2 ⇒ 3). As $f \in I$, $f \xrightarrow{F}_+ 0$ and the result follows from Theorem 8.6.6.

(3 ⇒ 4). Clearly $\mathrm{Lt}(F) \subseteq \mathrm{Lt}(I)$. Taking $f \in I$, we have

$$\mathrm{lt}(f) = \sum_i \mathrm{lt}(h_i)\mathrm{lt}(f_i),$$

for all i where $\mathrm{lp}(f) = \mathrm{lp}(h_i)\mathrm{lp}(f_i)$, hence $\mathrm{lt}(f) \in \mathrm{Lt}(F)$.

(4 ⇒ 1). For any $f \in I$, $\mathrm{lt}(f) \in \mathrm{Lt}(F)$, so

$$\mathrm{lt}(f) = \sum_{i=1}^{t} h_i \mathrm{lt}(f_i), \qquad (\dagger)$$

for some $h_i \in \mathbb{F}[x_1, \ldots, x_n]$. Each term in the right hand side of (†) is divisible by some $\mathrm{lp}(f_i)$, and thus so too is $\mathrm{lt}(f)$, as required. □

From Theorem 8.6.9 it follows immediately that the elements of a Gröbner basis of an ideal I generate I.

Now that we know what exactly a Gröbner basis *is*, we turn our attention to an efficient method of finding them, Buchberger's Algorithm, and how we can use them to determine when two systems of equations have the same variety. For now on we concentrate only on the theory that is applicable to our problems; the theory of Gröbner bases is more expansive than the narrow view on offer here.

Definition 8.6.10. *Let* $0 \neq f, g \in \mathbb{F}[x_1, \ldots, x_n]$ *and* $L = \mathrm{lcm}(\mathrm{lp}(f), \mathrm{lp}(g))$. *The* ***S-polynomial*** *of* *f and g is*

$$S(f, g) = \frac{L}{\mathrm{lt}(f)} f - \frac{L}{\mathrm{lt}(g)} g.$$

Theorem 8.6.11 (Buchberger). *Let $F = \{f_1, \dots, f_t\}$ be a set of non-zero polynomials in $\mathbb{F}[x_1, \dots, x_n]$. Then F is a Gröbner basis for the ideal $I = \langle f_1, \dots, f_t \rangle$ if and only if for all $i \neq j$,*

$$S(f_i, f_j) \xrightarrow{F}_{+} 0.$$

The proof of Theorem 8.6.11 requires further definitions and lemmata and can be found in Adams and Loustaunau (1994, pg. 41).

Algorithm 8.6.2 — *Buchberger's Algorithm*

Input: $P = \{p_1, \dots, p_t\} \in \mathbb{F}[x_1, \dots, x_n]$ **with** $p_i \neq 0$ $(1 \leq i \leq t)$**.**

Output: $F = \{f_1, \dots, f_s\}$**, a Gröbner basis for** $\langle P \rangle$

initialisation: $F := P$**,** $\mathcal{F} := \{\{p_i, p_j\} \mid p_i \neq p_j \in F\}$

While $\mathcal{F} \neq \emptyset$ **Do**

Choose any $\{f, g\} \in \mathcal{F}$

$\mathcal{F} := \mathcal{F} - \{\{f, g\}\}$

$S(f, g) \xrightarrow{F}_{+} h$**, where** h **is reduced with respect to** F

If $h \neq 0$ **Then**

$\mathcal{F} := \mathcal{F} \cup \{\{j, h\} \mid$ **for all** $j \in F\}$

$F := F \cup \{h\}$

Loop

The choice of polynomials in $\mathcal{F}$ means that the output of Buchberger's algorithm is not unique for a given ideal. By placing restrictions on the properties of a Gröbner basis we do achieve uniqueness, however.

Definition 8.6.12. *A Gröbner basis $G = \{g_1, \ldots, g_t\}$ is called a* ***reduced Gröbner basis*** *if, for all i, the leading coefficient of g_i is 1 and g_i is reduced with respect to $G - \{g\}$. Equivalently, for all i, $\mathrm{lc}(g_i) = 1$ and $\mathrm{lp}(g_i) \nmid \mathrm{lp}(g_j)\ \forall i \neq j$.*

Buchberger showed that, for a given term order, an ideal I has a unique reduced Gröbner basis. Further, the map $\phi : I \longrightarrow G$ which takes an ideal to its reduced Gröbner basis is injective; two ideals are equal exactly when their reduced Gröbner bases are equal. It is this fact that is crucial in determining when two sets of equations have the same variety.

8.6.4 Comparing Systems of Equations

We are now in a position to answer the three questions that interest us:

1. Do two given systems of equations have the same solutions?
2. Given polynomial f, is its variety a superset of a given system's variety?
3. Given polynomial f, is it a member of the ideal generated by the system?

Given two systems of equations, $F = \{f_1, \ldots, f_s\}$ and $G = \{g_1, \ldots, g_t\}$, the first question is answered computationally as follows: from Lemma 8.5.7, we know that $V(F) = V(\langle F \rangle)$. Further we know that, as reduced Gröbner bases are unique to their ideal, $\langle F\rangle = \langle G\rangle$ if and only if F and G have the same reduced Gröbner basis. Thus, if two systems of multivariate polynomial equations generate the same reduced Gröbner basis then their varieties are the same.

If the Gröbner basis of the teacher's system contains that of the students's system, but is not equal to it, we know immediately that the student is missing equations: the Gröbner basis of the student's system will generate an ideal that contains that of the teacher's system. Thus its variety will be larger: as each individual equation must be satisfied, adding equations to the system *decreases* the size of its variety (save for equations already in its ideal).

Finally, the *Ideal Membership Problem*, solved by reducing a given polynomial expression modulo the Gröbner basis of the teacher's system, allows us to check to see whether individual polynomials fit the requirements of a task.

Using Gröbner bases in this manner yields a further assistance to teachers; though one that is unlikely ever to arise: if a student models a situation incorrectly, and yet somehow creates a system with the correct variety, the system will reject it because its Gröbner basis will not be the same as that of the teacher's system. Thus, modelling a situation incorrectly but finding the correct variety does not result in a correct answer; something that would be missed were we to consider only student's solutions to systems. As a trivial example, were a teacher's model equivalent to the equation $(x-3)^2(x+3)^2$ and the student's model equivalent to $(x-3)(x+3)$, their systems would certainly have the same solutions, however the student's model is missing a lot of information inherent in the teacher's system besides its variety.

8.6.5 Limitations of the System

When modelling a situation using a system of equations, there are usually a number of choices that one must make. In almost all cases, for example, choices of variables' letters will have to be made; in others, an appropriate co-ordinate system will have to be chosen. The theory of Gröbner Bases allows us to create a system to assess questions given to students that result in systems of multivariate polynomials, so naturally any situation that cannot be expressed as such cannot be assessed using this method. Further, if a student is presented with a situation that may be expressed using a variety of co-ordinate systems, this method will be unable to determine between them. However, a teacher can get around this problem by prescribing which system to use.

More problematic is the choice of a point of origin. When a point of origin needs to be chosen, for example when calculating moments, correct systems with different Gröbner bases may be created. This is not something which is easy to get around, as creating possible answers for each potential point of origin is likely impracticable.

It is therefore up to the teacher to ensure that there are enough *implicit* and *explicit* indications of what choices to make where the system is unable to cope. Implicit indicators may be creating a question which has an obvious choice of co-ordinate system, so a student effectively has that 'choice' removed from them. Explicit indicators are where students are told what choice to make: "Use x for the length of the pipe".

The final possible limitation of the system is the speed at which calculations of Gröbner bases can be made. A number of advances in the speed of calculation have been made over the last two decades, notably in the work of Faugère (1999, 2002) who developed algorithms orders of magnitude faster than Buchberger's originals. The examples we give in the next chapter take mere milliseconds to generate responses, though were a teacher to use five or more variables with a large system of equations there is a potential for slowdown that makes the system unsuitable for assessment. For the purposes for which our system is intended, speed is not an issue.

8.7 Conclusion

The theory of Gröbner Bases is a robust and efficient method for calculating whether to systems of multivariate polynomial equations are the same. It allows us to determine further properties of the relationship between two such systems, in a manner that enables us to create a system for assessment that gives worthwhile feedback to students. By creating a system of assessment using Gröbner Bases we open up a range of questions that would be otherwise inassessible, or assessable only in a highly limited fashion by comparing solutions.

The following section covers the creation of an *answer test* for the computer-aided assessment software STACK, which implements the theory covered in this chapter. We also demonstrate the assessment of questions using the system.

Chapter 9

Applications of Gröbner Bases

In this chapter we demonstrate the application of the theory discussed in Chapter 8 to the CAA software STACK. We explain, with reference to the Maxima code written for this project, the creation of a new *answer test* that allows for the authoring of questions whose answers are systems of multivariate polynomials.

All the code in this project was written by the author, and reviewed by Christopher Sangwin. We discuss only the code of the answer test itself, and not the other changes implemented in the system.

We provide an example question that has been implemented using the new system to demonstrate the range of responses that the answer test is able to give. We then give futher examples of questions that can be assessed using the new answer test.

9.1 Potential Response Trees

The structure of a 'question' in STACK is rather more than a simple one question, one answer model. Each question may include a number of part-questions, where a part-question constitutes an answer that a student is required to give. These part-questions are included on the same page, and may be related to one-another mathematically, but from the point of view of the system they are independent. From here, when we refer to a 'question', we mean one such part-question.

Each question requires the student to enter an answer, often in a text box. Once validated and resubmitted, this answer is then processed by STACK via a *potential response tree*. A P.R.T. tells STACK which answer tests to apply to the answer, once its validity has been confirmed. Each node of the P.R.T. consists of the student's expression `SA`, the teacher's expression `TA` and the required answer test. The outcome of the answer test when applied to `SA` and `TA` determines whether `SA` is sent to another node of the answer test, or if the answer test should halt. Each of these options includes the possibility of further feedback or a note chosen on a per-node basis.

Figure 9.1 shows a node of a P.R.T. as it appears to someone authoring a question. The student's answer is a variable, *veqn*, while the teacher's answer has been entered directly (it too could be a variable or Maxima function). The answer test being applied is algebraic equivalence (AlgEquiv). For both `true` and `false` responses the node is terminal, as shown by *Next PR: -1*, a pseudonode that exits the tree. There are options for further feedback and notes to be given, and a 'Quiet' option which suppresses answer test feedback from being presented to the student. Notes and feedback from multiple tree nodes are concatenated by STACK.

The strength of the tree-based approach is as follows. Firstly, a teacher can look for application of particular buggy-/mal-rules (Young and O'Shea, 1981), by creating one answer test to compare `SA` with an *incorrect* `TA`, and provide specific feedback if the answer test returns

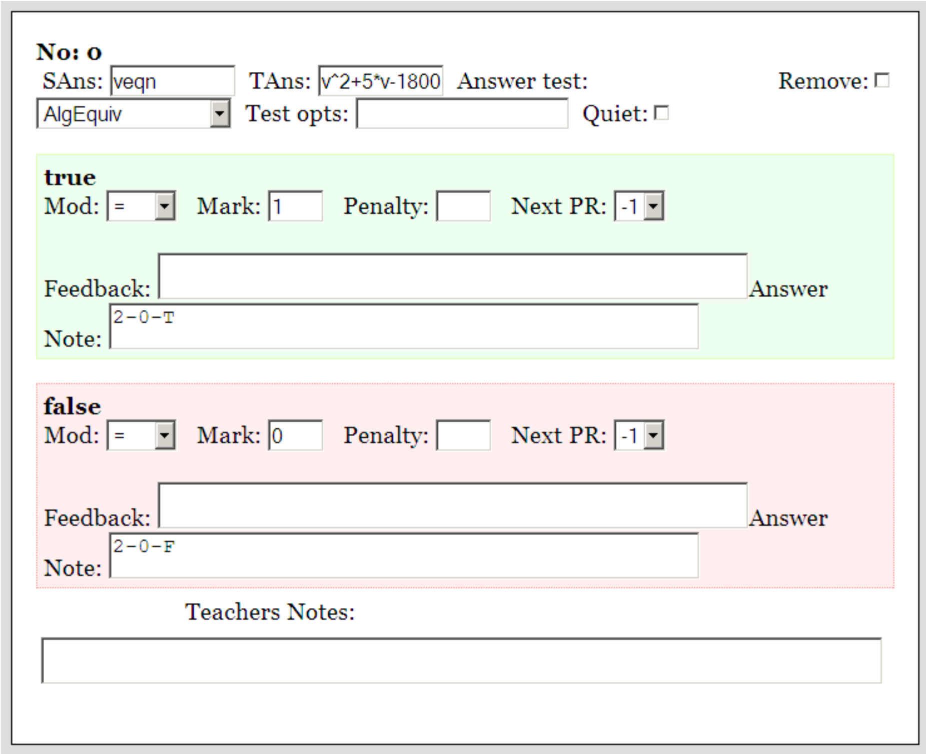


Figure 9.1: Potential Response Tree node

`true`. If the answer test returns `false`, either the student's answer is correct or they have made some other error in their answer. The answer test returning `false` will lead to another node of the P.R.T. which compares `SA` with the correct `TA`. The following section includes an example of this technique.

Using a potential response tree enables a teacher to establish separate properties of an answer, by comparing a student's answer with different tests, or to compare the student's answer to a range of possible correct answers.

9.2 Using Gröbner Bases to assess questions in STACK

The goal of this work is to implement an answer test in STACK that assesses systems of equations and provides useful feedback to students. Given a question that results in a system of multivariate polynomials, we want to answer two questions:

1. Does the student's answer have the same variety as the teacher's? That is to say, has the student correctly described the situation?

2. If the student's answer is not correct, what useful information can we determine about their answer to generate feedback?

The answer to our first question is, in principle, straightforward. Given a student's and a teacher's systems of equations, we compute the reduced Gröbner bases of the ideals that each generates, and if the reduced Gröbner bases are the same, then so too are the ideals and hence their varieties.

The answer to the second question is more involved. At the heart of Buchberger's Algorithm lies the Multivariate Division Algorithm, which has the same relationship to Buchberger's Algorithm as does polynomial division to Euclid's Algorithm. From Theorem 8.6.9 we know that an expression f is an element of an ideal I if an only if f can be reduced to zero by the elements of I's Gröbner basis using the Multivariate Division Algorithm.

This is the appeal of Gröbner bases even when applied to systems of univariate polynomials or simple linear equations; by being able to glean information from a student's answer a CAA system is able to give them meaningful feedback to promote learning. Given a student's answer as a system of polynomials, these techniques can compare this answer to the teacher's, telling us when:

- The student's system is inconsistent and so its variety is the empty set.
- The student's system is underdetermined; its variety contains the teacher's.
- The student's system is overdetermined; at least one equation should not be there.
- Which equations are the cause of the above problems.

The result of the work discussed here is an additional answer test, SysEquiv, added in STACK 2.2. The complete code of the SysEquiv answer test is included in Appendix B. The algorithms described in Chapter 8 are implemented by the existing `grobner` Maxima package.

To demonstrate the application of Gröbner bases in STACK we shall work with the following word problem:

▼ **Example 9.2.1.**

In a railway journey of 90 kilometres an increase of 5 kilometres per hour in the speed decreases the time taken by 15 minutes.

Write a system of equations to represent this situation using v as the speed of the train and t as the journey time.

The solution of Example 9.2.1 is the system of equations

$$\{90 = vt,\ 90 = (v+5)(t-15/60)\}. \tag{9.1}$$

On being handed a student's answer and a teacher's answer, the SysEquiv answer test first performs a number of tests to check that each answer is a system of equations. The initial checks only confirm the *validity* of the student's answer, not its correctness. The checks that the answer test performs are as follows:

1. Are both answers lists?
2. Are all list elements not atoms?
3. Are all list elements equations?
4. Are all list elements *polynomial* equations?
5. Do the lists of equations share the same variables?

By this point the answer test will have converted each answer to a list of polynomial *expressions*, or have returned `false` with appropriate feedback. In the case that the student has not used the same variables as the teacher, the answer test can tell the student whether they have too many, too few, or the correct number but incorrect variables.

Were the student to enter the system in Equations 9.1, theirs and the teacher's answers will have been converted to the system

$$\{90 - vt,\ 90 - (v + 5)(t - \frac{1}{4})\}. \tag{9.2}$$

The answer test then calculates the Gröbner bases (though not necessarily the *reduced* Gröbner bases) of the two systems. In the case of Equation 9.2 the Gröbner basis returned is

$$\{90 - tv,\ -4tv + v - 20t + 365,\ v - 20t + 5,\ -4t^2 + t + 18\}. \tag{9.3}$$

The two Gröbner bases in this case are the same, the Maxima function `poly_grobner_equal` returns true and the answer test terminates with a positive result. Figure 9.2 shows the question in Example 9.2.1 implemented in STACK and the output when the student's answer is as in System 9.1.

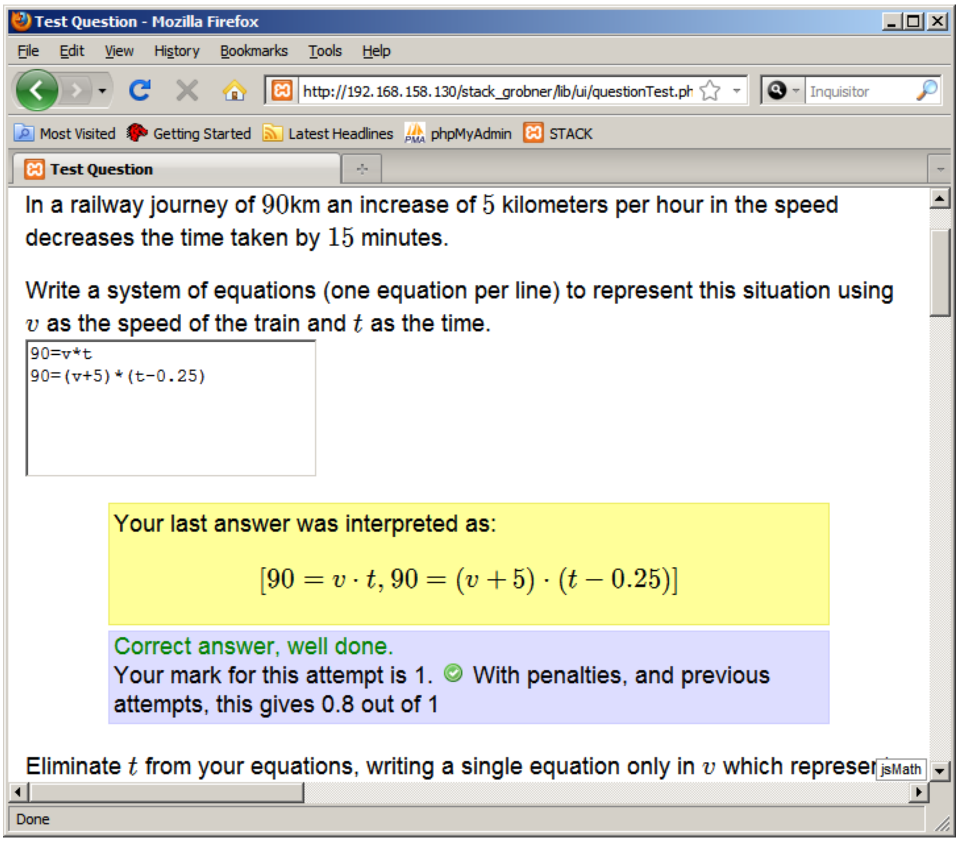


Figure 9.2: Example 9.2.1 implemented in STACK.

9.2.1 Alternate Answers

If a student enters an answer to a question that generates the same ideal as the teacher's answer, it is important that we mark this too as correct. For example, the system of equations

$$\{t = (90/v),\ v^2 + 5v - 1800 = 0\}, \tag{9.4}$$

is the same system as that in Equation 9.1, despite the fact that neither individual equation from each of the systems is equivalent to any other equation. This is a contrived example, it is unlikely that a student would settle upon (9.4) given the question we have posed here, but it is conceivable that with a more complicated situation a student may find a different but equivalent system.

Once the system has been converted to a list of expressions, the Gröbner basis returned by the answer test is

$$\{tv - 90,\ v + 5v - 1800,\ -v + 20t - 5,\ -4t^2 + t + 18\}, \tag{9.5}$$

which bears some resemblance to system 9.3 but is still not the same. If we look at the *reduced* Gröbner bases of the two systems we see that they are

$$\{v - 20t + 5,\ -4t^2 + t + 18\}, \text{ and}$$
$$\{-v + 20t - 5,\ -4t^2 + t + 18\},$$

respectively. These systems are not identical, the first equation of the first system is -1 times the first equation of the second system; though they clearly have the same roots and `poly_grobner_equal` will once again return true and the answer test will terminate. Figure 9.3 shows STACK's response when the student's answer is System 9.4.

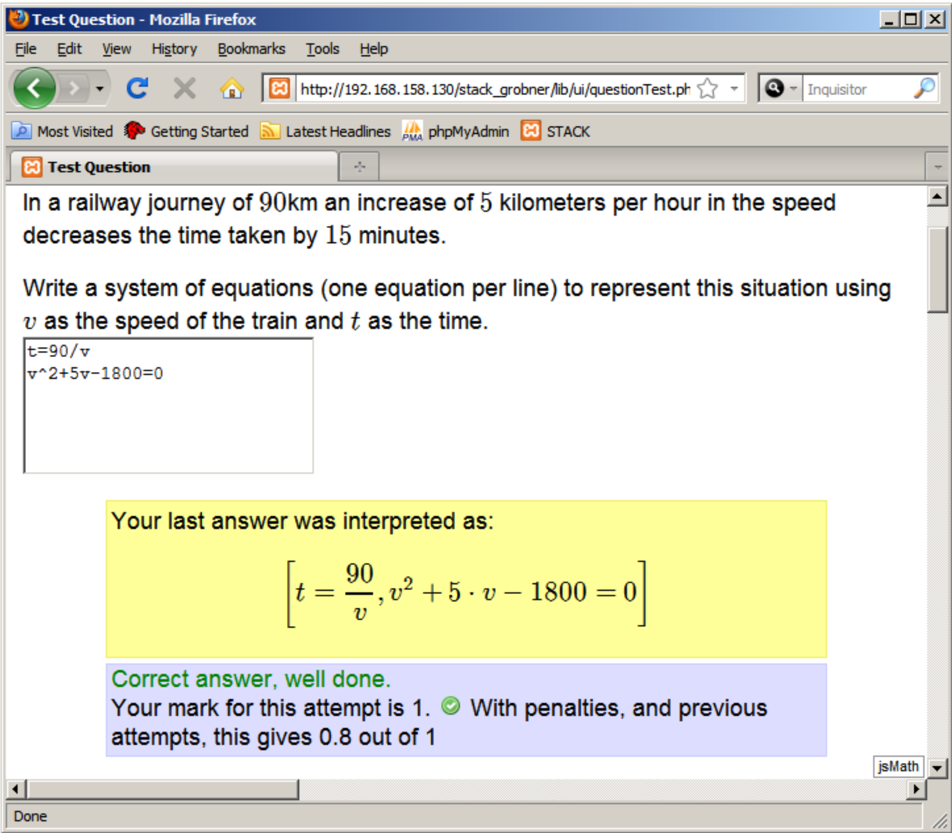


Figure 9.3: STACK accepting an alternative answer to Example 9.2.1.

9.2.2 Incorrect Responses

There are many ways in which a student's answer may be deficient in some way, here we consider just two. The first example uses the potential response tree to give a particular response where a predefined mistake has been made. In Example 9.2.1 the unit of distance is kilometres and the unit of time is hours. As all distances are measured in kilometres, it is unlikely that any errors of units will occur there, however the inclusion of a time given in minutes may mislead the inattentive student.

Were a student to create a system of equations where they mistakenly reduced the time t by 15 minutes instead of $\frac{1}{4}$ hour, their answer would be

$$\{90 = vt,\ 90 = (v + 5)(t - 15)\}. \tag{9.6}$$

In the case of Example 9.2.1, the first node on the potential response tree uses the SysEquiv answer test to compare to the incorrect system 9.6, and if the answer test returns `true`, specific feedback is returned to the student as shown in Figure 9.4.

If the answer test returns `false` in this instance, it does not necessarily mean that the student's answer is incorrect, only that it is not one possible wrong answer. Here the student's answer is passed to the next node of the potential response tree, which compares the student's answer with the correct system. Thus, in the examples given in Figures 9.2 and 9.3 both students' answers have returned false on the first answer test that has been applied.

For our second example of the answer test's response to an incorrect answer, we use the incorrect system

$$\{90 = vt,\ 90 = (v + 5)(t - 0.5)\}, \tag{9.7}$$

where 15 minutes has been mistakenly understood to be half an hour. This system will reach the second node on the potential response tree and be compared to the teacher's answer using the SysEquiv answer test. Unlike the two correct examples we have considered, the

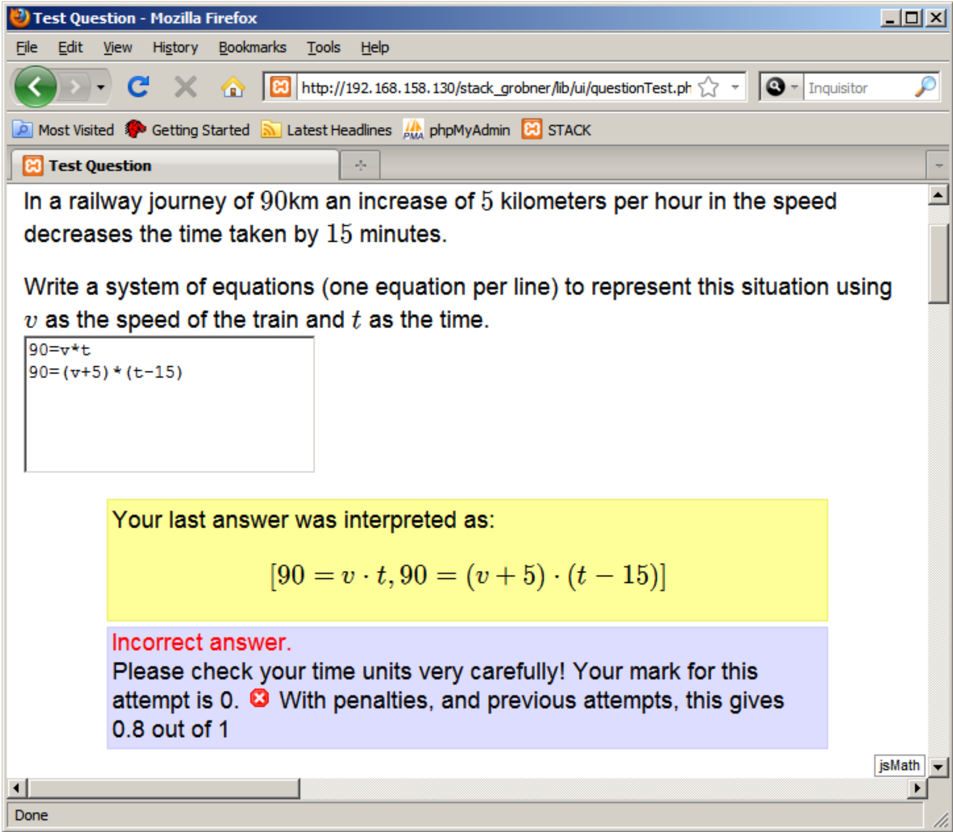


Figure 9.4: STACK giving specific feedback to a particular wrong answer.

function `poly_grobner_equal` will return false and the answer test will have to do more work to determine what is incorrect about the student's answer. Firstly the answer test will check whether the teacher's answer is a subset of the student's answer; if this is the case then the student has an *underdetermined* system and they need to place greater restriction on the set of results, possibly by including extra equations.

In this example the student's system is not a superset of the teacher's system, so each expression in the student's system is checked individually to see whether it reduces to zero modulo the Gröbner basis generated by the teacher's system. From Theorem 8.6.9 we know that expressions which do not reduce to zero do not fit the ideal generated by the teacher's

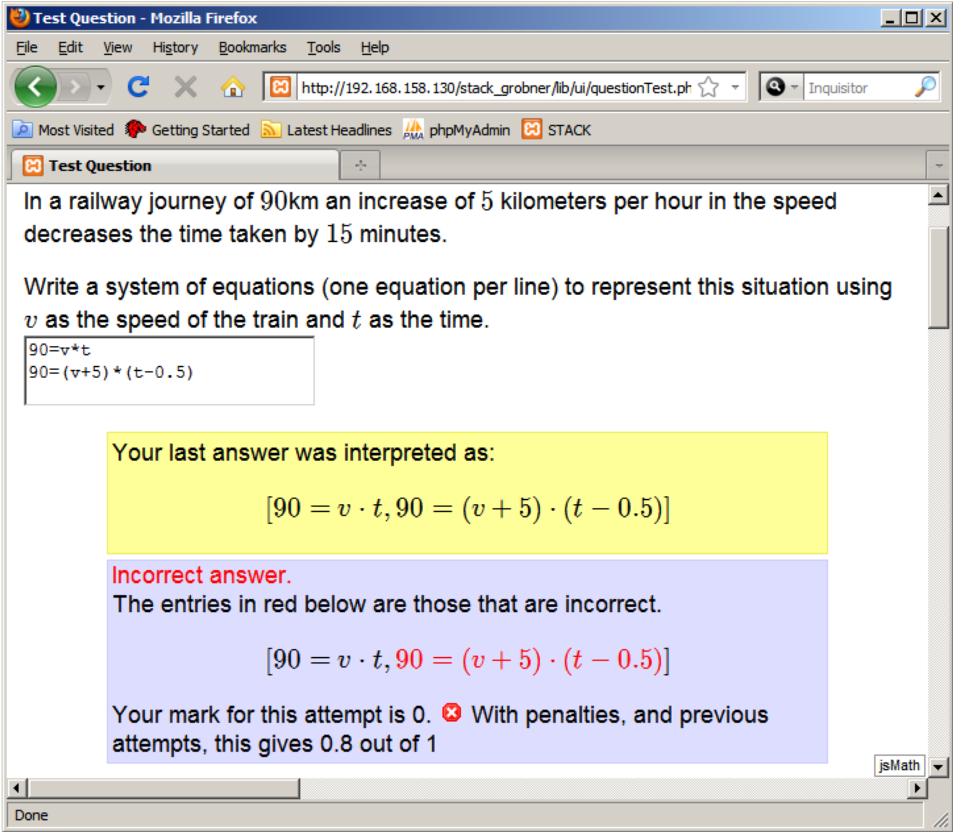


Figure 9.5: Generic response to an incorrect system of equations.

system, and so these expressions' equivalent equations in the student's system are coloured red[1]. Figure 9.5 shows the response given by the system when the student enters Equation 9.7 in answer to the example question.

9.3 Testing and Distribution

The system equivalence answer test was piloted with 91 students at the University of Birmingham in March 2011. Six questions were used in testing, including the *Students and Professors* problem. There were too few responses for analysis (most students answering one or two

[1]We acknowledge that this solution raises accessibility issues, and will be replaced in future with a better interface.

questions each), but answers were checked to ensure that the test had worked as expected, which it had.

Having passed testing, the answer test was included in the 2.2 release of STACK in late 2011. Since then, STACK has undergone significant developmental changes; it is no-longer a standalone system but provides a question type to the virtual learning environment Moodle. The feedback responses of the answer test have been translated into Finnish (by Matti Pauna at the University of Helsinki) and Swedish (by Mikael Kurula at Åbo Akademi University, Finland).

9.4 Answer Test Code

The following code was written by the author, with minor revisions by Dr. Sangwin, to implement the system equivalence answer test in STACK. It is also in Listing B.2 in Appendix B, without comments.

The main function of the answer test is `ATSysEquiv`, which takes two inputs (the student's and teacher's systems, respectively), and returns either `true` or `false`, depending on whether the first input defines the same system as the second. The function begins with a number of tests to ensure that unhandled errors will not be thrown by the system. All tests are applied to the teacher's system as well as the student's.

9.4.1 The `ATSysEquiv` function

The first section of the `ATSysEquiv` function begins by turning on error catching (`errcatch`) and simplifies the two answers (`simp`). The `simp` function replaces any variable defined entirely by numbers with those numbers, so for instance the system $[c = 0,\ y = 4x + c]$ is immediately reduced to $[y = 4x]$:

```
ATSysEquiv(SA,SB):=block([keepfloat,RawMark,FeedBack,AnswerNote,SAA,SAB,S1,S2,varlist,GA,GB,ret],
    RawMark:0, FeedBack:"", AnswerNote:"",
    keepfloat:true,               /* See pg 23 */

    /* Turn on simplification and error catch */
    SAA:errcatch(ev(SA,simp,fullratsimp,nouns)),
    if (is(SAA=[STACKERROR]) or is(SAA=[])) then return(StackReturnOb("0","ATSysEquiv_STACKERROR_SAns","")),
    SAB:errcatch(ev(SB,simp,fullratsimp,nouns)),
    if (is(SAB=[STACKERROR]) or is(SAB=[])) then return(StackReturnOb("0","ATSysEquiv_STACKERROR_TAns","")),
```

The next section checks firstly that the answers submitted are lists with the `listp` function and, if they are, that none of the list elements is just an atom (i.e. a number or variable). This uses the `maplist` function, which maps a given function (in this case `atom`, a predicate function that returns `true` if its input is an atom) to every element of a list:

```
    /* Are both answers lists? */
    if not listp(SA) then
        return(StackReturnOb("0","ATSysEquiv_SA_not_list",StackAddFeedback("","ATSysEquiv_SA_not_list"))),
    if not listp(SB) then
        return(StackReturnOb("0","ATSysEquiv_SB_not_list",StackAddFeedback("","ATSysEquiv_SB_not_list"))),

    /* Are all list elements not atoms? */
    if subsetp({true}, setify(maplist(atom, SA))) then
        return(StackReturnOb("0","ATSysEquiv_SA_not_eq_list",StackAddFeedback("","ATSysEquiv_SA_not_eq_list"))),
    if subsetp({true}, setify(maplist(atom, SB))) then
        return(StackReturnOb("0","ATSysEquiv_SB_not_eq_list",StackAddFeedback("","ATSysEquiv_SB_not_eq_list"))),
```

Next we check that every list element is an equation, using `maplist` again to apply the `op` function. `op` returns the highest operator in the expression tree as demonstrated in Figure 8.2.1—for an equation this will always be =.

```
/* Are all list elements equations? */
if is({"="}#setify(maplist(op,SA))) then
    return(StackReturnOb("0","ATSysEquiv_SA_not_eq_list",StackAddFeedback(""
        ,"ATSysEquiv_SA_not_eq_list"))),
if is({"="}#setify(maplist(op,SB))) then
    return(StackReturnOb("0","ATSysEquiv_SB_not_eq_list",StackAddFeedback(""
        ,"ATSysEquiv_SB_not_eq_list"))),
```

The final step before we can apply the Gröbner basis tests to the inputs is to check that they are both polynomials with the `polynomialpsimp` function (a helper function included in Listing B.2 but not here). This function cannot be applied to equations however, so we first convert our equations to expressions using the STACK-specific `stack_eqnprepare` function:

```
/* Turn our equations into expressions */
S1: maplist(stack_eqnprepare,stack_eval_assignments(exdowncase(SA))),
S2: maplist(stack_eqnprepare,stack_eval_assignments(exdowncase(SB))),
kill(SB),

/* Is each expression a polynomial? */
if subsetp({false}, setify(maplist(polynomialpsimp, S1))) then
    return(StackReturnOb("0","ATSysEquiv_SA_not_poly_eq_list",
        StackAddFeedback("","ATSysEquiv_SA_not_poly_eq_list"))),
if subsetp({false}, setify(maplist(polynomialpsimp, S2))) then
    return(StackReturnOb("0","ATSysEquiv_SB_not_poly_eq_list",
        StackAddFeedback("","ATSysEquiv_SB_not_poly_eq_list"))),
```

At this point we have two systems of multivariate polynomial expressions. First we check that the list of variables for each match and, if they do, determine their Gröbner bases with the `poly_buchberger` function. If the variables are not the same, we apply the `ATSysEquivVars` function described in Section 9.4.2 to determine if there are missing or extra variables.

Assuming there was no problem with the variables, we apply the `poly_grobner_equal` function to the pair of Gröbner bases to check whether they have the same reduced Gröbner basis. If they do, we return `true` from the `ATSysEquiv` function. If they do not, we apply the

ATSysEquivGrob function to determine how they differ.

```
    varlist: listofvars(S2),
    if not is(listofvars(S1)=varlist) then
        return(ATSysEquivVars(S1,S2)),

    GA: poly_buchberger(S1,varlist),
    GB: poly_buchberger(S2,varlist),
    kill(S1,S2),

    /* Determine whether our two lists of polynomials have the same Grobner Bases */
    if poly_grobner_equal(GA, GB, varlist) then
        return(StackReturnOb("1","","")),

    /*
        We now know the student's answer is in the correct form but there is
        something wrong with it. From here we use the grobner package to
        determine which, if any, of their equations is correct.
    */

    return(ATSysEquivGrob(GA, GB, SA, varlist))
```

9.4.2 The ATSysEquivVars Function

This function is called when the variables of two systems of polynomial expressions are not the same. It finds the set of variables of each of the two systems, and determines whether the student's system is a subset of the teacher's system (i.e., missing variables), a superset of the teacher's system (i.e., with superfluous varibles), or incorrect in some other way (neither a subset nor a superset of the teacher's system):

```
ATSysEquivVars(S1,S2):=block([XA,XB],
    XA: setify(listofvars(S1)),
    XB: setify(listofvars(S2)),
    if subsetp(XA,XB) then
        return(StackReturnOb("0","ATSysEquiv_SA_missing_variables",
            StackAddFeedback("","ATSysEquiv_SA_missing_variables"))),
    if subsetp(XB,XA) then
        return(StackReturnOb("0","ATSysEquiv_SA_extra_variables",
            StackAddFeedback("","ATSysEquiv_SA_extra_variables"))),
    return(StackReturnOb("0","ATSysEquiv_SA_wrong_variables",StackAddFeedback(""
        ,"ATSysEquiv_SA_wrong_variables"))),
)$
```

9.4.3 The `ATSysEquivGrob` Function

The `ATSysEquivGrob` function takes two Gröbner bases known to be different, and determines how they vary. It first checks whether the student's Gröbner basis is a subset of the teacher's—that is to say, the student's system is underdetermined. If it is not underdetermined, at least one of the expressions in the student's system will fail the membership test—some f will not reduce to zero modulo the Gröbner basis of the teacher's system. This is determined by the `poly_grobner_member` function and applied to each expression of the student's system in turn.

```
ATSysEquivGrob(GA,GB,S1,varlist):=block([retl,ret],

    /* Is the student's system underdetermined? */

    if poly_grobner_subsetp(GA,GB,varlist) then
        return(StackReturnOb("0","ATSysEquiv_SA_system_underdetermined",
            StackAddFeedback("","ATSysEquiv_SA_system_underdetermined"))),

    /*
        Given that the student's system is neither underdetermined nor equal to
        the teacher's, we need to find which equations do not belong in the
            system.
    */

    ret:[],

    for k:1 thru length(S1) do block([],
        if poly_grobner_member(stack_eqnprepare(stack_eval_assignments(
            exdowncase(S1[k]))), GB, varlist) then
            ret:append(ret,[S1[k]])
        else
            ret:append(ret,[texcolor("red", S1[k])])),

    return(StackReturnOb("0","ATSysEquiv_SA_system_overdetermined",
        StackAddFeedback("","ATSysEquiv_SA_system_overdetermined", StackDISP(ret,
         "\$\$"))))

)$
```

9.5 Example Questions

The principle inspiration for this work was Mayer (1981), first mentioned in Section 2.2. Mayer classified algebra story problems from U.S. high-school textbooks, dividing them first into *families* (based on their underlying source formulae, such as TIME RATE), *categories* (by the form of their story line, such as MOTION, CURRENT or WORK), and finally *templates* (from their specific propositional structure). Example 9.2.1 is one such problem from the time-rate family, and we present others here that have been implemented in STACK. (Some questions are altered from their originals to fit the form of the answer test and units appropriate to the United Kingdom.) The screenshots in this section are from Moodle, using STACK as a question type, and as such they appear different to those above.

9.5.1 Simple Pythagorean

This is adapted from a question in the *geometry* family and *triangle* category (Mayer did not number his questions).

▼ Example 9.5.1 — *Parisian Triangle*.

In central Paris, the Rue de la Paix and the Rue de Rivoli meet besides the Jardin des Tuileries at a right angle. At its other end the Rue de la Paix meets the Avenue de l'Opéra at the Palais Garnier. The Avenue meets the Rue de Rivoli at the Louvre.

Walking at 4 *kph it takes* 36 *minutes to complete the journey from the Louvre to the Palais Garnier and back via the Jardin des Tuileries. Each of the roads is straight, and the area enclosed by them is* $0.24\ km^2$.

What is the distance in metres of the journey described?

Using a for the length of the Avenue de l'Opéra, p for Rue de la Paix and r for Rue de Rivoli, and working in metres, write a system of equations which describes:

- *The relationship between the lengths of the three roads.*
- *The total distance around the roads.*
- *The area enclosed by the roads.*

By solving the above system of equations, what is the distance in metres from the Palais Garnier to the Louvre?

In this question students are required to create and solve a system of equations which represents mathematically the situation described to them. As with Example 9.2.1, the system itself is a relatively simple one, yet not one that could be assessed with existing CAA systems. In this question a system with incorrectly interpreted distnaces (mixtures of metres and kilometres) is included to catch students who apply buggy-/mal-rules. Prior to the implementation of the answer test, a teacher would have to work backwards from an incorrect answer to the final part of the question.

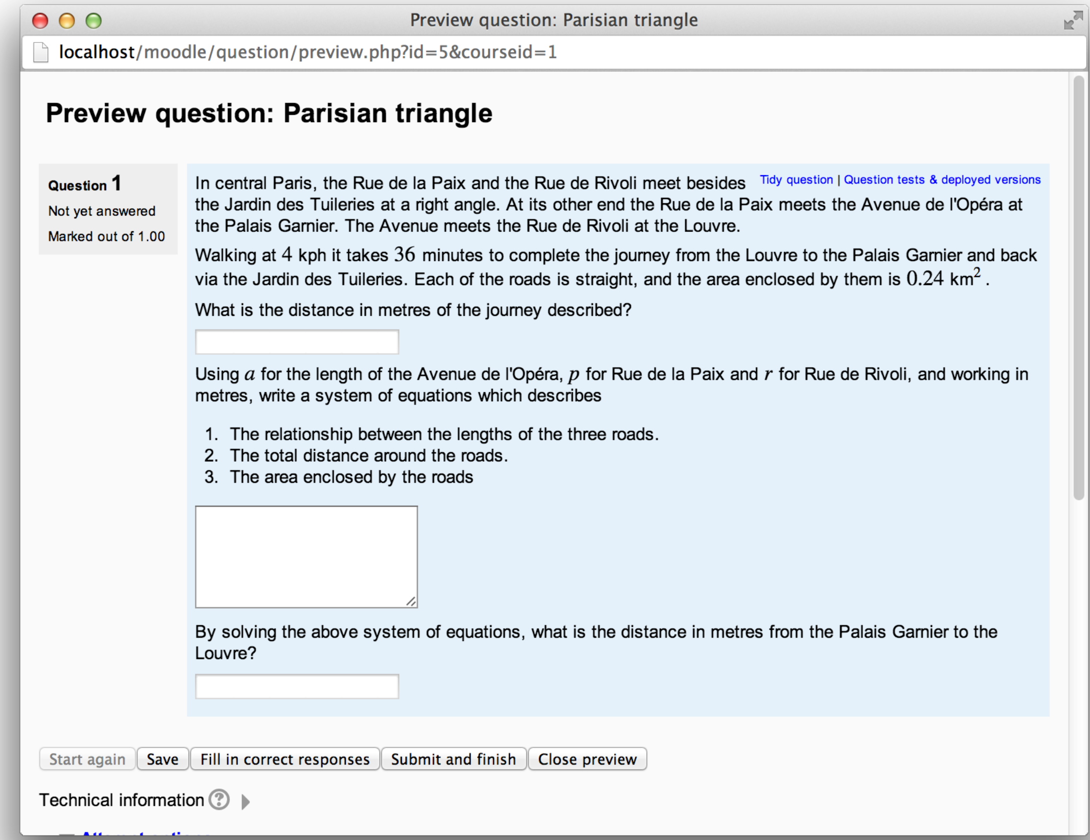


Figure 9.6: Example 9.5.1 implemented in STACK.

9.5.2 Equal Time

Taken from the *Current* category in the *Amount-Per-Time* family, this question is again relatively simple, but perhaps deceptively so. Some students answering this question in testing struggled with the lack of an explicit time period.

▼ **Example 9.5.2 — *Equal Time*.**

A riverboat travels at a rate of 15 *km/h in still water. On the river Avon, however, it travels* 60 *km upstream in the same time that it travels* 90 *km downstream.*

Write a system of equations to represent this, using t as the time and c as the speed of the current. Solve the system of equations to determine the speed of the current on the Avon, c.

9.5.3 Frame Absolute

Another from the *geometry* family, but the *rectangle* category.

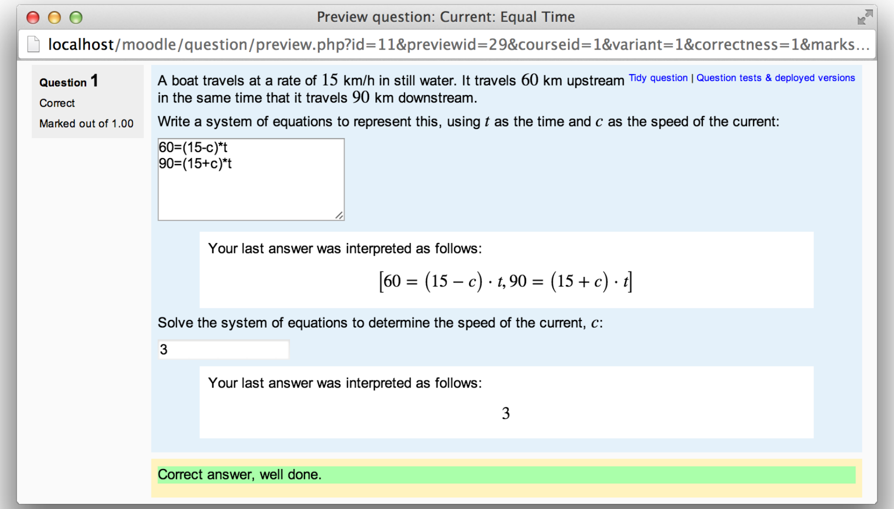


Figure 9.7: STACK accepting an answer to Example 9.5.2.

▼ Example 9.5.3 — *Frame Absolute.*

A framed mirror is 40 *cm wide and* 55 *cm tall. The area of the mirror is* 1924 *cm*^{2}. *Using* w *and* h *for the width and height of the mirror, respectively, and* f *for the width of the frame, write a system of equations to represent the width and height of the framed mirror, and the area of the mirror alone.*

Solve the system of equations to find f.

One possible buggy rule in this question is to incorrectly ascribe the width and height variables to be the total width and height of the framed mirror, for which a node has been added to the question's potential response tree. Because the SysEquiv answer test begins by comparing variables, another node must be created for the correct equation $(55 - f)(40 - f) = 1924$. It would be possible to account for this with further development work, however.

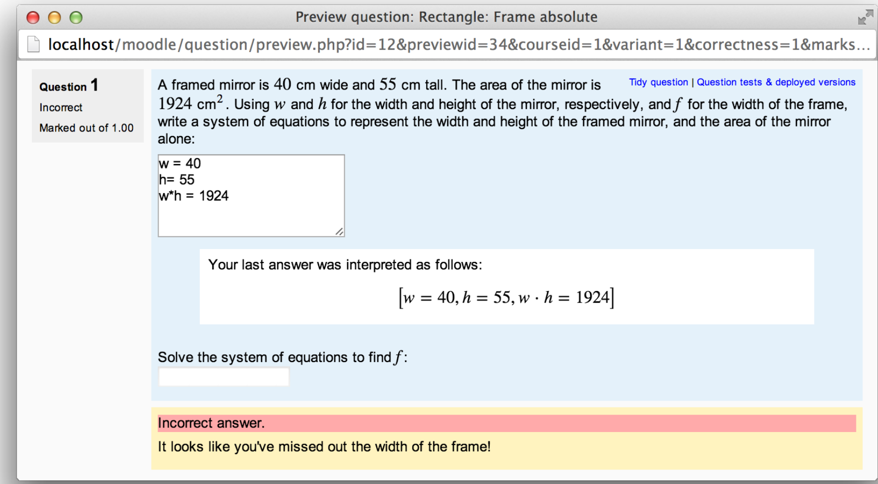


Figure 9.8: STACK identifying a buggy rule in an answer to Example 9.5.3.

9.5.4 Three Relative Amounts

This is from the *cost-per-unit* family and the *dry mixture* category.

▼ **Example 9.5.4.**

Three chemicals a, b, and c cost 60*p,* 40*p, and* 80*p respectively. They are mixed so that the number of grams of b is twice the number of grams of a and is 3 less than the number of grams of c. The mixture is worth £12.40. Write a system of equations to represent this (use a, b, and c as the number of grams of each chemical).*

This algebra story problem is one of those from Mayer (1981) which rather stretches the definition of 'word question', given the lack of story. Parsing such simple language into mathematics still causes problems for student unaccustomed to the process, however.

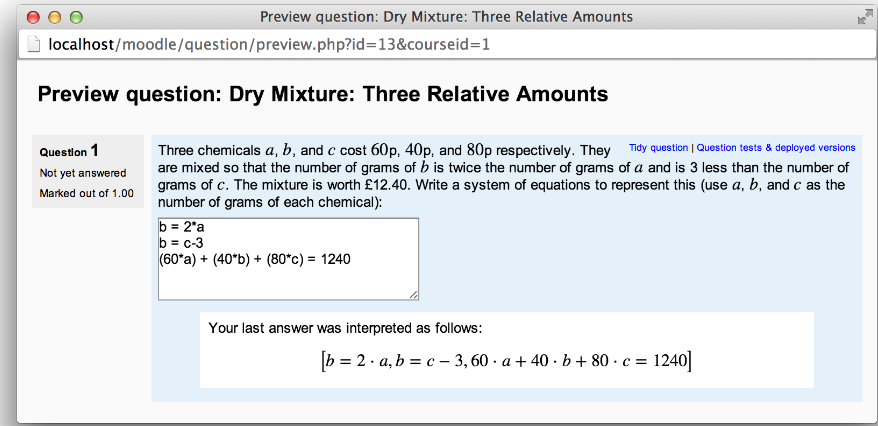


Figure 9.9: STACK validating an input for Example 9.5.3.

Chapter 10

Conclusions and Future Opportunities

In the first section of this chapter we summarise the conclusions we have made over the course of this thesis. In the second section, we note the limits of the work, and the opportunities that exist for expanding on it.

10.1 Conclusions

10.1.1 Part I

The first part of this thesis covered problem-solving, and concluded with a review of problem-solving teaching in six universities in England and Wales. In Chapter 2 we considered what it meant to *think mathematically*, and what *problem-solving*—a frequently-used but rarely-defined term—means. Defining mathematical thinking is difficult because many thought processes that we engage in every day may be included under the banner of 'mathematical thinking', and so there is no clear dividing line between mathematical and non-mathematical thinking. Thinking mathematically may be as much about enculturation into the community of mathematicians as it is about a particular thought process.

The work on problem-solving was prompted by the report of the HE Mathematics Curriculum Summit (Rowlett, 2011), held in January 2011 at the University of Birmingham, which acknowledged the importance of problem-solving but never explicitly defined it. We deferred to George Pólya's definition:

> To search consciously for some action appropriate to attain a clearly conceived but not immediately attainable aim. — Pólya, 1962, p. 1.117

This *concious searching* demonstrates that a 'problem' needs both a question being asked and a person being asked to answer it: a problem for one person need not be a problem for another.

We ended the chapter by considering the efforts of educators to teach problem-solving to students. It was Pólya who sparked the modern movement of teaching problem-solving with *How to Solve It*, (Pólya, 1945), in which he spoke in general terms about the steps of solving problems: understanding the problem, making a plan, carrying out the plan, looking back. Pólya talked not in terms of heuristics or algorithms, but in approaches and ways of thinking; he wanted not to teach 'problem-solving' but mathematical thinking.

Following Pólya, Schoenfeld (1992) demonstrated three strategies useful in attacking problems relating to the roots of polynomials. The issue with this approach, as Schoenfeld acknowledged, is that it does little to develop students' general problem-solving skills, but moves a particular class of questions from problems to exercises. Besides giving students problems to solve, and encouraging them to think about the process they are engaged in (as advocated by Mason et al. (1982)), explicitly teaching problem-solving continues to cause disagreement among mathematics educators.

In Chapter 3 we contrasted traditional *directed learning* (teaching by lectures and example classes) with *discovery learning* (a range of student-centred pedagogies developed since the 1960s). Lectures are popular in universities because they offer a relatively low-cost approach to teaching, scale well with numbers of students, and are effective at transmitting information to students (Bligh, 1998). The ratio of lectures to example classes in Russell Group universities' mathematics departments showed that directed learning dominated the teaching of first-year undergraduate mathematics.

The type of discovery learning we looked more closely at, due to its similarity to the Moore Method discussed in Part II, was *problem-based learning* (PBL). This is not a specific pedagogy, but is characterised by students solving problems with a minimum of exposition on the part of the teacher. It is also known as *enquiry-based learning* (EBL). The role of the teacher is to choose problems for students to work on that offer the right amount of assistance where it is necessary. The elements controlled by the teacher in assisting students are termed *scaffolding*, and form an integral part of the teaching process. This scaffolding clearly defines PBL as different from pure discovery learning, in which students are given no assistance; pure discovery learning is therefore termed *minimally-guided*. Chapter 3 ends with a discussion of arguments for and against PBL, in particular those of Kirschner et al. (2006) and Hmelo-Silver et al. (2007). Kirschner et al attack all discovery learning as minimally-guided, and claim it is only effective where teachers construct suitable scaffolding to assist students. Hmelo-Silver et al counter by arguing that there is great diversity amongst discovery learning pedagogies, and that scaffolding is not an admission of failure but and integral part of the process.

In Chapter 4, the final chapter of this part of the thesis, we reported the results of a questionnaire and six case-studies on problem-solving in England and Wales. While there were some problems with participation, in the majority of mathematics departments in England and Wales who responded to our questionnaire did not report a module specifically dedicated to or taught using problem-solving. Others interpreted our definition of problem-solving differently, seeing examples of problem-solving among numerical analysis or project work.

Six departments' replies to the questionnaire cited modules in their undergraduate mathematics programme that we agreed contained a considerable amount of problem-solving. These six modules were then investigated in case studies (two were not included in this thesis due to their having separate, dedicated chapters in the second part), which were the result of interviews with teaching staff and access to course literature and materials. The staff interviewed were generally enthusiastic about their teaching, but we were only able to report on their perceptions of the modules. Students were not available for interview for these case studies.

10.1.2 Part II

The second part of this thesis was about the *Moore Method*, a problem-based learning pedagogy for university mathematics teaching, developed in the United States in the first half of the 20$^{\text{th}}$ Century. Chapter 5 discussed the history of the Method, which was first taught by Robert Lee Moore at the University of Pennsylvania from 1911 onwards, and described its fundamental aspects. The Moore Method is today recognised as a type of PBL, but was developed largely in isolation from other pedagogies. The key aspects of the Method which define it from other PBL pedagogies are the progression of problems, and the focus on communication, both written and oral.

While the Moore Method is popular in some universities in the United States, particularly at the University of Texas at Austin and the University of Chicago, it has not been extensively analysed. Those who teach with the Moore Method at these universities are overwhelmingly mathematical descendants of Moore, themselves taught by the Method, and as such they are convinced of its efficacy. The second and third chapters of this part of the thesis are reports on Moore Method teaching in two mathematics departments in the United Kingdom, and as such contribute to what little research exists.

Chapter 6 sought to answer the research question: "Does the Moore Method as implemented at the University of Birmingham improve students' performances in other modules in the School of Mathematics?" 1Y is an optional first-year module, introduced in 2004/05 and taught using the Moore Method. Students are taught either naïve set theory or geometry.

Multivariate regression analysis demonstrated a correlation between 1Y participation and higher performance in some modules in the School of Mathematics. Regression analyses of first-year modules used an early Class Test mark to remove original differences as much as possible, and in first-year modules the unstandardised coefficient of the 1Y variable (being 1 for 1Y students and 0 for the rest of the cohort) was between 4.58% and 9.92%.

Regression analyses for second-year modules used students' first-year mean mark in place of the Class Test mark, and showed a distinction between three topics in pure mathematics and programming, where the 1Y coefficient was statistically significant at 5%, and three other modules with almost no effect. The differences here were smaller than for first-year modules, but the first-year mean of 1Y students was already influenced by their 1Y participation: when the first-year mean was modelled as a function of the Class Test mark and 1Y coefficient, the 1Y coefficient had an unstandardised coefficient of 8.11%.

Third year analyses showed no relationship between 1Y attendance and module performance in any of the seven modules considered. There are more third year modules than in the earlier years of the programme, and with only five years' data (as opposed to six for the second-year and seven for the first-year analyses) the number of data points per module—especially of 1Y students—was far smaller than in the first- and second-year analyses.

There was no clear difference in models when genders were considered separately, however with fewer female students taking 1Y the error bounds were wider. Overall, the results showed a strong correlation between 1Y participation and performance in some first- and second-year modules, though we are unable to assert a causal relationship.

Chapter 7 reported on staff experiences while teaching a newly-introduced Moore Method module at another university in England. Its aims were to record teachers' perceptions of:

- What the module objectives were
- What the Moore Method was
- Student attitudes and progress
- Module efficacy

This report was the result of thematic analysis of six semi-structured interviews with staff who had taught ten groups of students over the two years that the module had been running. Semi-structured interviews were chosen for their balance of specificity and freedom for elaboration by interviewees, and a thematic analysis was created from independent codings of interviews by the two researchers. Given the nature of the data being analysed, we cannot draw conclusions on the efficacy of the module or its groups, only report on the experiences of the staff.

Titled *New Investigations*, the central aims of the module were to improve students' problem-solving skills, their group work, and their written and oral communication. As such the module was not a traditional Moore Method course, group work and research outside classes were encouraged, however the central pillars of a Moore class, namely written and oral communication, and a progression of problems, were present. The extent to which each class could be said to be a modified Moore Method varied. Some classes appeared very similar to a Moore Method module, in spite of the addition of group work and external references. One appeared to be something altogether different, being based on a paper in communication theory and not a progression of individual problems.

The general reaction from students, as reported by our interviewees, was positive, however there was surprise at the speed of students' progress through the problems, and a gap between the expectation and reality of the mathematical knowledge that students had at their disposal. Groups progressed through the same problem sets at dramatically different speeds, and in some cases lecturers had to modify either the questions or the approach to bear positive results. In general staff were positive about the module, though as an addition to the undergraduate programme, and not a fundamental change to teaching practice.

10.1.3 Part III

The third part of this thesis moved from problem-solving to computer-aided assessment. Assessing solutions to problems is well beyond the abilities of existing CAA systems, but we can make small steps in the direction of problem-solving if we can assess types of systems of equations. The work presented in this chapter focused on the assessment of systems of multivariate polynomials in the CAA software STACK, which has been under development at Birmingham since 2005.

In Part III we covered the mathematics and implementation of the theory of Gröbner Bases in STACK, to assess students working with systems of multivariate polynomial equations. Here we give a brief overview of the mathematics involved and the application of the system in STACK.

The central component to assessing a student in STACK is an *answer test.* Answer tests take two inputs—a student's and teacher's answers—and determine whether the student's answer is sufficiently alike the teacher's answer. If they are it returns `true`. If they are not, it returns `false`, and a string detailing where the differences lie. This 'sufficient alikeness' is determined based on the type of question that the teacher is asking. In the most basic case this is algebraic equivalence of two expressions.

Before the work discussed in the chapter was completed, no CAA system could assess systems of multivariate polynomials. Systems of multivariate polynomials can arise from very simple situations—if we wish to find the points of intersection of a line and a circle, we have to solve a (very simple) system of multivariate polynomials (systems of univariate polynomials and linear equations are special cases). The principle issue when comparing two such systems is that they may look completely unlike one-another, indeed they may share no two equal polynomials, yet have the same algebraic structure. It is this underlying structure—that is to say they generate the same ideal—that defines our 'sufficient alikeness' for these systems, and so this forms our basis for comparison. The problem then is just one place removed, however: we cannot look at two sets of equations and say immediately whether they generate the same ideal. The solution to this is to compare their Gröbner bases. The theory of Gröbner Bases was developed by Bruno Buchberger in the 1960s, and the reduced (i.e., monic) Gröbner basis of an ideal generated by a system of multivariate polynomials is unique. Furthermore, Buchberger's Algorithm allows us to compute the reduced Gröbner basis of an ideal from any generating set quickly.

Using the `grobner` Maxima package (Maxima is the CAS STACK uses) we wrote an answer test for systems of multivariate polynomials in STACK. Other aspects of Gröbner Bases, in particular the ability to solve the Ideal Membership Problem (reducing a polynomial modulo the Gröbner basis), allow us to give worthwhile feedback to students whose answers are not correct.

The answer test *SysEquiv* was trialled with first-year students using algebra story problems, and introduced in the 2.2 release of STACK. Systems of polynomials and linear equations, being special cases of systems of multivariate polynomials, are also assessable with the system.

10.2 Omissions and Opportunities

In this section we give a brief overview of the possibilities for further work in the topics each part covered.

10.2.1 Part I

Chapter 2 gave a brief synopsis of mathematical thinking and problem-solving. On both of these topics much has been written in the two decades since Schoenfled's review (Schoenfeld, 1992), yet there is none more recent. The issue of the definition of 'problem-solving' itself remains unresolved. Responses to the questionnaire in Chapter 4 demonstrated the differing interpretations of lecturing mathematicians as to what constitutes problem-solving. An extended review of the literature would help to clarify the current state of problem-solving in mathematics teaching and research, though this would be a significant undertaking.

Chapter 3 briefly summarised the differences between *directed learning* and *discovery learning*. Directed learning is by no-means uniform, in particular in the use of assessments and lecturing style. Insofar as lecture notes are concerned, the amount that students are expected to write in any particular module varies widely, even within a department. It would be interesting to discover how these variables affect students' learning experiences, and serve as a more recent, mathematics-specific, alternative to Bligh (1998), itself now fifteen years old. Better would be an expanded taxonomy of mathematical pedagogies, including not only directed and discovery learning, but other alternative pedagogies besides.

In our review of the ratio of lectures to example classes, we looked at an unrepresentative sample of universities—namely those in the Russell Group. Furthermore, we were limited by the information available online and in module booklets. These limitations were made necessary by the amount of work that an exhaustive review would entail, and may have resulted in a misleading view of example classes being dominated by lectures. Some types of interaction are difficult to categorise: is a computer lab a lecture or an exercise class, a mixture or something entirely different (we suggest the latter)? Core first-year modules served as our basis for comparison; had we looked at second- or third-year modules, or those outside the main discipline, results would likely be different. Ideally, we would wish to review the ratio of lectures to example classes over a significant number of modules from all universities offering a G100 programme.

The principle issue faced during the work that contributed to Chapter 4 was the lack of responses to questionnaires that were sent to departments. These were addressed to individual heads of department or directors of teaching and, where there was no response, attempts were made to contact them by telephone. 35 of 59 departments completed the questionnaire and as such it is not comprehensive, though over 50% is a reasonably good response rate. To draw better conclusions on the state of problem-solving teaching in U.K. mathematics departments, a more representative sample of departments would be necessary.

Another issue with the questionnaire was the definition of *problem-solving*. It appeared that many respondents either misinterpreted our definition or deferred to their own view of what 'problem-solving' was. Perhaps as a result of this, only one non-Russell Group university was represented in our case studies, which was by no means our intention. It may be the case that, from the information given in the questionnaire responses and course literature available online, we incorrectly interpreted modules as *not* containing problem-solving, despite not having seen them being taught. This was also the case with the modules we *did* class as being taught by methods including problem-solving—reviews of the teaching are the lecturer's perceptions of the modules. We did not have access to students and the reality may be different, and we were unable to draw conclusions on the efficacy of those modules we studied.

Finally, it was not possible to determine the effects of varying class size, resource allocation, module level, or assessment criteria, with such a small sample and only secondary sources on which to base conclusions. A larger study of problem-solving teaching in U.K. mathematics departments that was able to take account of these elements would be highly valuable.

10.2.2 Part II

In Chapter 6 our methods for data analysis were largely determined by those data we had at our disposal. The 1Aa Class Test, while relatively wide-ranging for a single module, was taken after students had being studying at Birmingham for several weeks. It would be preferable to have a diagnostic test that assessed students over a range of topics, returning individual marks for each, and which was taken during the first two weeks of the undergraduate programme. STACK has been in use for diagnostic testing since 2010 (the SysEquiv answer test was trialled during the diagnostic testing in 2011), but using this test would have meant discarding most of the data available from previous years.

Similarly, additional diagnostic testing of students after their first, second, and third years may give further insights than individual module marks, though naturally performance in other modules is of fundamental importance.

Even with these alterations, and the existing inclusion of the Class Test mark to remove as much as possible the effects of 'better' students taking the module, we cannot discount the curiosity of students willing to take a module outside the norm taught by a non-standard method. From a research point of view it would be better to divide each year's cohort into two halves with similar representative statistics from a diagnostic test, teaching one half using the Moore Method and one without it. Being able to divide the cohort arbitrarily would enable us to fit the groups to a particular testing methodology, rather than finding the best fit for the available data. Gaining ethics approval for such a study may prove difficult, however.

No further testing is available at Birmingham as 1Y and 1X have been replaced with a first-year module on modelling and problem-solving. Parts of the new module incorporate the oral presentations and written work associated with 1Y, though the module is not taught using the Moore Method. Because all students now take this module, testing its effects within a cohort is not possible, and normalising results over different examinations from standard modules would be unlikely to produce worthwhile results.

Chapter 7 centred on the experiences of staff teaching a Moore Method module at a Russell Group university. It was interesting to record these experiences, due to the rarity of having so many staff teaching a particular module, though there were several problems that would need overcoming before we could draw conclusions on the efficacy of the module or the experiences of students taking it. As with the study conducted at Birmingham, research was conducted in a working mathematics department, and as such had to take what opportunities arose.

While we had access to teaching materials and completed three hours of interviews with staff, we were not able to observe teaching directly, and so cannot say to what extent the individual groups' classes varied, except as reported by staff. As such it is not possible to determine what variables may affect results. Even with direct observation this would likely be difficult, however, given the complex of variables in the situation (teacher, students, class size, topic, material, exact teaching method). Since interviews were conducted the number of staff teaching the module each year has fallen from five or six to three, so work to consider the range of effects of the module could now be more tightly controlled, however.

Only the author was able to visit the university to interview staff, which took a considerable amount of time, and most interviews took place during the Easter break when students were not around and staff had more time as a result. As such, students were not interviewed for this study. Even in groups, this would have required significant investments of time and money, and the sample would be skewed by those willing to talk about the module. Furthermore, interviewing students would likely require the consent of another university's ethics committee, a potentially lengthy process. Could these hurdles be overcome, much useful research would likely result.

Further research questions on the Moore Method arise from the discussions in Chapter 5. In particular, do Moore Method modules distract from other modules as reported by Dancis and Davidson (1970), and is the Method is capable of creating unhealthy levels of competition between students (Cohen, 1982, p. 474)?

10.2.3 Part III

The principle development for the System Equivalence answer test happened relatively early during this work, but due to time constraints was not developed further. While the answer test has now been in the release version of STACK for over two years, there are many opportunities to develop it further, either to assess a wider range of questions, or to give more useful feedback for the range of questions it does currently assess. Reporting on students' performance when answering the types of questions assessed by the answer test would be interesting, too. In this section we consider a few such possibilities.

In its current form the answer test requires teachers to be explicit in the variables they require students to use (though capitalisation does not matter). For small systems with three or fewer variables it would not be impractical to consider all permutations of a student's variables and compare these systems to the teacher's, though this does not preclude a student getting variables entirely backwards when creating their system. This is highly unlikely with complex systems, but of fundamental importance when assessing questions such as the *Students and Professors problem* (Example 2.2.3).

The answer test can tell a student if their system has too many or too few equations, or too many or to few variables, or when particular equations are incorrect. One fundamental problem with the feedback in the latter case is that incorrect equations are highlighted in red, which causes a problem for screen-readers used by visually-impaired people. The user interface for this feedback could therefore be improved. In the cases of incorrect equations, determining whether a particular equation 'should be' a correct equation from the teacher's system, and the difference between these two equations would improve feedback for students. This could be done by determining whether correct equations from the student's system generate the same ideal as a subset of the equations of the teacher's system, and what the differences between the student's incorrect and the teacher's remaining equations are. The algebraic equivalence answer test in STACK checks for the application of a number of buggy rules, and would prove valuable in this instance. In the case of the system described in Equation 9.6 in Section 9.2.2, it would be straightforward to determine that the the incorrect equation differs from the remaining teacher's equation only by a constant.

Extending the system to deal with differential equations involving multivariate polynomials would be a difficult, but not insurmountable problem, and open up a wider range of questions for assessment in STACK. STACK already includes an answer test for individual differential equations, and much its code would be directly applicable to updating the SysEquiv answer test.

One of the six systems used in initial testing of the answer test was the 'system' from the *Students and Professors problem*, $6S = P$. About a third of students failed to answer the question correctly, in agreement with the results of Clement et al. (1981). Other problems came from Mayer (1981), and while including only basic algebra these too were answered incorrectly by a large fraction of the students who completed the trial. STACK is now included as a question type in Moodle, a mature virtual learning environment with well-developed reporting features. It would be interesting to use these reporting features with a large set of Mayer's story problems and a greater number of students to determine whether students do indeed struggle to answer such problems correctly.

10.3 Final Word

The work discussed in this thesis, in particular the analysis of the Moore Method in Chapter 6, and the novel implementation of Gröbner Bases in Computer-Aided Assement in Chapter 9, makes modest but very useful contributions in two distinct areas of mathematics education. There are exciting opportunities to expand on the work in both the analysis of problem-based learning and the assessment of students with computers, and both these areas will be of increasing importance as computers—and the mathematics that underpins them—become increasingly valuable to our society.

> “ **Moriarty:** How are you with mathematics?
>
> **Seagoon:** I speak it like a native.
>
> — Milligan and Sykes, 1954

Appendix A

Results

		Levene's Test for Equality of Variances		t-test for Equality of Means						
	Equal Variances	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval Lower	Upper
MSM1Aa	Assumed	4.73	0.03	-5.21	1311	0.00	-12.85	2.47	-17.68	-8.01
Class Test	Not assumed			-6.13	76.57	0.00	-12.85	2.10	-17.02	-8.67
MSM1Aa	Assumed	2.46	0.12	-6.75	1327	0.00	-14.73	2.18	-19.01	-10.44
	Not assumed			-8.53	78.33	0.00	-14.73	1.73	-18.16	-11.29
MSM1Ab	Assumed	4.09	0.04	-6.10	1319	0.00	-15.63	2.56	-20.66	-10.61
	Not assumed			-7.22	76.62	0.00	-15.63	2.17	-19.95	-11.32
MSM1B	Assumed	3.52	0.06	-6.25	849	0.00	-15.09	2.42	-19.83	-10.35
	Not assumed			-7.66	85.50	0.00	-15.09	1.97	-19.01	-11.17
MSM1C	Assumed	3.67	0.06	-5.28	915	0.00	-13.72	2.60	-18.81	-8.62
	Not assumed			-6.22	76.07	0.00	-13.72	2.20	-18.10	-9.33
MSM1D	Assumed	1.57	0.21	-5.41	1080	0.00	-13.21	2.44	-18.00	-8.42
	Not assumed			-6.00	77.46	0.00	-13.21	2.20	-17.59	-8.82
MSM201	Assumed	1.12	0.29	-2.89	570	0.00	-8.43	2.92	-14.16	-2.69
	Not assumed			-2.50	58.38	0.02	-8.43	3.37	-15.18	-1.67
MSM203	Assumed	0.11	0.74	-4.73	306	0.00	-17.61	3.72	-24.93	-10.29
	Not assumed			-4.65	54.03	0.00	-17.61	3.79	-25.21	-10.01
MSM2A	Assumed	0.17	0.68	-4.64	982	0.00	-11.99	2.58	-17.06	-6.92
	Not assumed			-4.50	65.08	0.00	-11.99	2.67	-17.31	-6.66
MSM2B	Assumed	0.42	0.52	-4.80	659	0.00	-12.12	2.53	-17.08	-7.16
	Not assumed			-4.34	67.20	0.00	-12.12	2.79	-17.70	-6.55
MSM2C	Assumed	0.17	0.68	-4.85	894	0.00	-12.60	2.60	-17.71	-7.50
	Not assumed			-4.55	65.27	0.00	-12.60	2.77	-18.13	-7.07
MSM2D	Assumed	1.15	0.28	-3.47	622	0.00	-8.04	2.32	-12.59	-3.49
	Not assumed			-3.09	67.43	0.00	-8.04	2.60	-13.23	-2.85

Table A.1: Independent t-test results for first and second year courses

		Levene's Test for Equality of Variances		t-test for Equality of Means						
	Equal Variances	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval Lower	95% Confidence Interval Upper
MSM3A03	Assumed	0.68	0.41	-1.30	293	0.20	-6.63	5.11	-16.69	3.44
	Not assumed			-1.44	18.57	0.17	-6.63	4.60	-16.28	3.03
MSM3A04	Assumed	0.48	0.49	-1.45	212	0.15	-8.48	5.86	-20.02	3.07
	Not assumed			-1.88	13.49	0.08	-8.48	4.50	-18.17	1.22
MSM3A05	Assumed	0.37	0.54	-0.25	136	0.80	-1.64	6.64	-14.77	11.49
	Not assumed			-0.29	11.15	0.78	-1.64	5.66	-14.08	10.80
MSM3P05	Assumed	2.64	0.11	-3.03	263	0.00	-10.55	3.48	-17.40	-3.69
	Not assumed			-3.88	35.97	0.00	-10.55	2.72	-16.06	-5.03
MSM3P08	Assumed	5.88	0.02	-2.52	65	0.01	-17.37	6.88	-31.11	-3.63
	Not assumed			-3.26	42.98	0.00	-17.37	5.33	-28.11	-6.62

Table A.2: Independent t-test results for third year courses

Appendix B

Code

Listing B.1: Code for organising student module results data in Excel

```
Public Sub Setify()

Application.Calculation = xlCalculationManual
Application.ScreenUpdating = False

Dim i, j As Integer
Dim dblMark As Double
Dim strCode, strSRN As String
Dim booNew As Boolean

i = 1
strSRN = ""

Do While Not IsEmpty(Range("srn").Offset(i, 0))

    'Check whether we're working with a new student SRN
    If Range("srn").Offset(i, 0) = strSRN Then
            booNew = False
        Else
            booNew = True
            strSRN = Range("srn").Offset(i, 0)
    End If

    'Get the course code
    strCode = Range("srn").Offset(i, 1)

    'Get the offset for the course
    j = 1
    Do While Not Range("mark").Offset(0, j) = strCode
        j = j + 1
    Loop

    If booNew = True Then
            Range("mark").Offset(i, j).Value = Range("mark").Offset(i, 0).Value
            i = i + 1
        Else
            Range("mark").Offset(i - 1, j).Value = Range("mark").Offset(i, 0).Value
            Rows(i + 1).EntireRow.Delete
    End If
Loop

Application.Calculation = xlCalculationAutomatic
Application.ScreenUpdating = True

End Sub
```

Listing B.2: The SysEquiv answer test in STACK

```
/***********************************************/
/*                                             */
/*          System Equivalence Test            */
/*                                             */
/*  An addition to STACK using Grobner Bases   */
/*                                             */
/***********************************************/


/*
    What these functions do:

    - Determine whether the student's and teacher's answers are systems of
      equations
    - Convert the two systems of equations into two systems of expressions
    - Determine whether both systems are systems of multivariate polynomials
    - Compare the variables in student's and teacher's answers, if they're not
      the same tell the student
    - Find their Buchberger polynomials of the two systems
    - Use the Buchberger polynomials to compare the Grobner bases of the two
      systems
    - If the Grobner bases are not equal, determine whether the student's is a
      subset of the teacher's
    - If student's system has equations which should not be there, tell them
      which ones.
*/



/*
    Main function of the System Equivalence test

    Takes two inputs, checks whether they are
    lists of polynomials and delegates everything
    else to other functions.

    Process:

    - Is each answer a list?
    - Is each list element not an atom?
    - Is each list element an equation?
    - Is each list element a polynomial?
*/

/* Edited files: SysEquiv.php, AnsTestcontroller.php, lang/en/stack.php */

ATSysEquiv(SA,SB):=block([keepfloat,RawMark,FeedBack,AnswerNote,SAA,SAB,S1,S2,
    varlist,GA,GB,ret],
    RawMark:0, FeedBack:"", AnswerNote:"",
    keepfloat:true,               /* See pg 23 */

    /* Turn on simplification and error catch */
    SAA:errcatch(ev(SA,simp,fullratsimp,nouns)),
    if (is(SAA=[STACKERROR]) or is(SAA=[])) then return(StackReturnOb("0","
        ATSysEquiv_STACKERROR_SAns","")),
    SAB:errcatch(ev(SB,simp,fullratsimp,nouns)),
    if (is(SAB=[STACKERROR]) or is(SAB=[])) then return(StackReturnOb("0","
        ATSysEquiv_STACKERROR_TAns","")),

```

```
    /* Are both answers lists? */
    if not listp(SA) then
        return(StackReturnOb("0","ATSysEquiv_SA_not_list",StackAddFeedback("","ATSysEquiv_SA_not_list"))),
    if not listp(SB) then
        return(StackReturnOb("0","ATSysEquiv_SB_not_list",StackAddFeedback("","ATSysEquiv_SB_not_list"))),

    /* Are all list elements not atoms? */
    if subsetp({true}, setify(maplist(atom, SA))) then
        return(StackReturnOb("0","ATSysEquiv_SA_not_eq_list",StackAddFeedback("","ATSysEquiv_SA_not_eq_list"))),
    if subsetp({true}, setify(maplist(atom, SB))) then
        return(StackReturnOb("0","ATSysEquiv_SB_not_eq_list",StackAddFeedback("","ATSysEquiv_SB_not_eq_list"))),

    /* Are all list elements equations? */
    if is({"="}#setify(maplist(op,SA))) then
        return(StackReturnOb("0","ATSysEquiv_SA_not_eq_list",StackAddFeedback("","ATSysEquiv_SA_not_eq_list"))),
    if is({"="}#setify(maplist(op,SB))) then
        return(StackReturnOb("0","ATSysEquiv_SB_not_eq_list",StackAddFeedback("","ATSysEquiv_SB_not_eq_list"))),

    /* Turn our equations into expressions */
    S1: maplist(stack_eqnprepare,stack_eval_assignments(exdowncase(SA))),
    S2: maplist(stack_eqnprepare,stack_eval_assignments(exdowncase(SB))),
    kill(SB),

    /* Is each expression a polynomial? */
    if subsetp({false}, setify(maplist(polynomialpsimp, S1))) then
        return(StackReturnOb("0","ATSysEquiv_SA_not_poly_eq_list",
            StackAddFeedback("","ATSysEquiv_SA_not_poly_eq_list"))),
    if subsetp({false}, setify(maplist(polynomialpsimp, S2))) then
        return(StackReturnOb("0","ATSysEquiv_SB_not_poly_eq_list",
            StackAddFeedback("","ATSysEquiv_SB_not_poly_eq_list"))),

    /*
        At this point have two lists of polynomials. We now check whether the
        student's and teacher's polynomials have the same variables. If they do,
        we find their Grobner bases and determine whether the systems of
        equations have the same solutions
    */

    varlist: listofvars(S2),
    if not is(listofvars(S1)=varlist) then
        return(ATSysEquivVars(S1,S2)),

    GA: poly_buchberger(S1,varlist),
    GB: poly_buchberger(S2,varlist),
    kill(S1,S2),

    /* Determine whether our two lists of polynomials have the same Grobner Bases */
    if poly_grobner_equal(GA, GB, varlist) then
        return(StackReturnOb("1","","")),

    /*
        We now know the student's answer is in the correct form but there is
        something wrong with it. From here we use the grobner package to
        determine which, if any, of their equations is correct.
    */
```

```

    return(ATSysEquivGrob(GA, GB, SA, varlist))


)$


/* Checks that an expression is a polynomial */

polynomialpsimp(e):=block([],
    return(polynomialp(e, listofvars(e)))
)$


/* Takes two lists of expressions and compares the variables in each */

ATSysEquivVars(S1,S2):=block([XA,XB],
    XA: setify(listofvars(S1)),
    XB: setify(listofvars(S2)),
    if subsetp(XA,XB) then
        return(StackReturnOb("0","ATSysEquiv_SA_missing_variables",
            StackAddFeedback("","ATSysEquiv_SA_missing_variables"))),
    if subsetp(XB,XA) then
        return(StackReturnOb("0","ATSysEquiv_SA_extra_variables",
            StackAddFeedback("","ATSysEquiv_SA_extra_variables")))
)$

/*
    Grobner basis comparison

    This function takes two Grobner bases and a set of variables and determines
    whether the student's system is underdetermined or overdetermined. It also
    takes the student's original system so that if it is overdetermined it can
    tell them which equations should not be there.
*/

ATSysEquivGrob(GA,GB,S1,varlist):=block([retl,ret],

    /* Is the student's system underdetermined? */

    if poly_grobner_subsetp(GA,GB,varlist) then
        return(StackReturnOb("0","ATSysEquiv_SA_system_underdetermined",
            StackAddFeedback("","ATSysEquiv_SA_system_underdetermined"))),

    /*
        Given that the student's system is neither underdetermined nor equal to
        the teacher's, we need to find which equations do not belong in the
            system.
    */

    ret:[],

    for k:1 thru length(S1) do block([],
        if poly_grobner_member(stack_eqnprepare(stack_eval_assignments(
            exdowncase(S1[k]))), GB, varlist) then
            ret:append(ret,[S1[k]])
        else
            ret:append(ret,[texcolor("red", S1[k])])),

    return(StackReturnOb("0","ATSysEquiv_SA_system_overdetermined",
        StackAddFeedback("","ATSysEquiv_SA_system_overdetermined", StackDISP(ret,
         "\$\$"))))
```

```
)$
```

Appendix C
Pólya's Advice

First
You have to understand the problem

Understanding the Problem
- What is the unknown? What are the data? What is the condition?
- Is it possible to satisfy the condition? Is the condition sufficient to determine the unknown? Or is it insufficient? Or redundant? Or contradictory?
- Draw a figure. Introduce suitable notation.
- Separate the various parts of the condition. Can you write them down?

Second
Find the connection between the data and the unknown. You may be obliged to consider auxiliary problems if an immediate connection cannot be found. You should obtain eventually a *plan* of the solution.

Devising a Plan
- Have you seen it before? Or have you seen the same problem in a slightly different form?
- *Do you know a related problem?* Do you know a theorem that could be useful?
- *Look at the unknown!* And try to think of a familiar problem having the same or a similar unknown.
- *Here is a problem related to yours and solved before. Could you use it?* Could you use its result? Could you use its method? Should you introduce some auxiliary element in order to make its use possible?
- Could you restate the problem? Could you restate it still differently? Go back to definitions.
- If you cannot solve the proposed problem try to solve first some related problem. Could you imagine a more accessible related problem? A more general problem? Am more special problem? An analogous problem? Could you solve a part of the problem? Keep only a part of the condition, drop the other part; how far is the unknown then determined, how can it vary? Could you derive something useful from the data? Could you think of other data appropriate to determine the unknown? Could you change the unknown or the data, or both if necessary, so that the new unknown and the new data are nearer to each other?
- Did you use all the data? Did you use the whole condition? Have you taken into account all essential notions involved in the problem?

Third
Carry out your plan.

Carrying out the Plan
- Carrying out your plan of the solution *check each step.* Can you see clearly that the step is correct? Can you prove that it is correct?

Fourth
Examine the solution obtained.

Looking Back
- Can you *check the result?* Can you check the argument?
- Can you derive the result differently? Can you see it at a glance?
- Can you use the result or the method, for some other problem?

List of References

Adams, W. W. and P. Loustaunau (1994). *An Introduction to Gröbner Bases.*

Alcock, A. and A. Simpson (2001). *The Teaching and Learning of Mathematics at University Level*, Volume 7 of *New ICMI Study Series*, Chapter The Warwick Analysis Project: Practice and Theory, pp. 99–111. Kluwer.

Aulls, M. W. (2002). The contributions of co-occurring forms of classroom discourse and academic activities to curriculum events and instruction. *Journal of Educational Psychology 94*(3), 520–538.

Badger, M., T. O. Hawkes, and C. J. Sangwin (2012). *Teaching Problem-solving in Undergraduate Mathematics.* Coventry University.

Begle, E. G. (1979). *Critical Variables in Mathematics Education: Findings from a Survey of the Empirical Literature.* National Council of Teachers of Mathematics.

Bereiter, C. and S. Engelmann (1966). *Teaching Disadvantaged Children in the Preschool.* Prentice-Hall Englewood Cliffs.

Bligh, D. A. (1998). *What's the Use of Lectures?* Intellect Books.

Blum, W. and M. Niss (1991, 2). Applied mathematical problem solving, modelling, applications, and links to other subjects – state, trends and issues in mathematics instruction. *Educational Studies in Mathematics 22*(1), 37–68.

Boyer, C. and U. C. Merzbach (1991). *A History of Mathematics* (2nd ed.). John Wiley & Sons.

Braun, V. and V. Clarke (2006). Using thematic analysis in psychology. *Qualitative research in psychology 3*(2), 77–101.

Brousseau, G. (1997). *Theory of Didactical Situations in Mathematics: Didactiques Des Mathématiques, 1970–1990.* Kluwer. Balacheff, N., Cooper, M., Sutherland, R., and Warfield, V. (Trans.).

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review 31*(1), 21–32.

Buchberger, B. (1965a). *Ein Algorithmus zum Auffinden der Basis Elemente des Restklassenrings nach einem nulldimensionalen Polynomideal.* Ph. D. thesis, Universität Innsbruck.

Buchberger, B. (1965b). Introduction to Gröbner bases. *Logic of Computation.*

Burn, R. P. (1992). *Numbers and Functions: Steps into Analysis.* Cambridge University Press.

Burton, L. (1984). Mathematical thinking: The struggle for meaning. *Journal for Research in Mathematics Education*, 35–49.

Cardiff University (2011). Cardiff school of mathematics - undergraduate degree programmes. `http://www.cardiff.ac.uk/maths/resources/ugbrochure2011.pdf`.

Chalice, D. R. (1995). How to teach a class by the modified Moore method. *American Mathematical Monthly*, 317–321.

Clark, D. M. (2007). Graphs & optimization. *Journal of Inquiry Based Learning in Mathematics 2.*

Clement, J., J. Lochhead, and G. S. Monk (1981). Translation difficulties in learning mathematics. *American Mathematical Monthly 88*(4), 286–290.

Cohen, D. W. (1982). A modified Moore method for teaching undergraduate mathematics. *The American Mathematical Monthly 89*(7), 473–490.

Coppin, C. A., W. T. Mahavier, G. E. Parker, and E. L. May (2009). *The Moore Method: A Pathway to Learner-Centered Instruction.* Mathematical Association of America.

Crisan, C., L. Fradkin, V. Zernov, and S. Lerman (2010). First year engineering mathematics: the London South Bank University experience. *MSOR Connections 10*(1), 35–39.

Dacke, M. and M. V. Srinivasan (2008). Evidence for counting in insects. *Animal cognition 11*(4), 683–689.

Dancis, J. and N. Davidson (1970). The Texas method and the small group discovery method. *The Legacy of RL Moore Project*.

Dean Jr., D. and D. Kuhn (2007). Direct instruction vs. discovery: The long view. *Science Education 91*(3), 384–397.

Dewey, J. (1910). *How We Think.* DC Heath.

Dewey, J. and A. Small (1897). *My pedagogic creed.* E.L. Kellogg & Co.

Durham University (2012). 1h mathematics for the academic year 2012/2013. `http://www.maths.dur.ac.uk/teaching/php/modules_overview.php?level=1&ayear=2012`.

Faugère, J. C. (1999). A new efficient algorithm for computing Gröbner bases. *Journal of pure and applied algebra 139*(1-3), 61–88.

Faugère, J. C. (2002). A new efficient algorithm for computing Gröbner bases without reduction to zero (F 5). In *Proceedings of the 2002 international symposium on Symbolic and algebraic computation*, pp. 83. ACM.

Field, A. P. (2005). *Discovering Statistics Using SPSS.* SAGE Publications Ltd.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. *The Nature of Intelligence 12*, 231–235.

Gerofsky, S. (1996). A linguistic and narrative view of word problems in mathematics education. *For the Learning of Mathematics 16*(2), 36–45.

Good, C. (2006). Teaching by the Moore method. *MSOR Connections 6*(2), 34–38.

Guest, G., K. M. MacQueen, and E. E. Namey (2012). *Applied Thematic Analysis.* Sage Publications Ltd.

Gutenmakher, V. L., N. B. Vasilev, M. Saul, and A. Kundu (2004). *Lines and Curves: A Practical Geometry Handbook.* Birkhauser.

Halmos, P. R. (1980). The heart of mathematics. *American Mathematical Monthly 87*(7), 519–524.

Hmelo-Silver, C. E., R. G. Duncan, and C. A. Chinn (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist 42*(2), 99–107.

Howitt, D. and D. Cramer (2011). *Introduction to research methods in psychology* (3 ed.). Prentice Hall.

Hunt, S., J. Low, and K. C. Burns (2008). Adaptive numerical competency in a food-hoarding songbird. *Proceedings of the Royal Society B: Biological Sciences 275*(1649), 2373–2379.

Iannone, P. and A. Simpson (2012a). *Mapping University Mathematics Assessment Practices.* National HE STEM Programme.

Iannone, P. and A. Simpson (2012b). *Mapping University Mathematics Assessment Practices*, Chapter A Survey of Current Assessment Practices, pp. 3–15. National HE STEM Programme.

Imperial College London (2011). Imperial College London - BSc mathematics programme specification. `https://workspace.imperial.ac.uk/mathematics/Public/students/ug/courseguides/MathsBScProgrammeSpecification.pdf`.

Jones, F. B. (1977). The moore method. *American Mathematical Monthly*, 273–278.

Kirschner, P. A., J. Sweller, and R. E. Clark (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist 41*(2), 75–86.

Klahr, D. and M. Nigam (2004). The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning. *Psychological Science 15*(10), 1–11.

Lakatos, I., J. Worrall, and E. Zahar (1976). *Proofs and refutations: The logic of mathematical discovery.* Cambridge University Press.

Lawson, D., P. Giblin, G. Goodall, M. Grove, S. Johns, N. MacKay, and N. Steele (2007). Mathematics, statistics and operational research. Technical report, The Quality Assurance Agency for Higher Education.

Lesh, R. and J. Zawojewski (2007). *Second Handbook of Research on Mathematics Teaching and Learning*, Volume 2, Chapter Problem Solving and Modeling, pp. 763–804. National Council of Teachers of Mathematics.

Lomax, R. (2001). *An Introduction to Statistical Concepts for Education and Behavioral Sciences.* Lawrence Erlbaum Associates.

London, K. C. (2011). Mathematics degree course details. `http://www.kcl.ac.uk/prospectus/undergraduate/mathematics/details`.

London School of Economics (2011). Studying for a degree in mathematics and/with economics. `http://www2.lse.ac.uk/maths/Degrees/Studying_Mathematics_and_or_with_Economics.aspx`.

Lortie, D. (1975). *Schoolteacher: A Sociological Study.* University of Chicago Press.

Mahavier, W. S. (1999). What is the Moore method? *Problems, Resources, and Issues in Mathematics Undergraduate Studies 9*(4), 339–354.

Márquez, G. G. (2003). *One Hundred Years of Solitude.* Harper.

Marx, R. W., P. C. Blumenfeld, J. S. Krajcik, B. Fishman, E. Soloway, R. Geier, and R. T. Tal (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching 41*(10), 1063–1080.

Mason, J. (2001). On the use and abuse of word problems for moving from arithmetic to algebra. In H. Chick, K. Stacey, and J. Vincent (Eds.), *Proceedings of the 12th ICMI Study Conference*, Volume 2, pp. 430–437.

Mason, J. (2012). *University Science and Mathematics Education in Transition.*, Chapter Having Good Ideas Come-To-Mind: Contemporary Pólya Based Advice for Students of Mathematics, pp. 33–51. MSOR.

Mason, J., L. Burton, and K. Stacey (1982). *Thinking Mathematically* (1st ed.). Addison-Wesley.

Mason, J., L. Burton, and K. Stacey (2010). *Thinking Mathematically* (2nd ed.). Pearson Education Limited.

Matthews, B. and L. Ross (2010). *Research Methods: A Practical Guide for the Social Sciences.* Pearson Education Limited.

Mayer, R. E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories, and templates. *Instructional Science 10*(2), 135–175.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning. *American Psychologist 59*(1), 14–19.

Milligan, S. and E. Sykes (1954). The Goon Show.

Moore, E. H. (1967). On the foundations of mathematics. *The Mathematics Teacher*, 360–374.

Neuberger, J. W. (2009). Analysis. *Journal of Inquiry-Based Learning in Mathematics 13.*

Palmer, S. (2002). Enquiry-based learning can maximise a student's potential. *Psychology Learning and Teaching 2*(2), 82–86.

Parker, J. (2005). *R. L. Moore: Mathematician and Teacher.* Mathematical Association of America.

Plewis, I. (1997). *Statistics in Education.* Arnold.

Pólya, G. (1945). *How to Solve It: A new aspect of mathematical model.* Princeton University Press.

Pólya, G. (1954a). *Mathematics and Plausible Reasoning. I. Induction and Analogy in Mathematics.*, Volume 1. Princeton University Press.

Pólya, G. (1954b). *Mathematics and Plausible Reasoning. II. Patterns of Plausible Inference.*, Volume 2. Princeton University Press.

Pólya, G. (1962). *Mathematical Discovery: On understanding, learning and teaching problem solving.* Wiley.

Pólya, G. and G. Szegö (1925). *Aufgaben Und Lehrsätze Aus Der Analysis*, Volume 1. Dover.

Qualifications and Curriculum Authority (2007a). Mathematics – programme of study for key stage 3 and attainment targets.

Qualifications and Curriculum Authority (2007b). Mathematics – programme of study for key stage 4.

Queen Mary, University of London (2012). Modules 2012–13. `http://www.maths.qmul.ac.uk/undergraduate/modules`.

Resnick, L. B. (1988). Treating mathematics as an ill-structured discipline. *The Teaching and Assessing of Mathematical Problem Solving 3*, 32.

Richardson, D. (1968). Some undecidable problems involving elementary functions of a real variable. *The Journal of Symbolic Logic 33*(4), 514–520.

Robson, E. (2008). *Mathematics in Ancient Iraq: A Social History.* Princeton University Press.

Rowlett, P. (Ed.) (2011, January). *HE Mathematics Curriculum Summit.*

Sangwin, C. J. (2011, November). Modelling the journey from elementary word problems to mathematical research. *Notices of the American Mathematical Society 58*(10), 1436–1445.

Sangwin, C. J. (2013). *Computer Aided Assessment of Mathematics.* Oxford University Press.

Schoenfeld, A. H. (1992). *Handbook of Research on Mathematics Teaching and Learning*, Chapter Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics, pp. 334–370. MacMillan.

Shannon, C. E., W. Weaver, R. E. Blahut, and B. Hajek (1949). *The Mathematical Theory of Communication*, Volume 117. University of Illinois Press.

Smith, J. C. (2006). A sense-making approach to proof: Strategies of students in traditional and problem-based number theory courses. *The Journal of Mathematical Behavior 25*(1), 73–90.

Snyder, B. (1973). *The Hidden Curriculum.* Alfred-Knopf.

Sternberg, R. J. and T. Ben-Zeev (1996). *The Nature of Mathematical Thinking*. Lawrence Erlbaum Associates, Inc.

Tall, D. and S. Vinner (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics 12*(2), 151–169.

The Educational Advancement Foundation (2006). The educational advancement foundation. `http://eduadvance.org/`.

University College London (2011). Undergraduate mathematics syllabuses. `http://www.ucl.ac.uk/Mathematics/Courses/index.html`.

University of Birmingham (2010). University of Birmingham School of Mathematics – Student handbook 2010-2011.

University of Bristol (2011). First-year (level 1) mathematics units 2010-11. `http://www.maths.bris.ac.uk/study/PDFdocs/Level1_overview.pdf`.

University of Cambridge (2011). Guide to the mathematical tripos. `http://www.maths.cam.ac.uk/undergrad/course/triposguide.pdf`.

University of Exeter (2012). Modules running in 2012/13. `http://intranet.exeter.ac.uk/emps/taughtprog/mathematics/modules/2012/`.

University of Glasgow (2011). Mathematics and statistics - undergraduate study 2010. `http://www.gla.ac.uk/media/media_126858_en.pdf`.

University of Liverpool (2011). University of Liverpool – mathematical sciences. `http://www.liv.ac.uk/study/undergraduate/brochures/mathematical_science.pdf`.

University of Manchester (2011). Undergraduate handbook - modes of study. http://www.maths.manchester.ac.uk/undergraduate/ugstudies/Handbook-Present/2-10Handbook-2011.pdf.

University of Newcastle (2011). Mathematics bsc honours - Course content. http://www.ncl.ac.uk/undergraduate/course/G100/mathematics.

University of Nottingham (2011). Mathematical sciences - undergraduate study. http://www.nottingham.ac.uk/mathematics/documents/brochures/ugmathsbrochureforweb.pdf.

University of Oxford (2011). Studying mathematics at Oxford. https://www.maths.ox.ac.uk/system/files/attachments/introbook11_0.pdf.

University of Sheffield (2011). Undergraduate degree structure. http://www.shef.ac.uk/maths/prospectiveug/courses/structure.html.

University of Warwick (2011). Teaching - mathematics undergraduate admissions. http://www2.warwick.ac.uk/fac/sci/maths/admissions/ug/teaching.

von Glasersfeld, E. (1990). Chapter 2: An exposition of constructivism: Why some like it radical. *Journal for Research in Mathematics Education. Monograph 4*, 19–29.

Vygotsky, L. S. (1978). *Mind in Society.* Harvard University Press.

Weber, K. and L. Alcock (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics 56*(2), 209–234.

Wood, D., J. S. Bruner, and G. Ross (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry 17*(2), 89–100.

Young, R. M. and T. O'Shea (1981). Errors in children's subtraction. *Cognitive Science 5*(2), 153–177.